

Chemical function-based pharmacophore generation of selective κ -opioid receptor agonists by catalyst and phase

Jing Zhang · Guixia Liu · Yun Tang

Received: 14 July 2008 / Accepted: 8 October 2008 / Published online: 11 February 2009
© Springer-Verlag 2008

Abstract Two chemical function-based pharmacophore models of selective κ -opioid receptor agonists were generated by using two different programs: Catalyst/HypoGen and Phase. The best output hypothesis (Hypo1) of HypoGen consisted of five features: one hydrogen-bond acceptor (HA), three hydrophobic points (HY), and one positive ionizable function (PI). The highest scoring model (Hypo2) produced by Phase comprised four features: one acceptor (A), one positive ionizable function (P), and two aromatic ring features (R). These two models (Hypo1 and Hypo2) were then validated by test set prediction and enrichment factors. They were shown to be able to identify highly potent κ -agonists within a certain range, and satisfactory enrichments were achieved. The features of these two pharmacophore models were similar and consistent with experiment data. The models produced here were also generally in accord with other reported models. Therefore, our pharmacophore models were considered as valuable tools for 3D virtual screening, and could be useful for designing novel κ -agonists.

Keywords Catalyst/HypoGen · κ -opioid receptor agonists · Pharmacophore models · Phase · QSAR

Introduction

G-Protein coupled receptors (GPCRs) represent an important type of target for drug discovery in the post-genomic era. GPCRs constitute a superfamily of membrane receptors of utmost importance in pharmaceutical research [1]. This superfamily is characterized by a common structural motif of seven transmembrane-spanning (7TM) helices connected by intracellular (IL) and extracellular loops (EL) [2, 3]. Opioid receptors belong to the rhodopsin subclass within this superfamily [4]. It is now widely accepted that there are at least three opioid receptor subtypes, μ , κ and δ [5].

Opioids are widely used in the treatment of moderate-to-severe pain. However, the clinical usefulness of opioids such as morphine, which exert their analgesic effect through agonism of the μ -opioid receptor, is limited by significant side effects such as physical dependence, respiratory depression, constipation, and addiction liabilities. Increasing evidence has accumulated during the past decade to support the hypothesis that a selective κ -opioid agonist would be a powerful analgesic agent without the clinically limiting side effects of selective μ -opioid analgesic drugs [6–8]. The most important selective κ -agonists developed so far are arylacetamide derivatives such as U-50488 [9] and ICI 199441 [10] (see Fig. 1), which are centrally acting κ -agonists.

However, these centrally acting κ -agonists produce their own set of central nervous system (CNS) side effects such as dysphoria and diuresis [6]. This drawback has provided impetus for the discovery of peripherally acting κ -agonists. Much recent work in this area has been carried out by the Adolor Corporation (<http://www.adolor.com/>). Based on ICI 199441, Adolor have developed a number of κ -agonist analogues in order to avoid the side effects associated with CNS penetration [11–17].

J. Zhang · G. Liu (✉) · Y. Tang
Laboratory of Molecular Modeling and Design,
School of Pharmacy,
East China University of Science and Technology,
Box 268, 130 Meilong Road,
Shanghai 200237, China
e-mail: gxliu@ecust.edu.cn

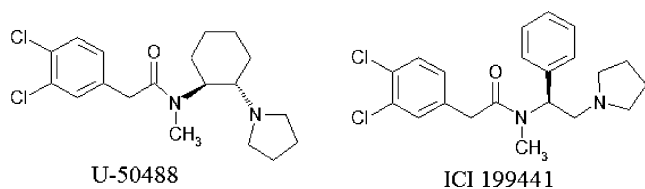


Fig. 1 Structures of κ -agonists U-50488 and ICI 199441

Discovering 3D pharmacophores that can explain the activity of a series of ligands is one of the most significant contributions of computational chemistry to drug discovery. In the present study, pharmacophore models for the κ -agonists mentioned above [11–17] were generated independently using two commonly used programs: Catalyst/HypoGen [18, 19] and Phase [20, 21]. These models were expected to provide a rational hypothetical picture of the primary chemical features responsible for activity, and thus to supply a useful knowledge for developing new active candidates targeting the κ -opioid receptor. Here, we report a comparison of the performance of these two programs.

Materials and methods

Inhibitory activity data (K_i) spanning over 5 orders of magnitude (from 0.043 to 2300 nM) for a set of 100 ICI 199441 analogs were collected from the literature (see Table 1) [11–17]. All activity data were measured using the same method by the same research group. Catalyst 4.10/HypoGen and Phase 2.0 were used to generate pharmacophore models from these compounds.

Catalyst

Biological activity data and training set selection

The compounds were divided into two sets: a training set and a test set. The selection of a suitable training set is critical for the quality of automatically generated pharmacophore models. To ensure the statistical relevance of the calculated models, the training set should contain a set of diverse compounds together with their activity values. These should originate from comparable binding assays and be spread equally over at least 4–5 orders of magnitude. Each selected compound should add some new information to the model while avoiding redundancy and bias, both in terms of structural features and activity range. The most active compounds should be included because they could provide critical information on pharmacophore requirements. On the basis of the above criteria, 25 compounds were selected as the training set, including ICI 199441 (compound 91) [10].

To validate our pharmacophore, the other 75 compounds were used as the test set. For the purpose of estimation (prediction), all compounds were classified by their activity as highly active ($K_i < 10$ nM, +++), moderately active (10 nM $\leq K_i \leq 100$ nM, ++), or inactive ($K_i > 100$ nM, +).

All 2D chemical structures were produced with the ISIS/Draw, version 2.5 drawing program, and the conformational analysis for each molecule was implemented using the Poling algorithm and CHARMM force field parameters within the Catalyst software package. A maximum number of 250 conformations for each compound were selected using “best conformer generation” option with a constraint of 20 kcal/mol energy thresholds above the minimum conformer searched to ensure an exhaustive characterization of conformational space. All other parameters were kept at their default settings.

