JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

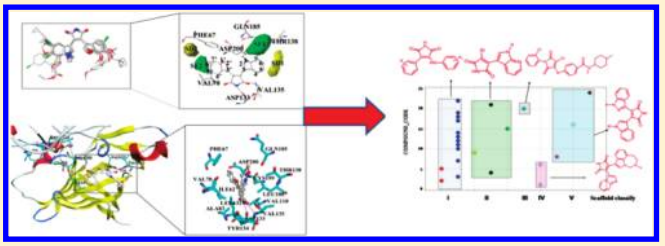
A New Protocol for Predicting Novel GSK-3β ATP Competitive Inhibitors

Jiansong Fang, Dane Huang, Wenxia Zhao, Hu Ge, Hai-Bin Luo,* and Jun Xu*

Research Center for Drug Discovery and Institute of Human Virology, School of Pharmaceutical Sciences, Sun Yat-Sen University, 132 East Circle at University City, Guangzhou, China 510006

Supporting Information

ABSTRACT: Glycogen synthase kinase 3β (GSK-3β) is a potential therapeutic target for cancer, type-2 diabetes, and Alzheimer's disease. This paper proposes a new lead identification protocol that predicts new GSK-3β ATP competitive inhibitors with topologically diverse scaffolds. First, three-dimensional quantitative structure−activity relationship (3D QSAR) models were built and validated. These models are based upon known GSK-3β inhibitors, benzofuran-3-yl-(indol-3-yl) maleimides, by means of comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA). Second, 28 826 maleimide derivatives were selected from the PubChem database. After filtration via Lipinski's rules, 10 429 maleimide derivatives were left. Third, the FlexX-dock program was employed to virtually screen the 10 429 compounds against GSK-3β. This resulted in 617 virtual hits. Fourth, the 3D QSAR models predicted that from the 617 virtual hits, 93 compounds would have GSK-3β inhibition values of less than 15 nM. Finally, from the 93 predicted active hits, 23 compounds were confirmed as GSK-3β inhibitors from literatures; their GSK-3β inhibition ranged from 1.3 to 480 nM. Therefore, the hits rate of our virtual screening protocol is greater than 25%. The protocol combines ligand- and structure-based approaches and therefore validates both approaches and is capable of identifying new hits with topologically diverse scaffolds.



1. INTRODUCTION

Glycogen synthase kinase 3 (GSK-3), a cytosolic serine/threonine protein kinase, is ubiquitously expressed in mammalian tissues.[1−5] GSK-3 has two isoforms, α and β, that are 97% homologous in their catalytic domains. It is an important enzyme that is responsible for various phosphorylation catalyzes. It is involved in signaling pathways, including cell differentiation, cellular growth and proliferation, metabolic processes, apoptosis, inflammation, and mechanisms involved in neuronal functions.[6−9] Thus, GSK-3 has recently emerged as a potential therapeutic target for several human diseases, such as cancer,[10,11] type-2 diabetes,[12,13] and Alzheimer's disease.[14,15]

GSK-3 inhibitors are classified into three types: ATP competitive inhibitors, non-ATP competitive inhibitors, and metal ion competitive inhibitors (in the Mg^{2+} binding site).[16] Usually, the activity of GSK-3β is regulated via phosphorylation, at the binding cavity, of the key residues, Arg96, Arg180, and Lys205. In the cavity, Lys85, Asp200, and Glu97 are important for ATP recognition and affinity, as Asp200 interacts with the phosphate hydroxyl group of ATP.[17] Many studies focus on the ATP competitive inhibitors that regulate GSK-3β. In this paper, we report a new protocol for identifying novel GSK-3β ATP competitive inhibitors.

Structural requirements for GSK-3 ATP competitive inhibitors have been explored by a number of research groups.[18−29] Most of the previous 3D QSAR studies modeling GSK-3 ATP competitive inhibitors were done by building comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) models, followed by docking validations.[18−26] These studies were based on homologue analogues, such as indirubin analogues, pyrazolopyrimidine derivatives, bisarylmaleimide series, 3-anilino-4-phenylmaleimides, thiazolo-[5,4-f]quinazolin-9-one derivatives, and 2,4-disubstituted thiadiazolidinones. The main purposes of those studies were to create,

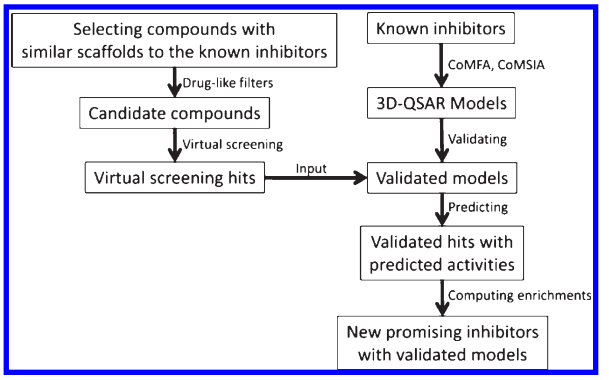
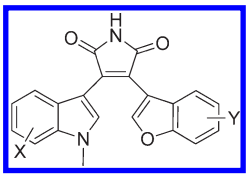


Figure 1. Protocol of combined ligand- and structure-based approaches.

Received: March 9, 2011
Published: May 26, 2011

Table 1. Chemical Structures and Biological Activities of 38 GSK-3β Ligands Used for 3D QSAR Analysis



