Journal of Computer-Aided Molecular Design, 9 (1995) 283–294
ESCOM

J-CAMD 293

Quantitative structure–agonist activity relationship of capsaicin analogues

Gilles Klopman* and Ju-Yun Li

Department of Chemistry, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-7078, U.S.A.

Received 1 November 1994
Accepted 17 February 1995

Keywords: QSAR; MULTICASE; Capsaicin; Pharmacophore

Summary

The MULTIple Computer Automated Structure Evaluation (MULTICASE) methodology has been used to study the quantitative structure–agonist activity relationship of a series of capsaicin agonists. A number of substructures and physicochemical properties of capsaicin analogues were identified as being responsible for high agonist potency. The optimal log P value for the agonist potency as estimated from QSAR analysis is 5.12. It was also found that a cluster of inactive molecules in the database have lipophilicity values below 2.94. Molecular modeling was employed to elucidate the detailed structural features of the pharmacophore of capsaicin analogues. Systematic conformational analysis has shown that the activity of capsaicin analogues strongly depends upon their ability to reach the required conformational profile. Based upon these observations, a three-dimensional pharmacophore model for the capsaicin–receptor interactions is proposed.

Introduction

Capsaicin (see Fig. 1) [1], the principal pungent constituent of hot peppers, one of the most heavily consumed spices [2], has been identified as a prototype for a novel class of analgesics [3–5]. The basic structure of capsaicin is that of an amide (B region), with a hydrophilic ring on one end (A region) and a lipophilic carbon chain (C region) on the other. Its structure is different from the other two known types of analgesic agents, i.e., the opioids and the non-steroidal anti-inflammatory agents (NSAIDs). Capsaicin is known to have a wide spectrum of biological effects on the cardiovascular and respiratory systems [1,6]. It induces an initial pain response by topical application to skin or mucous membrane, followed by desensitization. It is therefore of interest to identify the structural features that are necessary for activation or blockage of the capsaicin receptor, in order to separate the desired analgesic effects from the algesic effects [7]. Previous studies on the molecular mechanism of capsaicin have suggested that its agonist activity is probably initiated by interaction with a specific receptor–ion channel complex in the membrane, inducing an increase in cytosolic Ca^{2+} concentration, which in turn causes selective excitation of some neurons and a period of desensitization [8–13].

The potential clinical use of the analgesic and peripheral anti-inflammatory effects of capsaicin [1,6,14], as well as the discovery of an ultra-potent capsaicin analogue, resiniferatoxin (see Fig. 2), has attracted much attention and prompted a number of investigators to study the relationship between the structures of capsaicin analogues and their agonist activities [6,15–21]. The existence of a specific 'receptor' peculiar to this subset of sensory neurons was not evident until it was demonstrated that [^3H]resiniferatoxin specifically binds to neurons that are sensitive to capsaicin [16,19,20,22–24]. Resiniferatoxin was found to exhibit the characteristic biological activities of capsaicin. Furthermore, the discovery of the first competitive antagonist of capsaicin and resiniferatoxin, capsazepine [17,18,25,26], has strengthened the suspicion of the existence of a specific receptor.

Early work done by Szolcsanyi and Jansco-Gabor established a structure–activity profile for a small set of capsaicin analogues [27,28]. The activities of these compounds were assessed based on behavioral tests, such as

*To whom correspondence should be addressed.

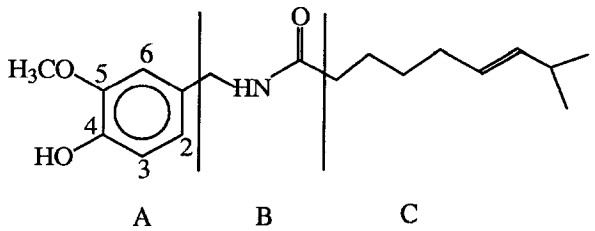


Fig. 1. Structure of capsaicin.

measurements of eye wipe in rat and pungency on the human tongue. Christopher et al. have recently published a series of papers dealing with the effects of structural variation on the agonist activity by modifying the functionalities of capsaicin in the three regions (see Fig. 1) [4,26,29,30]. They established a comprehensive structure–activity relationship (SAR) for the capsaicin analogues, indicating that (i) a parent 3-methoxy-4-hydroxybenzyl ring in the A region is required; (ii) a dipolar amide or thiourea in the B region is beneficial; and (iii) a lipophilic octanyl or *p*-chlorophenethyl is the most beneficial group in the C region. A two-dimensional schematic model was proposed to rationalize the structure–activity profile of compounds which differ in the B region, based on consideration of multiple H-binding interactions.

Until now, however, no quantitative structure–activity relationship (QSAR) analysis of the capsaicin analogues has been reported in the literature. Therefore, we have applied the MULTICASE program, an artificial intelligence system, to evaluate whether such a relationship could be found. Our program was able to uncover several crucial substructures and physicochemical properties relevant to activity, and QSAR equations correlating the agonist activity were indeed obtained.

Molecular modeling and conformational analysis were performed to gain a better understanding of the interactions between capsaicin and its receptor, in an attempt to identify the conformation required for activity. The molecular modeling study showed that the activities of capsaicin analogues strongly depend upon their conformational profiles. It was found that for some inactive compounds, their lack of activity is primarily due to their inability to achieve an energetically favorable conformation shared by the active compounds. Based upon this study, a three-dimensional pharmacophore model for the capsaicin–receptor interactions has been proposed that can be used to design new and hopefully more active compounds.

Materials and Methodologies

Experimental data

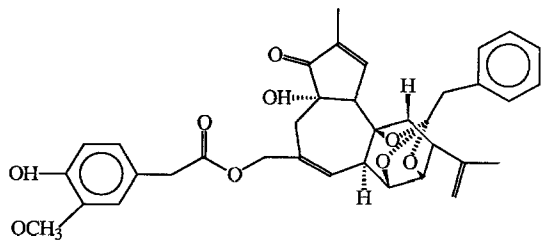
A database of 123 capsaicin analogues was compiled from recent publications by Walpole et al. [4,26,29–31]. These compounds were tested as potential agonists in an assay in which the $^{45}Ca^{2+}$ influx activity was measured [10,19,26]. Experimental values were expressed as the concentration needed for 50% (EC_{50}) of maximal Ca^{2+} influx induced into neonatal rat dorsal root ganglia (DRG). The structures and experimental activities of these analogues are listed in Table 1.

MULTICASE methodology

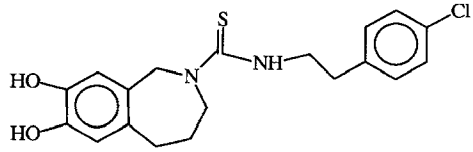
The MULTIple Computer Automated Structure Evaluation (MULTICASE) and its predecessor, the Computer Automated Structure Evaluation (CASE), have been described in various publications [32,33]. Basically, both the CASE and MULTICASE programs are artificial intelligence systems, capable of identifying structural descriptors that may be associated with the properties of the molecules examined. Molecular structure and experimental activity are the input required by the program. A breakpoint activity is chosen between the active and inactive molecules of the learning set, so as to ensure a suitable distribution between active and inactive molecules in the data set for subsequent discriminant analysis.