Generation of pharmacophore model

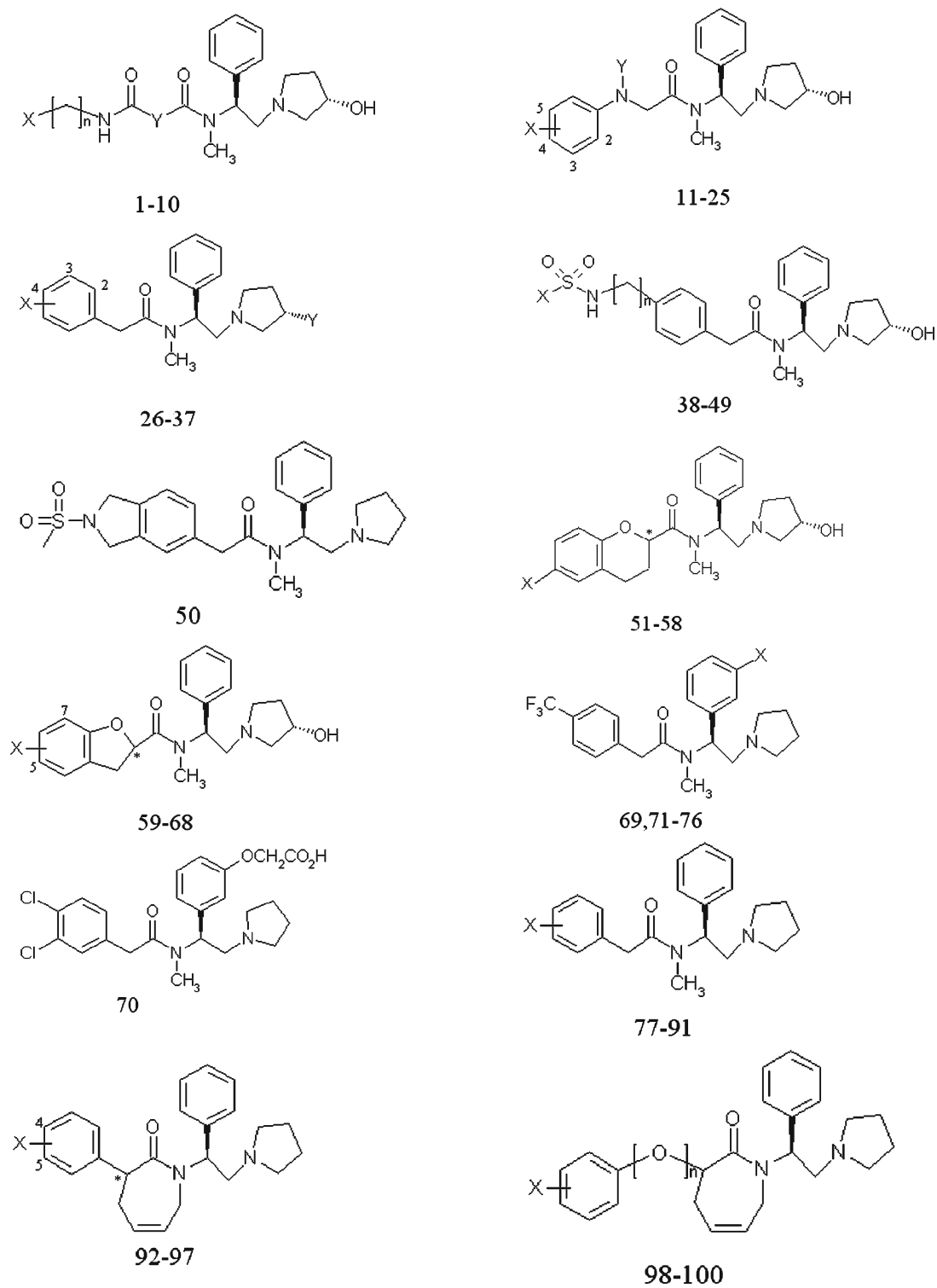
The aim of the HypoGen module in Catalyst is to find hypotheses that are common among the active compounds of the training set. Pharmacophores that best correlate the three-dimensional arrangement of features in a given set of training compounds with the corresponding pharmacological activities (IC_{50} or K_i) are constructed and ranked. An initial analysis revealed that hydrogen bond acceptor (HA), hydrophobic group (H), and positive ionizable (PI) features could effectively map all critical chemical/structural features of all the training set molecules. These features were therefore selected to form the essential information in this hypotheses generation process.

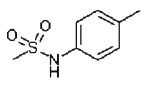
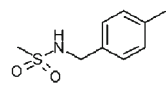




Hypothesis generation in Catalyst has three consecutive steps: constructive phase, subtractive phase, and optimization phase. In the constructive phase, Catalyst identifies active compounds. It then identifies inactive compounds in the subtractive phase. Finally, in the optimization phase, Catalyst attempts 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 generation process stops when optimization no longer improves the score.





Assessment of the quality of pharmacophore hypotheses

After the generation of pharmacophore hypotheses, Catalyst provides two approaches, cost analysis and cross-validation test, to help assess the quality of the hypotheses.

Cost analysis In addition to generating hypotheses, Catalyst also performs two important theoretical cost calculations (represented in bit units) that determine the success of any pharmacophore hypothesis. The first is the cost of an

Table 1 Structures and binding affinities of compounds in the training set and test set

Compound	X	Y	n	*	K_i (nM)	pK_i
1*#	Ph	CH ₂	0		0.27	9.569
2#	Ph	CH ₂	1		15	7.824
3#		CH ₂	0		29	7.538
4#		CH ₂	0		96	7.018
5	Ph	CH ₂ CH ₂	0		6.0	8.222
6	Ph	NHCH ₂	0		6.5	8.187
7#		CH ₂	0		0.53	9.276
8*#		CH ₂ CH ₂	0		23	7.638
9*#		CH ₂	0		1.3	8.886
10*!		CH ₂ CH ₂	0		280	6.553
11*!	H	H			0.17	9.770
12	4-CF ₃	H			0.39	9.409
13!	3,4-Cl	H			0.11	9.959
14#	4-NO ₂	H			0.21	9.678
15	4-CN	H			0.66	9.180
16	3-CN	H			0.50	9.301
17#	2-CN	H			0.39	9.409
18	H	CH ₃			0.32	9.495
19*#	H	CH ₃ CO			3.8	8.420
20	4-CN	CH ₃			1	9.000
21	4-CH ₃ SO ₂ NHCH ₂	H			16	7.796
22#	3-CH ₃ SO ₂ NHCH ₂	H			3.9	8.409
23*!	2-CH ₃ SO ₂ NHCH ₂	H			0.78	9.108
24	4-CH ₃ SO ₂ NH	H			5.7	8.244
25	4-CH ₃ CH ₂ CH ₂ SO ₂ NH	H			1.5	8.824

26#	H	H		5.8	8.237
27	2-CH ₂ NH ₂	H		10	8.000
28	3-CH ₂ NH ₂	H		8.7	8.060
29*#	4-CH ₂ NH ₂	H		57	7.244
30!	2-CH ₂ NHCOCH ₃	H		1.9	8.721
31	3-CH ₂ NHCOCH ₃	H		23	7.638
32#	4-CH ₂ NHCOCH ₃	H		47	7.328
33#	4--CH ₂ NHCONHCH ₃	H		6.1	8.215
34	2-CH ₂ NHSO ₂ CH ₃	H		12	7.921
35#	3-CH ₂ NHSO ₂ CH ₃	H		10	8.000
36	4-CH ₂ NHSO ₂ CH ₃	H		2.9	8.538
37#!	4-CH ₂ NHSO ₂ CH ₃	OH		0.6	9.222
38	CH ₂ CH ₃		1	1.3	8.886
39*	CH(CH ₃) ₂		1	2.7	8.569
40	(CH ₂) ₂ CH ₃		1	1.3	8.886
41#!	(CH ₂) ₃ CH ₃		1	0.8	9.097
42	Ph		1	5.9	8.229
43#			1	5.9	8.229
44#			1	1.7	8.770
45			1	1.3	8.886
46			1	2.8	8.553
47*#!	CH ₃		0	0.4	9.398
48#	CH ₃		2	2.8	8.553
49	CH ₃		3	1.7	8.770
50#!				1.9	8.721
51	H		R	1.6	8.796
52#	H		S	30.5	7.516
53#!	NHSO ₂ (CH ₂) ₂ CH ₃		R	1.6	8.796
54	NHSO ₂ (CH ₂) ₂ CH ₃		S	14.2	7.848