				predicted pIC_{50}	
no.	X	Y	actual pIC_{50}	CoMFA	CoMSIA
1	H	H	7.46	7.66	7.50
2	5-F	6-CH_2OH	9.46	9.37	9.50
3	5-F	6-CH_2OCH_3	7.62	7.53	7.52
4[a]	5-F	6-OH	8.46	8.16	8.30
5	5-F, 6-I	7-OCH_3	6.61	6.49	6.41
6	5-F, 6-Cl	H	6.74	7.06	6.96
7	5-F, 6-Cl	6-CH_2OH	9.02	8.77	8.77
8	5-F, 6-Cl	6-OCH_3	6.06	5.92	5.86
9	5-F, 6-Cl	6-cyclopropylmethoxy	5.89	5.82	5.85
10	5-F, 6-Cl	6-cyclobutylmethoxy	5.39	5.46	5.62
11	5-F, 6-Cl	7-OCH_3	6.59	6.75	6.82
12	5-F, 6-*p*-Cl-Ph	7-OCH_3	5.15	5.22	5.19
13[a]	5-Br	H	8.15	7.77	7.87
14	5-Br	7-OCH_3	8.12	7.89	7.87
15	5-Br	6-CH_2OH	9.29	9.60	9.60
16	5-Br	6-prop-2-ynyloxy	7.60	7.49	7.52
17[a]	5-Br	6-allyloxy	7.32	6.95	6.97
18	5-Br	6-O-(p-CH_3O)-Bn	6.48	6.50	6.41
19	5-Cl	5-F	7.38	7.11	7.05
20	5,7-dibromo	7-OCH_3	7.05	7.05	7.01
21	5-I	H	7.46	7.33	7.34
22	5-I	5-F	6.74	7.02	6.97
23[a]	5-CN	6-CH_2OH	7.88	7.60	6.80
24	5-cyclopropyl	H	6.63	6.50	6.73
25	5-cyclopropylethynyl	5-F	7.79	7.82	7.79
26	5,6-methylenedioxy	5-F	6.15	6.21	6.24
27	5-OCH_3, 6-Cl	H	6.36	6.41	6.44
28[a]	5-OCH_3, 6-I	H	6.65	6.38	6.13
29	6-OH	H	7.82	7.70	7.72
30[a]	6-CF_3	7-OCH_3	6.08	5.79	6.18
31	7-CH_2OH	H	8.27	8.24	8.21
32[a]	7-CH_2OH	6-CH_2OH	8.29	8.33	9.25
33	7-CH_2OCH_3	H	9.64	9.63	9.66
34	7-CH_2OCH_3	6-CH_2OH	9.14	9.22	9.11
35	7-CH_2CH_2COOEt	H	7.99	7.89	7.94
36	7-CH_2CH_2COOH	H	8.92	8.89	8.98
37[a]	7-OH	H	7.26	7.98	7.98
38	1H-benzo[g]	5,6-difluoro	6.50	6.51	6.51

[a] Molecules in the test set.

for given homologue scaffolds, predictive 3D QSAR models for lead optimizations and to prove, through docking validations, that the ligand-based 3D models were consistent with the structure-based binding models.

In our opinion, 3D QSAR models do not always need to be consistent with the ligand−receptor binding models in order to possess predictive power. However, the predictive power of a 3D QSAR model can be used to enhance structure-based virtual screening for lead identification.

The purposes of this work are to explore, with 3D QSAR approaches, more structurally diverse scaffolds that are based upon known GSK-3β inhibitors and to create a new lead identification

dx.doi.org/10.1021/ci2001154 |*J. Chem. Inf. Model.* 2011, 51, 1431–1438

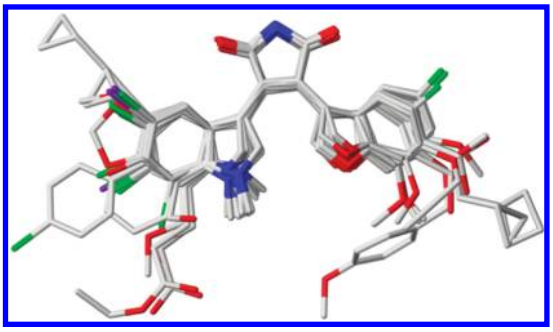


Figure 2. The superimposition of the 30 compounds in the training set.

Table 2. Statistical Results of the CoMFA and CoMSIA Models[a]

		CoMSIA				
	CoMFA	SHE	SED	SEA	SEHA	SEHDA
			Training Set			
r^2	0.984	0.983	0.983	0.982	0.983	0.982
q^2	0.602	0.587	0.560	0.585	0.665	0.654
ONC	4	4	2	2	2	2
F	241.443	183.916	177.616	171.593	186.868	170.914
			Contribution			
s	0.171	0.181	0.185	0.188	0.180	0.188
S	0.569	0.179	0.304	0.252	0.135	0.130
E	0.431	0.428	0.478	0.393	0.289	0.260
H	–	0.393	–	–	0.303	0.287
D	–	–	0.218	–	–	0.110
A	–	–	–	0.353	0.273	0.212
			Test Set 1			
r^2_{test}	0.905	0.804	0.703	0.780	0.761	0.718
q^2_{test}	0.872	0.673	0.500	0.631	0.587	0.518
k	0.989	0.885	1.088	1.187	1.043	1.059

[a] r^2: Conventional correlation coefficient of the training set; q^2: cross-validated correlation coefficient of the training set; r^2_{test}: conventional correlation coefficient of the test set; q^2_{test}: cross-validated correlation coefficient of the test set; ONC: optimal number of the principal components; F: statistical squared deviation ratio; S: standard error of estimate; and k: slope of the regression line derived from the test set.

protocol, by combining ligand-based 3D QSAR methods and structure-based docking methods, in order to improve the enrichment in virtual screening.

The flow of the protocol is illustrated in Figure 1: (1) 3D QSAR models were generated from known inhibitors (a training set)[30] and validated through docking approaches; (2) extra compounds were selected, based on a substructure search for the maleimide core, from the PubChem[31] database and were then filtered by Lipinski's rules; (3) the FlexX-dock program was employed to virtually screen the selected compounds against GSK-3β, so as to produce virtual hits; (4) the 3D QSAR models predicted the GSK-3β inhibition of the virtual hits; and (5) the compounds with the top scores were confirmed with experimental results. This protocol allowed us to refine structure-based virtual screening results and to identify new topologically diverse drug leads. The protocol may be applied to other drug lead identifications.

Table 3. CoMFA and CoMSIA GSK-3β Inhibition Prediction for 27 Maleimides[a]

compound ID[b]	experimental pIC_{50}	CoMSIA predicted pIC_{50}	CoMFA predicted pIC_{50}
M1	6.94	7.17	8.14
M2	6.81	7.31	8.38
M3	6.85	7.21	8.04
M4	7.13	7.18	7.98
M5	7.12	7.24	7.96
M6	7.07	7.35	7.92
M7	6.96	7.329	8.255
M8	6.96	7.38	8.02
M9	7.28	7.2	6.96
M10	7.94	7.21	8.04
M11	7.31	7.11	8.06
M12	7.44	7.37	7.91
M13	6.78	7.057	7.69
M14	5.92	6.38	6.94
M15	6.89	6.37	6.61
M16	6.4	6.65	7.31
M17	6.12	6.29	6.56
M18	7.74	7.49	8.12
M19	6.92	7.28	7.89
M20	7.31	6.92	7.45
M21	8.69	8.31	8.42
M22	8.33	8.64	10.79
M23	8.72	8.55	10.15
M24	8.46	9.08	10.53
M25	8.89	9.44	10.56
M26	8.89	8.76	10.11
M27	8.7	8.55	9.8

[a] From refs 24, 42, and 43. [b] The IDs are taken from Figure 3.

