Bioorganic & Medicinal Chemistry 17 (2009) 2598–2606

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

Quantitative structure–activity relationship (QSAR) for a series of novel cannabinoid derivatives using descriptors derived from semi-empirical quantum-chemical calculations

Antonio M. Ferreira [a,*], Mathangi Krishnamurthy [b], Bob M. Moore II [b], David Finkelstein [a], Donald Bashford [a]

[a] Department of Structural Biology, St. Jude Children's Research Hospital, Mail Stop 311, 332 N. Lauderdale St., Memphis, TN 38105, USA
[b] Department of Pharmaceutical Sciences, College of Pharmacy, University of Tennessee Health Science Center, Memphis, TN 38103, USA

ARTICLE INFO

Article history:
Received 2 May 2008
Revised 20 November 2008
Accepted 21 November 2008
Available online 3 December 2008

Keywords:
Synthetic cannabinoids
CB1 receptor
CB2 receptor
Semi-empirical quantum mechanics
QSAR
Molecular dipole moment

ABSTRACT

Recent work implicating the cannabinoid receptors in a wide range of human pathologies has intensified the need for reliable QSAR models for drug discovery and lead optimization. Predicting the ligand selectivity of the cannabinoid CB_1 and CB_2 receptors in the absence of generally accepted models for their structures requires a ligand-based approach, which makes such studies ideally suited for quantum-chemical treatments. We present a QSAR model for ligand–receptor interactions based on quantum-chemical descriptors (an eQSAR) obtained from PM3 semi-empirical calculations for a series of phenyl-substituted cannabinoids based on a ligand with known in vivo activity against glioma [Duntsch, C.; Divi, M. K.; Jones, T.; Zhou, Q.; Krishnamurthy, M.; Boehm, P.; Wood, G.; Sills, A.; Moore. B. M., II. *J. Neuro-Oncol.*, **2006**, *77*, 143] and a set of structurally similar adamantyl-substituted cannabinoids. A good model for CB_2 inhibition ($R^2 = 0.78$) has been developed requiring only four explanatory variables derived from semi-empirical results. The role of the ligand dipole moment is discussed and we propose that the CB_2 binding pocket likely possesses a significant electric field. Describing the affinities with respect to the CB_1 receptor was not possible with the current set of ligands and descriptors, although the attempt highlighted some important points regarding the development of QSAR models.

1. Introduction

Cannabinoids represent an exciting class of compounds for the development of novel therapeutic agents.[1–5] These compounds are generally divided into three structural classes; endocannabinoids (e.g., anandamide and 2-arachidonylglycerol), nonclassical cannabinoids (e.g., CP 55,940 and WIN-55,212-2), and classical cannabinoids. The classical cannabinoids (see Fig. 1) are defined by the natural products of the *Cannabis sativa* from which the two active components Δ^9- and Δ^8-tetrahydrocannabinol (THC) are derived, with Δ^9-THC being the major constituent. These compounds primarily target the CB_1 and CB_2 receptors, though interactions with other receptor types are known.[5–8] The CB_2 receptors had been thought to reside exclusively in smooth muscle and immune cells but have been recently identified in the brain.[9,10] The CB_1 receptors are located primarily in the central nervous system with action at CB_1 being responsible for the psychotropic properties.[11–13]

The range of human diseases in which cannabinoids have been implicated is truly remarkable, including glioma,[14–18,7,19,20] pancreatic tumors,[21,22] Alzheimer's disease,[23] inflammation,[24–26] amyotrophic lateral sclerosis,[27] obesity,[28] colon cancer,[29] liver fibrosis,[30] leukemia,[31] prostate cancer,[32] and glaucoma.[33,34] Of particular interest to medicinal chemistry is the ability of THC to cross the blood brain barrier. Because of this property, the central tricyclic moiety of classical cannabinoids makes an attractive scaffold for development of novel lead compounds with applications in the treatment of neurological disorders.[35,36]

Due to the wide range of potential applications for these compounds, a great deal of effort has been expended in the development of compounds targeted at the two known cannabinoid receptors (CB_1 and CB_2).[1,37,38,3,39–47] In the absence of good struc-

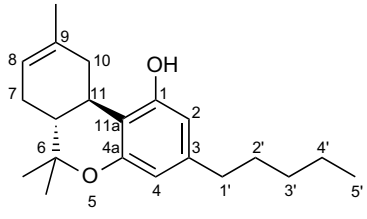


Figure 1. Numbering scheme for Δ^8-THC. The ring systems are labeled A, B and C from right to left with Ring A containing the C1 hydroxyl group and Ring B containing the cyclic oxygen atom. All substituents included in the present study (see Table 1) are substitutions at the C3-position.

* Corresponding author. Tel.: +1 901 495 4624; fax: +1 901 495 2945.
E-mail address: Antonio.Ferreira@stjude.org (A.M. Ferreira).

0968-0896/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2008.11.059

tural data on the receptor binding pockets QSAR studies are particularly valuable. Despite the potential utility of such models there have been surprisingly few studies to accompany the wealth of experimental investigations.[48] Previous quantitative structure–activity relationship (QSAR) models[49–51,48,52–54] have used 3D CoMFA and CoMSIA analyses to develop models of receptor binding and inhibition based on steric and electrostatic considerations.

We have previously proposed that the electronic structure of the phenyl-substituted cannabinoids plays an important role in determining binding affinity with respect to the CB_2 receptor.[55] Combined with the existing body of knowledge based on more traditional QSAR approaches such insights could be a powerful tool for the design of novel cannabinoids with improved potency and selectivity. The present work strives to develop QSAR models based on quantum-chemical descriptors (which we term eQSAR) easily derived from semi-empirical calculations. The application of such methods is well-established[56–63] and an excellent review on the subject is available from Karelson et al.[64] Previous computational work on cannabinoid ligands[65–67] has used a combination of methods although only Honório and da Silva[68] have attempted to develop an eQSAR using semi-empirical calculations. However, that work simply classified compounds as psychoactive or psychoinactive and did not include any specific binding data for the ligands.

2. Results and discussion

Selection of the compounds for inclusion in the present study was based on several criteria. Of primary importance was the availability of binding data for both receptor subtypes. Compounds without specific data for both CB_1 and CB_2 were removed from consideration. Similar structures were desired to minimize steric effects in order to focus on the relevant electronic structure parameters. Lastly, we desired compounds with systematic differences in substitution position and bonding patterns. The present set of compounds, given in Table 1, allows us to study potentially subtle electronic effects through related derivatives. Another attractive feature of these compounds is that the scaffold molecule, KM-233 (compound **6**), has been shown to have efficacy versus glioma with in vivo xenograft experiments.[17] The synthesis, characterization and binding assays for Compounds **6** through **22** have been described elsewhere.[85] Compounds **1** through **5** were described in a recent paper by Lu et al.[65] and offer structural features similar to the KM-233 series with subtle variations in bonding patterns.