55#	CH ₂ NHSO ₂ CH ₃	R	4.5	8.347
56	CH ₂ NHSO ₂ CH ₃	S	99	7.004
57*#	CH ₂ NHCOOCH ₃	R	2.6	8.585
58*#!	CH ₂ NHCOOCH ₃	S	130	6.886
59*!	H	R	0.8	9.097
60#	H	S	2.4	8.620
61*	5-NHCOOCH ₃	R	14.7	7.833
62	5-NHCOOCH ₃	S	5.8	8.237
63#	5-NHSO ₂ CH ₃	R	1.3	8.886
64#	5-NHSO ₂ CH ₃	S	1.2	8.921
65	7- NHSO ₂ CH ₃	R	3.9	8.409
66	7- NHSO ₂ CH ₃	S	1.3	8.886
67#!	5- CH ₂ NHSO ₂ (CH ₂) ₂ CH ₃	R	3.6	8.444
68!	5- CH ₂ NHSO ₂ (CH ₂) ₂ CH ₃	S	3.3	8.481
69#!	H		0.058	10.237
70*			2.69	8.570
71#	NO ₂		0.63	9.201
72!	NH ₂		0.31	9.509
73#	NHSO ₂ CH ₃		0.36	9.444
74*#	NHSO ₂ NH ₂		0.48	9.319
75	NHCO(CH ₂) ₂ CO ₂ H		9.5	8.022
76#!	N(CH ₂ CO ₂ H) ₂		2300	5.638
77#	2-NO ₂		0.41	9.387
78#	3-NO ₂		0.065	10.187
79	4-NO ₂		1.1	8.959
80	2-NO ₂ , 3,4-Cl		0.091	10.041
81	2-NH ₂ , 3,4-Cl		0.086	10.066
82	2-NH ₂		0.58	9.237
83#	3-NH ₂		0.93	9.032

84	4-NH ₂		0.93	9.032
85	2-N(SO ₂ CH ₃) ₂		8.6	8.066
86*#!	2-N(SO ₂ CH ₃) ₂ , 3,4-Cl		0.096	10.018
87*#	3-N(SO ₂ CH ₃) ₂		6.0	8.222
88#	4-N(SO ₂ CH ₃) ₂		25.0	7.602
89*	2-CF ₃		0.13	9.886
90	3-CF ₃		0.064	10.194
91*#!	4,5-Cl		0.043	10.398
92#	4,5-Cl	R	0.34	9.469
93#	4,5-Cl	S	17	7.770
94	4-Cl	R	2.0	8.699
95	4-Cl	S	35	7.456
96*	4-OCH ₃	R	22	7.658
97*!	4-OCH ₃	S	800	6.097
98*	4-CF ₃	0	0.72	9.143
99*#!	5-SO ₂ NC ₄ H ₈	0	1900	5.721
100*	4,5-Cl	1	2.6	8.585

* Training set molecules that were applied to HypoGen pharmacophore generation

! Compounds that were applied to Phase pharmacophore generation

Molecules that were included in the construction of 3D-QSAR models in Phase program

ideal hypothesis (fixed cost), which represents the simplest model that fits all data perfectly. The second is the cost of the null hypothesis (null cost), which represents the highest cost of a pharmacophore without features and whose estimated activity is the average of the activity data of the training set molecules. These calculations represent the upper and lower bounds for the hypotheses generated. The greater the difference between these two cost values, and the closer the total cost of the generated hypothesis is to the fixed cost, the more statistically significant the hypothesis is thought to be. According to randomized studies, a cost difference of 40–60 between the total cost and the null cost indicates a 75–90% chance of representing a true correlation in the data.

Cross-validation test To further assess the statistical significance of the pharmacophore hypotheses, a validation procedure based on Fischer's randomization test was applied [22]. The activity values of the training set

molecules are scrambled randomly using the CatScramble technique, available in the Catalyst/HypoGen module, and new spreadsheets are created. The number of spreadsheets generated depends on the level of statistical significance one wants to achieve, e.g., 19, 49, or 99 random spreadsheets have to be generated if you want to achieve confidence levels of 95%, 98%, or 99%, respectively [23–26]. In our validation test, we selected the 95% confidence level, and 19 spreadsheets were created by the CatScramble command.

Validation of pharmacophores

Activity prediction A performance prediction of the activity of new compounds is vital for a generated hypothesis. Therefore, a set of 75 ICI 199441 analogs, which were not included in the training set, were taken as a test set to be predicted by the hypothesis. These molecules cover a wide range of activities from 0.058 to 2,300 nM.

Enrichment factor In lead-discovery studies, the pharmacophore model should identify active leads targeting the κ -opioid receptor in the database screening. An enrichment factor could indicate the ability of a program to identify active compounds.

The best hypothesis was used as a query in the screening of 3D conformational molecular structure databases. A refined SPECS database (<http://www.specs.net>) of 43,423 molecules, seeded with 25 known agonists was taken for the database search to validate whether the pharmacophore model could identify active compounds or not. The original SPECS database, totally 196,665 molecules, was refined with some basic criteria: molecular weight from 250 to 550; ClogP from 0 to 5; rotatable bonds from 3 to 8; hydrogen donors from 1 to 5; hydrogen acceptors from 3 to 7 and LogSw from -5 to 0.

A database search in Catalyst involves two algorithms. The Fast Flexible Search Database/Spreadsheets command computes already existing conformers of the database, and the Best Flexible Search Databases/Spreadsheets is able to change the conformation of a molecule during computation. This spiked database (containing 43,448 molecules that comprised 43,423 molecules from SPECS database and 25 known agonists) was screened with the pharmacophore model using the “Fast Flexible Search Database/Spreadsheets” option, and a maximum of 100 conformers per compound were generated. For both procedures, only those structures that map all features of the pharmacophore template are retrieved. The enrichment factor (E) [19] was calculated using Eq. 1

$$E = \text{Ha/Ht} \div \text{A/D} \quad (1)$$

where Ht = the number of hits retrieved, Ha = the number of active molecules in the hit list, A = the number of active molecules present in the database, and D = the total number of molecules in the database. This method not only validates our pharmacophore models, but could also assess how good the pharmacophore is in selective retrieval of known agonists.

Phase

Phase [20, 21]—a highly flexible system for pharmacophore perception, structure alignment, activity prediction, and 3D database searching—is a more recently developed pharmacophore modeling package.

Active and inactive sets selection

All 2D chemical structures were again generated with ISIS/Draw. The Maestro graphical interface was used to build 3D models of the 100 ICI 199441 analogs. Phase incorporates a structure-cleaning step utilizing LigPrep [20], which attaches

hydrogens, converts 2D structures to 3D, generates stereoisomers, and optionally neutralizes charged structures or determines the most probable ionization state.

A maximum of 500 conformations were generated for each molecule using MacroModel torsional sampling with OPLS_2005 post-processing [20]. Each minimized conformer was filtered through a relative energy window of 10 kcal/mol and a redundancy check of 1 Å in the heavy atom positions.

The set of active compounds should contain as much structural diversity as possible, so that the resulting pharmacophore models are applicable across different chemical families. Because the sketches of the most active molecules in this study were very similar, the DiverseSolutions module of SYBYL7.0 [27] was used to ensure their diversity. DiverseSolutions is a package of programs designed to address a wide variety of tasks associated with the concept of chemical diversity. All molecules were assigned to different cells, which represent different chemical spaces, and the most active molecule of each cell would be selected as the active set.

