

## Pharmacophore Models of Paclitaxel- and Epothilone-Based Microtubule Stabilizing Agents

Sangbae Lee,<sup>†</sup> Yuno Lee,<sup>‡</sup> James M. Briggs,<sup>§,\*</sup> and Keun Woo Lee<sup>‡,\*</sup>

<sup>†</sup>Department of Immunology, Beckman Research Institute of the City of Hope, CA 91010, USA

<sup>‡</sup>Division of Applied Life Science (BK21 Program), Systems and Synthetic Agrobiotech Center (SSAC),

Plant Molecular Biology and Biotechnology Research Center (PMBBRC), Research Institute of Natural Science (RINS),

Gyeongsang National University (GNU), Jinju 660-701, Korea. \*E-mail: kwlee@gnu.ac.kr

<sup>§</sup>Department of Biology and Biochemistry, University of Houston, TX 77204-5001, USA. \*E-mail: jbriggs@uh.edu

Received February 25, 2013, Accepted April 1, 2013

Microtubules play an important role in intracellular transport, mobility, and particularly mitosis. Paclitaxel (Taxol™) and paclitaxel-like compounds have been shown to be anti-tumor agents useful for various human tumors. Paclitaxel-like compounds operate by stabilizing microtubules through interface binding at the interface between two  $\beta$ -tubulin monomers in adjacent protofilaments. In this paper we present the elucidation of the structural features of paclitaxel and paclitaxel-like compounds (*e.g.*, epothilones) with microtubule stabilizing activities, and relate their activities to spatial and chemical features of the molecules. CATALYST program was used to generate three-dimensional quantitative structure activity relationships (3D-QSARs) resulting in 3D pharmacophore models of epothilone- and paclitaxel-derivatives. Pharmacophore models were generated from diverse conformers of these compounds resulting in a high correlation between experimental and predicted biological activities ( $r = 0.83$  and  $0.91$  for epothilone and paclitaxel derivatives, respectively). On the basis of biological activities of the training sets, five- and four-feature pharmacophore hypotheses were generated in the epothilone and paclitaxel series. The validation of generated hypotheses was achieved by using twelve epothilones and ten paclitaxels, respectively, which are not in the training sets. The clustering (grouping) and merging techniques were used in order to supplement spatial restrictions of each of hypothesis and to develop more comprehensive models. This approach may be of use in developing novel inhibitor candidates as well as contributing a better understanding of structural characters of many compounds useful as anticancer agents targeting microtubules.

**Key Words :** Microtubule, Epothilone, Paclitaxel, Pharmacophore, Drug design

### Introduction

Tubulin, a heterodimeric protein composed of globular  $\alpha$ - and  $\beta$ -tubulin subunits, represents the monomeric building block of microtubules. Tubulin polymerizes end to end creating a protofilament comprised of non-covalently interacting  $\alpha$ - and  $\beta$ -tubulin subunits (... $\alpha/\beta\alpha/\beta\alpha/\beta...$ ). Thirteen protofilaments come together laterally forming a closed tube, called a microtubule that is essentially crucial for many cellular processes, such as rigidifying the cell and translocation of vesicles, granules, organelles, and chromosomes through special attachment proteins.<sup>2-6</sup> The process of cell division makes use of microtubules through their ability to form the mitotic spindle. This results from the dynamic (dis)assembly process of depolymerization and repolymerization.<sup>7</sup> Microtubules are “dynamically unstable” – tubulin heterodimers continuously attach and detach from the protofilament termini. Detached tubulin heterodimers can be used to form the mitotic spindle, required for mitosis. Rapidly dividing cells, *e.g.*, bone marrow, skin, and carcinogenic, are most sensitive to interference of the dynamic instability of microtubules.

With blocking microtubule assembly, many natural prod-

ucts such as colchicines,<sup>8</sup> rhizoxin,<sup>9</sup> vinblastine,<sup>10</sup> and vincristine<sup>11</sup> bind tubulin, thereby interfering with the production of the mitotic spindle by destabilizing microtubules. In contrast, paclitaxel<sup>12-14</sup> and epothilones<sup>15-17</sup> hyperstabilize microtubules by binding to  $\beta$ -tubulin/ $\beta$ -tubulin interfaces in adjacent protofilaments resulting in a reduced ability to form the mitotic spindle due to a diminished pool of disassociated tubulin heterodimers.<sup>1,18</sup> The most well known tubulin-stabilizing agent is paclitaxel (Taxol™),<sup>19</sup> which is currently a front-line anticancer agent. However, paclitaxel is not an ideal drug because of its susceptibility to Multi-Drug Resistance (MDR) and its low aqueous solubility, leading to difficulties in administration/formulation and toleration.<sup>20-24</sup>

Recently, several non-taxoid compounds including epothilones have been shown to stabilize microtubules. Epothilone A and B, which are more soluble than paclitaxel, can tightly bind to the  $\beta$ -tubulin subunits and inhibit cell division by hyperstabilizing microtubules so that they can not readily dissociate to form the mitotic spindle.<sup>25-31</sup> Even though a number of selective microtubule-stabilizing agents have recently described and their high-affinity binding sites have been detected on the tubulin protein, their complexed molecular structures are not fully understood except X-ray crystal

structure of taxotere.<sup>32</sup>

Therefore, in order to better understand the pharmacophore elements for epothilones and paclitaxels, and to provide guidelines for creating new drug candidates, a three-dimensional quantitative structure-activity relationship (3D-QSAR) approach, as implemented in CATALYST,<sup>33</sup> was employed in this study. The aim of this study is the generation of hypotheses (3D-pharmacophore models) in which experimental biological activities are correlated with the 3D-arrangement of chemical features of the compounds. CATALYST is used to generate hypotheses using the 3D distribution and combination of structural features. These hypotheses are used to predict the biological activities of test sets of compounds not included in the training sets in order to allow the rank ordering of synthetic priorities. In addition, they were used as a 3D-query to search compound databases, such as, NCI and ACD in order to identify new candidate drugs.<sup>34</sup>

## Materials and Methods

**Software.** In order to build a 3D-QSAR model and to generate pharmacophore, computational studies were carried out using Accelrys/CATALYST version 4.6.<sup>33</sup>

**Selecting Training Sets for Epothilone and Paclitaxel Derivatives.** In order to generate pharmacophore models of these microtubule anticancer agents, training sets were divided into two representative series: one for epothilones and one for paclitaxels. For the epothilone derivatives, a training set of 37 compounds, evaluated for cytotoxicity against an ovarian cancer cell line (1A9), was constructed using the SKETCHER module in CATALYST. Each member of the training set was energy minimized to the closest local minimum using the CHARMM forcefield.<sup>35</sup> For the paclitaxel derivatives, the  $\beta$ -tubulin complex with docetaxel (PDB code: ITUB)<sup>36</sup> was used as the initial template structure for the training set of 25 compounds targeting non-small cell lung carcinoma (A549). The structure of the  $\beta$ -tubulin/docetaxel complex was minimized using the CVFF (Consistent Valence Forcefield) forcefield with DISCOVER\_3/InsightII.<sup>37</sup> The minimized docetaxel structure was replaced with paclitaxel and minimized as described above.

**Generating Pharmacophores with CATALYST.** In the CATALYST program, all conformers of each member of the training sets were generated using the “Best generation method” of the ConFirm module, which created a maximum of 255 conformations within an internal energy range of 20 kcal/mol for each molecule.<sup>35,38</sup> These conformers were built by using the Poling algorithm.<sup>38-40</sup> Hypotheses, ten in total, were generated after considering features (chemical functions) such as hydrogen-bond donors (HBDs), hydrogen-bond acceptors (HBAs), and aliphatic and aromatic hydrophobic (*ali-P* and *aro-P*).<sup>41-44</sup>

The predictive 3D-pharmacophore models were developed with the HypoGen module in CATALYST, which generated top ten scoring hypothesis and consisted of three phases: (1) the constructive phase in which the hypothesis is generated

with the most active compounds, (2) the subtractive phase, in which some of the hypotheses fitting the inactive or weakly active compounds are removed, and (3) the optimization phase, in which all hypotheses are subjected to perturbations such as adding a new feature or deleting an existing feature, and rotating the vectors attached to the features. HypoGen to investigate the trends and variations in biological activity with the chemical structure through investigating the 3D arrangement of the chemical features in the training set compounds.

The parameters for generating hypotheses were kept at the following values: weight variation, 0.302; mapping coefficient, 0; spacing, 2.97 Å; activity uncertainty, 3; and tolerance factor, 1. The weight variation represents the expected standard deviation of the feature weight for good hypothesis. The mapping coefficient is a penalty for hypotheses for which training set molecules that are topologically similar map in different ways to the hypothesis. The default value of 0 means that it is not used in hypothesis generation. The spacing indicates the minimum distance between actual feature of hypothesis. An uncertainty of 3 means that biological activity is existed between activity/3 and activity\*3. The tolerance factor is used to select the variable tolerance mode of HypoGen. The default value of 0 means that it uses standard HypoGen mode with no variable tolerance. Setting this value to 1 allows individual feature tolerance during optimization.

**Analysis and Refinement of Generated Pharmacophores.** The quality of the generated hypotheses was evaluated by the cost factor analysis to determine the best hypothesis with the appropriate conformers after generation of all ten hypotheses. The lowest cost hypothesis (1<sup>st</sup> hypothesis) was considered to be the best due to the nature of the cost factor analysis method (*i.e.* the highest cost difference, lowest RMS error, and the best correlation coefficient) and is acknowledged to possess the features representative of all of the hypotheses. After generating hypotheses, three-term cost factors were analyzed.<sup>45</sup> All cost factors of a hypothesis were calculated by summing these three component cost factors. The units of cost are binary bits. The hypothesis costs were calculated according to the number of bits to describe a hypothesis. Simpler hypotheses require fewer bits for a complete description, and this assumption is made that simpler hypotheses are better.

$$\text{cost} = eE + wW + cC \quad (1)$$

The weight cost (*W*) represented in a Gaussian form is expressed by the feature weight deviated from an idealized value of 2. The configuration cost (*C*) is described by the complexity of the hypothesis in the form of the exponent of the entropy. The error cost (*E*), which acts major role in decision of overall cost of a hypothesis among three costs, is computed from the deviation between the predicted and experimental activities of the training set compounds.