Results from the receptor binding assays[85,65] reveal some intriguing relationships for these two classes of compounds. Within the KM-233 series, the dependence of $K_i(CB_2)$ on substitution position within the halogen series is particularly interesting. The effect is most pronounced with the fluoro-substituted derivatives (**8** and **15**) where substitution at the *para* position versus the *meta* position yields a nearly 14-fold change (12.40 vs 0.90 nM). Considering the close similarity of these structures and the free rotation about the C1′–C2′ bond, it is unlikely that this effect is due to steric considerations and suggests an electronic contribution. Furthermore, the results for CB_1 in the halogen-substituted series (compounds **15**, **16** and **17**) suggest a role for the electronic structure as the K_i decreases with the decreasing electron withdrawing power of the substituents, ranging from 76.1 nM for the fluoro-substituted derivative to 5.03 nM for the bromo-substituted derivative. Results for the adamantyl compounds show a similar dramatic change in $K_i(CB_2)$ between AM755 and AM757 (compounds **4** and **5**), which differ only by a double-bond at the C1′ position. This small perturbation in the chemical bonding induces a nearly ninefold increase in activity, from 76.0 nM to 8.90 nM, providing further evidence for the importance of subtle differences in the electronic structure.

As a validation of the semi-empirical method for determining the ligand geometries the τ_1 (rotation about the C3–C1′ bond) and τ_2 (rotation about the C1′–C2′ bond) angles calculated for three of the derivatives have been compared against available NMR data (see Table 2). This analysis shows differences of less than 20 degrees for all but one of the dihedral angles with respective average absolute differences in τ_1 and τ_2 angles of 13.01 and 8.18 degrees for the *p*-fluoro derivative, 5.83 and 3.24 degrees for the *m*-fluoro derivative, and 6.94 and 5.02 for the thiophene derivative. These results are consistent with our previous analysis on structurally similar compounds.[69]

Lu et al.[65] performed semiempircal calculations similar to those reported here (see Section 3) using the AM1 model Hamiltonian yielding structures for the individual conformers that are extremely similar to our PM3 results. Based on these results, they have argued that steric differences arising from the addition of the double bond (see Fig. 4 in their paper) can explain this observation. Our own comparison of the set of conformers for these two compounds reveals that the difference in the spatial distributions for these two compounds is minimal. The calculated RMSD values for the sets of conformers obtained for these two derivatives are reported in Table 3 and demonstrate significant overlap between the two sets indicating that the two compounds are capable of occupying similar spatial regions. Additionally, we would point out that an analysis of the conformational space available to each is unlikely to account for the differences associated with the biologically active conformations. Such an analysis might be more convincing if rotation about the τ_1 and τ_2 angles were hindered but this is not the case for these molecules. In the case of AM757, the adamantyl sidechain does occupy slightly more space near Ring A but we believe that these small structural differences are unlikely candidates for explaining the experimental data. The nearly order-of-magnitude difference in binding for these two compounds is likely due to electronic effects caused by interactions of the π-electrons of the double bond with the aromatic π-electron density of Ring A.

To investigate the role of electronic structure factors in the binding of these compounds we undertook a statistical analysis to construct an eQSAR model for predicting K_i. Such an analysis of the role electronic properties play in receptor binding requires that either a particular conformer be selected to represent each compound in the analysis or that we employ average quantities, such as those obtained from Boltzmann averaging. For the present study, we have chosen the first approach as we feel a proper description of the biologically active conformer is more directly relevant to predictions of inhibition.

The absence of accepted crystal structures for the cannabinoid receptors precludes a traditional structure-based analysis to determine the appropriate set of conformers. We instead focus on the properties of the ligands to make inferences about the ligand binding pocket. A comparison of the lowest-energy conformers (in terms of calculated free energy) is given in Figure 2 a for KM-233 (the scaffold molecule) and the other derivatives with $K_i(CB_2) <$ 1.0 nM. In combination with the high affinities seen in the receptor binding assays, the remarkable similarity between the structures suggests a potential biologically active conformation for these cannabinoids. Using KM-233 as a reference molecule, the RMSD values for the structures in Figure 2 were calculated revealing nearly identical structures for the *m*-bromo derivative (**10**), the *m*-cyano derivative (**12**) and the *m*-fluoro derivative (**8**), with RMSD values of 0.006, 0.015, and 0.007, respectively. The RMSD values for compounds with $K_i(CB_2) > 1.0$ nM were all 0.2370 or larger. Based on the structural similarities and the low K_i values for these, we hypothesize that the biologically active conformers for the CB_2 receptor will likely resemble the lowest-energy conformations for these four molecules. Hypothetical biologically active conformers were therefore selected based on molecular alignment with

Table 1
Cannabinoid compounds included in the present study

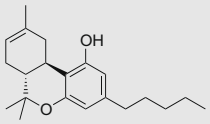
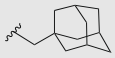
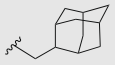
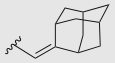
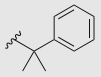
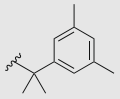
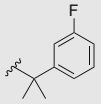
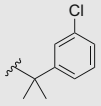
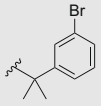
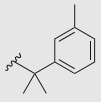
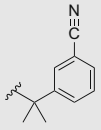
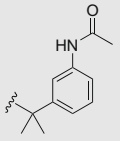
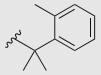
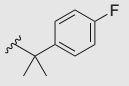
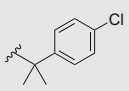
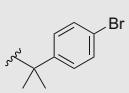
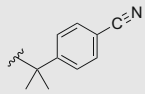
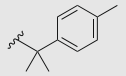
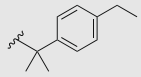
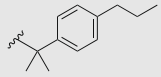
	Compound	Structure	K_i(CB$_1$)(nM)	K_i(CB$_2$) (nm)
	Δ^8–THC		28.5 ± 3.30	25.0 ± 4.80
1	AM411		6.80 ± 0.80 [35.7]	52.00 ± 5.00 [26.0]
2	AM744		34.90 ± 5.90 [40.13]	14.00 ± 2.30 [40.86]
3	AM729		29.3 ± 2.9 [51.6]	26.9 ± 4.2 [45.5]
4	AM757		79.7 ± 11.40 [43.2]	76.0 ± 12.20 [32.8]
5	AM755		48.6 ± 3.80 [16.4]	8.90 ± 0.70 [9.21]
6	KM-233		12.3 ± 0.61 [4.06]	0.91 ± 0.08 [1.07]
7	KM-233-3,5-dimethyl		11.0 ± 1.67 [4.98]	7.45 ± 0.38 [2.68]
8	KM-233-*m*-fluoro		5.26 ± 0.94 [22.2]	0.90 ± 0.02 [4.18]
9	KM-233-*m*-chloro		2.80 ± 0.05 [6.92]	3.54 ± 0.71 [1.51]
10	KM-233-*m*-bromo		1.59 ± 0.16 [4.68]	0.54 ± 0.03 [1.19]
11	KM-233-*m*-methyl		2.53 ± 0.54 [6.44]	1.13 ± 0.02 [1.66]
12	KM-233-*m*-cyano		2.72 ± 0.29 [2.69]	0.91 ± 0.05 [0.84]
13	KM-233-*m*-acetamide		13.7 ± 1.88 [11.0]	13.6 ± 1.62 [18.8]
14	KM-233-*o*-methyl		34.4 ± 2.84 [14.4]	10.65 ± 1.27 [7.87]
15	KM-233-*p*-fluoro		76.1 ± 1.55 [34.7]	12.40 ± 0.24 [6.05]
16	KM-233-*p*-chloro		18.8 ± 1.39 [6.60]	1.68 ± 0.20 [1.45]
17	KM-233-*p*-bromo		5.03 ± 0.39 [3.96]	1.54 ± 0.16 [1.04]