Specifically, the program takes each molecule of the learning set and breaks it into all possible linear fragments of 2 to 10 connected heavy atoms. These fragments are labeled as active or inactive, depending on their origin. The occurrence of these fragments is then submitted to a series of statistical analyses, to identify those whose distribution is significantly skewed towards either active or inactive in the entire database. A fragment with significant (> 80%) binomial distribution probability between active and inactive molecules is believed to be activating (biophore) or inactivating (biophobe), respectively. Fragments encountered randomly in both active and inactive molecules are regarded as irrelevant to activity.

MULTICASE performs the analysis in a hierarchical way, rather than in a one-step regression as in CASE.



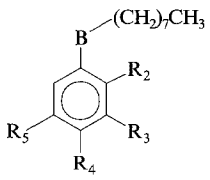
Resiniferatoxin



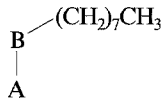
Capsazepine

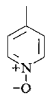
Fig. 2. Structures of resiniferatoxin and capsazepine.

TABLE 1
EXPERIMENTAL ACTIVITIES AND STRUCTURES OF 123 CAPSAICIN ANALOGUES

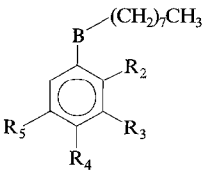


Compound no.	R_2	R_3	R_4	R_5	B	Ca^{2+} influx EC_{50} (μM)
1	H	H	H	H	CH_2NHCO	>100
2	H	OCH_3	OH	H	CH_2NHCO	0.55
3	H	OCH_3	OCH_3	H	CH_2NHCO	6.41
4	H	OCH_3	H	H	CH_2NHCO	>100
5	H	H	OCH_3	H	CH_2NHCO	>100
6	OCH_3	H	H	H	CH_2NHCO	>100
7	H	OH	H	H	CH_2NHCO	>100
8	H	$-OCH_2O-$		H	CH_2NHCO	>100
9	H	OH	OH	H	CH_2NHCO	0.63
10	H	OCH_3	H	OH	CH_2NHCO	>100
11	OCH_3	H	OH	OCH_3	CH_2NHCO	>100
12	OCH_3	H	OH	H	CH_2NHCO	>100
13	H	OCH_3	OCH_3	OCH_3	CH_2NHCO	>100
14	OH	OH	OH	H	CH_2NHCO	7.64
15	H	OCH_3	SH	H	CH_2NHCO	>100
16	H	OCH_3	NO_2	H	CH_2NHCO	7.91



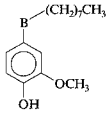
	A	B	Ca^{2+} influx EC_{50} (μM)
17		CH_2NHCO	28.0
18		CH_2NHCO	>100
19		CH_2NHCO	31.5
20		CH_2NHCO	>100

	A	B	Ca^{2+} influx EC_{50} (μM)
21		CH_2NHCO	>100
22		CH_2NHCO	>100
23		CH_2NHCO	>100

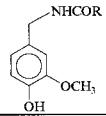


	R_2	R_3	R_4	R_5	B	Ca^{2+} influx EC_{50} (μM)
24	H	OCH_3	H	H	CH_2CONH	>100
25	H	OCH_3	OH	H	CH_2CONH	0.30
26	H	H	OH	H	CH_2CONH	6.50
27	H	$-OCH_2O-$		H	CH_2CONH	>100
28	H	OH	OH	H	CH_2CONH	0.41
29	H	CH_3	OCH_3	H	CH_2CONH	>100
30	H	CH_3	OH	H	CH_2CONH	>100
31	H	NO_2	OH	H	CH_2CONH	1.04
32	H	NH_2	OH	H	CH_2CONH	>100
33	H	$NHCOCH_3$	OH	H	CH_2CONH	>100
34	H	OC_2H_5	OH	H	CH_2CONH	4.34
35	H	OH	OCH_3	H	CH_2CONH	1.82
36	H	$-CH=CH-CH=N-$		H	CH_2CONH	>100
37	H	OCH_3	OH	OCH_3	CH_2CONH	>100
38	$-OCH_2CH_2O-$		OH	H	CH_2CONH	>100
39	OCH_3	OCH_3	OH	H	CH_2CONH	>100

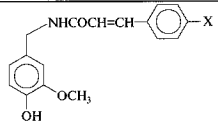
TABLE 1 (continued)



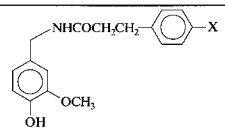
	B	Ca^{2+} influx EC_{50} (μM)
40	NHCO	4.48
41	$(CH_2)_2NHCO$	18.30
42	$CH_2N(CH_3)CO$	>100
43	CH_2OCO	14.20
44	CH_2NHCS	0.28
45	CH_2NHSO_2	1.32
46	CONH-	>100
47	$(CH_2)_2CONH$	2.32
48	$CH_2CON(CH_3)$	6.29
49	CH(OH)CONH	1.16
50	CH_2COO	0.67
51	CH_2COS	1.17
52	CH=CHCONH (*E*)	>100
53	CH=CHCONH (*Z*)	17.90
54	$CH_2NHCONH$	0.36
55	$CH_2NHCSNH$	0.06
56	NHCSNH	2.57
57	$CH_2N(CH_3)CSNH$	>100
58	$CH_2NHCSN(CH_3)$	0.53
59	$CH_2NHCSNHCO$	>100
60	$CH_2NHC(=NCN)NH$	3.28
61	$CH_2NHC(=CH\text{-}NO_2)NH$	>100
62	CH=CHCO (*E*)	>100
63	CH_2CH_2CO	2.13
64	CH_2COCH_2	3.78
65	$CH_2CH_2CH_2$	>100

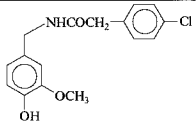
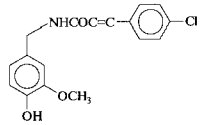
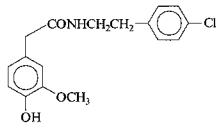


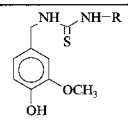
	R	Ca^{2+} influx EC_{50} (μM)
66	$(CH_2)_4CH=CHCH(CH_3)_2$ (*E*)	0.30
67	$(CH_2)_6CH(CH_3)_2$	0.19
68	$(CH_2)_{10}CO_2H$-	100
69	$CH_2O(CH_2)_2OCH_3$	>100
70	$CH_2O(CH_2)_2O(CH_2)_2OCH_3$	>100
71	$(CH_2)_{16}CH_3$	>100
72	$(CH_2)_7CH=CH(CH_2)_7CH_3$ (*E*)	0.36
73	$(CH_2)_7CH=CH(CH_2)_7CH_3$ (*Z*)	0.17
74	CH_2Br	>100
75	CH_2Cl	>100



	X	Ca^{2+} influx EC_{50} (μM)
76	4-H (*E*)	11.80
77	4-Cl (*E*)	1.24
78	4-Cl (*Z*)	50.10
79	4-NO_2 (*E*)	4.58
80	4-CN (*E*)	26.50
81	4-Ph (*E*)	0.24
82	4-$N(CH_3)_2$ (*E*)	4.39
83	4-I (*E*)	0.35
84	4-NHCHO (*E*)	1000
85	2,4-Cl,Cl	0.62
86	3-Cl	2.58



