

# Glycogen synthase kinase-3 inhibition by 3-anilino-4-phenylmaleimides: insights from 3D-QSAR and docking

Sivaprakasam Prasanna · Pankaj R. Daga ·  
Aihua Xie · Robert J. Doerksen

Received: 16 May 2008 / Accepted: 6 September 2008 / Published online: 7 October 2008  
© Springer Science+Business Media B.V. 2008

**Abstract** Glycogen synthase kinase-3, a serine/threonine kinase, has been implicated in a wide variety of pathological conditions such as diabetes, Alzheimer's disease, stroke, bipolar disorder, malaria and cancer. Herein we report 3D-QSAR analyses using CoMFA and CoMSIA and molecular docking studies on 3-anilino-4-phenylmaleimides as GSK-3 $\alpha$  inhibitors, in order to better understand the mechanism of action and structure-activity relationship of these compounds. Comparison of the active site residues of GSK-3 $\alpha$  and GSK-3 $\beta$  isoforms shows that all the key amino acids involved in polar interactions with the maleimides for the  $\beta$  isoform are the same in the  $\alpha$  isoform, except that Asp133 in the  $\beta$  isoform is replaced by Glu196 in the  $\alpha$  isoform. We prepared a homology model for GSK-3 $\alpha$ , and showed that the change from Asp to Glu should not affect maleimide binding significantly. Docking studies revealed the binding poses of three subclasses of these ligands, namely anilino, N-methylanilino and indoline derivatives, within the active site of the  $\beta$  isoform, and helped to explain the difference in their inhibitory activity.

**Keywords** 3D-QSAR · CoMFA · CoMSIA · Docking · GSK-3 · Maleimides

## Introduction

Glycogen synthase kinase-3 (GSK-3), a serine/threonine kinase [1], is involved in many key biochemical processes [2] such as phosphorylation of glycogen synthase in the rate-limiting step of glycogen biosynthesis and tau hyperphosphorylation. Inhibition of GSK-3 is being considered as a potential intervention to treat such devastating diseases as type-2 diabetes [3, 4], cancer [5–7], malaria [8–10], cardiovascular disease [11, 12], Alzheimer's disease [13–17], bipolar disorder [13, 18, 19], and neuroinflammation [19–21]. The thought that GSK-3 inhibition could simultaneously arrest two or more diseases is compelling but naturally concerns over selectivity are crucial. Against this backdrop we feel strongly that there is a great need for rational drug design and have commenced steps in this direction [22].

Three isoforms of GSK-3,  $\alpha$ ,  $\beta$  and  $\beta 2$ , have been identified in mammalian cells and highly homologous GSK-3's have been found in other species [10]. Much of the early work on human GSK-3 did not distinguish between isoforms or looked exclusively at GSK-3 $\beta$ . Recent reports have presented strong evidence that GSK-3 $\alpha$  and GSK-3 $\beta$  play distinct roles in normal and abnormal function [17, 21] and hence there could be different therapeutic value in developing isoform-nonspecific and isoform-selective inhibitors. The GSK-3 $\alpha$  and GSK-3 $\beta$  isoforms have 97% sequence similarity [16] in their kinase cores ( $\sim 308$  residues), but the  $\alpha$  isoform is 63 residues longer at its N-terminus than the  $\beta$  isoform. The

S. Prasanna · P. R. Daga · A. Xie · R. J. Doerksen (✉)  
Department of Medicinal Chemistry, School of Pharmacy,  
University of Mississippi, University, MS 38677-1848, USA  
e-mail: rjd@olemiss.edu

R. J. Doerksen  
Research Institute of Pharmaceutical Sciences,  
School of Pharmacy, University of Mississippi,  
University, MS 38677-1848, USA

N-terminal phosphorylation site is Ser21 for GSK-3 $\alpha$  and Ser9 for GSK-3 $\beta$ . This site is embedded in a conserved seven residue motif in both isoforms. Ali et al. [2] have reviewed the biochemical and functional properties of GSK-3. The first X-ray crystal structure of GSK-3 $\beta$  was reported in 2001 [23], and several X-ray structures containing co-crystallized ligands have also been reported, including one containing a 3-anilino-4-arylmaleimide [24], but no X-ray structure has been reported for GSK-3 $\alpha$ .