2. MATERIALS AND METHODS

2.1. Data Set. A data set of 38 benzofuran-3-yl-(indol-3-yl) maleimides, whose in vitro biological activity data were reported as IC_{50} (from the literature),[30] was used in the current study. The data set was divided into a training set of 30 and a test set of 8 compounds. Both of the sets were randomly selected. The IC_{50} values were converted, using the formula ($pIC_{50} = -\log IC_{50}$), to the corresponding pIC_{50} values. The chemical scaffold, structures, and biological activities of the 38 compounds are listed in Table 1. The 28 826 compounds were selected from the PubChem database[31] through a substructure search for the maleimide core. The data were further filtered based upon Lipinski's rules. The filtering resulted in 10 429 compounds, which would be screened against GSK-3β protein.

2.2. Geometry Optimization and Molecular Alignment. The modeling studies were performed using the molecular modeling package, SYBYL7.3.5.[32] CoMFA[33] and CoMSIA[34] approaches are required to align structures in the training set. The energy-minimized conformation of the most active compound, **33**, was chosen as the putative bioactive conformation. Gasteiger–Huckel atomic partial charges[35] were computed for

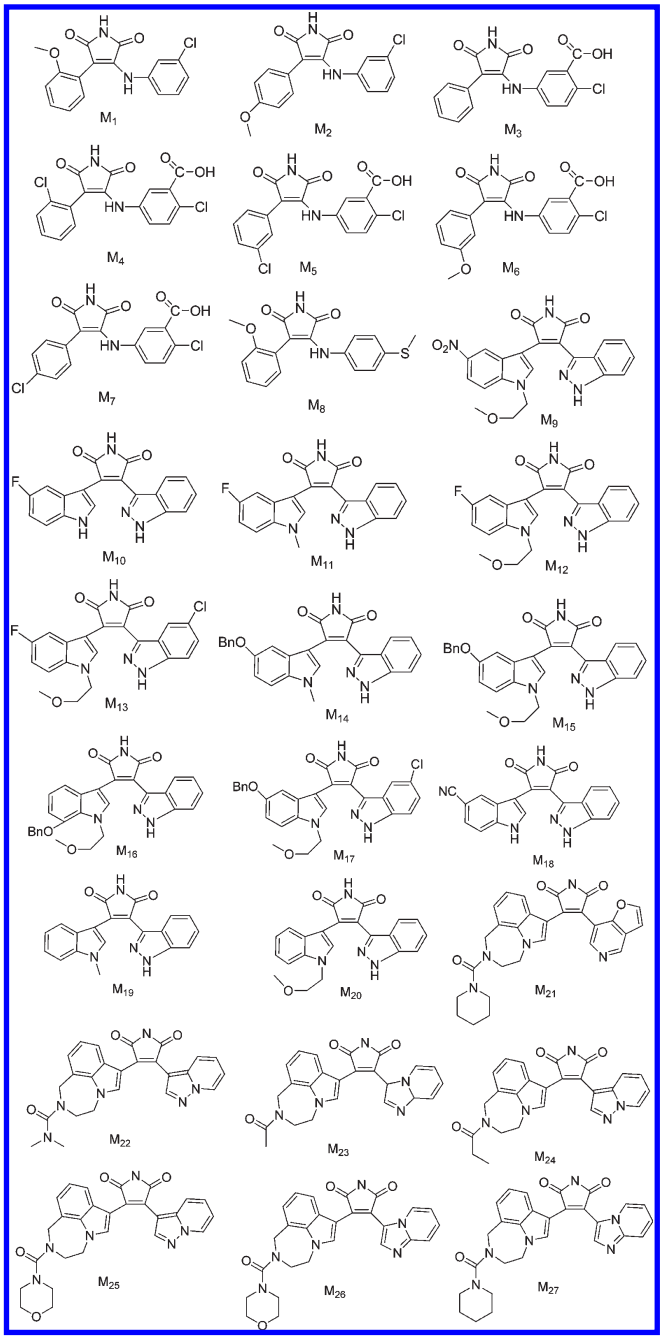


Figure 3. The 27 GSK-3β inhibitors from refs 24, 42, and 43.

all the chemical structures in the training set. The energy minimizations for the training set were performed using a distance-dependent dielectric function from the Tripos force field. The structures in the training set were aligned to the energy-minimized conformation of compound **33**, as shown in Figure 2.

2.3. 3D QSAR Models. In order to derive the CoMFA fields from the training set, a 3D cubic lattice was generated from the space covering the 30 aligned compounds. The steric and electrostatic parameters were calculated at each lattice intersection of a 3D grid that is spaced at 1.0 Å intervals. The steric and electrostatic fields were calculated from standard Tripos force fields. An sp^3 carbon atom with one formal positive charge was used as a probe to generate steric (Lennard-Jones potential) and electrostatic (Coulombic potential) field energies. The steric and electrostatic fields were truncated at 30.00 kcal/mol. For CoMSIA analysis, combinations among the five field descriptors—steric, electrostatic, hydrophobic, hydrogen-bond donor, and hydrogen-bond acceptor—were calculated. Partial least-squares (PLS)[36] regression analysis was used to establish quantitative relationships between molecular descriptors and biological activities. The leave-one-out (LOO) method was applied for cross-validation analysis. The cross-validated q^2 that resulted in the optimal number of components and the lowest standard error of prediction was taken. With the CoMFA_STD scaling option, equal weights were assigned to steric and electrostatic fields. In order to productively build models and reduce noise, a minimal filter value of 2 kcal/mol was applied. The cross-validated regression coefficient,[37] q^2, was determined to measure the robustness and the predictive ability of the derived models. Finally, non-cross-validation was performed to derive, with the conventional correlation coefficient r^2 and estimated standard error, the PLS regression models. For the CoMFA model, the region focusing method[38] was performed to enhance the predictive ability. Region focusing was an iterative procedure, which refined a model by increasing the weight for those lattice points that were pertinent to the model.