Table 3 (*continued*)

	Compound	Structure	$K_i(CB_1)$(nM)	$K_i(CB_2)$ (nm)
18	KM-233-*p*-cyano		9.25 ± 0.23 [7.80]	2.53 ± 0.23 [2.04]
19	KM-233-*p*-methyl		3.13 ± 0.37 [6.11]	0.88 ± 0.05 [1.59]
20	KM-233-*p*-ethyl		1.85 ± 0.16 [4.35]	0.67 ± 0.05 [2.38]
21	KM-233-*p*-propyl		1.77 ± 0.20 [2.03]	7.83 ± 0.79 [5.78]
22	KM-233-thiophenyl		1.08 ± 0.04 [1.40]	0.27 ± 0.01 [0.30]

Side chains represent the substitution at the C3 position as indicated in Figure 1. The data for Compounds **1** through **5** were taken from Lu et al.[65] Predicted values are given in square brackes and were obtained from Eq. 1 for $K_i(CB_2)$ and from the rejected model in the text for $K_i(CB_1)$.

Table 2
Calculated versus experimental torsion angles (in degrees) for selected cannabinoid derivatives

Compound	NMR angles		Structure	PM3 results	
	τ_1	τ_2		τ_1	τ_2
Thiophene	330	125	A	324	128
	140	130	I	144	128
	40	225	L	36	233
	235	225	H	221	232
p-Fluoro	310	320	H	320	320
	240	260	E	224	224
	210	40	B	225	40
	150	140	F	137	136
	30	40	G	41	40
m-Fluoro	30	40	M	48	40
	130	130	J	137	136
	40	220	K	41	223
	220	220	I	224	223

The τ_1 angle is defined as the C2–C3–C1′–C2′ angle and the τ_2 angle is defined as the C3–C1′–C2′–C3′ angle. In each case, the C2′ atom is taken as the bridgehead carbon, with the C3′ atom adjacent and belonging to the associated ring system.

Table 3
Calculated RMSD values for the conformers of AM755 and AM757

AM757 conf.	AM755 conf.			
	A	B	C	D
A	2.956	1.617	2.832	1.755
B	2.889	1.610	2.767	1.789
C	2.803	1.519	2.705	0.681
D	1.582	2.727	1.517	3.200
E	1.629	2.976	1.616	3.052
F	1.651	2.747	1.608	2.799
G	2.924	1.640	2.796	1.723
H	1.638	2.717	1.571	3.221
I	2.742	1.580	2.643	0.649
J	1.621	3.471	1.708	3.503
K	3.541	1.662	3.420	1.724
L	1.679	3.493	1.760	3.565
M	3.585	1.791	3.462	1.723

KM-233 (as shown in Fig. 2b) using the largest possible subset of atoms. The selected conformers for all 22 compounds are given in Table 4 and the accompanying Supplementary data contains the complete set of PM3 results for all conformers of each compound. Similar alignments were obtained for the compounds with $K_i(CB_1) < 1.0$ nM (as shown in Fig. 3) with RMSD values for the structures of 0.010 (*p*-ethyl), 0.006 (*p*-propyl) and 4.278 (thiophenyl) using the *m*-bromo derivative as the reference structure. In the worst case, the differences between compounds **10** and **22** are attributable to different torsion angles for the thiophene ring as compared to the phenyl substituent. However, the figure clearly shows that these two groups occupy the same region of space. The similarity between the best CB_2 and CB_1 binders suggests that using a distinct set of conformers for developing the QSAR for $K_i(CB_1)$ would be unnecessary.

For this set of biologically active conformers, a set of 29 molecular descriptors was selected from the semi-empirical results. The complete set of descriptors (see Table 5) and the definition of the derived quantities is given in Section 3. An examination of Figs. 2 and 3 suggests that the dipole moment of the ligands may play a role in determining binding affinity, which prompted the inclusion of these terms in the QSAR models.

As many of the molecular descriptors are related explicitly through an algebraic form or phenomenologically (e.g., ΔH_f and E_{aq}, ΔG° and $\Delta S_{Vib.}$), issues of multicollinearity were a concern in attempting multiple linear regression. Multicollinearity is a property of linear regression models containing explanatory variables that are strongly correlated with each other. Variance inflation factors[70] are a formal method for detecting the presence of multicollinearity in a linear regression model and are defined as

$$\mathrm{VIF}_i = (1 - R_i^2)^{-1}$$

where R_i^2 is the coefficient of multiple determination (or multiple correlation coefficient) obtained when the *i*th explanatory variable is regressed against the remaining variables in the model. Values of the variance inflation factors (VIFs) greater than 10 are indicative of unacceptably large collinearity among the explanatory variables, with an ideal value of unity indicating that there is no such correlation.

We created several categories of explanatory variables and required potential models to contain no more than one variable from each of these categories as a means to avoid potential multicollinearity. Categories for total energy (ΔH_f, E_{solv} and E_{aq}), shape (V, A and O), orbital energy terms (ε_{HOMO}, ε_{LUMO}, χ, η, $1/\eta$, ω and $\langle\alpha\rangle$) and vibrational energy terms (ΔG°, ΔH, $\Delta H_{Vib.}$, ΔS, $\Delta S_{Vib.}$, $\Delta S_{Rot.}$, $\Delta S_{Trans.}$) were used. The remaining descriptors were omitted from the categorical selection rules. This effectively reduced the number of potential explanatory variables while removing terms that were explicitly covariant from potential models. With the limited number of parameters and small data set an all-possible-regressions analysis (with all first- and second-order terms) was possible and we undertook a comparison of the best possible models with the results of a stepwise linear regression analysis using the reduced set of potential explanatory variables and the $\log(K_i)$ values as the response variables. Potential models were

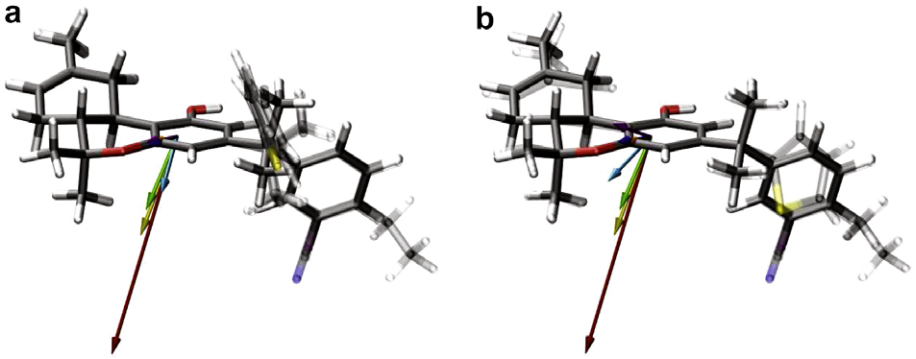


Figure 2. Superposition of the lowest-energy conformers in terms of calculated free energy (a) and the hypothesized biologically active conformers (b) for the high-affinity CB_2 compounds **6**, **8**, **10**, **12**, **19**, **20**, and **22**. The structures are drawn with transparent surfaces to emphasize the small differences between structures. Carbon atoms are drawn in gray, oxygens in red, and nitrogens in blue with the halogens fluorine and bromine in purple and green, respectively. Vectors representing the calculated dipole moments have been scaled by a factor of two in order to make the small dipole of compounds **6**, **10** and **19** more visible. Coloring of the dipole representations is by compound according to the following key; **6**: orange, **8**: yellow, **10**: green,**12**: red, **19**: blue, **20**: purple and **22**: cyan. It is interesting to note the similarity among the dipole moments of the lowest-energy structures (a) and the hypothesized biologically active conformers (b).