Though many classes of GSK-3 inhibitors have been reported [25], in this work we chose to study 74 maleimides from Smith et al. [26] which were shown to be selective towards GSK-3, from amongst a wide range of kinases. Smith et al. [26] focused on GSK-3 $\alpha$  and discovered a maleimide hit through high throughput screening of the SmithKline Beecham compound collection [27]. They used in vitro binding to rabbit GSK-3 $\alpha$ , which shows 95% homology to human GSK-3 $\alpha$  and is >99% identical in the kinase domain [26]. The maleimides were some of the first inhibitors to be assayed specifically against the  $\alpha$  isoform. Smith et al. [26] found that the compounds were approximately equally effective at inhibiting human GSK-3 $\beta$  (IC<sub>50</sub> data not reported) and both human and rabbit GSK-3 $\alpha$  [26, 27].

Herein we report 3D-QSAR (quantitative structure-activity relationship) analysis using CoMFA (comparative molecular field analysis) and CoMSIA (comparative molecular similarity indices analysis) [28]; sequence alignment to compare the active site amino acids between the two isoforms, GSK-3 $\alpha$  and 3 $\beta$ ; comparison of the 3D conformation of GSK-3 $\alpha$  and GSK-3 $\beta$  using a homology model for GSK-3 $\alpha$  which we prepared based on an X-ray crystal structure of GSK-3 $\beta$ ; and molecular docking to obtain putative ligand-protein binding poses of arylmaleimides in order to understand better the physicochemical requirements for inhibition of GSK-3. Previously we published a 2D-QSAR analysis using classical Hansch [29] and Fujita-Ban methods for 67 of the 74 Smith et al. maleimides, including 3-anilino and 3-N-methylanilino-maleimides [22]. Some 3D-QSAR and other molecular modeling studies have been reported on inhibitors of GSK-3 [30–40]. Bharatam et al. [31] reported pharmacophore models of some compounds assayed against the  $\beta$  isoform and included the GSK-3 $\alpha$  data for maleimide **52** (our numbering) in their training sets. Lescot et al. [35] reported a CoMFA [41] and docking study of the Smith et al. [26] compounds. Important shortcomings of the CoMFA report of Lescot et al. [35] and our thorough and systematic CoMFA and CoMSIA analyses together with the novel insights we gained will be discussed below.

## Materials and methods

### 3D-QSAR (CoMFA and CoMSIA)

#### *Training and test set*

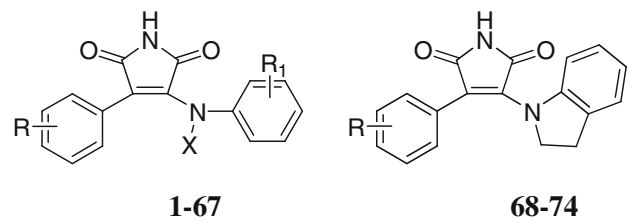
The series of 4-arylmaleimide derivatives included 64 with 3-anilino-, four compounds with 3-N-methylanilino- (one of these had no reported activity so we left it out) and seven compounds with 3-indolino- substituents. In Table 1, the structures and activities of the 74 ligands are listed. The data set was divided into training ( $n = 56$ ) and test sets, using great care to ensure a uniform distribution of structurally different compounds from each class with a wide range of pIC<sub>50</sub> in both sets. The resulting mean log activity of training and test set compounds matched well: 6.75 and 6.61, respectively.

#### *Geometry optimization*

The chemical structures were drawn in Sybyl and the geometries were optimized using the MMFF94 force field using a distance-dependent dielectric function and convergence criterion of 0.001 kcal/mol. In order to validate the use of MMFF94, we performed Hartree-Fock and B3LYP 6-31G\*\* geometry optimizations [42] of the minimum energy conformation of **38** and of the bioactive conformation of **52**.

#### *Alignment*

CoMFA and CoMSIA require an alignment protocol for the ligands. We investigated two templates to which all the ligands were aligned. For the first template, we chose the minimum energy conformation of the most active compound **38** as the putative bioactive conformation (“ME template”). To prepare the minimum energy conformation, we performed a systematic conformational search in vacuum for **38**, using the MMFF94 force field in Sybyl, followed by minimization of the lowest energy conformation generated. For the second template, we extracted **52** from its published X-ray co-crystal structure with GSK-3 $\beta$  (PDB code: 1Q4L) [24] and then used MMFF94 to optimize it to a local minimum, with no significant conformational change (“XR template”). We performed two separate alignments, one for the ME template and one for the XR template, using Sybyl’s automated database alignment procedure. The common scaffold fragment of the template molecule used for alignment is shown in red in Fig. 1.