Two important theoretical hypotheses, the fixed cost hypothesis and the null hypothesis, were computed to assist in the evaluation of each of the ten hypotheses. The fixed cost

hypothesis ( $H_{\text{fixed}}$ ), in which the error cost is minimal and the slope of the activity correlation line was one, was a lower bound on the cost of the simplest possible hypothesis that still fitted to the data perfectly.

$$\text{fixed cost} = eE(x=0) + wW(x=0) + cC \quad (2)$$

The null cost hypothesis ( $H_{\text{null}}$ ) presumes that there is no statistically significant structure in the data and that the experimental activities are normally distributed about their mean. The error cost in the null hypothesis is higher than that in the fixed cost hypothesis and the slope of the activity correlation line in the null hypothesis is zero.  $H_{\text{null}}$  represents the highest cost of a pharmacophore model with no features and estimates the activity to be the average activity of the training set molecules.

$$\text{null cost} = eE(x_{\text{est}} = \bar{x}) \quad (3)$$

The measure of the difference between a null and a fixed cost hypothesis is important in the analysis of the results of the hypothesis generation. The greater the difference in the cost between two hypotheses, the higher the probability for finding useful models. One point that has to be emphasized is that the features of the generated hypothesis may be edited manually by distance constraint, angle constraint, etc. A meaningful hypothesis may result when the difference between two values is large; a value of 40-60 bits for a hypothesis may indicate that it has 75-90% probability for correlating the data.<sup>33</sup>

**Validation of Generated Pharmacophores.** In order to validate pharmacophore models generated from the training set of compounds, we applied the models to test sets of compounds (*i.e.* ones that were not included in the initial training set) in order to judge whether the ranked order of activities is accurately predicted for the test sets of compounds. To test how well the compounds matched a proposed hypothesis, we used the ‘Compare/Fit’ function that evaluates the geometry of the conformers of the compounds. The Fit function is expressed by a numerical value that quantifies the fitting capacity between each compound and generated hypothesis. Perfect mapping indicates that a fit value can be expressed by the sum of the weight of the features in the hypothesis. A higher fit value represents a better fit between the functional groups of the molecule to the appropriate features of the pharmacophore.

$$\text{Fit} = \sum_{\substack{\text{mapped} \\ \text{Hypofunctions}}} w \left[ 1 - \sum_{\text{spheres}} \left( \frac{\text{Disp}}{\text{Tol}} \right)^2 \right] \quad (4)$$

where  $w$  = adaptively determined weight of the hypothesis function,  $\text{Disp}$  = the displacement of the feature from the center of the location constraint, and  $\text{Tol}$  = the radius of the location constraint sphere for the feature (tolerance). Using fast and best fit procedures of Compare/Fit, the test set was fitted to a CATALYST hypothesis to predict the biological activities. The test set consisted of 12 compounds for the epothilone derivatives and of 10 compounds for the paclitaxel derivatives. The fast fit algorithm identifies the fit of

the compound to a proposed hypothesis without minimizing the conformers, whereas the best fit algorithm starts with the fast fit and then minimizes with the CHARMM forcefield. The best fit algorithm, while more time consuming, may sample a larger range of conformational space due to the minimization of the distance between the hypothesis features and the atoms to which they are meant to fit. The predictions obtained from the fast and best fit algorithms were compared *via* a residual that was computed from the difference (in log units) between the predicted and experimental activities. In general, a predicted activity within 1 log unit of the experimental activity is considered to be a valid prediction.<sup>46</sup>

CATALYST supports clustering and merging of generated hypotheses. By clustering all of the hypotheses, it is possible to divide the hypotheses into several groups (clusters) based on common features. It can also be used to obtain a new hypothesis by merging the analogous series of hypotheses.

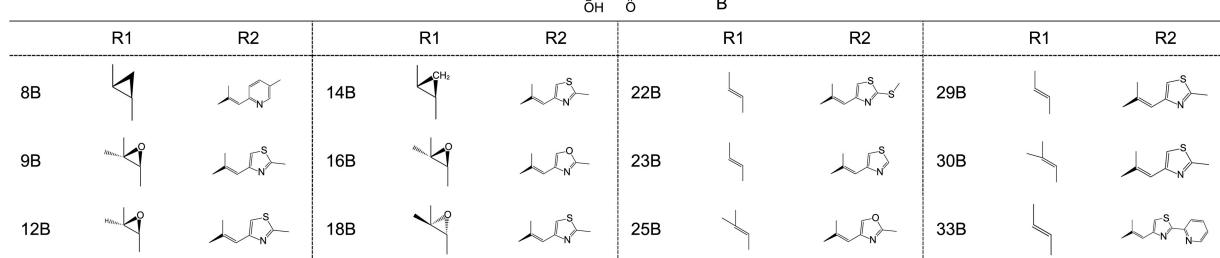
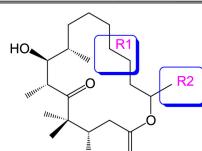
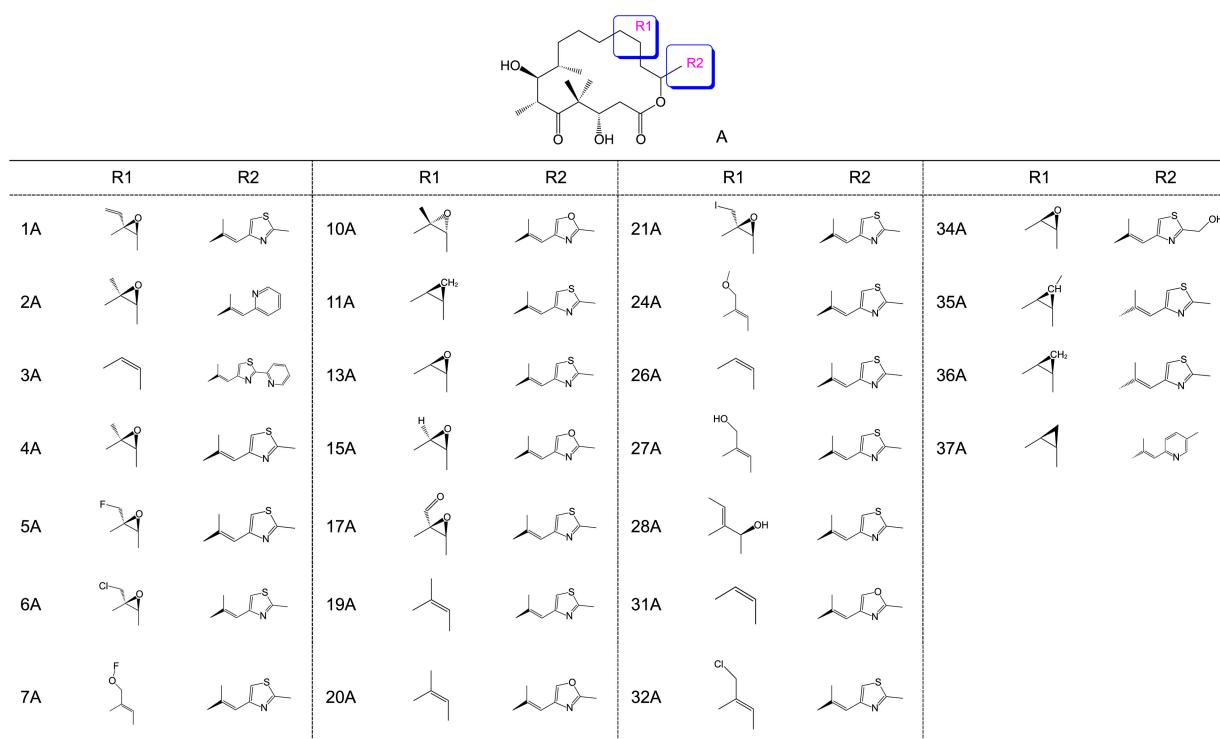
These pharmacophore models may be used to search 3D databases for structurally novel candidate molecules. There seems to be two main objects to screen such databases: first is for validating generated hypotheses, and second is for discovering more potential hit compounds than compounds of training set or test set. Two algorithms can be selected for database search: Fast and Best Flexible Search algorithms.<sup>33</sup> The Fast Flexible Search algorithm considers existing conformation precalculated during the search. While on the other, the Best Flexible Search algorithm takes transforming the conformation into account within internal energy threshold of 9.5 kcal/mol as default. Our database searching strategy was performed using the “Best Flexible Search” method that manipulates the conformers to minimize the distances between pharmacophore features and mapped atoms in the model. The lowest cost hypothesis and newly merged hypothesis were used as the query, respectively, against three databases in order to discover and compare epothilone-like hit lists: Maybridge, NCI, and ACD. These 3D databases are supplied with CATALYST version 4.6. In a CATALYST-formatted database, compounds are stored as multiple conformers that are representative of their available conformational space.<sup>37,38</sup> In our study, ACD, NCI and Maybridge database screening were used for designing new hit lists of microtubule stabilizing agents.