Table 4
The hypothesized biologically active conformers for the compounds included in the present study

Compound	Biologically active conformer	Total number of conformers	$\Delta\Delta G^{\circ}$ (kcal/mol)
1	B	2	0.00
2	A	2	0.34
3	D	4	0.13
4	I	13	0.00
5	D	4	0.00
6	E	7	0.00
7	L	18	0.00
8	J	14	0.00
9	J	14	0.00
10	J	14	0.00
11	N	21	0.00
12	J	14	0.00
13	N	57	0.84
14	J	14	0.33
15	F	13	1.86
16	J	14	0.33
17	F	8	0.02
18	F	8	0.09
19	F	8	0.00
20	J	14	0.01
21	L	16	0.00
22	AM	48	0.00
23	F	8	0.01

The conformer designation corresponds to the entries in Supplementary data. The total number of conformers represents the number of conformers identified by the conformational search procedure described in the text.

identified using a consensus ranking of R^2 values, PRESS scores[71] and Akaike Information Criterion (AIC)[72] values. Quantile-comparison plots, which plot the standardized residuals for a particular model versus the theoretical residuals, were used as a diagnostic tool to assess models identified with the consensus rankings. The finer details of the statistical analysis are provided in the Supplementary data.

Results for the CB_2 receptor data provided two very similar models with nearly identical R^2 values and quantile-comparison plots. Given these similarities, the variance inflation factors became the deciding factor when selecting the final model. Our model describes the CB_2 binding as a function of the aqueous heat of formation (E_{aq}), the *x*-component (see Section 3) of the dipole moment (μ_x), the ovality (O) and the polarizability ($\langle\alpha\rangle$), with VIF values for all variables of < 5.0. The fitted regression model is

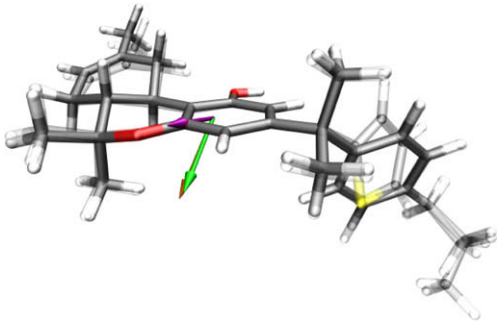


Figure 3. Superposition of the lowest-energy conformers in terms of calculated free energy for the high-affinity CB_1 compounds **10**, **20**, **21** and **22**. As with Figure 2, the structures are drawn with transparent surfaces and the dipoles have been scaled by a factor of two. Calculated dipole moments are colored according to the following key orange: **10**, purple: **20**, red: **21** and green: **22**.

Table 5
Symbols and definitions of the molecular descriptors used in the present study

Descriptor	Definition	Units
ΔH_f	Heat of formation	kcal/mol
E_{solv}	Solvation energy[78]	kcal/mol
E_{aq}	$\Delta H_f + E_{solv}$	kcal/mol
A	Van der Waals surface area	Å^2
V	Van der Waals volume	Å^3
MW	Molecular weight	g/mol
ε_{HOMO}	Energy of the HOMO	eV
ε_{LUMO}	Energy of the LUMO	eV
ZPE	Zero-point energy	kcal/mol
$\log(P)$	Octanol–water partition coefficient[79]	
ΔG°	Gibbs free energy of formation	kcal/mol
ΔH	Total enthalpy	kcal/mol
$\Delta H_{Vib.}$	Vibrational enthalpy	kcal/mol
ΔS	Total entropy	kcal/mol
$\Delta S_{Vib.}$	Vibrational entropy	kcal/mol
$\Delta S_{Rot.}$	Rotational entropy	kcal/mol
$\Delta S_{Trans.}$	Translational entropy	kcal/mol
μ	Total dipole moment	D
μ_x,μ_y,μ_z	Dipole moment components in the molecular reference frame	D
q^+	Maximum positive charge on any H atom	e
q^-	Maximum negative charge on any H atom	e
O	Ovality (see Eq. 4)	
χ	Electronegativity (see Eq. 5)	eV
η	Hardness (see Eq. 6)	eV
η^{-1}	Softness ($\frac{1}{\eta}$)	eV^{-1}
ω	Electrophilicity (see Eq. 7)	eV
$\langle\alpha\rangle$	Isotropic polarizability (see Eq. 8)	Å^3

$$\log(K_i(CB_2)) = -0.0296E_{aq} + 0.4325\mu_x - 61.5026O + 1.3562\langle\alpha\rangle - 4.7348 \quad (1)$$

with $R^2 = 0.775$, a p-value of 1.18×10^{-5} and an F-statistic of 15.5. The p-value, or significance, represents the probability that the variation in the model occured by chance. Thus, the smaller the p-value the more significant the model. The F-statistic is directly related to the p-value and significant results are obtained when the F-statistic assumes a relatively high value. An F-statistic of 1.0 would indicate a model with no statistical significance. Cross-validation of the model using the leave-one-out method as implemented in the `CVlm` function from the `DAAG` library[73] in R demonstrated that the overall error mean square for the predicted values was 1.12. When comparing a set of linear models, the error mean square provides a measure of the variance between the models when ignoring a subset of the data. Since our error mean square is on the order of the smallest value in the data set, this indicates a reliable regression model. A plot of the predicted binding versus the experimental results is given in Figure 4. The figure demonstrates the limitations of the model, particularly for compounds with binding affinities in the nM range, but in light of the limited size of the data set and the good model diagnostics we feel these shortcomings are within acceptable limits.

Developing a model for CB_1 binding was more problematic. None of the models investigated for $K_i(CB_1)$ were deemed suitable for publication, however the attempt highlighted some important general aspects of model development. We provide an account of our attempts to develop a model for CB_1 in an effort to add our voice to the debate on the pitfalls associated with the use of automated methods.

Whereas the stepwise and all-possible-regressions procedures identified a number of models with R^2 values of 0.70 or greater for $K_i(CB_2)$, the same methods produced only two models with $R^2 > 0.50$ for $K_i(CB_1)$. There were, however, some interesting similarities among all of the models with moderate predictive power. Each of the best performing models for $K_i(CB_1)$ contained either an ovality term (O) or a vibrational entropy term (ΔS_{Vib}) and an interaction term between two of the explanatory variables, indicating nonlinear behavior. Generally, the models lacked terms involving the electronic properties of the molecules, although dipole terms did appear in several models. The absence of electronic terms from all of the best-performing models suggests that binding at the CB_1 receptor is not significantly influenced by the more subtle electronic structure of the molecules. In contrast, the QSAR model for CB_2 binding is highly dependent on the electronic structure of the ligands with significant contributions from μ_x and $\langle\alpha\rangle$. In fact, the p-value for the polarizability term (3.72×10^{-4}) is the most significant of any of the explanatory variables.