	X	Ca^{2+} influx EC_{50} (μM)
87	H	45.26
88	4-Cl	3.09
89	4-NO_2	21.0
90	4-OH	>100
91		7.77
92		4.10
93		0.66



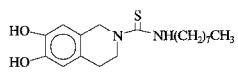
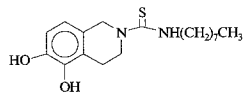
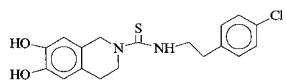
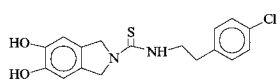
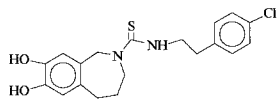
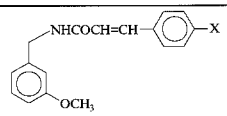
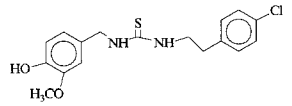
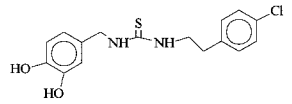
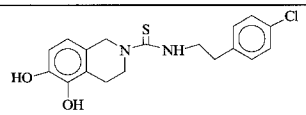
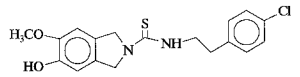
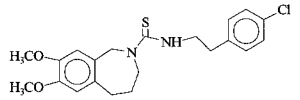
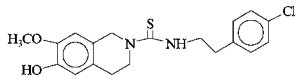
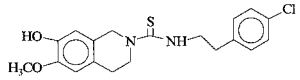
	R	Ca^{2+} influx EC_{50} (μM)
94	H	>100
95	$(CH_2)_3CH_3$	9.48
96	$(CH_2)_4CH_3$	1.96
97	$(CH_2)_5CH_3$	0.18
98	$(CH_2)_6CH_3$	0.29
99	$(CH_2)_8CH_3$	0.10
100	$(CH_2)_9CH_3$	0.11
101	$(CH_2)_{11}CH_3$	0.16
102	$(CH_2)_{15}CH_3$	>100
103	$(CH_2)_{17}CH_3$	8.1
104	$(CH_2)_7CH=CH(CH_2)_7CH_3$ (*Z*)	>100
105	adamantyl	1.08
106	$CH(Ph)_2$	0.27
107	$C(Ph)_3$	>100
108		0.74
109		>100
110		0.3
111		0.5
112		>100; Antagonist IC_{50}=0.5 μM

TABLE 1 (continued)



	X	Ca^{2+} influx EC_{50} (μM)
113	4-NHCHO (*E*)	>100
114	4-NO_2 (*E*)	4.58
115	4-H (*E*)	11.80
116	4-$N(CH_3)_2$ (*E*)	4.39
117	4-Cl (*E*)	1.24
118	4-I (*E*)	0.35
119		0.06
120		0.1
121		>100
122		>100
123		>100
124		1.09; Partial agonist; 18% efficacy, not included in the learning set
125		1.23; Partial agonist; 41% efficacy, not included in the learning set

Initially, the program identifies the substructure with the highest probability of being responsible for activity, so that it can explain as many active molecules in the learning set as possible. These molecules are removed from the data set, and the remaining molecules are searched for the next significant substructure. This procedure is repeated until either enough biophores or biophobes have been found to account for the activity of the entire data set, or the remaining data cannot be explained by statistically significant descriptors.

The molecules in the database are then divided into subclasses, based on the presence of each biophore. For each subclass characterized by a common biophore, a local QSAR is performed to identify modulators that are capable of modifying the activity due to the biophore. The modulators might be substructural descriptor or molecular physicochemical properties, such as the partition coefficient (log P) [34], solubility [35], molecular weight or quantum mechanics parameters (HOMO and LUMO) [36].

Once the learning set has been analyzed, the resulting descriptors relevant to the observed activity can be used to predict the activity of new molecules and as a basis for speculation about possible underlying mechanisms.

Molecular modeling and conformational analysis

All molecular modeling manipulations and conformational analyses were performed using the Insight II (version 2.2.0) [37] molecular modeling package on a Silicon Graphics Indigo R3000. The three-dimensional structure of each compound was built using the builder module. All energies were calculated in the Discover module (version 2.9/3.1) using the Consistent Valence Force Field (CVFF) [38] until the rms (root-mean-square) derivative of the energy was below 0.0001 kcal/mol/Å. The resulting minimized conformation for each compound was then used as the starting conformation in a systematic conformational search.

The systematic conformational search was performed in the Search_Compare module by varying the defined torsional angles with a step size of 30°. A conformation was accepted only if no steric clashes occurred between van der Waals radii of the constituent atoms in the compound and the conformational energy was within 10 kcal/mol of the most stable conformation generated from this search. The accepted conformations were then submitted to an energy minimization routine.

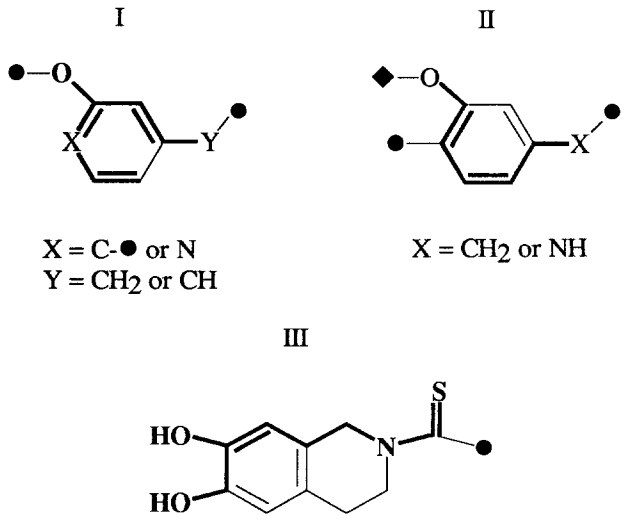


Fig. 3. Biophores identified by the MULTICASE program for Ca^{2+} influx agonist activity. They are shown in boldface lines and symbols. • refers to non-hydrogen atoms or groups; ♦ refers to any atoms or groups.