## Results and Discussion

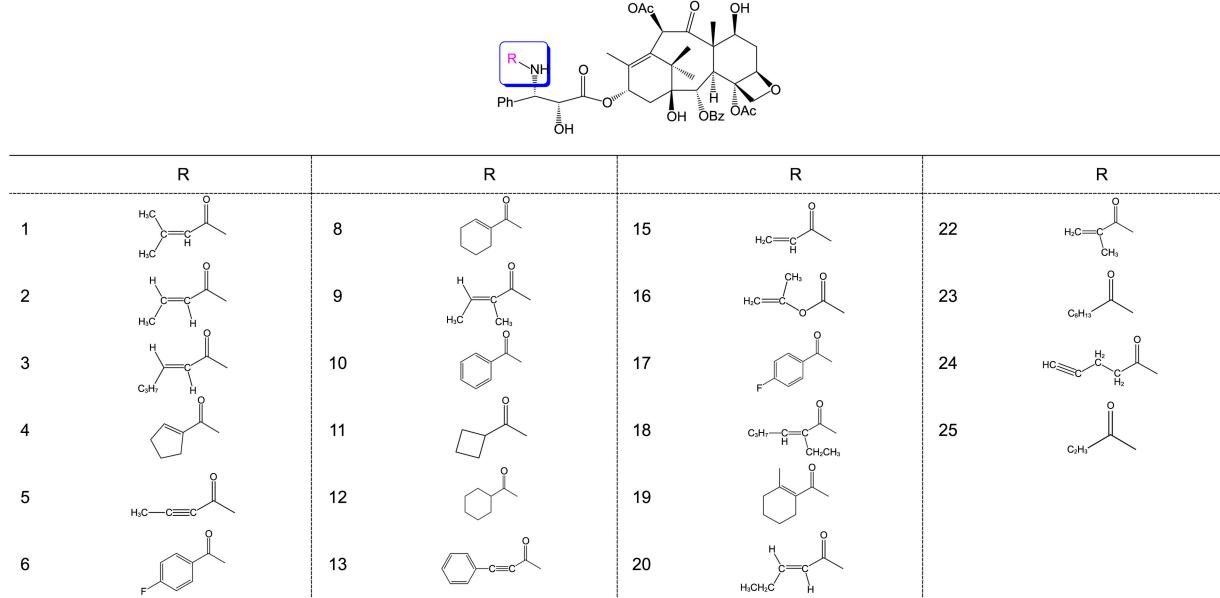
The ConFirm and HypoGen modules in CATALYST v. 4.6 were used to develop predictive models correlating biological activity and 3D structures for epothilone and paclitaxel derivatives. These results will be used to construct training sets and test sets and to generate their pharmacophore models.

**Pharmacophore Generation for the Epothilone and Paclitaxel Derivatives.** In order to define the pharmacophores that aim to identify the best 3D arrangement of chemical features explaining the activity variation among derivatives, two groups of training sets based on epothilone- and paclitaxel moiety were selected. One containing 37

(a)



(b)

**Figure 1.** (a) Chemical structures of training set for the epothilone derivatives. (b) Chemical structures of training set for the paclitaxel derivatives.

**Table 1.** (a) Experimental and predicted activities of training set calculated on the lowest cost hypothesis ( $H_1$ ) for the epothilone derivatives

Cmpd. no.	Fit	Activities ( $IC_{50}$ , nM)		Activity Scale		
		EA <sup>a</sup>	PA <sup>b</sup>	EA	PA	
1A	10.56	0.05	0.023	+++	+++	(-0.34) <sup>c</sup>
2A	8.68	0.1	1.7	+++	+++	(1.23)
3A	9.39	0.12	0.33	+++	+++	(0.44)
4A	9.75	0.13	0.15	+++	+++	(0.06)
5A	9.01	0.2	0.8	+++	+++	(0.60)
6A	8.52	0.4	2.4	+++	+++	(0.78)
7A	7.91	0.5	10	+++	++	(1.30)
8B	8.39	0.62	3.4	+++	+++	(0.74)
9B	9.47	1	0.28	+++	+++	(-0.55)
10A	9.39	1.5	0.33	+++	+++	(-0.66)
11A	8.38	1.6	3.4	+++	+++	(0.33)
12B	7.95	2	9.2	+++	+++	(0.66)
13A	8.67	2.2	1.8	+++	+++	(-0.09)
14B	8.43	2.7	3	+++	+++	(0.05)
15A	8.09	3	6.6	+++	+++	(0.34)
16B	7.83	3.5	12	+++	++	(0.54)
17A	8.25	5	4.5	+++	+++	(-0.05)
18B	7.91	6	10	+++	++	(0.22)
19A	8.30	6.5	4.1	+++	+++	(-0.20)
20A	7.92	8	9.7	+++	+++	(0.06)
21A	7.90	10	10	++	++	(0)
22B	7.83	15	12	++	++	(-0.10)
23B	6.81	20	130	++	+	(0.81)
24A	7.48	30	27	++	++	(-0.05)
25B	7.91	32	10	++	++	(-0.51)
26A	7.50	32	26	++	++	(-0.09)
27A	7.84	40	12	++	++	(-0.52)
28A	7.37	50	35	++	++	(-0.16)
29B	7.51	60	25	++	++	(-0.38)
30B	7.72	61	16	++	++	(-0.58)
31A	7.81	68	13	++	++	(-0.72)
32A	7.98	80	8.5	++	+++	++
33B	6.86	82	110	++	+	(0.13)
34A	7.87	100	11	+	++	(-0.96)
35A	6.69	180	170	+	+	(-0.03)
36A	7.41	230	32	+	++	(0.86)
37A	6.92	300	97	+	++	(0.51)

<sup>a</sup>+++: highly active compounds<sup>++</sup>: moderately active compounds<sup>+</sup>: weakly active compounds<sup>a</sup> and <sup>b</sup> represent experimental activity and predicted activity, respectively. The value of parenthesis. <sup>c</sup>means residual calculated by the difference of log unit of predicted activity and experimental activity.

epothilone derivatives and another containing 25 paclitaxel derivatives were prepared as shown in Figure 1. The activity values for the two training sets enabled the construction of pharmacophores for these classes of microtubule stabilizing agents. The activities of the compounds in the epothilone training set are  $IC_{50}$  and those of paclitaxel derivatives are  $ED_{50}/ED_{50}(\text{paclitaxel})$  and range five and six orders of magnitude, respectively, as shown in Table 1.

**Table 1.** (b) Experimental and predicted activities of training set calculated on the lowest cost hypothesis ( $H_1$ ) for the paclitaxel derivatives

Cmpd. no.	Fit	Activities ( $ED_{50}/ED_{50}(\text{paclitaxel})$ )		Activity Scale		
		EA <sup>a</sup>	PA <sup>b</sup>	EA	PA	
1	10.61	0.008	0.043	+++	+++	(0.73) <sup>c</sup>
2	10.60	0.008	0.044	+++	+++	(0.74)
3	10.47	0.0083	0.06	+++	+++	(0.86)
4	9.32	0.05	0.84	+++	+++	(1.23)
5	10.54	0.067	0.051	+++	+++	(-0.12)
6	9.08	0.39	1.5	+++	+++	(0.59)
7	9.24	0.4	1	+++	+++	(0.40)
8	8.76	0.56	3.1	+++	+++	(0.74)
9	10.56	1	0.048	+++	+++	(-1.32)
10	9.93	1	0.2	+++	+++	(-0.70)
11	8.73	1.2	3.3	+++	+++	(0.44)
12	8.71	4.4	3.5	+++	+++	(-0.10)
13	7.60	6	45	+++	++	(0.88)
14	8.43	10	6.5	++	+++	(-0.19)
15	8.50	11	5.6	++	+++	(-0.29)
16	8.09	11	14	++	++	(0.11)
17	7.68	29	37	++	++	(0.11)
18	8.04	39	16	++	++	(-0.39)
19	7.91	79	21	++	++	(-0.58)
20	7.67	100	38	+	++	(-0.42)
21	7.03	180	160	+	+	(-0.05)
22	7.55	200	50	+	++	(-0.60)
23	7.72	340	34	+	++	(-1.00)
24	7.71	400	35	+	++	(-1.06)
25	6.40	740	690	+	+	(-0.03)

After generating all ten hypotheses for epothilone and paclitaxel derivatives, the first hypothesis ( $H_1$ ) was selected as the best one as a result of cost factor analysis. The results of this cost factor analysis are given in Table 2. In the epothilone derivatives, the value of the total cost of the hypothesis generated ( $H_1$ ), the null hypothesis, and the fixed hypothesis were 170.30, 210.33, and 143.50 bits, respectively. The corresponding values for the paclitaxel derivatives were, however, 134.68, 208.61, and 102.40 bits. As a result, the difference between the cost of  $H_1$  and  $H_{\text{null}}$  ( $\Delta(H_{\text{null}} - H_1)$ ) and the difference between  $H_{\text{fixed}}$  and  $H_{\text{null}}$  ( $\Delta(H_{\text{null}} - H_{\text{fixed}})$ ) were 40.03 and 66.83 bits for the epothilone derivatives and 73.93 and 106.21 for the paclitaxels. The predicted activity values along with the experimental values for the paclitaxel derivatives demonstrate better correlation ( $r = 0.91$ ) than that of the epothilone derivatives ( $r = 0.83$ ), as presented in Table 2. Their correlation coefficient ( $r$ ) measurement supports the assessment of the quality of model. The regressions between experimental and predicted activities are plotted in log-scale in Figure 2.