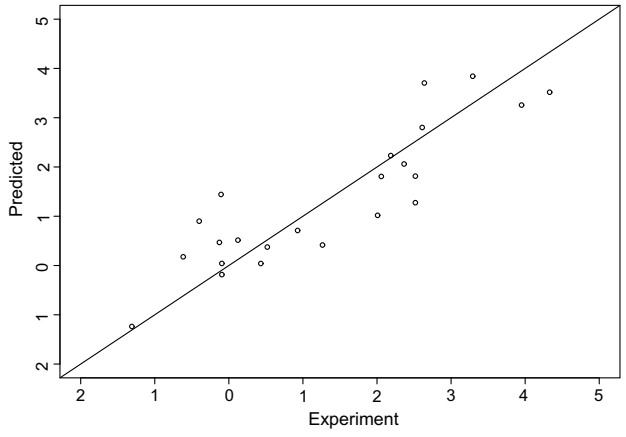


Figure 4. Calculated versus experimental values for $\log(K_i(CB_2))$ using the linear regression model of Eq. 1.

The best of the two linear models for $K_i(CB_1)$, with $R^2 = 0.597$, had E_{aq}, μ_x, q^-, $\log(P)$ and the product $\mu_x q^-$ as the explanatory variables. The other, with $R^2 = 0.562$, included E_{aq}, μ_x, A, $\langle\alpha\rangle$ and the product $A\langle\alpha\rangle$. An analysis of the quantile-comparison plots revealed that the model with the higher R^2 suffered from significant deviations from normality, indicating nonrandom errors in the model. The multiple linear regression fit for the other was given by

$$\log(K_i(CB_1)) = -0.0328E_{aq} + 0.5176\mu_x + 0.7765A + 6.2492\langle\alpha\rangle - 0.0120A\langle\alpha\rangle - 413.6835$$

with $R^2 = 0.563$, a p-value of 9.75×10^{-3} and an F-statistic of 4.36. Cross-validation showed the overall mean squared of the error in predicted values was an acceptable 1.44. To this point, the model appears to provide a respectable description for CB_1 binding. For many automated procedures, this set of diagnostics for a multiple linear regression would be acceptable, especially given that the quantile-comparison plot showed no egregious variations from normality. However, examination of the variance inflation factors revealed terms with values greater than 25.0, clearly indicating problems stemming from multicollinearity. One remedy for such complications is to normalize the data using a Box-Cox transformation, defined in R as

$$y(\lambda) = \frac{y^{\lambda}-1}{\lambda}, \quad \text{if } \lambda \neq 0,$$
$$y(\lambda) = \log(y), \quad \text{if } \lambda = 0.$$

Optimization of the Box-Cox λ values (powers) yielded some remarkably large values, with $|\lambda| >= 5$ for nine of the explanatory variables and 4 having $|\lambda| > 15$. As a general rule, one would like to have values of $|\lambda| <= 3$ in all but the most extreme circumstances. Several of the lower-order transformations were tried and a new model was identified with $R^2 = 0.593$, a p-value of 5.62×10^{-3} and an F-statistic of 4.95 containing terms up to $\lambda = -7$. Unfortunately, the VIF values for this model increased with respect to the untransformed data.

Given the failure of the variable transformations, we decided to attempt a principle component analysis (PCA). The goal in PCA analysis is to identify a minimal number of new explanatory variables (principle components) constructed as a linear combination of the original variables. The principle components are by construction orthogonal and therefore noncollinear. A minimum of seven principle components were required to obtain a model with predictive accuracy comparable to that obtained with the all-possible-regressions. Furthermore, an examination of the coefficients for the principle components revealed no dominant terms.

Our next attempt to obtain a suitable model employed a ridge regression procedure. With ridge regression, a biasing constant (c) is employed to reduce the total mean squared error using

$$\boldsymbol{b}^r = (\boldsymbol{r}_{xx} + c\boldsymbol{I})^{-1}\boldsymbol{r}_{YX} \quad (2)$$

as opposed to the standard regression model where

$$\boldsymbol{b} = \boldsymbol{r}_{xx}^{-1}\boldsymbol{r}_{YX}. \quad (3)$$

In these expressions, $\boldsymbol{b}$ is the vector of standardized regression coefficients, $\boldsymbol{r}_{xx}$ is the correlation matrix of the explanatory variables and $\boldsymbol{r}_{YX}$ is the correlation matrix between the dependent variables and each explanatory variable. Plots of the regression coefficients as a function of the biasing constant are used as a diagnostic tool and variables demonstrating instabilities in these plots are recommended for removal from the model. Instability in all but two of the explanatory variables was seen.

From these analyses it is clear that a simple linear model will fail to provide an adequate description of CB_1 binding within the

constraints of the current data and set of descriptors. We therefore attempted a sophisticated nonlinear regression model called a generalized additive model. The need to resort to nonlinear model of this type can be justified by the failure of the linear regressions using unmodified data, the nonlinearity seen in the quantile-comparison plots for the best linear models (see Supplementary data), the large λ values returned by the Box-Cox procedures, and the instabilities in the explanatory variables identified with the ridge regression procedure. Generalized additive models are relatively new in statistics and have the general form

$$y_i = f_1(x_i) + f_2(z_i, v_i) + f_3(u_i) + \cdots$$

where x_i, z_i, v_i and u_i are explanatory variables. Spline functions (g_i) are used to estimate the f_i by finding the set of functions, $\boldsymbol{g}$, that minimizes the expression

$$\|\boldsymbol{y} - \boldsymbol{g}\|^2 + \lambda J_{\mathrm{md}}(g)$$

where $J_{\mathrm{md}}(g)$ is a penalty functional for the spline functions and the parameter λ controls the balance between data fitting and smoothness of g. We constructed a series of generalized additive models employing a thin plate regression spline with shrinkage[74] via the `mgcv` library available for R. Given the limited number of data points in the present study, a sixth-order penalty for including additional terms in the spline regression was employed to minimize the potential for over-fitting. Such over-fitting is manifested as high-degree polynomials in the explanatory variables or very low degrees of freedom for the resulting model. Several models were generated with excellent R^2 values but all suffered from severe over-fitting of the data (d.f. < 3) despite the penalty.

3. Methods

The NMR structures for compounds **8**, **15**, and **22** were obtained from minimization of the average conformations obtained from NOESY spectra using distance constraints. All spectra were acquired at 23 °C and 500 MHz on a Varian Inova-500 spectrometer using a 5-mm HCN triple resonance probe. Both proton and carbon chemical shifts were referenced to the residual solvent peak of DMSO (2.49 ppm for proton and 40 ppm for carbon). For two-dimensional NOESY measurements, a total of 512 fids were recorded for the indirect dimension, with a 2 second recycle delay. The TRIAD NMR package within the Sybyl software package[75] was used for data processing and analysis. Peaks in the NOESY spectra were assigned and integrated using TRIAD standard functions and subsequently used to generate distance constraints with MARDIGRAS. Results from each of the five mixing times gave very similar distance constraints, hence each distance constraint was averaged over the five mixing times to get the final set of distance constraints for the molecule. The resulting constraints were then examined to ensure that the error in distances conformed to established errors for NOE constraints wherein; $x < 2.5$ Å was ±0.1 Å; $x \leqslant 3.0$ Å was ±0.2 Å; $x \leqslant 3.5$ Å was ±0.3 Å; and $x \geqslant 3.5$ Å was ±0.4 Å.