2.4. External Cross-Validation Criteria for 3D QSAR Models. A high LOO cross-validated q^2 value does not necessarily mean that the model exhibits a good predictive ability for a test set. Other criteria[39,40] ($r^2_{test} \geq 0.6$, $q^2_{test} \geq 0.5$, and $0.85 \leq k \leq 1.15$) were used for external cross-validation. Where the r^2_{test} is the conventional correlation coefficient of the test set, the q^2_{test} is the cross-validated correlation coefficient, and k is the slope of the regression line derived from the test set.

2.5. Docking Studies. One way to validate the ligand-based models is to dock the putative conformations of the models to the active site of a target (if its structural data is available). The docking study can reveal if the models are consistent with the active poses of the ligands in the binding site. With the FlexX-dock program,[41] a cocrystal structure of GSK-3β (PDB code: 1R0E) was chosen for the docking study . In order to validate the docking reliability of the FlexX-dock program, the ligand (3-indolyl-4-arylmaleimide) of 1R0E was docked back to the binding pocket in GSK-3β, and the root-mean-square deviation (RMSD) between the docked and crystal poses of the ligand was calculated. The RMSDs for the top 10 docked poses ranged from 0.7 to 1.7 Å, which confirmed that the FlexX-dock program could produce reliable docking results for 1R0E. All 38 benzofuran-3-yl-(indol-3-yl) maleimides were docked into the ATP binding site of GSK-3β. After each ligand was docked by the FlexX-dock program, the energy was minimized with the force field of CHARMM27.

3. RESULTS AND DISCUSSION

3.1. Statistical Analyses for CoMFA and CoMSIA Models. One CoMFA and five CoMSIA models were built using the same training set of 30 compounds. The CoMFA model was chosen after rigorous cycles of model development and validation, which were based on internal predictions of the training set and the external predictions of the test set of 8 compounds. The statistical parameters for the model developed are shown in Table 2. The CoMFA model shows that the cross-validated $q^2 = 0.602$ (with four components) and the non-cross-validated $r^2 = 0.984$.

The five CoMSIA models were SHE, SED, SEA, SEHA, and SEHDA, where S, H, E, A, and D stand for steric, hydrophobic,

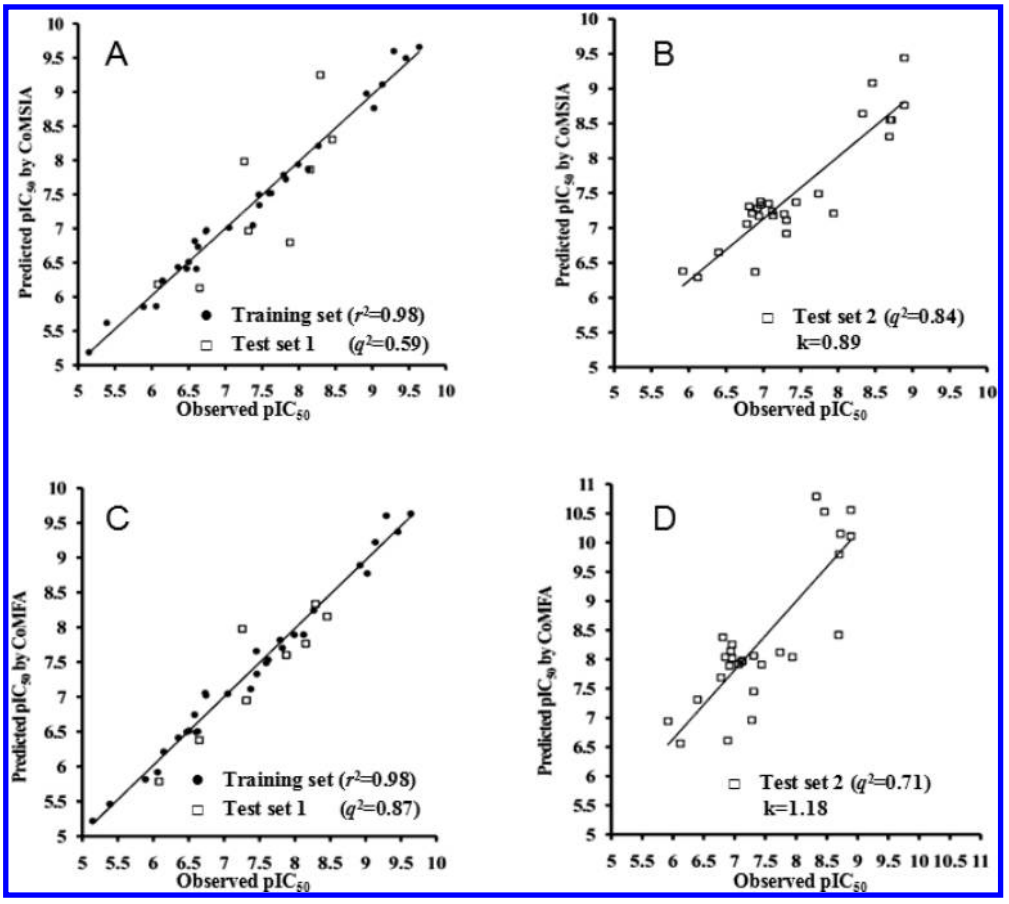


Figure 4. 3D QSAR models were tested with test sets 1 and 2. Structures in the test set 2 were more structurally diverse than the structures in the test 1. (A and C) 3D QSAR models being tested against the training set and test set 1. (B and D) 3D QSAR models being tested against the test set 2.

electrostatic, hydrogen-bond acceptor, and hydrogen-bond donor field, respectively. As shown in Table 2, the best CoMSIA model is SEHA with the cross-validated q^2 = 0.665 (with two components) and the non-cross-validated r^2 = 0.983. Therefore, the model's statistical parameters also deny any possibility of chance correlation.

Compounds in the same training set were subjected to the leave-four-out tests. A Y-randomization PLS analysis[39,40] was performed to further evaluate the extent of chance correlation in these 3D QSAR models. In these tests, the means, from 50 runs for the CoMFA and CoMSIA-SEHA models, of the q^2 (0.594 and 0.662) values indicated no significant difference between the LOO (0.602 and 0.665) and the leave-four-out cases. In addition, none of the 20 Y-randomization tests gave q^2 values of more than 0.4. This further verified that the above 3D QSAR models were not established by chance correlation. These models are robust.