Inactive compounds can be used to eliminate hypotheses that do not provide a good explanation of activity on the basis of the pharmacophores alone, while only the active set is used for developing common pharmacophore hypotheses. The inactive set may be used subsequently to assign adjusted scores that reflect the degree to which the models distinguish active from inactive compounds. This is particularly useful if everything in the active set is built on a common scaffold, which can give rise to a number of spurious pharmacophore models that have nothing to do with ligand binding. In this study, $\text{p}K_i$ values were used as the criteria to select the inactive set.

Generation of pharmacophore model

Phase provides six built-in types of pharmacophore features: hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobic group (H), negatively charged group (N), positively charged group (P), and aromatic ring (R). The default pharmacophore feature definitions were used in site generation.

Common pharmacophores are identified using a tree-based partitioning technique, and they were generated by a systematic variation of the number of sites (n_{sites}) and the number of matching active compounds (n_{act}). With $n_{\text{act}} = n_{\text{act_tot}}$ initially ($n_{\text{act_tot}}$ is the total number of active compounds), n_{sites} was varied from seven to three until at least one hypothesis was found and scored successfully. If this failed, then n_{act} was decreased by one and the n_{sites} cycle repeated.

Scoring with respect to active compounds was conducted using default parameters for site, vector, and volume terms. Hypotheses that emerged from this process were subsequently scored with five inactive compounds, using a

weight of 1.0. The scoring procedure provides a ranking of different hypotheses.

Validation of pharmacophores

Activity prediction All hypotheses produced in the last step were then used to build 3D QSAR (quantitative structure-activity relationship) models. The activity data of the whole set was evaluated by generated QSAR models, to assess the quality of the pharmacophore hypotheses. In these 3D QSAR models, chemical features of ligand structures are mapped to a cubic 3D grid. The accuracy of the models increases with increasing number of PLS (partial least squares) factors until over-fitting starts to occur.

Phase offers two choices for the structural components that form the basis of the QSAR model. One is atom-based, in which all atoms are taken into account, and the other is pharmacophore-based, in which merely the pharmacophore sites that can be matched to the hypothesis are considered. In this study, atom-based QSAR models with 1–3 PLS factors were employed.

Enrichment factor These pharmacophore hypotheses were further validated by the database searching approach in Phase, to retrieve known active compounds from the database mentioned above. Each molecule was represented by a maximum of 100 conformations, and the tolerance on matching the pharmacophore was ± 1 Å, applied to each of the six inter-feature distances. The search process is normally performed in two steps: finding and fetching. In the finding step, the database is searched for geometric arrangements of pharmacophore sites that match the site types and intersite distances of the chosen hypothesis. In the fetching step, the match file is used as a lookup table to rapidly retrieve the relevant conformers from the database and align them to the hypothesis. The finding step is the center of the search, and is the most time-consuming part of the process. The enrichment factor was calculated using the method mentioned above.

Comparison of the two methods

Several programs, such as Catalyst, Phase, DISCO [28], GASP [29], and Galahad [30], have been developed for the automatic identification of pharmacophore models. The main differences between the first two of these programs lie in the algorithms used for alignment and in the way in which conformational flexibility is handled.

Algorithms

The HypoGen algorithm tries to find hypotheses that are common among the active compounds of the training set but

do not reflect inactive compounds. It observes the principle of Ockham's razor [31], "plurality should not be posited without necessity", thus constructing a model that best correlates with measured activities and consists of as few features as possible.

In Phase, common pharmacophores are identified using a tree-based partitioning technique that groups together similar pharmacophores according to their intersite distances, i.e., the distances between pairs of sites in the pharmacophore.

Conformer generation

It should be noted that the number of conformations generated by Phase is larger than that generated by Catalyst, which has an upper limit of 255; Phase can generate at most 1,000 conformers.

In Phase, we can generate conformers for each ligand by one of two methods: a ligand torsion search, which involves systematic sampling around rotatable bonds, or a mixed Monte-Carlo multiple minimization/low-mode search (MCM/MLMOD). Both methods can be followed by MacroModel minimization and filtering. In Catalyst, conformers are generated using the Poling algorithm, which penalizes any newly generated conformer if it is close to an already formed conformer in the set.

Furthermore, Phase provides a higher degree of flexibility and feedback than Catalyst, emphasizing the user as an integral part of the pharmacophore development process. The ultimate goal of Phase is to suggest a set of plausible models, while Catalyst provides only the top 10 hypotheses. In assessing the quality of generated pharmacophore models, Catalyst provides more approaches than Phase. In addition to approaches to activity prediction and enrichment factor that are similar to those of Phase, Catalyst also has its own specific methods, namely, cost analysis and cross-validation test.

Results

Catalyst

Pharmacophore generation

A set of ten pharmacophore hypotheses were generated using the training set listed in Table 1. The results of the hypotheses, which include different cost values calculated during hypotheses generation along with root mean squares (RMS), correlation (r), and pharmacophore features, were listed in Table 2.

For good hypotheses, the value of the total cost of the hypothesis is expected to be close to the fixed cost values. In this study, the total cost of the best hypothesis was 119.581 and the fixed cost of the run was 100.482. The cost of the null hypothesis for all ten hypotheses was 164.540.

The difference between the null cost and the total cost was 44.96 bits, which is within the 40–60 bit range, thus indicating that the top-ranked hypothesis, Hypo1, had a 75–90% probability of correlating the data.

Hypo1 had the best values in terms of total cost (119.581) and error cost (99.979), and had the lowest RMS deviations (1.127) and highest correlation coefficient (0.900). Therefore, Hypo1 was selected as the best pharmacophore. This model (Fig. 2) consisted of a spatial arrangement of five chemical features: one hydrogen-bond acceptor (HA), three hydrophobic points (HY), and one positive ionizable function (PI). Table 3 lists the actual and estimated K_i values of training set compounds calculated on the basis of Hypo1.

For the 25 molecules in the training set, all the active compounds were predicted as active (+++), two moderately active compounds were predicted as inactive (+), and one inactive compound was predicted as moderately active (++) . The difference between the actual and the estimated activity observed for the three compounds was only about 1 order of magnitude, which might be an artifact of the program, which uses a different number of degrees of freedom for these compounds to mismatch the algorithm.

The error factor (also listed in Table 3) shows that 24 out of the 25 molecules in the training set have errors less than 10, which means that the activity prediction of these compounds falls between 10-fold greater and 1/10 of the actual activity, while the remaining compound has an error of not higher than 14. The hypothesis can discriminate closely between stereoisomers (compounds 57, 58 and 96, 97).