For each of the three compounds, a four-step simulated annealing using 1 fs time steps and the constraints generated by MARDIGRAS was performed as follows: (1) 1 ns dynamics at 300 K; (2) 1 ns heating to 500 K; (3) another 1 ns heating phase to 700 K; (4) a 1 ns equilibration to 500 K. Additional parameters included the Tripos force field with Gasteiger-Hückel charges, an 8 Å nonbonding cutoff, and distance dependent dielectric constant function. The experimentally obtained NOE distance constraints were applied during all steps of the dynamics runs, and the aromatic carbons were defined as aggregates to maintain the ring geometry. The molecular geometry was sampled at 1000 fs intervals during phase (1) of the dynamics runs and once during the heating and cooling periods. A total of 1007 conformations were collected over the course of the dynamics for further analysis and these were subjected to 20 dynamics simulations each to obtain average conformations. These average conformations were then minimized with a gradient tolerance of 0.005 kcal/(mol Å) without defined aggregates or experimental NOE distance constraints to obtain the final average conformations.

Systematic conformer searches were performed using SPARTAN'02[76] at the PM3 level of theory with an energy window of 5000 kcal/mol and a gradient tolerance of 10^{-4} kcal/mol. Methyl group rotations were allowed and considered distinct conformers. The resultant structures were then re-optimized with a gradient tolerance of 10^{-6} kcal/mol. Numerical Hessians of the local minima were calculated and only those conformers confirmed as local minima were retained. The statistical analyses for the development of the QSAR model were performed using the open-source R software.[77]

The complete set of quantum-chemical properties used in the present study is given in Table 5 with all results coming from the Spartan'02 software. The solvation energy (E_{solv}) was calculated with the SM5.4 method of Cramer and Truhlar[78] and the octanol–water partition coefficients ($\log(P)$) are from the method of Ghose and Crippen.[79] Natural population analysis[80] was used to determine the so-called TLSER parameters[57] describing the maximum positive and negative charges on the hydrogen atoms (q^+, q^-). Components of the dipole moment were transformed into a right-handed molecular reference frame defined by the plane of Ring A, with the x-direction being from C1 to C4a and the y-direction being from C11a to C3. Using this fixed portion of the scaffold structure to define the dipole coordinate system allows for direct comparison of the individual dipole components between compounds.

Several derived quantities were also included, namely the ovality

$$O = \frac{A}{4\pi\left(\frac{3V}{4\pi}\right)^{\frac{2}{3}}} \tag{4}$$

the electronegativity,

$$\chi = \frac{\varepsilon_{\mathrm{HOMO}} - \varepsilon_{\mathrm{LUMO}}}{2} \tag{5}$$

the hardness,

$$\eta = \frac{\varepsilon_{\mathrm{LUMO}} - \varepsilon_{\mathrm{HOMO}}}{2} \tag{6}$$

the electrophilicity

$$\omega = \frac{\chi^2}{2\eta} \tag{7}$$

and the polarizability ($\langle\alpha\rangle$). The polarizability is estimated from an empirical fit[81] using the van der Waals volume (V) and the hardness (η) according to

$$\langle\alpha\rangle = 0.979920\eta^2 - 13.0352\eta + 0.08V + 41.3791. \tag{8}$$

Procedures for the experimental K_i determinations are reported elsewhere.[85] Briefly, the binding affinities were determined using membrane preparations of the human receptors transfected into HEK293 EBNA cells. Binding assays were performed with CP55,940 as the competing radioactive ligand and 10 μM WIN 55212-2 was used for determining nonspecific binding.

4. Conclusions

We have demonstrated that the semi-empirical modeling yields excellent agreement with the structures determined using NMR

spectroscopy, lending credence to the calculated electronic properties and therefore the subsequent analyses. Our statistical analysis gives excellent results based on four quantum-chemical descriptors for these ligands in the case of CB_2 binding. The strong influence of the dipole moment terms in the statistical model combined with the structural analysis from Figure 2 suggests that the ligand binding pocket in the CB_2 receptor likely possesses a local electrostatic field. Such a field could be generated by several charged residues and would exert a significant effect only on molecules with a significant dipole moment or a high polarizability. Furthermore, it appears that the component of the dipole moment parallel to the plane of Ring A and directed along the vector from C1 to C4a (away from the OH group) is the most significant. When combined with the structural information, this may provide some insight into which residues are likely involved, but such an analysis is beyond the scope of the present manuscript and must wait until more reliable receptor structures are made available.

In the case of the adamantyl cannabinoids, we believe we have made a strong argument for the differences in binding for AM757 and AM755 based on the electronic structure of these molecules as opposed to the original suggestion of Lu et al. that sterics are the predominating factor. Indeed, the ovality difference between these two ligands is only $1.528516 - 1.530760 \approx -0.002$ (see Supplementary data), which in the context of the QSAR equation accounts for only 11% of the difference in $\log(K_i)$ values for these compounds. Combined with the structural similarities discussed previously (Table 3) this provides a strong argument against predominance by sterics. In comparison, the polarizability difference of $73.36867 - 73.08220 \approx 0.286$ accounts for 30% of the total variation. We conclude that influence of the C1′ double-bond on the electronic structure is likely more important than the small conformational differences between these two ligands. We believe that the role of the electronic structure in determining binding at the CB_2 receptor has been clearly demonstrated and such considerations should be employed in the development of novel CB_2-selective compounds. In particular, structures should be optimized with respect to the polarizability and the dipole moment. The important component of the molecular dipole will be that which is in the plane of Ring A and directed away from the hydroxyl group at the C1 position.

However, prediction of CB_1 binding is not possible with the current combination of data and descriptors. The difficulties encountered in the attempt to develop a model for CB_1 binding highlight the need for rigorous characterization of all QSAR models (a point which has been raised previously[82,83]). As was seen with the multiple linear regression model, it is possible for automated procedures to generate models that appear valid unless careful attention is paid to the appropriate diagnostic statistics. Variance inflation factors are not usually reported in the literature and we would advocate their disclosure as a standard measure of model performance in future work as a means to avoid such pitfalls.

The insensitivity seen for $\log(K_i(CB_1))$ with respect to the electronic structure descriptors is interesting in itself. If this is truly the case, it would offer a particularly attractive route to designing novel cannabinoids as substitutions affecting the electronic structure can be made to optimize CB_2 characteristics without significantly impacting CB_1 binding. Furthermore, one would expect that more traditional QSAR analysis, such as CoMFA and CoMSIA, would be successful in predicting CB_1 binding. Indeed, preliminary work[84] suggests that CoMFA modeling works reasonably well in the CB_1 case, but fails for CB_2.

Acknowledgments

The authors wish to acknowledge the support of the American-Lebanese-Syrian Associated Charities (ALSAC). We wish to thank Jing Ma (SJCRH) for helpful discussions regarding the PCA analyses. AMF wishes to thank The University of Memphis for support as a visiting professor during a portion of this work. The NMR analysis was supported by the College of Pharmacy at the University of Tennessee Health Science Center.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2008.11.059.