TABLE 2
CALCULATED log P VALUES [34] OF THE COMPOUNDS EVALUATED

Compound	Calculated log P	Compound	Calculated log P	Compound	Calculated log P
1	5.11	42	5.22	83	4.31
2	4.95	43	5.51	**84**	**2.56**
3	5.06	44	5.86	85	4.47
4	5.09	45	5.09	86	3.82
5	5.09	46	4.55	87	3.38
6	5.09	47	5.35	88	4.03
7	4.97	48	5.22	89	3.37
8	5.29	49	4.04	90	3.24
9	4.83	50	5.51	91	3.63
10	4.95	51	5.93	92	3.55
11	4.92	52	5.13	93	4.03
12	4.95	53	5.13	**94**	**1.65**
13	5.04	54	4.94	95	3.42
14	4.69	55	5.89	96	3.82
15	5.64	56	5.49	97	5.08
16	5.08	57	5.52	98	5.49
17	4.48	58	5.52	99	6.29
18	3.19	59	4.83	100	6.69
19	4.46	60	7.32	101	7.50
20	4.48	61	6.93	102	9.11
21	5.38	62	5.64	103	9.91
22	3.34	63	5.86	104	9.29
23	5.14	64	5.86	105	4.36
24	5.09	65	6.78	106	5.53
25	4.95	66	4.18	107	7.28
26	4.97	67	4.40	108	5.26
27	5.29	68	4.31	109	5.26
28	4.83	**69**	**1.24**	110	4.34
29	5.47	**70**	**1.42**	111	4.10
30	5.36	71	8.57	112	4.59
31	4.97	72	8.35	**113**	**2.70**
32	4.30	73	8.35	114	3.30
33	4.33	**74**	**2.01**	115	3.30
34	5.35	**75**	**1.72**	116	3.35
35	4.95	76	3.16	117	3.96
36	5.59	77	3.82	118	4.45
37	4.92	78	3.82	119	4.97
38	5.40	79	3.16	120	4.86
39	4.92	80	2.85	121	4.34
40	4.2	81	4.85	122	4.21
41	5.35	82	3.21	123	4.82

Compounds **69**, **70**, **74**, **75**, **84**, **94** and **113**, shown in boldface, are the inactive compounds whose log P values are below 2.94. Of these, only compound **113** does not contain the top biophore.

Results and Discussion

MULTICASE analysis

The database of 123 capsaicin analogues was submitted to MULTICASE analysis. The experimental data and structures of these compounds are listed in Table 1. Compounds were considered to be inactive when their EC_{50} was above 100 μM. As a result, of the 123 compounds in the database, 66 were classified as being active to a varying degree and 57 as inactive.

MULTICASE first performs a cluster analysis, and the results showed that log P can be used as a criterion to identify a cluster from the entire database. Indeed, no compound exhibiting a log P value [34] below 2.94 was found to be active. Thus, seven compounds were considered to be inactive because of their low lipophilicity, in spite of the fact that some of them contain a moiety recognized by the program as being relevant to activity (see below). These compounds are probably not able to effectively pass through the lipophilic cell membranes to reach the active site of the receptor [39,40]. As a result, it is anticipated that log P will be a significant contributor to the QSAR that fits the activity of the capsaicin analogues.

After the cluster analysis, MULTICASE identified three significant fragments as biophores with a high probability of relevance (Fig. 3). Upon first examination, it

appears that the potency of a capsaicin analogue is highly related to the nature of the A region. Biophore I is the most significant fragment, since this fragment alone could account for 76% of the active compounds in the database. Biophore I shows that compounds containing a 3-alkoxy-4-substituted benzyl group in the A region have a high probability of being active. Attempts to substitute any of the three hydrogens at positions 2, 5 and 6 (Fig. 3) with other groups yielded compounds with a dramatic decrease in potency (compounds **11**, **13**, **14**, **37**, **38** and **39**). Although biophore I did not specify the substituent requirement at the 4 position, it was found that 66 of the 70 compounds containing this biophore have a hydroxyl group at this position. This is highly consistent with observations made by other groups [4]. Furthermore, it has been noticed that the removal of the phenolic OH group at the 4 position of the benzene ring resulted in abolition of the agonist activity (compounds **4**, **24**, **113–118**) and such compounds were found to exhibit, at best, weak activity as capsaicin antagonists.

Biophore I occurs in 70 compounds, 52 of which are active. Among the 18 inactive compounds, six have been found in the cluster of molecules having lipophilicity values below a certain value (2.94), as shown in Table 2. Compounds containing biophore I were then submitted for a further CASE analysis, in order to determine the properties that modulate the activity of a compound due to biophore I. The following QSAR, Eq. 1, was obtained:

$$\log (1/EC_{50}) = -3.63 + 1.29\, n_1M_1 + 0.40\, n_2M_2 + 2.45\, n_3M_3 - 1.08\, n_4M_4 - 1.77\, n_5M_5 - 0.12 \log^2 P + 1.23 \log P \quad (1)$$

$$n = 70;\ SD = 0.77;\ r = 0.72;\ F\ (7, 62, 0.05) = 9.40$$

where EC_{50} (in μM) is the concentration of a drug necessary to produce 50% of the maximal response, M_i is the *i*th modulator, n_i is the occurring frequency of M_i in a molecule, and log P is the partition coefficient between *n*-octanol and water, calculated by a procedure described in a previous paper [34]. The calculated log P values of the compounds examined are listed in Table 2. The five structural modulators identified by this QSAR analysis are shown in Fig. 4. The optimal log P value [41] for agonist potency, as estimated from Eq. 1, is 5.12.

With this equation, 51 out of the 52 active compounds are correctly classified. Of the 18 inactive compounds, three were correctly classified with Eq. 1 and six were classified as being inactive (compounds **69**, **70**, **74**, **75**, **84** and **94**) because of their low log P values (see above). The incorrect assignment of the remaining nine inactive compounds, falsely positive, is due to the fact that no deactivating modulator was identified for these compounds. It suggests that other unidentified factors may be detrimental to activity.

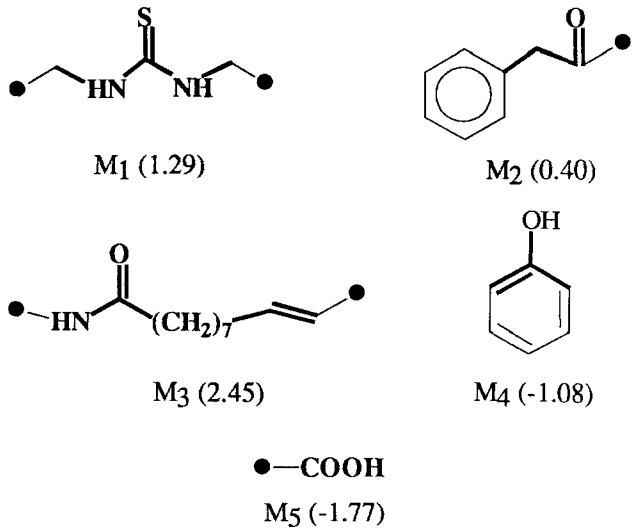


Fig. 4. Modulators of biophore I. Numbers in parentheses refer to their activity contribution. The larger the number, the more potent the drug that contains the corresponding modulator.

Of the five substructural modulators, M_1–M_3 are seen as activating fragments whereas M_4 and M_5 are deactivating fragments. Close examination of the compounds containing these modulators revealed that M_1–M_3 are related to the dipolar B region, while M_4 and M_5 are related to the nonpolar groups in the C region. Comparisons between M_1–M_3 showed that the presence of an H-bond acceptor in these modulators, such as a C=S in M_1 and a C=O in M_2 and in M_3, are common. Thus, compounds **63** and **64**, which contain a carbonyl group in the B region, have EC_{50} values of 2.13 and 3.78 μM, respectively, while compound **65**, lacking the carbonyl group in the B region, is devoid of activity ($EC_{50} > 100$ μM).