In the epothilone derivatives, the lowest cost hypothesis ( $H_1$ ) possessed five features including three *ali*-Ps and two HBAs. On the other hand, four features with an *ali*-P, a HBA, and two HBDs were found in the paclitaxel derivatives as

**Table 2.** (a) Cost factor analysis measured in bits for 10 hypotheses of the epothilone derivatives

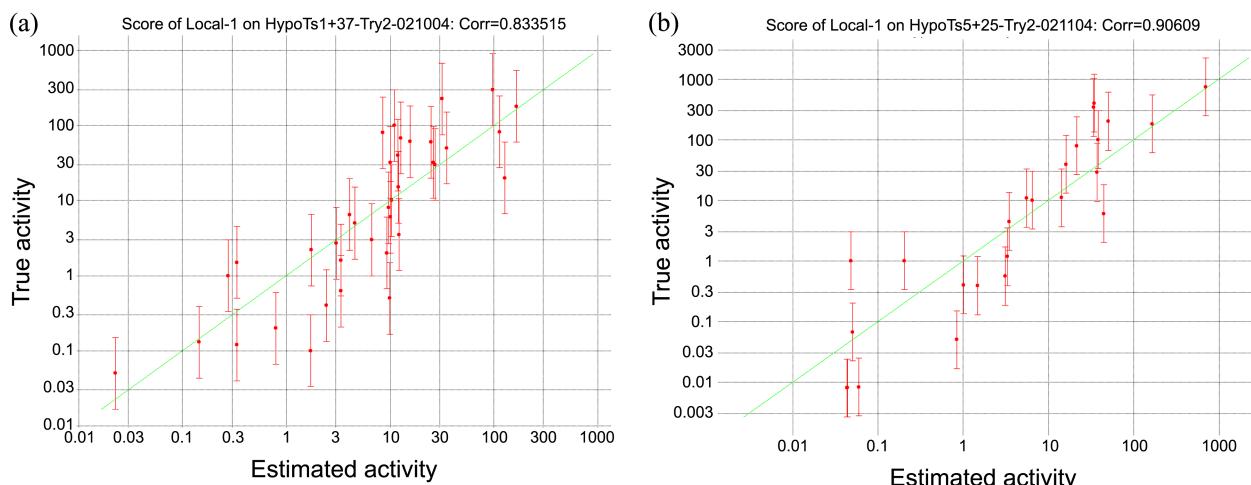
Hypothesis No. <sup>a</sup>	Total cost	$H_{\text{null}} - H_{\text{gen}}$	Error cost	RMS	Correlation (r)	Features <sup>b</sup>			
1	<b>170.30</b>	<b>40.03</b>	<b>150.73</b>	<b>1.19</b>	<b>0.83</b>	<b>HBA,</b>	<b>HBA,</b>	<b>ali-P,</b>	<b>ali-P,</b>
2	178.82	31.51	159.12	1.37	0.77	HBA,	HBA,	ali-P,	ali-P,
3	179.26	31.07	159.08	1.37	0.77	HBA,	HBD,	HBD,	ali-P
4	179.57	30.76	160.43	1.39	0.76	HBA,	HBA,	ali-P,	ali-P,
5	179.67	30.66	158.26	1.35	0.78	HBA,	HBD,	HBD,	ali-P
6	180.96	29.37	161.67	1.42	0.75	HBA,	HBA,	ali-P,	ali-P,
7	182.73	27.60	163.55	1.45	0.74	HBA,	HBA,	ali-P,	ali-P,
8	183.91	26.42	164.36	1.47	0.73	HBA,	HBA,	HBA,	ali-P
9	185.50	24.83	164.93	1.48	0.73	HBA,	HBA,	HBD,	ali-P
10	185.86	24.47	165.84	1.50	0.72	HBA,	HBA,	HBA,	ali-P

<sup>a</sup>Null cost = 210.33, Fixed cost = 143.50, Configuration cost = 17.93. All cost units are in bits. <sup>b</sup>HBA, hydrogen-bond acceptor; HBD, hydrogen-bond donor; ali-P, aliphatic hydrophobic region.

**Table 2.** (b) Cost factor analysis measured in bits for 10 hypotheses of the paclitaxel derivatives

Hypothesis No. <sup>a</sup>	Total cost	$H_{\text{null}} - H_{\text{gen}}$	Error cost	RMS	Correlation (r)	Features <sup>b</sup>			
1	<b>134.68</b>	<b>73.93</b>	<b>107.92</b>	<b>1.38</b>	<b>0.91</b>	<b>HBA,</b>	<b>HBD,</b>	<b>HBD,</b>	<b>ali-P</b>
2	140.97	67.64	117.89	1.64	0.86	HBA,	HBD,	HBD,	ali-P
3	141.38	67.23	116.47	1.61	0.87	HBA,	HBD,	HBD,	ali-P
4	141.69	66.92	119.44	1.68	0.85	HBA,	HBD,	HBD,	ali-P
5	142.38	66.22	116.55	1.61	0.87	HBA,	HBD,	HBD,	ali-P
6	143.07	65.54	120.54	1.71	0.84	HBA,	HBD,	HBD,	ali-P
7	143.22	65.39	118.95	1.67	0.85	HBA,	HBD,	HBD,	ali-P
8	146.78	61.83	125.81	1.83	0.82	HBD,	HBD,	ali-P,	aro-P
9	146.97	61.64	124.94	1.81	0.82	HBA,	HBD,	HBD,	ali-P
10	150.80	57.81	129.75	1.91	0.80	HBA,	HBD,	HBD,	ali-P

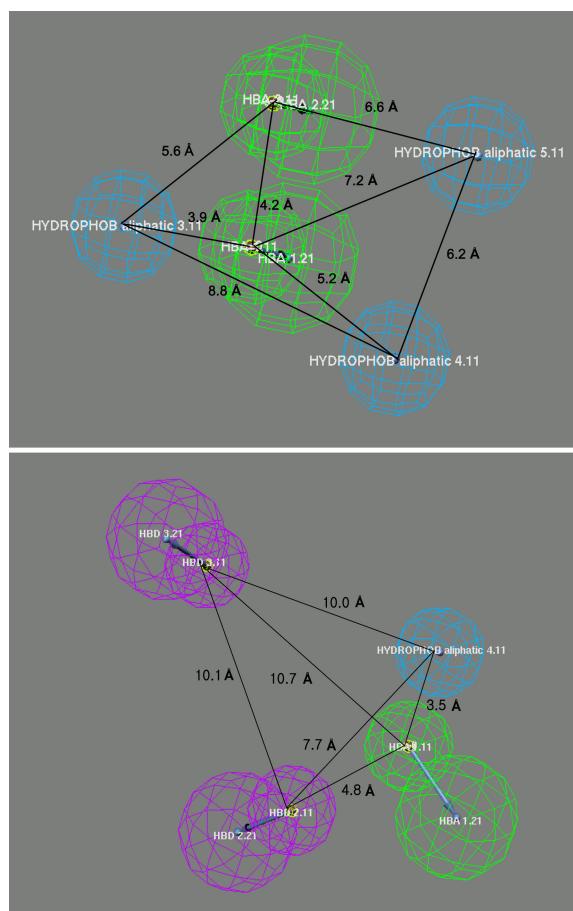
<sup>a</sup>Null cost = 208.61, Fixed cost = 102.40, Configuration cost = 17.19. All cost units are in bits. <sup>b</sup>HBA, hydrogen-bond acceptor; HBD, hydrogen-bond donor; ali-P, aliphatic hydrophobic region; aro-P, aromatic hydrophobic region.

**Figure 2.** Graph regression displaying the experimental versus predicted activity ( $IC_{50}$ ) obtained with best significant hypothesis for the epothilone and for the paclitaxel derivatives.

shown in Figure 3. Three dimensional coordinates and inter-feature distances of  $H_1$  for the epothilone and paclitaxel derivatives are given in Table 3. Three ali-P regions and a second HBA are located at distances of 3.9 Å, 5.2 Å, 7.2 Å, and 4.2 Å from the first HBA in the pharmacophore model derived from epothilone derivatives. Similarly, an ali-P region,

and two HBDs are located at distances of 3.5 Å, 4.8 Å, and 10.7 Å, respectively, from the first HBA region for the paclitaxel derivatives.

As a further verification of the generated pharmacophore models (Fig. 4(a)), the parent compound for each class of compounds was fitted to  $H_1$ . Epothilone B ( $IC_{50} = 0.13$  nM)



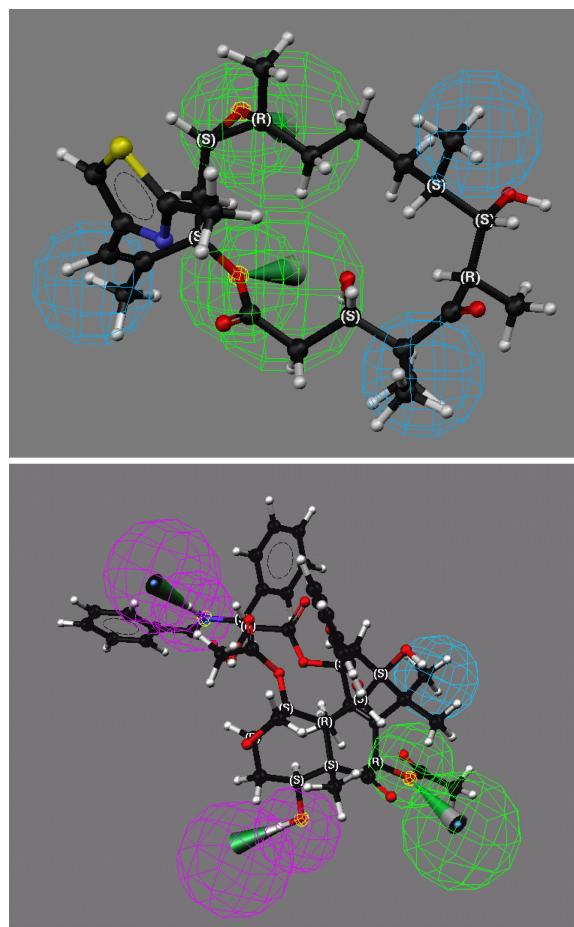
**Figure 3.** The distances between pharmacophore features. (a) for epothilone derivatives targeting cell line 1A9 and (b) for paclitaxel derivatives targeting cell line A549. Shown are an aliphatic hydrophobic sphere (cyan), a hydrogen-bond acceptor (green), and two hydrogen-bond donors (magenta).

and paclitaxel ( $ED_{50}/ED_{50(\text{paclitaxel})} = 1$ ), typical microtubule stabilizing agents, were individually fitted to each lowest cost hypothesis ( $H_1$ ) on behalf of epothilone and paclitaxel derivatives, as shown in Figure 4.