3.2. Test-Sets Validations. Both the CoMFA and CoMSIA-SEHA models were built from the same training set, with almost the same r^2 values (about 0.98, see Table 2). The CoMFA model k (the slopes of the regression line) value was 0.989. Among the five CoMSIA models, the SEHA model produced the best results. For the CoMSIA-SEHA model, its k value was 1.043. According to the 3D QSAR statistical criteria: $r^2_{test} \geq 0.6$, $q^2_{test} \geq 0.5$, and $0.85 \leq k \leq 1.15$, the CoMSIA-SEHA and CoMFA models were qualified for predicting GSK-3β inhibitions (Table 3).

To validate the predictive power of the 3D QSAR models, the two test sets were prepared. The test set 1 contained 8 compounds, which shared the same scaffold (Table 1). These compounds were aligned with the energy-minimized conformation of compound **33** (Table 1 and Figure 2). Then, the CoMFA and CoMSIA models predicted the pIC_{50} for each aligned conformation of a compound. For the CoMFA model, q^2_{test}, r^2_{test}, and k values were 0.872, 0.905, and 0.989, respectively. For the CoMSIA-SEHA model, q^2_{test}, r^2_{test}, and k values were 0.587, 0.761, and 1.043, respectively. These results indicated that both the CoMFA and the CoMSIA-SEHA models had reasonable predictive power for test set 1. Since the CoMFA model q^2_{test} value was 0.872, and the CoMSIA-SEHA model q^2_{test} value was 0.587; the former model was better than the latter (Figure 4A and C).

The test set 2 contained 27 GSK-3β inhibitors that were selected from other references.[24,42,43] The inhibitors were structurally more diverse than the ones in test set 1. The test set 2 was purposely prepared to further test the predictive powers of the CoMFA and CoMSIA-SEHA models. For this test set, the CoMFA model k value was 1.18 (Figure 4D), which is out of the reliability range ($0.85 \leq k \leq 1.15$). Therefore, the CoMFA model was not reliable for test set 2 because the compounds in this test set had more diverse scaffolds (see Figures 3 and 6). On the other hand, the CoMSIA-SEHA model k value was 0.89 (Figure 4B) for test set 2, and the q^2_{test} was 0.84. Hence, we concluded that the CoMSIA-SEHA model is predictive for test set 2.

3.3. Contour Analysis. The effects of all field descriptors contributing to bioactivity can be partitioned and viewed as 3D

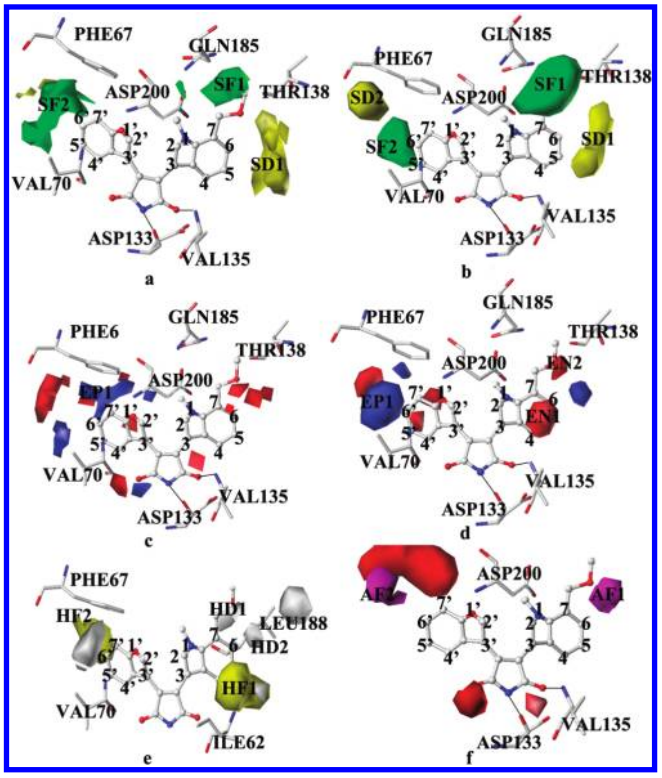


Figure 5. The CoMFA (steric, a and electrostatic, c) and CoMSIA (steric, b, electrostatic, d, hydrophobic, e, and hydrogen-bond acceptor, f) contour maps with molecule 33 as the reference. Refer to the online version for colors. In (a, b), green (labeled as "SF1" and "SF2") and yellow ("SD1" and "SD2") contours refer to the sterically favored and disfavored areas, respectively, whereas blue ("EP1") and red ("EN1" and "EN2") contours refer to regions where electropositive substituents are favorable and unfavorable in (c and d), respectively. In (e), yellow ("HF1" and "HF2") or gray ("HD1" and "HD2") contours indicate regions where hydrophobic or hydrophilic groups are favorable. In (f), magenta ("AF1" and "AF2") and red contours indicate regions where hydrogen-bond acceptor groups are favorable and unfavorable, respectively.

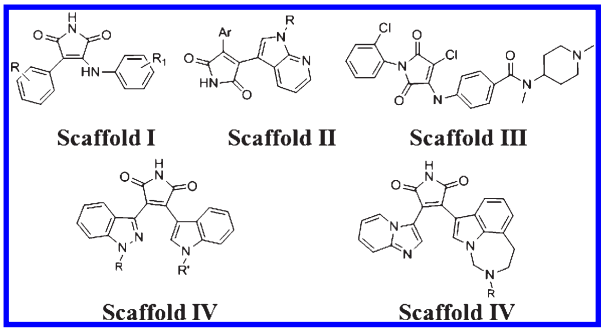


Figure 6. The scaffold classification of 23 confirmed hits of virtual screening.

coefficient contour plots.[33] The contour maps of the CoMFA and CoMSIA-SEHA models, along with the molecular docking models, are represented in Figure 5.