Modulator 1, consisting of a thiourea moiety with methylene groups at each end, is the most significant and strongest activating modulator associated with biophore I. The reason for its beneficial effect on potency is still not clear. It is of interest, though, to find that through incorporation of the bridging methylene, the planar thiourea moiety in the B region introduces some rigidity into the A region.

N-methylation of the amide and the thiourea in the B region yields less active (**48** and **58**) or even inactive (**42** and **57**) compounds as compared to compounds **2**, **25** and **55**. Methylation on the nitrogen atom adjacent to the A region, as seen in compounds **42** and **57**, leads to total inactivity, while methylation on the distal nitrogen in compounds **48** and **58** only marginally reduces the activity. We thus hypothesize that the spatial relationship between the thiourea plane and the 3-methoxy-4-hydroxy phenyl ring might play a crucial role in determining the activity. Interestingly, Walpole et al. have proposed that the angle between the plane of the thiourea and the plane of the benzene ring is critical for agonist activity [26], which is highly consistent with our hypothesis.

To investigate this hypothesis further, we have performed an extensive molecular modeling and conformational analysis of the relevant compounds. Details are described below.

A hydrophilic moiety, such as a carboxylic group in M_4 and a phenolic hydroxyl group in M_5, totally destroys the agonist activity of a capsaicin analogue. Both of these modulators were found to appear in the C region. On the other hand, lipophilic substituents, such as octanyl or *p*-chloro phenethyl, are found to be beneficial in this region. Therefore, the lipophilicity of the C region appears to play a critical role in agonist activity. This observation leads to two possible assumptions: (i) the receptor cavity is spatially restricted and highly lipophilic for accommodating the C region moiety of a capsaicin analogue, or (ii) the polar group at the end of the C region interacts with the π-cloud of the aromatic ring in the A region, through a so-called folding effect. A molecular dipole ($Ar^{\delta+}$-$X^{\delta-}$) would be present which may reduce the lipophilicity of the molecule and consequently change the nature of its conformational profile.

The next significant activating fragment, biophore II, is similar to biophore I. In biophore II, a hydroxyl group at the 3 position and an NH-group next to the phenyl ring are acceptable compared with biophore I. No specific substituent is required at the 5 position. Seven active compounds were explained by this biophore.

The last significant activating fragment, biophore III, indicates that an α,α-catechol moiety in the A region of molecules where the A and B regions are constrained by the existence of a fusing ring is beneficial as well (compounds **108**, **110**, **111** and **112**). The size of the fused ring was found to have a dramatic impact on the agonist activity. For example, compounds **108**, **110** and **111** (five- and six-membered rings) are potent agonists, while compound **112** (seven-membered ring) has no agonist activity at a concentration of 100 μM, but evokes a moderately antagonist activity. Molecular modeling of these compounds suggests that the relative spatial disposition of the α,α-catechol and the thiourea planes is crucial. Further details are described in the next section.

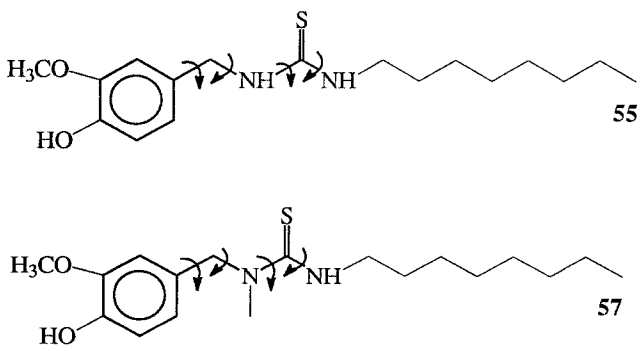


Fig. 5. Defined torsional angles in some typical analogues.

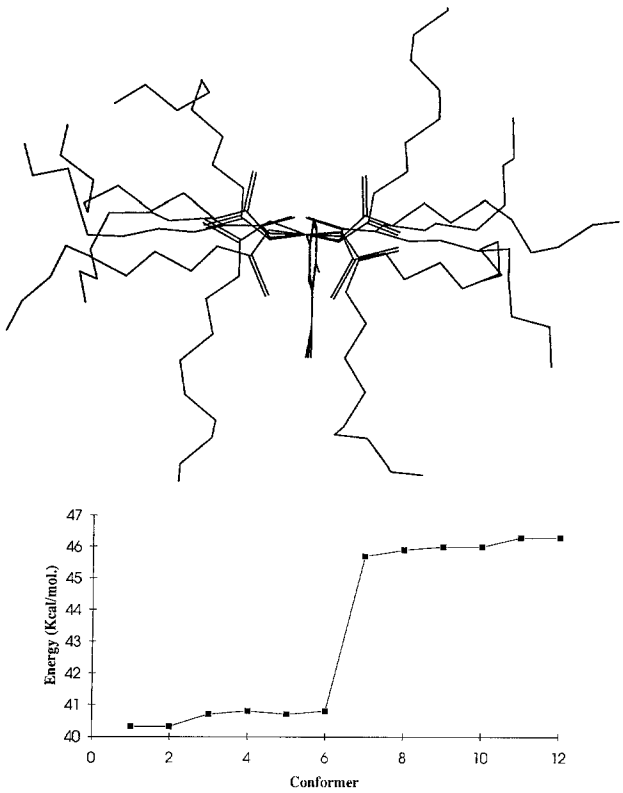


Fig. 6. Low-energy conformations of the most active compound, **55** (top), and graph of the conformational energy versus each conformer (bottom). The configurations of thiourea in conformers 1–6 are trans,trans, while those in conformers 7–12 are trans,cis.

Molecular modeling and conformational analysis

Since detailed knowledge of the capsaicin receptor has not yet been obtained, information regarding the three-dimensional structure of the pharmacophore of capsaicin analogues can only be derived indirectly from the binding affinity data of a congeneric set of molecules. Therefore, the capsaicin analogues which contain the parent 3-methoxy-4-hydroxy benzyl group, the top biophore mentioned previously, and several similar compounds were selected for further molecular modeling and conformational analysis.

Our initial conformational analysis used the mapping of distance constraints between essential atoms identified by MULTICASE, such as the oxygen atoms on the phenyl ring, the center of the phenyl ring and the oxygen/sulfur atom in the dipolar B region. The results showed that the loss in biological activity of some compounds might be due to the unattainability of the active conformation, since the conformational space that contains sterically allowed conformations common to the set of active compounds is absent in the inactive compounds. Furthermore, in the course of our SAR study it was noticed that alkylation of the nitrogen atom of thiourea in the B region results in a dramatic reduction of the agonist activity (compounds **55**, **56**, **57**, **122** and **124**).

Indeed, the change in agonist activity upon addition of a methyl group to the nitrogen atom (adjacent to the A region) in **55**, to give **57**, was remarkable. A rationale has been sought for the different biological activities of these two compounds. To this end, we performed a systematic conformation search on the two compounds. The torsional angles (Fig. 5) defining the respective positions of the A and B regions were varied with 30° increments.