In the pharmacophore model derived from epothilone

**Table 3.** (a) Coordinates and inter-feature distances of the lowest cost hypothesis ( $H_1$ ) for the epothilone derivatives

	HBA →	HBA →	ali-P	ali-P	ali-P
Weights	2.31	2.31	2.31	2.31	2.31
Tolerances	1.60	2.20	1.60	2.20	1.60
Coords X	-0.82	0.35	-0.99	0.17	-4.63
Coords Y	-1.47	-1.02	2.69	3.44	-1.45
Coords Z	-0.91	-3.63	-0.77	-3.44	-1.75
HBA ↓					
	3.0 Å				
HBA ↓	4.2 Å	4.9 Å			
	5.6 Å	4.5 Å	3.0 Å		
ali-P	3.9 Å	5.3 Å	5.6 Å	7.0 Å	
ali-P	5.2 Å	5.2 Å	8.0 Å	8.4 Å	8.8 Å
ali-P	7.2 Å	7.7 Å	6.6 Å	7.4 Å	10.8 Å
					6.2 Å

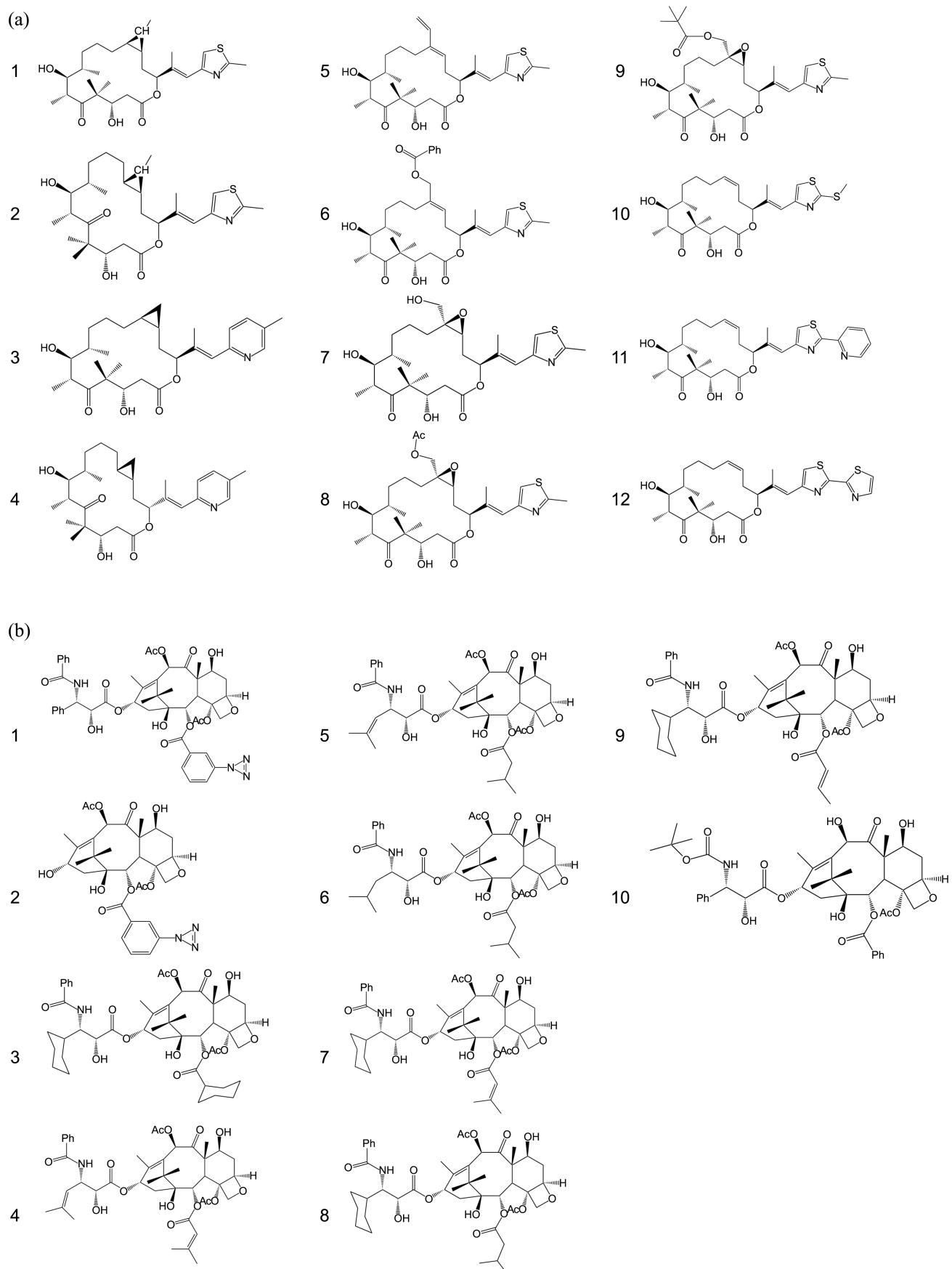


**Figure 4.** Epothilone B and Paclitaxel fitted to CATALYST pharmacophore generated from Table 1 and Figure 2. Shown are three aliphatic hydrophobic spheres (cyan), and two hydrogen-bond acceptors (green).

derivatives, the two HBA regions match the epoxide group of C12-C13, which is not essential for tubulin polymerization but is implicated in cytotoxicity and the ether group between C1 and C15 whose stereochemistry is strongly correlated to its activity.<sup>1,47</sup> Two of the three ali-P regions are

**Table 3.** (b) Coordinates and inter-feature distances for the lowest cost hypothesis ( $H_1$ ) for the paclitaxel derivatives

	HBA →	HBD →	HBD →	ali-P
Weights	3.24	3.24	3.24	3.24
Tolerances	1.60	2.20	1.60	2.20
Coords X	4.81	7.45	2.82	0.85
Coords Y	0.01	-1.38	-3.20	-5.47
Coords Z	-1.96	-2.28	0.96	0.97
HBA ↓				
	3.0 Å			
HBD ↓	4.8 Å	5.9 Å		
	7.4 Å	8.4 Å	3.0 Å	
HBD ↓	10.7 Å	13.6 Å	10.1 Å	9.9 Å
	11.6 Å	14.5 Å	10.7 Å	10.8 Å
ali-P	3.5 Å	5.8 Å	7.7 Å	10.0 Å
				10.8 Å



**Figure 5.** (a) Chemical structures of epothilone-based test set for cell line of 1A9. (b) Chemical structures of paclitaxel-based test set for cell line of A549.

located at C8 and C4 (4,4-ethano group) and the third feature is located at C16 (bulky group) in the side chain.

The pharmacophore model for paclitaxel derivatives consists of two HBDs, a HBA, and an *ali-P* region (Fig. 4(b)). Two HBDs are located at the amide group near the C3'-phenyl of the side chain and the hydroxyl group at position C7. The HBA matches the acetyl group at position C10 and an *ali-P* region occupies position C15 (15, 15-ethano group).

**Validation of the Pharmacophore Models Using Test Sets.** After constructing 3D-QSAR models from the training sets, test sets containing twelve and ten compounds, respectively, for epothilone<sup>-18,48,49</sup> and paclitaxel-derivatives,<sup>50-56</sup> were generated and used to predict activity values (Fig. 5). The methods for generating conformational models of the test set compounds were the same as those used in generating the training sets. The predicted activities for the test set compounds were then compared with experimental activities using the fast and best fit methods in the Compare/Fit module in CATALYST (Table 4).

The validation methods used on the training and test sets can be divided into two categories, as shown in Tables 1 and Table 4. One uses the differences in activity values (residuals in log units) between experimental and predicted values, as shown in the parentheses in Table 1 and Table 4. Predicted activities within 1 log unit can be considered to be accurate to within experimental error for data obtained from the same biological experiments.<sup>44</sup> In the other method, we classified the activity scale of compounds into highly active compounds (+++, lower than 10 nM in activity), moderately active compounds (++, 100 nM > activity ≥ 10 nM), and weakly active compounds (+, more than 100 nM).<sup>57,58</sup> In this study, we consider a classification method using the residual of the difference in the activity values as having greater weight than a method directly using an activity scale because we have identified some flaws in the activity scaling method. For example, in the activity scaling method, five compounds in the training set of paclitaxel derivatives (13<sup>th</sup>, 14<sup>th</sup>, 15<sup>th</sup>, 20<sup>th</sup>, and 22<sup>nd</sup> compounds) were judged to be beyond the active range by a simple classification (Table 1(b)) even though they are within 1 log unit.