References and notes

1. Amar, M. B. *J. Ethnopharmacol.* **2006**, *105*, 1.
2. Jonsson, K.-O.; Holt, S.; Fowler, C. J. *Basic Clin. Pharmacol. Toxicol.* **2006**, *98*, 124.
3. Pacher, P.; Bátkai, S.; Kunos, G. *Pharmacol. Rev.* **2006**, *58*, 389.
4. Pertwee, R. G. *Br. J. Pharmacol.* **2006**, *147*, 163.
5. Pertwee, R. G. *Curr. Neuropharmacol.* **2004**, *2*, 9.
6. Contassot, E.; Tenan, M.; Schnüriger, V.; Pelte, M.-F.; Dietrich, P.-Y. *Gynecol. Oncol.* **2004**, *93*, 182.
7. Contassot, E.; Wilmotte, R.; Tenan, M.; Belkouch, M.; Schnuriger, V.; de Tribolet, N.; Burkhardt, K.; Dietrich, P. *J. Neuropathol. Exp. Neurol.* **2004**, *63*, 956.
8. Wiley, J. L.; Martin, B. R. *Chem. Phys. Lipids* **2002**, *121*, 57.
9. Onaivi, E. S.; Ishiguro, H.; Gong, J.-P.; Patel, S.; Perchuk, A.; Meozzi, P. A.; Myers, L.; Mora, Z.; Tagliagerro, P.; Gardner, E.; Brusco, A.; Akinshola, B. E.; Liu, Q.-R.; Teasenfitz, L.; Uhl, G. R. *Ann. N.Y. Acad. Sci.* **2006**, *1074*, 514.
10. Van Sickle, M. D.; Duncan, M.; Kingsley, P. J.; Mouihate, A.; Urbani, P.; Mackie, K.; Stella, N.; Makriyannis, A.; Piomelli, D.; Davison, J. S.; Marnett, L. J.; Di Marzo, V.; Pittman, Q. J.; Patel, K. D.; Sharkey, K. A. *Science* **2005**, *310*, 329.
11. Felder, C. C. *Annu. Rev. Pharmacol. Toxicol.* **1988**, *38*, 179.
12. Pertwee, R. G. *Curr. Med. Chem.* **1999**, *6*, 635.
13. D'Souza, D. C. *Int. Rev. Neurobiol.* **2007**, *78*, 289.
14. Ellert-Miklaszewska, A.; Grajkowska, W.; Gabrusiewicz, K.; Kaminska, B.; Konarska, L. *Brain Res.* **2007**, *1137*, 161.
15. Velasco, G.; Galve-Roperh, I.; Sánchez, C.; Blázquez, C.; Guzmán, M. *Neuropharmacology* **2004**, *47*, 315.
16. McAllister, S. D.; Chan, C.; Taft, R. J.; Luu, T.; Abood, M. E.; Moore, D. H.; Aldape, K.; Yount, G. *J. Neuro-Oncol.* **2005**, *74*, 31.
17. Duntsch, C.; Divi, M. K.; Jones, T.; Zhou, Q.; Krishnamurthy, M.; Boehm, P.; Wood, G.; Sills, A.; Moore, B. M., II *J. Neuro-Oncol.* **2006**, *77*, 143.
18. Sánchez, C.; de Ceballos, M. L.; del Pulgar, T. G.; Rueda, D.; Corbacho, C.; Velasco, G.; Galve-Roperh, I.; Huffman, J. W.; y Cajal, S. R.; Gúzman, M. *Cancer Res.* **2001**, *61*, 5784.
19. Hinz, B.; Ramer, R.; Eichele, K.; Weinzierl, U.; Brune, K. *Mol. Pharmacol.* **2004**, *66*, 1643.
20. Massi, P.; Vaccani, A.; Ceruti, S.; Colombo, A.; Abbracchio, M. P.; Parolaro, D. *J. Pharmacol. Exp. Ther.* **2004**, *308*, 838.
21. Carracedo, A.; Gironella, M.; Lorente, M.; Garcia, S.; Guzmán, M.; Velasco, G.; Iovanna, J. L. *Cancer Res.* **2006**, *66*, 6748.
22. Fogli, S.; Nieri, P.; Chicca, A.; Adinolfi, B.; Mariotti, V.; Iacopetti, P.; Breschi, M. C.; Pellegrini, S. *FEBS Lett.* **2006**, *580*, 1733.
23. Eubanks, L. M.; Rogers, C. J.; Beuscher, A. E., IV; Koob, G. F.; Olson, A. J.; Dickerson, T. J.; Janda, K. D. *Mol. Pharmacol.* **2006**, *3*, 773.
24. Gutierrez, T.; Farthing, J. N.; Zvonok, A. M.; Makriyannis, A.; Hohmann, A. G. *Br. J. Pharmacol.* **2007**, *150*, 153.
25. Hohmann, A. G. *Pain* **2005**, *118*, 3.
26. Ueda, Y.; Miyagawa, N.; Matsui, T.; Kaya, T.; Iwamura, H. *Eur. J. Pharmacol.* **2005**, *520*, 164.
27. Kim, K.; Moore, D. H.; Makriyannis, A.; Abood, M. E. *Eur. J. Pharmacol.* **2006**, *542*, 100.
28. McLaughlin, P. J.; Qian, L.; Wood, J. T.; Wisniecki, A.; Winston, K. M.; Swezey, L. A.; Ishiwari, K.; Betz, A. J.; Pandarinathan, L.; Xu, W.; Makriyannis, A.; Salamone, J. D. *Pharmacol. Biochem. Behav.* **2006**, *83*, 396.
29. Perna, F.; Bechi, P.; Perigli, G.; Manera, C.; Schiavone, N.; Papucci, L.; Magnelli, L.; Masini, M.; Lulli, M.; Donnini, M.; Capaccioli, S.; Cianchi, F. *Dig. Liver Dis.* **2006**, *38*, S107.
30. Teixeira-Clerc, F.; Julien, B.; Grenard, P.; Nhieu, J. T. V.; Deveaux, V.; Li, L.; Serriere-Lanneau, V.; Ledent, C.; Mallat, A.; Lotersztajn, S. *Nature Med.* **2006**, *12*, 671.
31. Powles, T.; te Poele, R.; Shamash, J.; Chaplin, T.; Propper, D.; Joel, S.; Oliver, T.; Liu, W. M. *Blood* **2005**, *105*, 1214.
32. Sarfaraz, S.; Afaq, F.; Adhami, V. M.; Mukhtar, H. *Cancer Res.* **2005**, *65*, 1635.
33. Tomida, I.; Pertwee, R. G.; Azuara-Blanco, A. *Br. J. Ophthalmol.* **2004**, *88*, 708.
34. Porcella, A.; Maxia, C.; Gessa, G. L.; Pani, L. *Eur. J. Neurosci.* **2001**, *13*, 409.
35. Thakur, G. A.; Duclos, R. I., Jr.; Makriyannis, A. *Life Sci.* **2005**, *78*, 454.
36. Bifulco, M.; Di Marzo, V. *Nat. Med.* **2002**, *8*, 547.
37. Di Marzo, V. *Trends Pharmacol. Sci.* **2006**, *27*, 134.
38. Howlett, A. C.; Mukhopadhyay, S.; Norford, D. C. *J. Neuroimmune Pharmacol.* **2006**, *1*, 305.
39. Pertwee, R. G. *AAPS J.* **2005**, *7*, E625.
40. Lambert, D. M.; Fowler, C. J. *J. Med. Chem.* **2005**, *48*, 5059.