Table 2 Results of top ten pharmacophore hypotheses generated using training set molecules^a by means of Catalyst/HypoGen

Hypothesis	Total cost	Error cost	RMS ^b	Correlation	Features ^c
1	119.581	99.979	1.127	0.900	HA,HY,HY,HY,PI
2	120.431	102.774	1.223	0.878	HA,HY,HY,PI,RA
3	129.518	108.112	1.386	0.849	HA,HY,HY,PI
4	131.443	111.582	1.483	0.818	HA,HY,HY,PI
5	135.406	118.971	1.671	0.753	HA,HY,HY,PI,RA
6	136.092	119.670	1.687	0.747	HA,HY,HY,PI,RA
7	137.598	119.699	1.688	0.750	HA,HY,PI,RA
8	137.603	121.041	1.719	0.736	HA,HY,HY,PI,RA
9	137.834	119.819	1.691	0.749	HA,HY,PI,RA
10	138.177	118.074	1.649	0.770	HA,HY,PI,RA

^a Null cost=164.540; Fixed cost = 100.482; Configuration cost =15.267. All cost units are in bits. Configuration cost: a fixed cost which depends on the complexity of the hypothesis space being optimized

^b RMS (root mean square), the deviation of the log (estimated activities) from the log (measured activities) normalized by the log (uncertainties)

^c HA Hydrogen-bond acceptor, HY hydrophobic feature, RA aromatic ring feature, PI positive ionizable feature

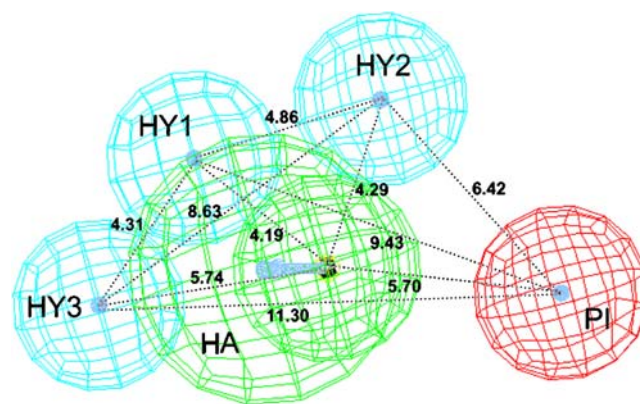


Fig. 2 The best hypothesis model, Hypo1, produced by the HypoGen module in the Catalyst4.10 software package. Pharmacophore features are color-coded: light-blue hydrophobic groups, red positive ionizable group, green hydrogen-bond acceptor. All distances between pharmacophore features are reported in Angstroms. HY1 hydrophobic group 1, HY2 hydrophobic group 2, HY3 hydrophobic group 3, HA hydrogen bond acceptor, PI positive ionizable group

In the training set, all highly active compounds map the features of hydrophobic point (HY2) and positive ionizable function (PI). All the compounds in the training set map the PI feature, which reveals that this feature could be largely responsible for the high molecular bioactivity, and thus should be taken into account in discovering or designing novel κ -opioid receptor agonists. The most active compound, compound 91, has a fitness score of 11.59 when mapped to Hypo1 (Fig. 3a) whereas the most inactive compound, compound 99, corresponds to a value of 8.46 (Fig. 3b).

The quality of the pharmacophore was further assessed using the CatScramble technique in Catalyst. This procedure was reiterated 19 times. The results of the randomization tests are listed in Table 4. None of the outcome hypotheses has a lower cost score than the initial hypothesis, which indicates that the original pharmacophore is reliable and that there is a 95% chance that the best hypothesis represents a true correlation in the training set activity data.

Pharmacophore validation

Activity prediction The predictive power of Hypo1 was validated with 75 test set compounds. Of these, 51 of 60 highly active compounds and 4 of 14 moderately active compounds were predicted correctly. Four highly active compounds were underestimated as moderately active and five highly active compounds were underestimated as inactive; nine moderately active compounds were overestimated as false positive and one was underestimated as

Table 3 Output of the score hypothesis process on the training set using Catalyst/HypoGen

No.	Compound	Actual K_i (nM)	Estimated K_i (nM)	Error factor ^a	Fit value ^b	Activity scale ^c	Estimated activity scale	Mapped features				
								HA	HY1	HY2	HY3	PI
1	91	0.043	0.1	2.4	11.59	+++	+++	+	+	+	+	+
2	86	0.096	0.11	1.2	11.54	+++	+++	+	+	+	+	+
3	89	0.13	0.15	1.2	11.42	+++	+++	+	+	+	+	+
4	11	0.17	0.84	4.9	10.68	+++	+++	+	+	+	-	+
5	1	0.27	0.47	1.7	10.93	+++	+++	+	-	+	+	+
6	47	0.4	2.2	5.5	10.26	+++	+++	+	+	+	-	+
7	74	0.48	1.4	3	10.44	+++	+++	+	+	+	-	+
8	98	0.72	3.3	4.6	10.08	+++	+++	-	+	+	+	+
9	23	0.78	0.63	-1.2	10.80	+++	+++	+	+	+	-	+
10	59	0.8	5.8	7.2	9.84	+++	+++	+	-	+	+	+
11	9	1.3	0.4	-3.2	10.99	+++	+++	+	-	+	+	+
12	57	2.6	2	-1.3	10.30	+++	+++	+	+	+	+	+
13	100	2.6	2.5	-1	10.20	+++	+++	+	+	+	-	+
14	70	2.7	0.74	-3.7	10.73	+++	+++	+	+	+	-	+
15	39	2.7	2.5	-1.1	10.20	+++	+++	+	+	+	-	+
16	19	3.8	1.4	-2.8	10.47	+++	+++	+	+	+	-	+
17	87	6	6.8	1.1	9.77	+++	+++	+	+	+	-	+
18	61	15	24	1.6	9.22	++	++	+	-	+	+	+
19	96	22	7.7	-2.9	9.71	++	+	-	+	+	+	+
20	8	23	210	9.3	8.27	++	+	+	-	+	-	+
21	29	57	13	-4.3	9.48	++	++	+	+	+	-	+
22	58	130	21	-6.1	9.27	+	++	+	+	-	+	+
23	10	280	210	-1.3	8.27	+	+	+	-	+	-	+
24	97	800	240	-3.4	8.22	+	+	-	+	-	+	+
25	99	1900	140	-14	8.46	+	+	+	+	+	-	+

^a The error factor is computed as the ratio of the measured activity to the activity estimated by the hypothesis or the inverse if estimated is greater than measured

^b Fit value indicates how well the features in the pharmacophore overlap the chemical features in the molecule

^c Activity scale: +++ <10 nM (highly active), ++ 10–100 nM (moderately active), + >100 nM (inactive)

inactive; one inactive compound was overestimated as highly active, since HypoGen is not yet able to deal with this kind of spatial problem. In summary, most of the compounds in the test set were predicted correctly according to their biological activity.

Enrichment factor Hypo1 was further validated by searching for active molecules from a database targeting κ -opioid receptors. For this validation experiment, when a spiked database having 43,448 compounds including 25 known inhibitors of κ -agonists was screened with Hypo1, 2003 molecules were retrieved as hits. Among these hits, 19 molecules were from the 25 known active compounds. Thus, the enrichment factor (Eq. 1) was found to be 16.467.

Phase

Pharmacophore generation

According to the diversity analysis using the DiverseSolutions module of SYBYL7.0, 17 active compounds were

identified. An activity threshold of 100 nM K_i values was applied to retrieve five inactive compounds from the datasets. In this study, pharmacophores with four features common to 17 active compounds were identified and scored according to the superposition of pharmacophore site points, alignment of vector characteristics, overlap of molecular volumes, and penalization of matches to inactive set molecules. The highest scoring pharmacophore model (Hypo2) contained one hydrogen bond acceptor (A), one positive ionizable group (P), and two aromatic ring (R) features, as shown in Fig. 4. The most active molecule was scored with the highest fitness, 3.0, and it was automatically selected as the reference ligand. The most active and the most inactive compounds mapped to Hypo2 are illustrated in Fig. 5.