The most highly active values in the test set were predicted to have residuals smaller than 1 log unit. We obtained this result using the fast fit method for the epothilone and paclitaxel derivatives. According to this result, ten of twelve epothilone derivatives (83%) and eight of ten paclitaxel derivatives (80%) had residuals smaller than 1 log unit, as shown in Table 4. The plots of correlation of predicted versus experimental activity for the training and test sets using the fast fit method are given in Figure 6. In contrast, fitting using the best fit method was very poor for eight out of twelve epothilone derivatives and for seven out of ten of the paclitaxel derivatives. It seems to be caused by the fact that the energy of the conformer used in the fitting with the best fit method is higher than that for the conformer used in the fast fit method because best fit starts from the fast fit conformation and then searches for a conformation with a better fit value (*See Materials and Methods*).<sup>44,59</sup>

**Table 4.** (a) Experimental and predicted activities of test set calculated on the lowest cost hypothesis ( $H_1$ ) for the epothilone derivatives

Cmpd. no.	EA <sup>a</sup>	PA <sup>b</sup>		Activity scale		
		Fast	Best	EA	PA <sub>fast</sub>	PA <sub>best</sub>
1	8.8	74	1.3	+++	++ (0.92)	+++ (-0.83) <sup>c</sup>
2	26	12	0.35	++	++ (-0.34)	+++ (-1.87)
3	1.4	3.4	2	+++	+++ (0.39)	+++ (0.15)
4	300	96	2	+	++ (-0.49)	+++ (2.18)
5	5	16	1.4	+++	++ (0.51)	+++ (-0.55)
6	32	100	0.037	++	+ (0.49)	+++ (-2.94)
7	50	0.43	0.017	++	+++ (-2.07)	+++ (-3.47)
8	20	12	1.7	++	++ (-0.22)	+++ (-1.07)
9	80	3	0.79	++	+++ (-1.43)	+++ (-2.01)
10	9	12	0.3	+++	++ (0.12)	+++ (-1.48)
11	32	45	2	++	++ (0.15)	+++ (-1.20)
12	10	19	2.2	++	++ (0.28)	+++ (0.66)

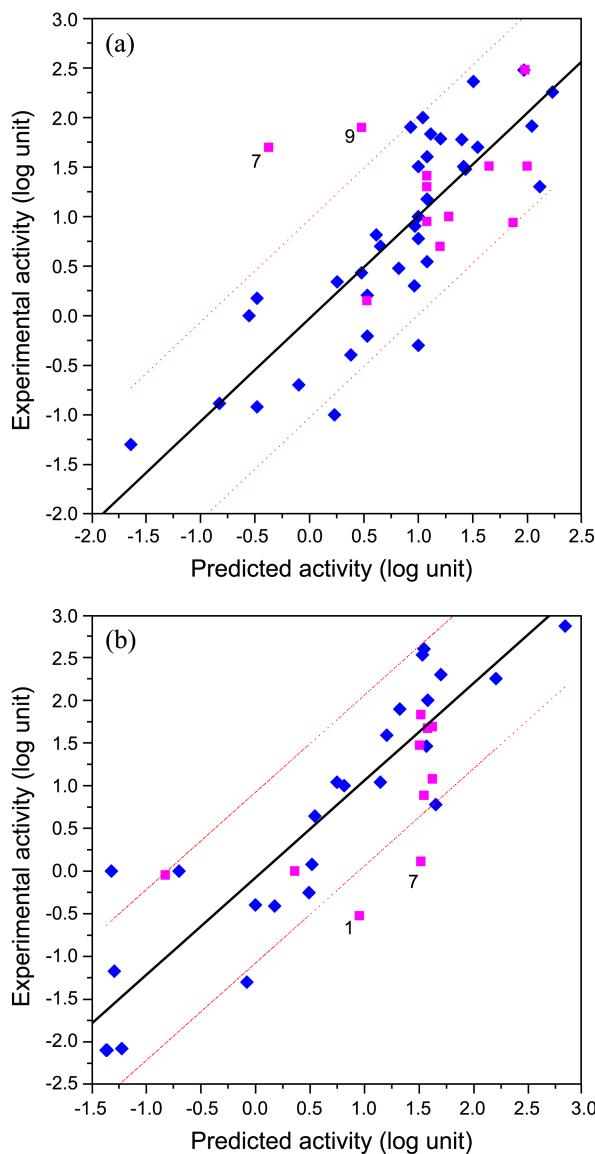
+++: highly active compounds. ++: moderately active compounds. +: weakly active compounds. <sup>a</sup> and <sup>b</sup> represent experimental activity and predicted activity, respectively. The value of parenthesis <sup>c</sup> means residual calculated by the difference of log unit of predicted activity and experimental activity

**Table 4.** (b) Experimental and predicted activities of test set calculated on the lowest cost hypothesis ( $H_1$ ) for the paclitaxel derivatives

Cmpd. no.	EA <sup>a</sup>	PA <sup>b</sup>		Activity Scale		
		Fast	Best	EA	PA <sub>fast</sub>	PA <sub>best</sub>
1	0.3	9	0.68	+++	+++ (1.48)	+++ (0.36) <sup>c</sup>
2	68	33	23	++	++ (-0.31)	++ (0.47)
3	49	42	0.67	++	++ (-0.07)	+++ (1.86)
4	0.9	0.15	0.016	+++	+++ (0.78)	+++ (1.75)
5	30	32	0.67	++	++ (0.03)	+++ (1.65)
6	47	38	0.47	++	++ (-0.09)	+++ (2.00)
7	1.3	33	0.4	+++	++ (1.40)	+++ (0.51)
8	7.7	35	0.58	+++	++ (0.66)	+++ (1.12)
9	12	42	0.56	++	++ (0.54)	+++ (1.33)
10	1.0	2.3	0.025	+++	+++ (0.36)	+++ (1.60)

The training sets used for building the pharmacophores generally covered a larger range of activity values (-1.3 to 2.5 log units for the epothilone derivatives and -2.1 to 2.9 for the paclitaxel derivatives) than that of test sets (0.1 to 2.5 log units for epothilone and -0.5 to 1.8 for paclitaxel derivatives). It was illustrated that the prediction of activity value for the test sets was an interpolation from the compounds in the training sets. Based on this judgment, it may be useful to appreciate how well the compound is able to extrapolate values beyond the range of training set. Through reiterative revision of the established pharmacophore model by adding a high quality test set of compounds to the training set, better predicted activity values should be obtained from the new models.

**Complementary Pharmacophores Using Clustering and Merging Methods.** After generation of pharmacophore models from epothilone derivatives, ten hypotheses were

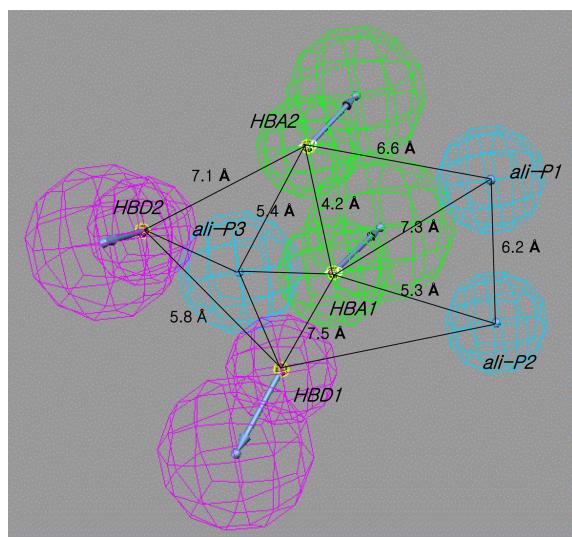


**Figure 6.** Correlation of CATALYST-predicted activity (logPA) values with experimental activity (logEA) values generated with the training and test set using the fast fit method for the epothilone derivatives (a) and for the paclitaxel derivatives (b). Compounds of training and test set are shown in blue and magenta, respectively. The outer red lines represent the 1 log residual boundary. This figure was generated with Origin v.6.0.

grouped with the case of two clusters, in which Hypo 1 and Hypo 5 that are two good quality hypotheses in each cluster were merged with a 0.8 Å distance tolerance (Table 5 and Fig. 7). The lowest cost hypothesis (Hypo 1) has five features with two HBAs and three *ali-P* regions, while the Hypo 5 has four features with two HBDs, one HBA, and one *ali-P* region. As a result of analysis of the distance tolerance between features of two hypotheses, it was demonstrated that these newly merged hypothesis contains two HBAs, two HBDs, and three *ali-Ps* (Fig. 7), and *ali-P3* region was proved to exist in almost same 3D spatial coordinate of the two pharmacophores. Clustering and merging may be used in the following cases: (1) in order to make a more selective

**Table 5.** Clustering result of ten hypotheses generated for the epothilone derivatives.

No. clusters	2	3	4	5	6	7	8	9
Hypo 1	1	1	1	1	1	1	1	1
Hypo 4	1	1	1	2	2	2	2	2
Hypo 2	1	1	1	2	2	2	3	3
Hypo 7	1	1	1	2	2	2	3	4
Hypo 6	1	1	1	2	2	3	4	5
Hypo 8	1	2	2	3	3	4	5	6
Hypo 10	1	2	2	3	4	5	6	7
Hypo 5	2	3	3	4	5	6	7	8
Hypo 3	2	3	3	4	5	6	7	8
Hypo 9	2	3	4	5	6	7	8	9



**Figure 7.** Merged hypothesis for epothilone derivatives from the 1<sup>st</sup> and 5<sup>th</sup> models using a distance tolerance of 0.8 Å.

model when it is used to search small molecule databases and (2) in order to generate a model that is more detailed.

**3D Database Search.** To find out new hit compounds for epothilone-like microtubule stabilizing agent, lowest cost hypothesis ( $H_1$ ) and newly merged hypothesis ( $H_m$ ) were used as the query against three databases: Catalyst/Maybridge2004 (about 55,000 compounds), and Catalyst/NCI2000 (about 239,000 compounds), and Catalyst/ACD2001 (about 267,000 compounds). In order to create hit compounds for epothilone-like microtubule stabilizing agent, first, the shape based query (a merged shape and hypothesis query) method for  $H_1$  was used because this search should give lower percentage of false positives than using generated pharmacophore only. The bound conformation of epothilone-B was used to create the shape query. Second, merged hypothesis ( $H_m$ ) with clustering and merging function of Catalyst program was considered as another query of 3D databases to find out more elevated hit compounds for microtubule stabilizing agent.