For compound **55**, 12 low-energy conformations were obtained (Fig. 6). The thiourea moiety adopts a trans, trans-planar configuration in the global minima [42]. The trans,cis-planar configuration was found in the local minima, but is about 5 kcal higher in energy than the trans,trans configuration. For compound **57**, 10 low-energy conformations were obtained (Fig. 7). The methylated thiourea moiety adopts a cis,trans-planar configuration in the global minimum, instead of the trans,trans configuration which is less stable due to strong repulsive nonbonded interactions between the bulky methyl group on the nitrogen and the phenyl ring. Moreover, a systematic conformational search showed that compounds **55** and **56** have a comparable conformational space, while that of compound **57** is much narrower. Apparently, the steric effect present in compound **57** has a dramatic impact on its agonist activity.

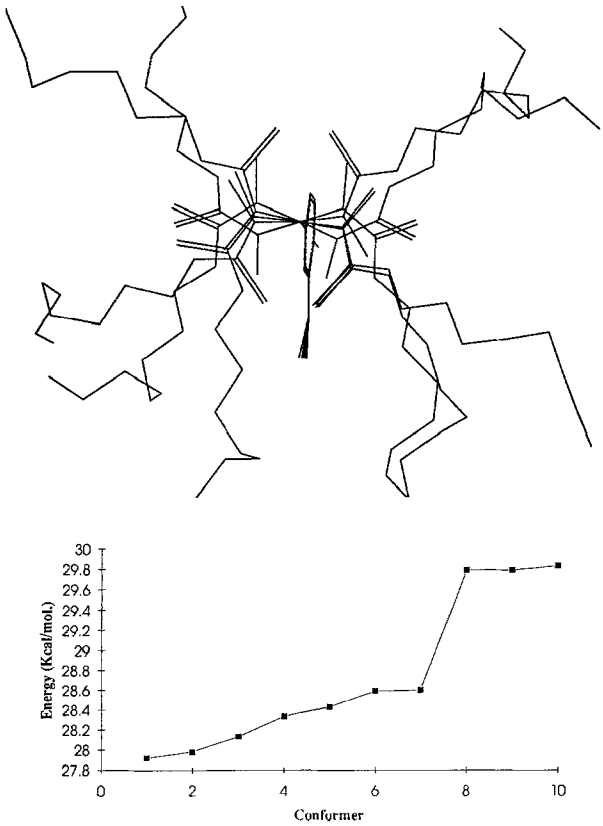


Fig. 7. Low-energy conformations of the inactive compound **57** (top) and graph of the conformational energy versus each conformer (bottom). The configurations of thiourea in conformers 1–6 are trans,cis, while those in conformers 7–10 are trans,trans or cis,cis.

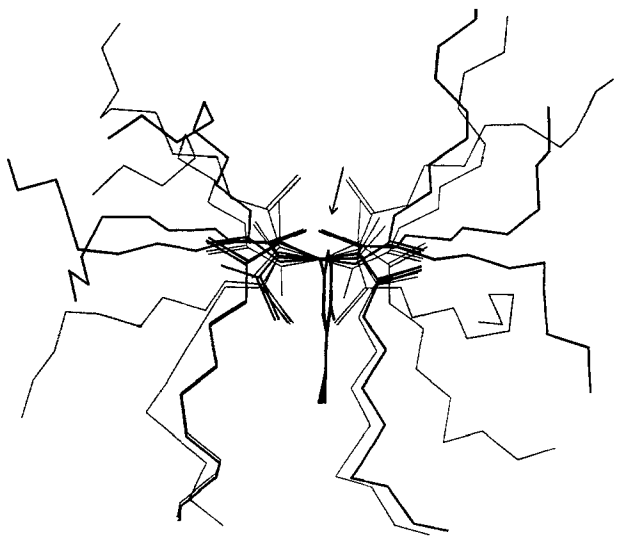


Fig. 8. Superimposition of the low-energy conformations of **55** (thick) onto **57** (thin). The global minima conformations of **55** are indicated by an arrow.

In an attempt to systematically evaluate the difference in the conformational space between these two compounds, we superimposed (see Fig. 8) all their low-energy conformations. As shown in Fig. 8, most of the conformations overlap well, except for two global minima of compound **55** that are not attainable by the inactive compound **57**. Indeed, a large conformational energy is required for compound **57** to adopt these conformations. Upon close examination of the two conformations, it was found that two torsional angles are crucial for defining the spatial relationship between the thiourea plane and the phenyl ring. In other words, the methylene unit acts as a joint to hold these two planes in an optimal spatial orientation. The two energetically equal minima were found to be symmetric with respect to the 3-methoxy-4-hydroxy phenyl ring. We therefore hypothesize that these might be the active conformations and the spatial relationship between the 3-methoxy-4-hydroxy phenyl ring and the thiourea in these conformations is the pharmacophore needed for capsaicin analogues to exert their agonist activity. The three-dimensional model for the assumed pharmacophore is shown in Fig. 9. The angle between the phenyl ring and the thiourea plane is about 84° and the distances between the sulfur atom and the two oxygen atoms on the phenyl ring are 7.78 and 8.71 Å, respectively.

It is well accepted that the global minimum conformation of a compound (non-receptor bound) is not necessarily the active (bound) conformation, as long as it can adopt the active conformation at a sufficiently low expense of energy. We therefore evaluated the energy required for a capsaicin analogue to mimic the assumed active conformation. A subset of 13 compounds, in which the substructures in the A and C regions were kept con-

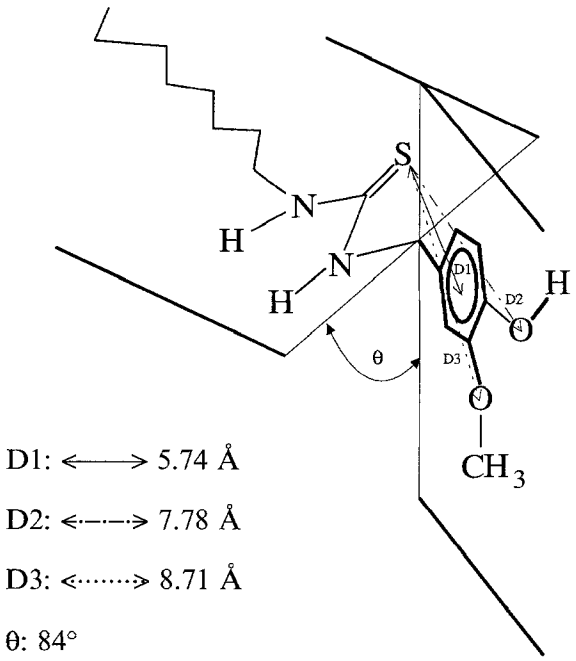


Fig. 9. Three-dimensional model of the proposed pharmacophore.

stant while the functionality in the B region was varied, was selected for this analysis. The structures are shown in Table 3 and the energies needed to reach the reference conformation are given in Table 4. The number of H-bond acceptors/donors matching the thiourea moiety of the reference compound **55** and the corresponding rms deviations (obtained from the superimposition process) are also listed. These numbers, along with the rms, are given to indicate the degree of the observed mapping between the reference conformation and each compound.