Figure 5a and b shows the steric fields of the CoMFA and CoMSIA-SEHA models, respectively. The green areas mean that larger bulky functional groups close to these areas increase the activities, whereas the yellow areas mean that larger bulky

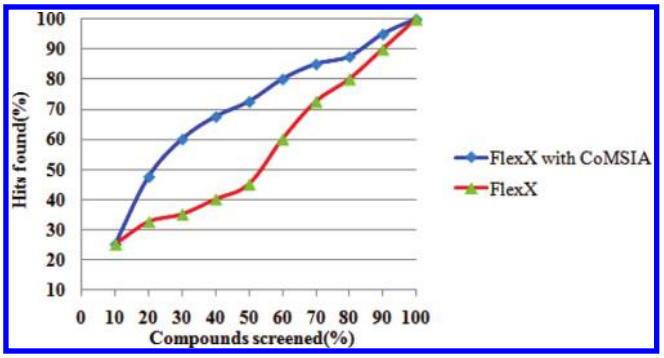


Figure 7. The enrichment curves for FlexX docking with and without combining CoMSIA model.

functional groups decrease the activities. Two green areas ("SF1" and "SF2") and one yellow area ("SD1") exist in the binding site pocket of GSK-3β. Because the green contour ("SF1") is located among the flexible regions of residues Thr138 and Gln185, which is close to the 7- position of the indazole ring. The yellow contour ("SD1") is located at the 5 and 6 positions of the indazole ring; thus, a bulky substituent is disfavored the activity enhancement in this region. Bulky groups at these positions may interfere with the hydrogen-bond interactions between ligands and Val135, which explains why the activity (5.15) of compound **12** is lower than that of **11**'s (6.59). The same trend is demonstrated when comparing compounds **19** (7.38) and **26** (6.15). The green contour ("SF2") and yellow contour ("SD2"), from the CoMSIA-SEHA model, are located near the 6′ position of the benzofuran ring. Too small or too bulky groups for these ligands may decrease their hydrophobic interactions with Phe67 and Val70. For example, compound **2**, with 6-CH_2OH, inhibits GSK-3β 10 times more than compound **4**, with 6-OH. However, the bulky group near this area reduces the biological activity; for example, compound **9** (5.89) is more active than compound **10** (5.39).

Electrostatic contour maps from the CoMFA and CoMSIA-SEHA models are shown in Figure 5c and d, respectively. Blue and red areas mean, respectively, that the electropositive groups are beneficial or detrimental to inhibitory activities. Two red contours ("EN1" and "EN2") exist at the 4, 5 and 6, 7 positions of the indazole ring in Figure 5d but are not observed in Figure 5c. Both the CoMSIA and CoMFA electrostatic maps show the presence of the blue contour (EP1) at the 6′ and 7′ positions of the benzofuran ring. These positions favor the presence of electropositive groups that enhance the ligands' inhibitory activities. According to the MMFF94 partial atomic charge calculations, the hydrogen (0.15 |e|) at the 5′ position of the benzofuran ring in compound **21** is replaced by 5-F (−0.081 |e|) in compound **22**. The 6-CH_2OH (0.144 |e|) at the 6′ position of the benzofuran ring in compound **7** is replaced by the 6-OCH_3 (−0.0825 |e|) in compound **8**, and the 6-CH_2OH (0.1435 |e|) at the same position in compound **2** substitutes for 6-OH (−0.0825 |e|) in compound **4**. Therefore, the activities of compounds **21**, **7**, and **2** are higher than those of molecules **22**, **8**, and **4**, respectively.

A hydrophobic contour map from the CoMSIA-SEHA model is shown in Figure 5e. It could be observed that one yellow contour ("HF1") exists near the 5 position of the indazole ring, and a medium yellow contour ("HF2") is located at the 6′ position of the benzofuran ring. This suggests that large hydrophobic interactions in the two regions enhance the inhibitory

activity. The 5 position of the indazole ring lies in cavity 1, which is composed of the hydrophobic residues, Ile62 and Leu188. The 6 position of the benzofuran ring was located in cavity 2, which is composed of the hydrophobic residues, Phe67 and Val70. Thus, hydrophobicity is preferred in these positions. Two gray contours, "HD1" and "HD2", indicate that hydrophilic groups at the 6 and 7 positions of the indazole ring will increase the biological activity.

Hydrogen-bond acceptor contour maps from the CoMSIA-SEHA model are shown in Figure 5f. Two hydrogen bonds, between ligands and GSK-3β, at Asp133 and Val135 are observed in docking studies. Two magenta contour maps, "AF1" and "AF2", are shown, respectively, at the 6 position of the benzofuran ring and at the 6 position of the indazole ring. These areas indicate that hydrogen donors will increase the activity. For example, compound **7** contains a hydrogen donor at 6-CH_2OH, while compound **6** does not have a hydrogen donor at the same place. Therefore, the activity of compound **7** is 100 times more active than compound **6**'s. The same explanation applies for the activity difference between compounds **29** and **1**.

3.4. Predicting GSK-3β Inhibitions for More Structurally Diverse Compounds. Our goal is to discover new GSK-3β inhibitor scaffolds. However, the molecular docking method is not completely reliable and, therefore, unsatisfactory. By combining the 3D QSAR approach with the molecular docking method, we hope to improve the molecular docking method. Hence, we selected 28 826 compounds containing maleimide cores from the PubChem database, and after filtration by Lipinski's rules, 10 429 compounds were left. Then, the FlexX-dock program was employed to virtually screen the 10 429 compounds against GSK-3β. This resulted in 617 virtual hits, which have docking score values of less than −28 kcal/mol. The CoMSIA-SEHA model predicted that 93 compounds from the 617 virtual hits had GSK-3β inhibition values of less than 15 nM (Table S1, Supporting Information). Finally, 23 compounds, from the 93 predicted active hits, were confirmed from references as GSK-3 β inhibitors; their GSK-3 β inhibition ranged from 1.3 to 480 nM. Therefore, the hits rate of our virtual screening protocol was greater than 25%. This result demonstrates that the virtual screening protocol, which combines structure- and ligand-based approaches, can produce better hit rates as well as more structurally diverse hits. As shown in Figure 6, the 23 confirmed GSK-3β inhibitors belong to 5 different scaffolds, which are topologically different from the scaffold of the test set.

3.5. Analyses of the Enrichment Curves. Two virtual screening experiments were executed in order to demonstrate the performance of the protocol. First, the FlexX-dock program was employed to virtually screen the 10 429 compounds against GSK-3β (Figure 7, the curve in red). Then, the same library was screened by the FlexX-dock program by combining the CoMSIA-SEHA prediction data (Figure 7, the curve in blue). With the information from the CoMSIA-SEHA prediction, the FlexX-dock program screened the 20% compounds of the virtual library and found the 47.5% real hits. If the information was not combined, then the same docking program screened the 20% compounds of the virtual library and only found 32.5% real hits. The hit enrichment was therefore improved approximately 1.5 times.