Pharmacophore validation

Activity prediction The same collection of 100 κ -agonists was used to define a 47-member training set and a 53-member test set for the generation and validation of 3D

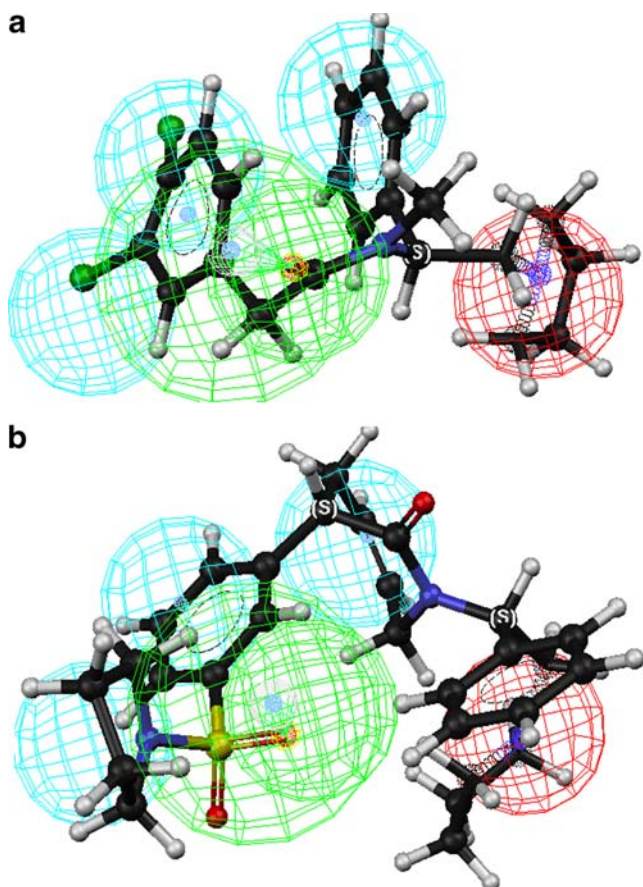


Fig. 3 Pharmacophore mapping of the most active (**a**) and most inactive (**b**) compounds on the best hypothesis model Hypo1 generated by Catalyst/HypoGen. Pharmacophore features are color-coded as in Fig. 2

pharmacophore models that rationalized the experimental activity data. Both the training set and the test set molecules performed best when characterized by three PLS factors. The training set achieved R^2 of 0.84, with an SD; the test set obtained Q^2 of 0.46, and Pearson- R of 0.692.

Enrichment factors This pharmacophore model was used to carry out a search of a combined database comprising 43,423 compounds from the SPECS database and 25 known high-affinity κ -agonists. This search recovered 690 hits including 24 of the 25 known active compounds, which corresponded to an enrichment factor of 60.436.

Discussion

Two pharmacophore models were successfully generated by Catalyst and Phase. We compared these two pharmacophore models in terms of their properties, and also compared our models with other reported pharmacophore models.

Table 4 Results from cross-validation run using CatScramble

Hypothesis	Total cost	RMS	Correlation(r)
1	148.414	1.890	0.671
2	145.968	1.833	0.698
3	143.763	1.802	0.705
4	156.258	2.059	0.585
5	152.709	2.011	0.614
6	150.853	1.875	0.690
7	143.236	1.764	0.732
8	154.723	1.993	0.624
9	156.563	2.050	0.594
10	144.903	1.836	0.690
11	149.564	1.916	0.662
12	149.702	1.942	0.644
13	148.541	1.983	0.624
14	147.501	1.912	0.668
15	139.380	1.599	0.795
16	155.547	2.142	0.536
17	139.288	1.692	0.760
18	142.562	1.779	0.719
19	140.924	1.656	0.763
Hypo1	119.581	1.127	0.900

Comparison of the two generated models

Pharmacophore features

The best pharmacophore model (Hypo1) produced by Catalyst/HypoGen consisted of five features: one hydrogen-bond acceptor (HA), three hydrophobic groups (HY), and one positive ionizable function (PI). The highest scoring model (Hypo2) obtained from Phase comprised four features: one acceptor (A), one positive ionizable group (P), and two aromatic ring features (R).

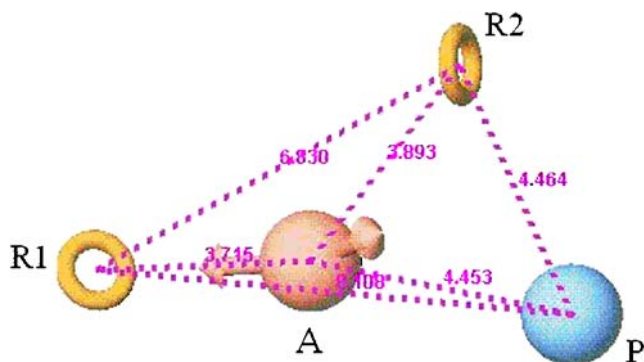


Fig. 4 The best pharmacophore model Hypo2 produced by Phase. Pharmacophore features are color-coded: orange aromatic ring, blue positive ionizable group, red hydrogen-bond acceptor. All distances between pharmacophore features are reported in Ångstroms. R1 Aromatic ring 1, R2 aromatic ring 2, A hydrogen bond acceptor, P positive ionizable group

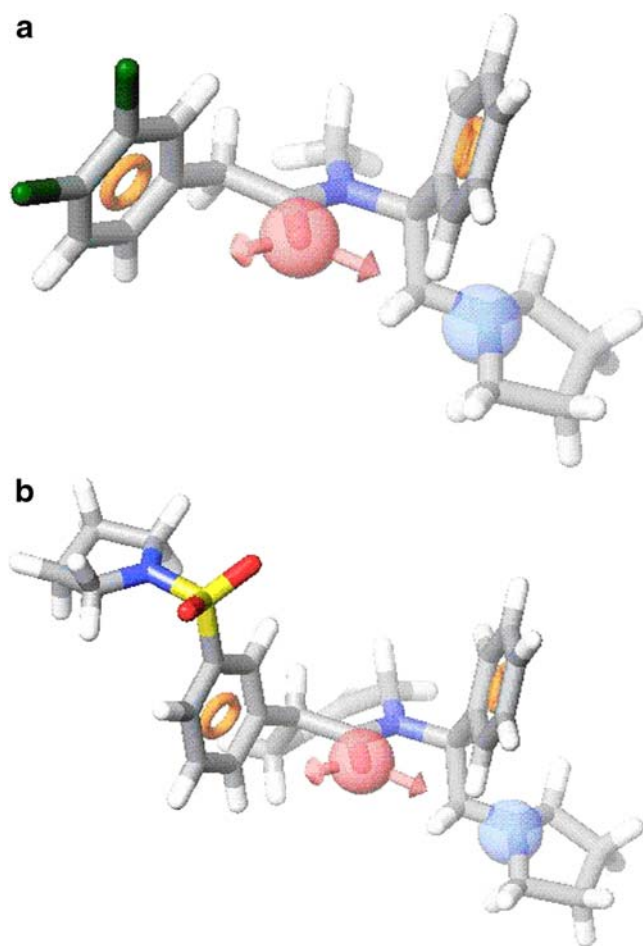


Fig. 5 Pharmacophore mapping of the most active (**a**) and most inactive (**b**) compounds on the best hypothesis model Hypo2 generated by Phase. Pharmacophore features are color-coded as in Fig. 4

The two models had two common features, the acceptor and the positive features. In addition, they also each had some unique features. Take the most active compound, compound **91** in the training set, for example. The fitness value was 11.59 when mapped to Hypo1 (Fig. 3a) and 3.0 when mapped to Hypo2 (Fig. 5a).