41. Lange, J. H. M.; Kruse, C. G. *Drug Discovery Today* **2005**, *10*, 693.
42. Govaerts, S.; Muccioli, G.; Hermans, E.; Lambert, D. *Eur. J. Pharmacol.* **2004**, *495*, 43.
43. Lange, J.; Kruse, C. *Curr. Opin. Drug Discov. Devel.* **2004**, *7*, 498.
44. Palmer, S.; Thakur, G.; Makriyannis, A. *Chem. Phys. Lipids* **2002**, *121*, 3.
45. Pertwee, R. G.; Ross, R. A. *Prosta. Leukotr. Ess. Fatty Acids* **2002**, *66*, 101.
46. Papahatjis, D. P.; Kourouli, T.; Makriyannis, A. *J. Heterocycl. Chem.* **1996**, *33*, 559.
47. Eissenstat, M. A.; Bell, M. R.; D'Ambra, T. E.; Alexander, E. J.; Daum, S. J.; Ackerman, J. H.; Gruett, M. D.; Kumar, V.; Estep, K. G.; Olefirowicz, E. M.; Wetzel, J. R.; Alexander, M. D.; Weaver, J. D., III; Haycock, D. A.; Luttinger, D. A.; Casiano, F. M.; Chippari, S. M.; Kuster, J. E.; Stevenson, J. I.; Ward, S. J. *J. Med. Chem.* **1995**, *38*, 3094.
48. Fichera, M.; Cruciani, G.; Bianchi, A.; Musumarra, G. *J. Med. Chem.* **2000**, *43*, 2300.
49. Durdagi, S.; Kapou, A.; Kourouli, T.; Andreou, T.; Nikas, S. P.; Nahmias, V. R.; Papahatjis, D. P.; Papadopoulos, M. G.; Mavromoustakos, T. *J. Med. Chem.* **2007**, *50*, 2875.
50. Chen, J.-Z.; Han, X.-W.; Liu, Q.; Makriyannis, A.; Wang, J.; Xie, X.-Q. *J. Med. Chem.* **2006**, *49*, 625.
51. Salo, O. M. H.; Savinainen, J. R.; Parkkari, T.; Nevalainen, T.; Lahtela-Kakkonen, M.; Gynther, J.; Laitinen, J. T.; Poso, A. *J. Med. Chem.* **2006**, *49*, 554.
52. Keimowitz, A. R.; Martin, B. R.; Razdan, R. K.; Crocker, P. J.; Mascarella, S. W.; Thomas, B. F. *J. Med. Chem.* **2000**, *43*, 59.
53. Shim, J.-Y.; Collantes, E. R.; Welsh, W. J.; Subramaniam, B.; Howlett, A. C.; Eissenstat, M. A.; Ward, S. J. *J. Med. Chem.* **1998**, *41*, 4521.
54. Huffman, J.; Zengin, G.; Wu, M.; Lu, J.; Hynd, G.; Bushell, K.; Thompson, A.; Bushell, S.; Tartal, C.; Hurst, D.; Reggio, P.; Selley, D.; Cassidy, M.; Wiley, J.; Martin, B. *Bioorg. Med. Chem.* **2005**, *13*, 89.
55. Ferreira, A. M.; Moore, B. M., II; Krishnamurthy, M. *J. Mol. Struct. (Theochem)* **2004**, *674*, 131.
56. Famini, G. R.; Wilson, L. Y. *J. Phys. Org. Chem.* **1999**, *12*, 645.
57. Wilson, L. Y.; Famini, G. R. *J. Med. Chem.* **1991**, *34*, 1668.
58. Gancia, E.; Montana, J. G.; Manallack, D. T. *J. Mol. Graphics Modell.* **2001**, *19*, 349.
59. Bhattacharjee, A. K.; Kyle, D. E.; Vennerstrom, J. L.; Milhous, W. K. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1212.
60. Dixon, S.; Merz, K. M., Jr.; Lauri, G.; Ianni, J. C. *J. Comput. Chem.* **2005**, *26*, 23.
61. Netzeva, T. I.; Aptula, A. O.; Benfenati, E.; Cronin, M. T. D.; Gini, G.; Lessigiarska, I.; Maran, U.; Vračko, M.; Schüürmann, G. *J. Chem. Inf. Modell.* **2005**, *45*, 106.
62. Loader, R. J.; Singh, N.; O'Malley, P. J.; Popelier, P. L. A. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 1249.
63. Peters, M. B.; Raha, K.; Merz, K. M., Jr. *Curr. Opin. Drug Discov. Devel.* **2006**, *9*, 370.
64. Karelson, M.; Lobanov, V. S.; Katritzky, A. R. *Chem. Rev.* **1996**, *96*, 1027.
65. Lu, D.; Meng, Z.; Thakur, G. A.; Fan, P.; Steed, J.; Tartal, C. L.; Hurst, D. P.; Reggio, P. H.; Deschamps, J. R.; Parrish, D. A.; George, C.; Järbe, T. U. C.; Lamb, R. J.; Makriyannis, A. *J. Med. Chem.* **2005**, *48*, 4576.
66. Raitio, K. H.; Salo, O. M. H.; Nevalainen, T.; Poso, A.; Järvinen, T. *Curr. Med. Chem.* **2005**, *12*, 1217.
67. Xie, X.-Q.; Chen, J.-Z.; Billings, E. M. *Prot. Struct. Funct. Genet.* **2003**, *53*, 307.
68. Honório, K. M.; da Silva, A. B. F. *J. Mol. Modell.* **2005**, *11*, 200.
69. Krishnamurthy, M.; Ferreira, A. M.; Moore, B. M., II *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3487.
70. Neter, J.; Wasserman, W.; Kutner, M. H. *Applied Linear Regression Models*, 2nd ed.; Richard D. Irwin: Homewood, IL, 1989.
71. Allen, D. M. *Technometrics* **1974**, *16*, 125.
72. Akaike, H. *IEEE Trans. Automat. Control* **1974**, *19*, 716.
73. Maindonald, J.; Braun, W. J. *Data Analysis and Graphis Using R: An Example-Based Approach*, 2nd ed.; Cambridge University Press, 2007.
74. Wood, S. N. *J. Roy. Stat. Soc. B* **2003**, *65*, 95.
75. SYBYL 7.3; Tripos International: 1699 South Hanley Road, St. Louis, Missouri, 63144, USA.
76. SPARTAN'*02*; Wavefunction: Irvine, California, 2002.
77. R Development Core Team. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2006.
78. Geisen, D. J.; Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chim. Acta* **1997**, *98*, 85.
79. Ghose, A. K.; Crippen, G. M. *Int. J. Quantum Chem.* **1985**, *28*, 315.
80. Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735.
81. Spartan computational FAQ page. http://www.wavefun.com/support/sp_compfaq/Properties_FAQ.html (June 2006).
82. Mager, P. P. *J. Chemometr.* **1995**, *9*, 211.
83. Winkler, D. A. *Briefings Bioinform.* **2002**, *3*, 73.
84. Battacheree, H.; Moore, B. M., II unpublished data.
85. Moore, B., II et al. Bioorg. Med. Chem., submitted for publication.