4. CONCLUSION

A new protocol for GSK-3β ATP competitive inhibitor lead identification has been proposed in this paper. The protocol combined ligand- (CoMFA and CoMSIA) and structure-based (FlexX-dock) approaches. Our experiments lead to the following conclusions: (1) CoMFA only has predictive ability for the compounds that share the same scaffold of the training set. However, CoMSIA can predict bioactivity for the compounds that have more diverse scaffolds (in comparison to the training set's). (2) 3D QSAR models are usually used for lead optimization, but the enrichment of a structure-based virtual screening approach can be improved when the 3D QSAR approaches are combined. This protocol may be useful for the lead identification of other drug targets.

■ ASSOCIATED CONTENT

Supporting Information. The structures of the 93 hits (Figure S1). Detailed information for the 93 hits is from the Pubchem database (Table S1). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: xujun9@mail.sysu.edu.cn or luohb77@mail.sysu.edu.cn.

■ ACKNOWLEDGMENT

This work was funded in part of the major scientific and technological special project by the ministry of Science and Technology of China (2010ZX09102-305), Guangdong Recruitment Program of Creative Research Groups, and the Fundamental Research Funds for the Central Universities (10ykjc01 and 10ykjc20).

■ REFERENCES

(1) Cross, D. A.; Alessi, D. R.; Cohen, P.; Andjelkovich, M.; Hemmings, B. A. Inhibition of glycogen synthase kinase-3 by insulin mediated by protein kinase B. *Nature* **1995**, *378*, 785–789.

(2) Hoeflich, K. P.; Luo, J.; Rubie, E. A.; Tsao, M. S.; Jin, O.; Woodgett, J. R. Requirement for glycogen synthase kinase-3beta in cell survival and NF-kappaB activation. *Nature* **2000**, *406*, 86–90.

(3) Woodgett, J. R. Molecular cloning and expression of glycogen synthase kinase-3/factor A. *EMBO J.* **1990**, *9*, 2431–2438.

(4) Ali, A.; Hoeflich, K. P.; Woodgett, J. R. Glycogen synthase kinase-3: properties, functions, and regulation. *Chem. Rev.* **2001**, *101*, 2527–2540.

(5) Eldar-Finkelman, H. Glycogen synthase kinase 3: an emerging therapeutic target. *Trends Mol. Med.* **2002**, *8*, 126–132.

(6) Doble, B. W.; Woodgett, J. R. GSK-3: tricks of the trade for a multi-tasking kinase. *J. Cell. Sci.* **2003**, *116*, 1175–1186.

(7) Martinez, A.; Castro, A.; Dorronsoro, I.; Alonso, M. Glycogen synthase kinase 3 (GSK-3) inhibitors as new promising drugs for diabetes, neurodegeneration, cancer, and inflammation. *Med. Res. Rev.* **2002**, *22*, 373–384.

(8) Chen, R. H.; Ding, W. V.; McCormick, F. Wnt signaling to beta-catenin involves two interactive components. Glycogen synthase kinase-3beta inhibition and activation of protein kinase C. *J. Biol. Chem.* **2000**, *275*, 17894–17899.

(9) Nikoulina, S. E.; Ciaraldi, T. P.; Mudaliar, S.; Mohideen, P.; Carter, L.; Henry, R. R. Potential role of glycogen synthase kinase-3 in skeletal muscle insulin resistance of type 2 diabetes. *Diabetes* **2000**, *49*, 263–271.

(10) Plyte, S. E.; Hughes, K.; Nikolakaki, E.; Pulverer, B. J.; Woodgett, J. R. Glycogen synthase kinase-3: functions in oncogenesis and development. *Biochim. Biophys. Acta* **1992**, *1114*, 147–162.

(11) Manoukian, A. S.; Woodgett, J. R. Role of glycogen synthase kinase-3 in cancer: regulation by Wnts and other signaling pathways. *Adv. Cancer Res.* **2002**, *84*, 203–229.

(12) MacAulay, K.; Woodgett, J. R. Targeting glycogen synthase kinase-3 (GSK-3) in the treatment of Type 2 diabetes. *Expert Opin. Ther. Targets* **2008**, *12*, 1265–1274.

(13) Summers, S. A.; Kao, A. W.; Kohn, A. D.; Backus, G. S.; Roth, R. A.; Pessin, J. E.; Birnbaum, M. J. The role of glycogen synthase kinase 3beta in insulin-stimulated glucose metabolism. *J. Biol. Chem.* **1999**, *274*, 17934–17940.

(14) Baum, L.; Hansen, L.; Masliah, E.; Saitoh, T. Glycogen synthase kinase 3 alteration in Alzheimer disease is related to neurofibrillary tangle formation. *Mol. Chem. Neuropathol.* **1996**, *29*, 253–261.

(15) Bhat, R. V.; Budd Haeberlein, S. L.; Avila, J. Glycogen synthase kinase 3: a drug target for CNS therapies. *J. Neurochem.* **2004**, *89*, 1313–1317.

(16) Van, W. J.; Haefner, B. Glycogen synthase kinase-3 as drug target: from wallflower to center of attention. *Drug News Perspect.* **2003**, *16*, 557–565.

(17) Ajay, B. P.; Chitti, S.; Rajesh, B.; Vishnu, P. V.; Radha Kishen, J. V.; Khadar Vali, R. In silico Based Ligand Design and Docking Studies of GSK-3β Inhibitors. *Chem-Bio Inf. J.* **2010**, *10*, 1–12.

(18) Dessalew, N.; Patel, D. S.; Bharatam, P. V. 3D QSAR and molecular docking studies on pyrazolopyrimidine derivatives as glycogen synthase kinase-3beta inhibitors. *J. Mol. Graphics Model.* **2007**, *25*, 885–895.

(19) Zhang, N.; Jiang, Y.; Zou, J.; Zhang, B.; Jin, H.; Wang, Y.; Yu, Q. 3D QSAR for GSK-3beta inhibition by indirubin analogues. *Eur. J. Med. Chem.* **2006**, *41*, 373–378.

(20) Kunick, C.; Lauenroth, K.; Wieking, K.; Xie, X.; Schultz, C.; Gussio, R.; Zaharevitz, D.; Leost, M.; Meijer, L.; Weber, A.; Jorgensen, F. S.; Lemcke, T. Evaluation and comparison of 3D QSAR CoMSIA models for CDK1, CDK5, and GSK-3 inhibition by paullones. *J. Med. Chem.* **2004**, *47*, 22–36.