As can be seen in Table 4, most of the active compounds were able to adopt the reference conformation at relatively little expense of energy. The global minima conformations of compounds **2**, **44**, **53** and **54** were found to be very close to the reference conformation and these compounds are all active. All the other active compounds were found to be able to adopt the reference conformation with an energy increase of less than 3 kcal/mol. However, for the inactive compounds **42**, **52**, **57** and **62**, the energy of the reference conformation was more than 20 kcal above that of their most stable conformation. Compounds **42** and **57** have similar conformational profiles and cannot achieve the required conformation at all, because of the strong nonbonded repulsion interaction between the bulky *N*-methyl group and the phenyl ring.

It has been previously observed from SAR studies performed by Walpole et al. that compounds where the atom next to the phenyl ring in the A region is sp^2 hybridized are either weakly active (compound **53**) or inactive (compounds **52** and **62**). An sp^3 carbon atom was found to be the optimal bridging group between the A and B regions. However, it is interesting to find that compound **53**, where the bridging group is *cis*-vinylene, retains some activity ($EC_{50} = 17.90$ μM), while compounds **52** and **62**, where the bridging group is *trans*-vinylene, are inactive ($EC_{50} > 100$ μM). The results from our molecular modeling study are consistent with these observations, as shown in Table 4. The global minimum conformation of compound **53** is very close to the required conformation for activity. The phenyl and thiourea planes are not coplanar and are oriented as needed by the required conformation. Indeed, the steric hindrance that originates from the cis configuration prevents the coplanar configuration from being stable. On the other hand, the conjugation energy in compounds **52** and **62** prevents them from reaching the required conformation. These observations lend support to the assumption that the coplanarity between the 3-methoxy-4-hydroxy phenyl ring and the thiourea plane is detrimental to agonist activity. Further support comes from the observation that compounds where the 3-methoxy-4-hydroxy phenyl ring is constrained with respect to the thiourea plane are either inactive (compound **122**) or only partial agonists (compounds **124** and **125**). In-

TABLE 3
STRUCTURES OF CAPSAICIN ANALOGUES STUDIED BY MOLECULAR MODELING

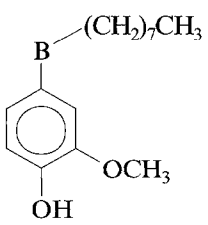
General structure	Compound no.	B	Ca^{2+} influx EC_{50} (μM)
	2	CH_2NHCO	0.55
	41	$(CH_2)_2NHCO$	18.30
	42	$CH_2N(CH_3)CO$	>100
	44	CH_2NHCS	0.28
	47	$(CH_2)_2CONH$	2.32
	52	CH=CHCONH (*E*)	>100
	53	CH=CHCONH (*Z*)	17.90
	54	$CH_2NHCONH$	0.36
	55	$CH_2NHCSNH$	0.06
	57	$CH_2N(CH_3)CSNH$	>100
	58	$CH_2NHCSN(CH_3)$	0.53
	62	CH=CHCO (*E*)	>100
	63	CH_2CH_2CO	2.13

TABLE 4
CONFORMATIONAL ENERGY FOR EACH COMPOUND EXAMINED

Compound	Conformational energy (kcal/mol)	No. of H-bond acceptors	No. of H-bond donors	Rms (Å)[a]	EC_{50} (μM)	MULTICASE[b] results (μM)
2	0	1	1	0.187	0.55	3.37
41	2.59	1	1	0.259	18.30	3.42
42	>20	1	1	0.158	>100	3.37
44	0	1	1	0.241	0.28	3.96
47	1.78	1	1	0.241	2.32	3.42
52	>20	1	1	0.165	>100	>100
53	0	1	0	0.644	17.90	>100
54	0	1	2	0.006	0.36	3.38
55	0	1	2	[c]	0.06	0.20
57	>20	1	1	0.073	>100	3.53
58	2.98	1	1	0.006	0.53	3.53
62	>20	1	0	0.180	>100	>100
63	0.91	1	0	0.540	2.13	3.96

The conformational energy is defined as the energy difference between the global minimum and the reference conformation.
[a] Rms obtained in the superimposition process.
[b] Values were obtained by applying the QSAR model to calculate the activity of the 13 compounds.
[c] The reference conformation.

deed, in compound **122**, containing a five-membered ring, coplanarity between the A and B regions was found, whereas in compound **124**, comprising a six-membered ring, a slight puckering between the thiourea and phenyl ring (an angle of about 25°) was observed. These findings have been confirmed from NMR spectroscopy and X-ray crystallography [26].

It is noteworthy that the replacement of a methoxy substituent by a hydroxyl group at the α position for compound **122** (EC_{50} > 100 μM) to give compound **111** (EC_{50} = 0.5 μM) and for compound **124** (partial agonist) to give compound **110** (EC_{50} = 0.3 μM), markedly increased the agonist activity. As a matter of fact, MULTICASE analysis, as mentioned above, has identified α,α-catechol attached to a thiourea moiety through a fused ring as an activating biophore (biophore III). Several possible reasons could account for this observation: (i) the receptor cavity for the hydrophilic binding site is rather constrictive, so that it may be able to accommodate the α,α-catechol analogue, but not the bulkier α,α-methoxy hydroxy phenyl analogue, which is crucial for further binding (narrow slot concept); (ii) the H-bond involving the methoxy group in the constrained configuration does not coincide with the unique directional vector of the hydrogen bond required for the interaction with the receptor; (iii) the extra H-bond donor functionality and the multiple directional vectors assumed by a freely rotating hydroxyl group may provide a favorable spatial disposition in such a way that good binding affinity can still be reached. From the above assumptions, (i) allows an explanation of the proposed pharmacophore in which a non-coplanar orientation between the 3-methoxy-4-hydroxy phenyl ring and the thiourea moiety is formed and by doing so, a smaller width of the hydrophilic moiety is attainable that can fit into the receptor slot.

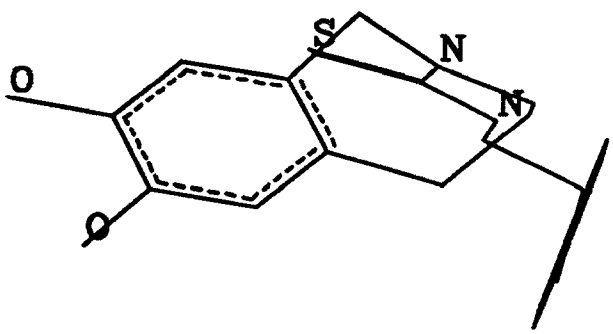


Fig. 10. The low-energy conformation of compound **112**.

On the other hand, compound **112**, where the size of the fused ring is seven, was found to be devoid of agonist activity, in spite of the presence of biophore III. It was, however, a moderately potent antagonist. Molecular modeling and experimental data for this compound have shown that the planes of the catechol ring and the thiourea were found to be approximately perpendicular in its low-energy conformations (Fig. 10). The perpendicular orientation between the A and B regions was also found in the low-energy conformations of inactive compound **57**. Based upon the observation presented above, it is likely that compound **57** or the catechol analogue of compound **57** are potentially potent antagonists. Further experimental data are required to confirm this hypothesis.

Conclusions

We have shown that MULTICASE is able to identify structural features and physicochemical properties in the compounds of a data set of 123 capsaicin analogues that account for the Ca^{2+} influx agonist activity. Both lipophilicity and structural features were shown to affect the agonist activity. The analysis clearly isolated the most im-

portant structural feature, a 3-methoxy-4-hydroxy benzyl group in the A region, as biophore I. Several associated modulators related to the B and C regions were also identified.