**Table 6.** Results of 3 Database searches for the epothilone derivatives.

Database	DB size	Total No. of Hits		% of database		Selected No. of Hits		% of database	
		H <sub>1</sub> <sup>a</sup>	H <sub>m</sub> <sup>b</sup>	H <sub>1</sub>	H <sub>m</sub>	H <sub>1</sub>	H <sub>m</sub>	H <sub>1</sub>	H <sub>m</sub>
Maybridge2004	55,000	159	14	0.29	0.03	11	0	0.02	0
NCI2000	239,000	626	649	0.26	0.27	34	10	0.01	0.004
ACD2001	267,000	432	720	0.16	0.27	15	0	0.006	0

<sup>a</sup>means lowest cost hypothesis for epothilone-like training set. <sup>b</sup>means merged hypothesis of 1<sup>st</sup> and 5<sup>th</sup> models for epothilone-like training set.

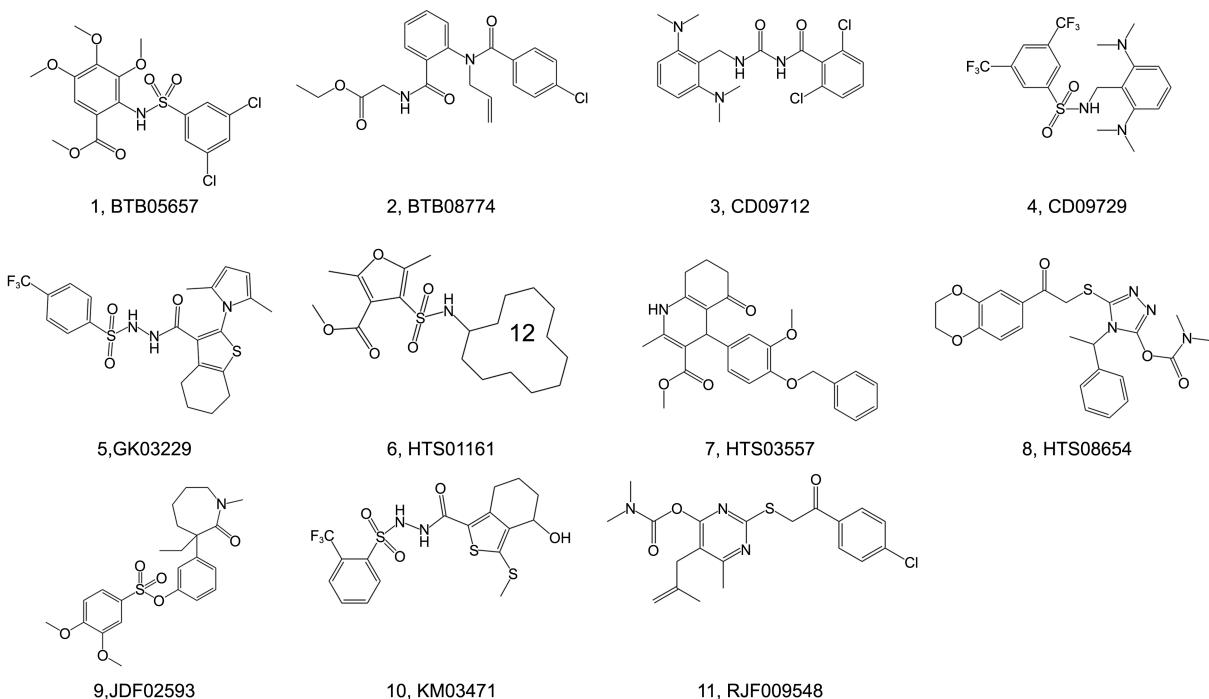
**Table 7.** Physical properties for hit lists using lowest cost hypothesis and Maybridge DB

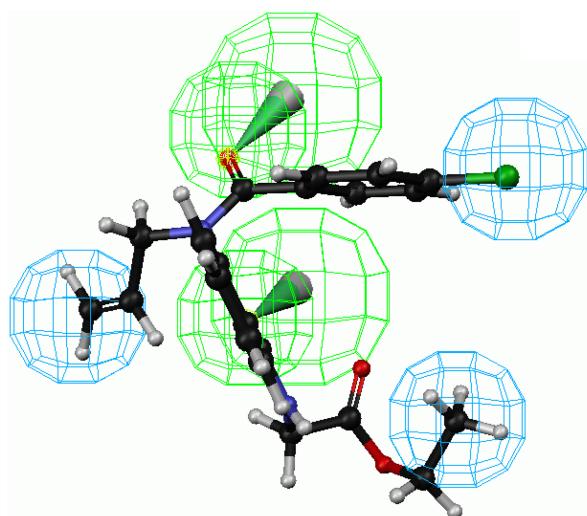
No.	Cmpds	MWT	No. Rotl	No. HBA	No. HBD	logP	BestFit Score
1	BTB05657	450.29	8	9	1	2.50	8.68
2	BTB08774	400.86	11	5	1	3.31	10.33
3	CD09712	409.31	8	4	2	4.20	8.48
4	CD09729	469.44	6	9	1	4.88	10.23
5	GK03229	497.52	6	8	2	4.73	7.32
6	HTS01161	399.50	5	6	1	3.58	9.58
7	HTS03557	433.50	7	5	1	2.27	6.92
8	HTS08654	468.53	9	8	0	3.81	8.07
9	JFD02593	447.55	7	6	0	3.83	8.98
10	KM03471	466.51	7	10	3	3.55	9.12
11	RJF00958	419.93	9	7	0	4.06	6.72

3D database for H<sub>1</sub> was searched using the “Best Flexible Search Databases/Spreadsheets” option to obtain as many hit as possible, and resulting in 159 hits total for Maybridge (0.29% of the whole Maybridge), 626 for NCI (0.26%), and 432 for ACD database (0.16%). On the other hand, total of

14 hit lists for Maybridge (0.03% of the whole Maybridge), 649 for NCI (0.27%), and 720 for ACD (0.27%) were produced for merged hypothesis (H<sub>m</sub>) query. The search results are shown in Table 6. To exclude those hit compounds that are unlikely to be epothilone-like microtubule stabilizing agent because of their high molecular weight, the hit lists were limited less than a 600 g/mol molecular weight cutoff and number of rotatable bonds was restricted less than 15 based on molecular weight (507.68) and number of rotatable bond (11) of epothilone B compound. The final hit lists were considered those with greater than 6.50 of fit value based on the lowest fit value in compounds of training set for lowest cost hypothesis in the Table 1(a) (fit values range from 6.69 to 10.56). For example, finally selected hit lists for lowest cost hypothesis from Maybridge2004 produce 30 compounds with fit value greater than 6.50, while 11 compounds for newly merged hypothesis (Table 7 and Fig. 8). The BTB08774 showing the highest fit value was finally selected as the best hit compound and mapped with lowest cost pharmacophore model of microtubule stabilizing agent (Fig. 9).

In summary, the 3D-QSAR method may be applied for elucidating structurally important groups for binding in the

**Figure 8.** Compounds identified by Maybridge Database with lowest cost pharmacophore model of Epothilone-based microtubule stabilizing agents.



**Figure 9.** Best conformation of BTB08774, compound 2, fit to lowest cost pharmacophore model of microtubule stabilizing agent. Shown are three aliphatic hydrophobic spheres (cyan), and two hydrogen-bond acceptors (green).

active site of a receptor. This study may also be applied to the case of other 3D- or 4D-QSAR methods that require more structural information, such as, descriptor-based inhibitor modeling.<sup>60,61</sup> Generated or merged pharmacophore models can be useful for database searching. Designed candidate ligands can be used for testing proposed binding site models.

A number of methods such as CoMFA<sup>13,62</sup> or WHIM (Weighted Holistic Invariant Molecular descriptor)<sup>63-66</sup> and many publications were available for carrying out 3D-QSAR studies of ligands or inhibitors against receptors including microtubules,<sup>67</sup> but the pharmacophore modeling of microtubule stabilizing agents using 3D-QSAR/CATALYST was new attempt. This approach makes it possible to structurally define unique properties of the active site in an unknown receptor by generating pharmacophore models from ligands and their activities.

**Acknowledgments.** This research was supported by Basic Science Research Program (2012R1A1A4A01013657) and Management of Climate Change Program (2010-0029084) through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (MEST) of Republic of Korea. And this work was also supported by the Next-Generation BioGreen 21 Program (PJ009486) from Rural Development Administration (RDA) of Republic of Korea.