In compound **91**, the acceptor feature corresponds to carbonyl oxygen. The amide carbonyl group is a crucial feature in the class of arylacetamides. In fact, replacement of the amide linkage with a reversed amide, reduced N-methyl amide, or ester functional group significantly decreases or abolishes affinity [32–34]. This is consistent with the hypothesis that the carbonyl oxygen forms a hydrogen bond with the receptor. A hydrogen bond is formed between the hydroxyl of Tyr312 (TMVII) and the carbonyl oxygen of arylacetamides. Moreover, the role of the Tyr312 side chain as a hydrogen-bond donor seems to be confirmed by the reduced affinity of arylacetamides toward the Tyr312/Ala mutant [35].

In compound **91**, the positive feature is attributed to the ammonium moiety, which, typical of all active opioid ligands, is believed to form a salt bridge with the Asp138 carboxylate group in TMIII domain of the κ -receptor [24, 36–38].

Hypo1 had another three hydrophobic features (HY1, HY2 and HY3), while Hypo2 had another two aromatic ring features (R1 and R2). In compound **91**, HY1 corresponds to the benzene ring, which is generally substituted with two o-chlorine atoms. This position also maps the R1 of Hypo2. A π -stacking interaction exists between the Tyr312 side chain and the benzene ring of arylacetamides [39]. HY2 corresponds to the benzene ring linked to the chiral carbon (S-configuration), and it also maps the R2 of Hypo2. HY3 corresponds to chlorine at the C4 position of the benzene ring. Hydrophobic interactions were formed between the dichlorophenyl moiety of arylacetamides and the surrounding side chains of Tyr312, Leu224, Leu295, and Ala298 [39]. From the discussion above, we can conclude that the difference in feature generation between Phase and Catalyst lies simply in their dissimilar definitions of the hydrophobic point and aromatic ring features.

Enrichment factor

Hypo1 and Hypo2 were both validated by retrieving active molecules from spiked databases, and enrichment factors were calculated. In the 3D database containing 25 high-affinity κ -agonists, Hypo1 and Hypo2 were able to retrieve 19 and 24 active molecules, respectively, corresponding to 76% and 96%. Results of this query can be interpreted as a good validation of the generated pharmacophore hypotheses because 76% and 96% of active κ -agonists were identified as potential candidates. In addition to these known active molecules, Phase and Catalyst also picked out 690 and 1,984 new molecules, respectively, of which 221 molecules were in common. These common molecules may have a higher probability of being κ -agonist candidates. The enrichment factors were 16.467 and 60.436 for Hypo1 and Hypo2, respectively, indicating that Phase was three times more likely to pick an active compound from the database than Catalyst. As Phase retrieved more known active molecules along with fewer new molecules, this might indicate that Phase was more precise in its database search than was Catalyst in our study.

Comparison with other reported models

Recently, a pharmacophore model developed from a novel automated training set selection protocol, was generated for classical kappa opioid agonists by Singh et al. [40]. Their model consisted of four features: a hydrogen bond acceptor

(HA), a hydrophobic interaction (HY), a ring aromatic (R) and a positive ionizable moiety (PI). These results were absolutely in accord with our models.

They also constructed a 3D pharmacophore model of salvinorin A derivatives using Catalyst software [41]. Their pharmacophore model consisted of two hydrogen bond acceptor (HA) and three hydrophobic features (HY). This is in agreement with our models except for one positive ionizable group feature. Opioid receptor ligands are known to require a protonated nitrogen for high affinity binding. However, salvinorin A is a structurally unique, non-nitrogenous κ -opioid receptor agonist; it does not agree with any of the currently accepted pharmacophores of κ -opioid receptor ligands, or opioid pharmacophores in general, and demonstrates a new structural class of κ -opioid receptor agonist.

A summary supported by computer modeling studies was reported by Filizola et al. [42], describing the chemical, structural and physicochemical properties of μ , δ and κ opioid agonists. From 12 selected opioid agonists (morphine, hydromorphone, nalbuphine, xorphanol, butorphanol, dezocine, etorphine, fentanyl, lofentanyl, carfentanyl, SIOM and COMP1), four chemical moieties were found to be common for all compounds: a protonated nitrogen atom, two generic hydrophobic groups and the centroid of aromatic ring. These may be regarded as non-specific recognition motifs engaged in non-specific 3D recognition pharmacophore at μ , δ and κ opioid receptors. The common proton acceptor moiety was also identified for κ -opioid agonists. Our pharmacophore models are in accord with these results.

In addition to the four recognition motifs, Filizola also showed that κ -agonists possessed an extra unique chemical center that corresponded to the hydrophobic moiety [42]. The presence of this key group constituted the unique requirement for selective activation of κ -opioid receptor. In our study, the HY3 feature of Hypo1 seems to validate this, but unfortunately no corresponding feature was found in Hypo2.

Conclusions

To discover new potent κ -opioid receptor agonists, a ligand-based computational approach by Catalyst and Phase programs was employed to identify the molecular structure requirements of active agonists. In this work, two different programs, Catalyst/HypoGen and Phase, were performed to describe the essential pharmacophore of κ -opioid receptor agonists. One highly predictive pharmacophore model, which consisted of one hydrogen-bond receptor, three hydrophobic points, and one positive ionizable feature, was generated based on 25 training set compounds by the HypoGen module of Catalyst. Another highly predictive pharmacophore model, which comprised one hydrogen-

bond receptor, one positive ionizable function, and two aromatic ring features, was produced by Phase. The pharmacophore models that we developed using the two programs were analogous, and coincided well with experimental data as well as with other reported pharmacophore models. The difference in feature generation between the two programs lies solely in their dissimilar definition of the hydrophobic point and aromatic ring features.

In conclusion, our pharmacophore models can be considered as valuable tools for 3D database searches, and can also be applied to evaluate how well any newly designed compound maps on the pharmacophore prior to further studies including synthesis. Both applications may help in identifying or designing novel κ -agonist lead compounds for further biological evaluation and optimization.

Acknowledgments The authors greatly appreciate the software support for Catalyst from Prof. Hualiang Jiang of the Shanghai Institute of Materia Medica, CAS. This work was supported by National Natural Science Foundation of China (Grant 30600785), Shanghai Rising-Star Program (Grant 07QA14016), and Key 863 High-Tech Program (Grant 2006AA020404).