A three-dimensional model of the pharmacophore for capsaicin–receptor interactions has been proposed. The energy required for a capsaicin analogue to mimic the assumed pharmacophore was evaluated and the result was found to be consistent with the experimental data.

MULTICASE, combined with the pharmacophore model, can be utilized for the rational development and discovery of novel agonists.

References

1 Virus, R.M. and Gebhart, G.F., Life Sci., 25 (1979) 1273.
2 Rozin, P., In Green, B.G., Mason, J.R. and Kare, M.R. (Eds.) Irritation, Vol. 2, Marcel Dekker, New York, NY, 1990, pp. 231–269.
3 Butterworth, J.F.I. and Strichartz, G.R., Anesthesiology, 72 (1990) 711.
4 Walpole, C.S.J., Wriggleworth, R., Bevan, S., Campbell, E.A., Dray, A., James, I.F., Masdin, K.J., Perkins, M.N. and Winter, J., J. Med. Chem., 36 (1993) 2362.
5 Szolcsanyi, J., In Green, B.G., Mason, J.R. and Kare, M.R. (Eds.) Irritation, Vol. 2, Marcel Dekker, New York, NY, 1990, pp. 141–168.
6 Yeh, J.-L., Lo, Y.-C., Wang, Y. and Chen, I.-J., Brain Res. Bull., 30 (1993) 641.
7 Hayes, A.G., Oxford, A., Reynolds, M., Shingler, A.H., Skingle, M., Smith, C. and Tyers, M.B., Life Sci., 34 (1984) 1241.
8 Cholewinski, A., Burgess, G.M. and Bevan, S., Neuroscience, 55 (1993) 1015.
9 Bevan, S. and Szolcsanyi, J., Trends Pharmacol. Sci., 11 (1990) 330.
10 Wood, J.N., Winter, J., James, I.F., Rang, H.P., Yeats, J. and Bevan, S., J. Neurosci., 8 (1988) 3208.
11 Fitzgerald, M., Pain, 15 (1983) 109.
12 James, I.F., Walpole, C.S.J., Hixon, J., Wood, J.N. and Wriggleworth, R., Mol. Pharmacol., 33 (1988) 643.
13 Szallasi, A. and Blumberg, P.M., Adv. Pharmacol., 24 (1991) 123.
14 Brand, L.M., Skare, K.L., Loomans, M.E., Reller, H.H., Schwen, R.J., Lade, D.A., Bohne, R.L., Maddin, C.S., Moorehead, D.P., Fanelli, R., Chiabrando, C., Castelli, M.G. and Tai, H.H., Agents Actions, 31 (1990) 229.
15 Janusz, J.M., Buckwalter, B.L., Young, P.A., Lahann, T.R., Farmer, R.W., Kasting, G.B., Loomans, M.E., Kerckaert, G.A., Maddin, C.S., Berman, E.F., Bohne, R.L., Cupps, T.L. and Milstein, J.R., J. Med. Chem., 36 (1993) 2595.
16 Szallasi, A. and Blumberg, P.M., Neuroscience, 30 (1989) 515.
17 Bevan, S., Hothi, S., Hughes, G., James, I.F., Rang, H.P., Shah, K., Walpole, C.S.J. and Yeats, J.C., Br. J. Pharmacol., 107 (1992) 544.
18 Szallasi, A., Goso, C., Blumberg, P.M. and Manzini, S., J. Pharmacol. Exp. Ther., 267 (1993) 728.
19 Szallasi, A. and Blumberg, P.M., Brain Res., 524 (1990) 106.
20 Winter, J., Dray, A., Wood, J.N., Yeats, J.C. and Bevan, S., Brain Res., 520 (1990) 131.
21 Wehmeyer, K.R., Kasting, G.B., Powell, J.H., Kuhlenbeck, D.L., Underwood, R.A. and Bowman, L.A., J. Pharm. Biomed. Anal., 8 (1990) 177.
22 Szallasi, A., Gen. Pharmacol., 25 (1994) 223.
23 Szallasi, A., Szolcsanyi, J., Szallasi, Z. and Blumberg, P.M., Naunyn-Schmiedeberg's Arch. Pharmacol., 344 (1991) 551.
24 Szallasi, A. and Blumberg, P.M., Life Sci., 47 (1990) 1399.
25 Santicioli, P., Bianco, E.D., Figini, M., Bevan, S. and Maggi, C.A., Br. J. Pharmacol., 110 (1993) 609.
26 Walpole, C.S.J., Bevan, S., Bovermann, G., Boelsterli, J.J., Breckenridge, R., Davies, J.W., Hughes, G.A., James, I., Oberer, L., James, I.F., Winter, J. and Wriggleworth, R., J. Med. Chem., 37 (1994) 1942.
27 Szolcsanyi, J. and Jancso-Gabor, A., Arzneim.-Forsch. Drug Res., 25 (1975) 1877.
28 Szolcsanyi, J. and Jancso-Gabor, A., Arzneim.-Forsch. Drug Res., 26 (1976) 33.
29 Walpole, C.S.J., Wriggleworth, R., Bevan, S., Campbell, E.A., Dray, A., James, I.F., Masdin, K.J., Perkins, M.N. and Winter, J., J. Med. Chem., 36 (1993) 2373.
30 Walpole, C.S.J., Wriggleworth, R., Bevan, S., Campbell, E.A., Dray, A., James, I.F., Masdin, K.J., Perkins, M.N. and Winter, J., J. Med. Chem., 36 (1993) 2381.
31 Walpole, C.S.J. and Wriggleworth, R., In Wood, J.N. (Ed.) Capsaicin in the Study of Pain, Harcourt Brace, San Diego, CA, 1993, pp. 63–81.
32 Klopman, G., J. Am. Chem. Soc., 106 (1984) 7315.
33 Klopman, G., Quant. Struct.–Act. Relatsh., 11 (1992) 176.
34 Klopman, G., Li, J.-Y., Wang, S. and Dimayuga, M., J. Chem. Inf. Comput. Sci., 34 (1994) 752.
35 Klopman, G., Wang, S. and Balthasar, D.M., J. Comput. Chem., 32 (1992) 474.
36 Van Catledge, F.A., J. Org. Chem., 45 (1980) 4801.
37 Insight II, Version 2.2.0, Biosym Technologies Inc., San Diego, CA, 1992.
38 Dauber-Osguthorpe, P., Roberts, V.A., Osguthorpe, D.J., Wolff, J., Genest, M. and Hagler, A.T., Protein Struct. Funct. Genet., 4 (1988) 31.
39 Strichartz, G.R., Sanchez, V., Arthur, G.R., Chafetz, R. and Martin, D., Anesth. Analg., 71 (1990) 158.
40 Bradley, D.J. and Richards, C.D., Br. J. Pharmacol., 81 (1984) 161.
41 Hansch, C., Bjorkroth, J.P. and Leo, A., J. Pharm. Sci., 76 (1987) 663.
42 Szallasi, A. and Blumberg, P.M., Life Sci., 48 (1991) 1863.