## References

- Nicolaou, K.C.; Roschangar, F.; Vourloumis, D. *Chem. Int. Ed.* **1998**, *37*, 2014-2045.
- Brinkley, B. R. *Cold Spring Harbor Symp. Quant. Biol.* **1982**, *46*, 1029-1038.
- Ball, E. H.; Singer, J. J. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 123-126.
- Cooper, M. S.; Cornell-Bell, A. H.; Chernjawsky, A.; Dani, J. W.; Smith, S. J. *Cell* **1990**, *61*, 135-145.
- Rasenick, M. M.; Wang, N.; Yan, K. *Adv. Second Messenger Phosphoprotein Res.* **1990**, *24*, 381-386.
- Scholey, J. M. *Nature* **1990**, *343*, 118-120.
- Schiff, P. B.; Fant, J.; Horwitz, S. B. *Nature* **1979**, *277*, 665-667.
- Chaudhuri, A. R.; Seetharamalu, P.; Schwarz, P. M.; Hausheer, F. H.; Ludueña, R. F. *J. Mol. Biol.* **2000**, *303*, 679-692.
- Li, Y.; Kobayashi, H.; Hashimoto, Y.; Shirai, R.; Hirata, A.; Hayashi, K.; Hamada, Y.; Shioiri, T.; Iwasaki, S. *Chemico-Biological Interactions* **1994**, *93*, 175-183.
- Jordan, M. A.; Toso, R. J.; Thrower, D.; Wilson, L. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 9552-9556.
- Kobayashi, J.; Hosoyama, H.; Wang, X.-X.; Shigemori, H.; Koiso, Y.; Iwasaki, S.; Sasaki, T.; Naito, M.; Tsuruo, T. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 393-398.
- Ojima, I.; Kuduk, S. D.; Pera, P.; Veith, J. M.; Bernacki, R. J. *J. Med. Chem.* **1997**, *40*, 279-285.
- Roh, E. J.; Kim, D.; Choi, J. Y.; Lee, B.-S.; Lee, C. O.; Song, C. E. *Bioorg. Med. Chem.* **2002**, *10*, 3135-3143.
- Hood, K. A.; West, L. M.; Northcote, P. T.; Berridge, M. V.; Wakefield, J.; Miller, J. H.; Peloruside, A. *Cancer Res.* **2002**, *62*, 3356-3360.
- Chou, T.-C.; Zhang, X.-G.; Harris, C. R.; Kuduk, S. D.; Balog, A.; Savin, K. A.; Bertino, J. R.; Danishefsky, J. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 15798-15802.
- McDaid, H. M.; Mani, S.; Shen, H.-J.; Muggia, F.; Sonnichsen, D.; Horwitz, S. B. *Clin. Cancer Res.* **2002**, *8*, 2035-2043.
- Arslanian, R. L.; Tang, L.; Blough, S.; Ma, W.; Qiu, R.-G.; Katz, L.; Carney, J. R. *J. Nat. Prod.* **2002**, *65*, 1061-1064.
- Buey, R. M.; Diaz, J. F.; Andreu, J. M.; O'Brate, A.; Giannakakou, P.; Nicolaou, K. C.; Sasman, P. K.; Ritzén, A.; Namoto, K. *Chem. Biol.* **2004**, *11*, 225-236.
- Mastropolo, D.; Camerman, A.; Luo, Y.; Brayer, G. D.; Camerman, N. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6920-6924.
- Georg, G. I.; Chen, T. T.; Ojima, I.; Vyas, D. M. *Taxane Anticancer Agents: Basic Science and Current Status, ACS Symp. Series 583*; The American Chemical Society: Washington, D. C., 1995.
- Nicolaou, K. C.; Winssinger, N.; Vourloumis, D.; Ohshima, T.; Kim, S.; Pfefferkorn, J.; Xu, J. Y.; Li, T. *J. Am. Chem. Soc.* **1998**, *120*, 10814-10826.
- Cowden, C. J.; Paterson, I. *Nature (London)* **1997**, *387*, 238-239.
- Service, R. F. *Science* **1996**, *274*, 2009-2010.
- Suffness, M., Ed.; In *Taxol®: Science and Applications*; CRC Press: Boca Raton, FL, 1995.
- Bollag, D. M.; McQueney, P. A.; Zhu, J.; Hensens, O.; Koupal, L.; Leisch, J.; Goetz, M.; Lazarides, E.; Woods, C. M. *Cancer Res.* **1995**, *55*, 2325-2333.
- Rivkin, A.; Njardarson, J. T.; Biswas, K.; Chou, T.-C.; Danishefsky, S. J. *J. Org. Chem.* **2002**, *67*, 7737-7740.
- Rovinsky, E. K.; Eisenhauer, E. A.; Chaudhry, V.; Arbuck, S. G.; Donehower, R. C. *Semin. Oncol.* **1993**, *20*, 1-15.
- Fletcher, B. S.; Kujubadu, D. A.; Perrin, D. M.; Herschman, H. R. *J. Biol. Chem.* **1992**, *267*, 4338-4344.
- Tsuji, M.; Dubois, R. N. *Cell* **1995**, *3*, 493-501.
- Essayan, D. M.; Kagey-Sobotka, A.; Colarusso, P. J.; Lichtenstein, L. M.; Ozols, R. F.; King, E. D. *J. Allergy Clin. Immunol.* **1996**, *97*, 42-46.
- Giannakakou, P.; Sackett, D. L.; Kang, Y.-K.; Zhan, Z.; Buters, J. T. M.; Fojo, T.; Poruchynsky, M. S. *J. Biol. Chem.* **1997**, *272*, 17118-17125.
- Snyder, J. P.; Nettles, J. H.; Cornett, B.; Downing, K. H.; Nogales, E. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5312-5316.
- CATALYST version 4.6; Accelrys, Inc., San Diego, CA, 2001, <http://www.accelrys.com>
- Clark, D. E.; Westhead, D. R.; Sykes, R. A.; Murray, C. W. J. *Comp. Aided Mol. Des.* **1996**, *10*, 397-416.
- Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187-

- 217.
36. Nogales, E.; Wolf, S. G.; Downing, K. H. *Nature* **1998**, *391*, 199-203.
  37. InsightII; Accelrys: San Diego, CA, 1998.
  38. Smellie, A.; Teig, S. L.; Towbin, P. J. *J. Compt. Chem.* **1995**, *16*, 171-187.
  39. Smellie, A.; Kahn, S. D.; Teig, S. L. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 285-294.
  40. Smellie, A.; Kahn, S. D.; Teig, S. L. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 295-304.
  41. Gund, P. *Progress in Molecular and Subcellular Biology*; Hahn, F. E., Ed.; Springer-Verlag: New York, 1997; Vol 5, pp 117-143.
  42. Greene, J.; Kahn, S.; Savoy, H.; Sprague, P.; Teig, S. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1297-1308.
  43. Barnum, D.; Greene, J.; Smellie, A.; Sprague, P. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 563-571.
  44. Andrea, T. A.; Dietrich, S. W.; Murray, W. J.; Kollman, P. A.; Jorgensen, E. C.; Rothenberg, S. A. *J. Med. Chem.* **1979**, *22*, 221-232.
  45. Güner, O. F. *Pharmacophore, Perception, Development, and Use in Drug Design*; IUL Biotechnology Series: 2000; p 504.
  46. Ekins, S.; Bravi, G.; Binkley, S.; Gillespie, J. S.; Ring, B. J.; Wrighton, S. A. *Drug Metab. Dispos.* **2000**, *28*, 994-1002.
  47. Nicolaou, K. C.; Vourloumis, D.; Li, T.; Pastor, J.; Winssinger, N.; He, Y.; Ninkovic, S.; Sarabia, F.; Vallberg, H.; Roschangar, F.; King, N. P.; Finlay, M. R. V.; Giannakakou, P.; Hamel, E. *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 2097-2103.
  48. Nicolaou, K. C.; Ritzén, A.; Namoto, K.; Buey, R. M.; Díaz, J. F.; Andreu, J. M.; Wartmann, M.; Altmann, K.-H.; O'Brate, A.; Giannakakou, P. *Tetrahedron* **2002**, *58*, 6413-6432.
  49. Nicolaou, K. C.; Sasmal, P. K.; Rassias, G.; Reddy, M. V.; Altmann, K.-H.; Wartmann, M.; O'Brate, A.; Giannakakou, P. *Angew. Chem. Int. Ed.* **2003**, *42*, 3515-3520.
  50. Ojima, I.; Slater, J. C.; Pera, P.; Veith, J. M.; Abouabdellah, A.; Begue, J.-P.; Bernacki, R. *J. Bioorg. Med. Chem. Lett.* **1997**, *7*, 133-138.
  51. Ojima, I.; Bounaud, P.-Y.; Ahern, D. G. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 1189-1194.
  52. Ojima, I.; Inoue, T.; Chakravarty, S. *J. Fluor. Chem.* **1999**, *97*, 3-10.
  53. Cheng, Q.; Oritani, T.; Horiguchi, T.; Yamada, T.; Mong, Y. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 517-521.
  54. Cheng, Q.; Oritani, T.; Horiguchi, T. *Tetrahedron* **2000**, *56*, 1667-1679.
  55. Cheng, Q.; Kiyota, H.; Yamaguchi, M.; Horiguchi, T.; Oritani, T. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1075-1077.
  56. Wrasiclo, W.; Gaedicke, G.; Guy, R. K.; Renaud, J.; Pitsinos, E.; Nicolaou, K. C.; Reisfeld, R. A.; Lode, H. N. *Bioconjugate Chem.* **2002**, *13*, 1093-1099.
  57. Mustata, G. I.; Brigo, A.; Briggs, J. M. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 1447-1454.
  58. Debnath, A. K. *J. Med. Chem.* **2003**, *46*, 4501-4515.
  59. Ekins, S.; Bravi, G.; Wikle, J. H.; Wrighton, S. A. *J. Pharmacol. Exp. Ther.* **1999**, *291*, 424-433.
  60. Bravi, G.; Gancia, E.; Mascani, P.; Pegna, M.; Todeschini, R.; Zaliani, A. *J. Comp. Aided Mol. Des.* **1997**, *11*, 79-82.
  61. Ekins, S.; Bravi, G.; Ring, B. J.; Gillespie, T. A.; Gillespie, J. S.; Vandenbranden, M.; Wrighton, S. A.; Wikle, J. H. *J. Pharmacol. Exp. Ther.* **1999**, *288*, 21-29.
  62. Lee, K. W.; Briggs, J. M. *J. Comp. Aided Mol. Des.* **2001**, *15*, 41-55.
  63. Todeschini, R.; Grammatica, P. *Quant. Struct. Act. Relat.* **1997**, *16*, 113-119.
  64. Todeschini, R.; Bettoli, C.; Giurin, G.; Grammatica, P.; Miana, P.; Argese, E. *Chemosphere* **1996**, *33*, 71-79.
  65. Todeschini, R.; Lasagni, M.; Marengo, E. *J. Chemom.* **1994**, *8*, 263-273.
  66. Todeschini, R.; Vighi, M.; Provenzani, R.; Finzio, A.; Grammatica, P. *Chemosphere* **1996**, *32*, 1527-1545.
  67. Bleckman, A.; Meiler, J. *QSAR Comb. Sci.* **2003**, *22*, 722-728.