References

- Schwalbe H, Wess G (2002) *Chem Bio Chem* 3:915–919. doi:10.1002/1439-7633(20021004)3:10<915::AID-CBIC915>3.0.CO;2-L
- Gether U (2000) *Endocr Rev* 21:90–113. doi:10.1210/er.21.1.90
- Ulloa-Aguirre A, Stanislaus D, Janovick JA, Conn PM (1999) *Arch Med Res* 30:420–435. doi:10.1016/S0188-0128(99)00041-X
- Kieffer BL (1995) *Cell Mol Neurobiol* 15:615–635. doi:10.1007/BF02071128
- Dhawan BN, Cesselin F, Raghubir R, Reisine T, Bradley PB, Portoghese PS, Hamon M (1996) *Pharmacol Rev* 48:567–592
- DeHaven-Hudkins DL, Dolle RE (2004) *Curr Pharm Des* 10:743–757. doi:10.2174/1381612043453036
- Holzgrabe U, Brandt W (2003) *J Med Chem* 46:1383–1389. doi:10.1021/jm0210360
- Soukara S, Maier CA, Predoiu U, Ehret A, Jackisch R, Wunsch B (2001) *J Med Chem* 44:2814–2826. doi:10.1021/jm0108395
- Szmuszkowicz J, Von Voigtlander PF (1982) *J Med Chem* 25:1125–1126. doi:10.1021/jm00352a005
- Costello GF, James R, Shaw JS, Slater AM, Stutchbury NCJ (1991) *J Med Chem* 34:181–189. doi:10.1021/jm00105a027
- Chu GH, Gu M, Cassel JA, Belanger S, Graczyk TM, DeHaven RN, Conway-James N, Koblish M, Little PJ, DeHaven-Hudkins DL, Dolle RE (2007) *Bioorg Med Chem Lett* 17:1951–1955. doi:10.1016/j.bmcl.2007.01.053
- Chu GH, Gu M, Cassel JA, Belanger S, Stabley GJ, DeHaven RN, Conway-James N, Koblish M, Little PJ, DeHaven-Hudkins DL, Dolle RE (2006) *Bioorg Med Chem Lett* 16:645–648. doi:10.1016/j.bmcl.2005.10.034
- Le Bourdonnec B, Ajello CW, Seida PR, Susnow RG, Cassel JA, Belanger S, Stabley GJ, DeHaven RN, DeHaven-Hudkins DL, Dolle RE (2005) *Bioorg Med Chem Lett* 15:2647–2652. doi:10.1016/j.bmcl.2005.03.020

14. Chu GH, Gu M, Cassel JA, Belanger S, Graczyk TM, DeHaven RN, Conway-James N, Koblisch M, Little PJ, DeHaven-Hudkins DL, Dolle RE (2005) *Bioorg Med Chem Lett* 15:5114–5119. doi:10.1016/j.bmcl.2005.08.094
15. Kumar V, Guo D, Cassel JA, Daubert JD, DeHaven RN, DeHaven-Hudkins DL, Gauntner EK, Gottshall SL, Greiner SL, Koblisch M, Little PJ, Mansson E, Maycock AL (2005) *Bioorg Med Chem Lett* 15:1091–1095. doi:10.1016/j.bmcl.2004.12.018
16. Tuthill PA, Seida PR, Barker W, Cassel JA, Belanger S, DeHaven RN, Koblisch M, Gottshall SL, Little PJ, DeHaven-Hudkins DL, Dolle RE (2004) *Bioorg Med Chem Lett* 14:5693–5697. doi:10.1016/j.bmcl.2004.08.041
17. Kumar V, Marella MA, Cortes-Burgos L, Chang AC, Cassel JA, Daubert JD, DeHaven RN, DeHaven-Hudkins DL, Gottshall SL, Mansson E, Maycock AL (2000) *Bioorg Med Chem Lett* 10:2567–2570. doi:10.1016/S0960-894X(00)00519-9
18. Kurogi Y, Guner OF (2001) *Curr Med Chem* 8:1035–1055
19. Li H, Sutter J, Hoffman R, HypoGen (2000) An automated system for generating 3D predictive pharmacophore models. In: Guner O (ed) *Pharmacophore perception, development and use in drug design*. International University Line, La Jolla, CA
20. Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, Friesner RA (2006) *J Comput Aided Mol Des* 20:647–671. doi:10.1007/s10822-006-9087-6
21. Phase 2.0, Schrodinger, LLC: New York
22. Fischer R (1966) *The principle of experimentation, illustrated by a PSYcho-Physical Experiment*. Hafner, New York, Chap. II
23. CATALYST 4.10, Accelrys, <http://www.accelrys.com>: San Diego, CA
24. Funk OF, Kettmann V, Drimal J, Langer T (2004) *J Med Chem* 47:2750–2760. doi:10.1021/jm031041j
25. Krovat EM, Langer T (2003) *J Med Chem* 46:716–726. doi:10.1021/jm021032v
26. Tafi A, Costi R, Botta M, Santo R, Corelli F, Massa S, Ciacci A, Manetti F, Artico M (2002) *J Med Chem* 45:2720–2732. doi:10.1021/jm011087h
27. Sybyl 7.0, Triops, St Louis, MO
28. Martin YC *Distance comparisons (DISCO): a new strategy for examining 3D structure-activity relationships*. American Chemical Society, Washington DC, 1995
29. Jones G, Willett P, Glen RC (1995) *J Comput Aided Mol Des* 9:532–549. doi:10.1007/BF00124324
30. Richmond NJ, Abrams CA, Wolohan PR, Abrahamian E, Willett P, Clark RD (2006) *J Comput Aided Mol Des* 20:567–587. doi:10.1007/s10822-006-9082-y
31. Ockham WO English philosopher and Franciscan Monk, 1285, see <http://www.britannica.com/EBchecked/topic/424706/Ockhams-razor>
32. de Costa BR, Bowen WD, Hellewell SB, George C, Rothman RB, Reid AA, Walker JM, Jacobson AE, Rice KC (1989) *J Med Chem* 32:1996–2002. doi:10.1021/jm00128a050
33. Halfpenny PR, Horwell DC, Hughes J, Hunter JC, Rees DC (1990) *J Med Chem* 33:286–291. doi:10.1021/jm00163a047
34. de Costa BR, Radesca L, Di Paolo L, Bowen WD (1992) *J Med Chem* 35:38–47. doi:10.1021/jm00079a004
35. Thirstrup K, Hjorth SA, Schwartz TW (1996) In: *Investigation of the binding pocket in the kappa opioid receptor by a combination of alanine substitutions and steric hindrance mutagenesis*. 27th Meeting of the International Narcotics Research Conference, poster M30, 21–26 July 1996, Long Beach, CA
36. Kong H, Raynor K, Reisine T (1994) *Reg Pept* 54:155–156. doi:10.1016/0167-0115(94)90437-5
37. Surratt CK, Johnson PS, Moriwaki A, Seidleck BK, Blaschak CJ, Wang JB, Uhl GR (1994) *J Biol Chem* 269:20548–20553
38. Uhl GR, Childers S, Pasternak G (1994) *Trends Neurosci* 17:89–93. doi:10.1016/0166-2236(94)90110-4
39. Lavecchia A, Greco G, Novellino E, Vittorio F, Ronsisvalle G (2000) *J Med Chem* 43:2124–2134. doi:10.1021/jm991161k
40. Singh N, Nolan TL, McCurdy CR (2008) *J Mol Graph Model*. doi:10.1016/j.jmglm.2008.03.007
41. Singh N, Cheve G, McCurdy CR, Ferguson DM (2006) *J Comput Aided Mol Des* 20:471–493. doi:10.1007/s10822-006-9067-x
42. Flizola M, Villar HO, Loew GH (2001) *J Comput Aided Mol Des* 15:297–307. doi:10.1023/A:1011187320095