(21) Martinez, A.; Alonso, M.; Castro, A.; Dorronsoro, I.; Gelpi, J. L.; Luque, F. J.; Perez, C.; Moreno, F. J. SAR and 3D QSAR studies on thiadiazolidinone derivatives: exploration of structural requirements for glycogen synthase kinase 3 inhibitors. *J. Med. Chem.* **2005**, *48*, 7103–7112.

(22) Zeng, M.; Jiang, Y.; Zhang, B.; Zheng, K.; Zhang, N.; Yu, Q. 3D QSAR studies on GSK-3 inhibition by aloisines. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 395–399.

(23) Lescot, E.; Bureau, R.; Sopkova-de Oliveira Santos, J.; Rochais, C.; Lisowski, V.; Lancelot, J. C.; Rault, S. 3D QSAR and docking studies of selective GSK-3beta inhibitors. Comparison with a thieno[2,3-b] pyrrolizinone derivative, a new potential lead for GSK-3beta ligands. *J. Chem. Inf. Model.* **2005**, *45*, 708–715.

(24) Prasanna, S.; Daga, P. R.; Xie, A. H.; Doerksen, R. J. Glycogen synthase kinase-3 inhibition by 3-anilino-4-phenylmaleimides: insights from 3D QSAR and docking. *J. Comput. Aided Mol. Des.* **2009**, *23*, 113–127.

(25) Patel, D. S.; Bharatam, P. V. Selectivity criterion for pyrazolo-[3,4-b]pyrid[az]ine derivatives as GSK-3 inhibitors: CoMFA and molecular docking studies. *Eur. J. Med. Chem.* **2008**, *43*, 949–957.

(26) Xiao, J.; Guo, Z.; Guo, Y.; Chu, F.; Sun, P. Inhibitory mode of N-phenyl-4-pyrazolo[1,5-b] pyridazin-3-ylpyrimidin-2-amine series derivatives against GSK-3: molecular docking and 3D QSAR analyses. *Protein Eng., Des. Sel.* **2006**, *19*, 47–54.

(27) Katritzky, A. R.; Pacureanu, L. M.; Dobchev, D. A.; Fara, D. C.; Duchowicz, P. R.; Karelson, M. QSAR modeling of the inhibition of glycogen synthase kinase-3. *Bioorg. Med. Chem.* **2006**, *14*, 4987–5002.

(28) Lather, V.; Kristam, R.; Saini, J. S.; Kristam, R.; Karthikeyan, N. A.; Balaji, V. N. QSAR models for prediction of Glycogen Synthase kinase-3 beta inhibitory activity of indirubin derivatives. *QSAR Comb. Sci.* **2008**, *27*, 718–728.

(29) Sivaprakasam, P.; Xie, A. H.; Doerksen, R. J. Probing the physicochemical and structural requirements for glycogen synthase kinase-3 alpha inhibition: 2D-QSAR for 3-anilino-4-phenylmaleimides. *Bioorg. Med. Chem.* **2006**, *14*, 8210–8218.

(30) Gaisina, I. N.; Gallier, F.; Ougolkov, A. V.; Kim, K. H.; Kurome, T.; Guo, S.; Holzle, D.; Luchini, D. N.; Blond, S. Y.; Billadeau, D. D.; Kozikowski, A. P. From a natural product lead to the identification of potent and selective benzofuran-3-yl-(indol-3-yl)maleimides as glycogen synthase kinase 3beta inhibitors that suppress proliferation and survival of pancreatic cancer cells. *J. Med. Chem.* **2009**, *52*, 1853–1863.

(31) *The PubChem Project*; National Center for Biotechnology Information: Bethesda, MD; http://pubchem.ncbi.nlm.nih.gov/. Accessed February 25, 2011).

(32) *Sybyl Molecular Modeling Software Package*, v. 7.3.5, Tripos Inc.: St. Louis, MO, 2008.

(33) Crame, R. D.; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.

(34) Klebe, G. Molecular similarity indices in a comparative molecular field analysis (CoMFA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **1994**, *37*, 4130–4146.

(35) Gasteigner, J.; Marsili, M. A new model for calculating atomic charges in molecules. *Tetrahedron Lett.* **1978**, *19*, 3181–3184.

(36) Stahle, L.; Wold, S. Multivariate data analysis and experimental design in biomedical research. *Prog. Med. Chem.* **1988**, *25*, 291–338.

(37) Bush, B. L.; Nachbar, R. B., Jr. Sample-distance partial least squares: PLS optimized for many variables, with application to CoMFA. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 587–619.

(38) Cho, S. J.; Tropsha, A. Cross-validated R^2-guided region selection for comparative molecular field analysis: a simple method to achieve consistent results. *J. Med. Chem.* **1995**, *38*, 1060–1066.

(39) Golbraikh, A.; Tropsha, A. Beware of q^2!. *J. Mol. Graphics Model.* **2002**, *20*, 269–276.

(40) Tropsha, A.; Gramatica, P.; Gombar, V. K. The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb. Sci.* **2003**, *22*, 69–77.

(41) *FlexX in LeadIT 1.0*; BioSolveIT: Sankt Augustin, Germany, 2009.

(42) Kozikowski, A. P.; Gaisina, I. N.; Petukhov, P. A.; Sridhar, J.; King, L. T.; Blond, S. Y.; Duka, T.; Rusnak, M.; Sidhu, A. Highly potent and specific GSK-3 beta inhibitors that block Tau phosphorylation and decrease alpha-synuclein protein expression in a cellular model of Parkinson's disease. *ChemMedChem* **2006**, *1*, 256–266.

(43) Engler, T. A.; Malhotra, S.; Burkholder, T. P.; Henry, J. R.; Mendel, D.; Porter, W. J.; Furness, K.; Diefenbacher, C.; Marquart, A.; Reel, J. K.; Li, Y.; Clayton, J.; Cunningham, B.; McLean, J.; O'Toole, J, C.; Brozinick, J.; Hawkins, E.; Misener, E.; Briere, D.; Brier, R. A.; Wagner, J. R.; Campbell, R. M.; Anderson, B. D.; Vaughn, R.; Bennett, D. B.; Meier, T. I.; Cook, J. A. The development of potent and selective bisarylmaleimide GSK3 inhibitors. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 899–903.