**Table 1** Observed, converted and best CoMFA and CoMSIA predicted activity of 3-anilino-4-arylmaleimides


No.	R	R <sub>1</sub>	X	pIC <sub>50</sub>		
				Expt.	CoMFA	CoMSIA
1 <sup>a</sup>	H	H	H	6.28	6.24	6.30
2	2-Cl	H	H	6.67	6.50	6.54
3 <sup>a</sup>	2-OCH <sub>3</sub>	H	H	6.67	6.74	6.75
4	3-NO <sub>2</sub>	H	H	6.85	6.92	6.97
5	4-Cl	H	H	6.29	6.29	6.30
6	4-OCH <sub>3</sub>	H	H	6.41	6.37	6.42
7	H	3-Cl	H	6.52	6.55	6.52
8	2-Cl	3-Cl	H	6.71	6.74	6.76
9 <sup>a</sup>	2-OCH <sub>3</sub>	3-Cl	H	6.94	6.96	6.91
10	2-NO <sub>2</sub>	3-Cl	H	6.98	7.03	6.97
11	3-OCH <sub>3</sub>	3-Cl	H	6.59	6.67	6.62
12 <sup>a</sup>	3-NO <sub>2</sub>	3-Cl	H	7.15	7.06	7.09
13	4-Cl	3-Cl	H	6.35	6.56	6.53
14	4-OCH <sub>3</sub>	3-Cl	H	6.81	6.70	6.64
15	H	3-OH	H	6.15	6.18	6.19
16	2-Cl	3-OH	H	6.43	6.44	6.42
17	2-OCH <sub>3</sub>	3-OH	H	6.59	6.67	6.58
18	2-NO <sub>2</sub>	3-OH	H	6.60	6.57	6.63
19 <sup>a</sup>	3-Cl	3-OH	H	5.83	6.37	6.35
20	3-OCH <sub>3</sub>	3-OH	H	6.33	6.34	6.30
21	3-NO <sub>2</sub>	3-OH	H	6.63	6.78	6.74
22 <sup>a</sup>	4-Cl	3-OH	H	6.39	6.29	6.19
23	4-OCH <sub>3</sub>	3-OH	H	6.32	6.31	6.31
24	3-NO <sub>2</sub>	4-OH	H	6.91	6.93	6.99
25 <sup>a</sup>	4-Cl	4-OH	H	6.50	6.42	6.44
26	2-Cl	3-Cl,4-OH	H	6.82	6.85	6.87
27 <sup>a</sup>	2-OCH <sub>3</sub>	3-Cl,4-OH	H	6.86	7.13	7.03
28	2-NO <sub>2</sub>	3-Cl,4-OH	H	6.98	6.88	6.92
29	3-Cl	3-Cl,4-OH	H	7.03	6.86	6.83
30	3-NO <sub>2</sub>	3-Cl,4-OH	H	7.23	7.22	7.22
31	4-Cl	3-Cl,4-OH	H	6.76	6.73	6.67
32 <sup>a</sup>	H	3,5-diCl,4-OH	H	6.83	6.85	6.90
33	2-Cl	3,5-diCl,4-OH	H	7.03	6.99	7.16
34	2-OCH <sub>3</sub>	3,5-diCl,4-OH	H	7.09	7.19	7.22
35	2-NO <sub>2</sub>	3,5-diCl,4-OH	H	7.28	7.46	7.44
36 <sup>a</sup>	3-Cl	3,5-diCl,4-OH	H	7.24	7.04	7.08
37	3-OCH <sub>3</sub>	3,5-diCl,4-OH	H	6.85	7.04	7.03
38	3-NO <sub>2</sub>	3,5-diCl,4-OH	H	7.70	7.53	7.50

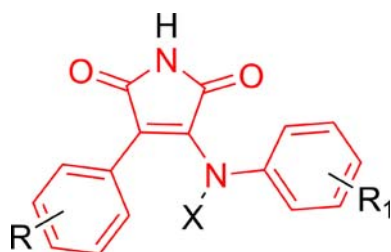
**Table 1** continued

No.	R	R <sub>1</sub>	X	pIC <sub>50</sub>		
				Expt.	CoMFA	CoMSIA
39	4-Cl	3,5-diCl,4-OH	H	7.04	6.91	6.91
40	4-OCH <sub>3</sub>	3,5-diCl,4-OH	H	7.08	6.99	7.02
41	4-NO <sub>2</sub>	3,5-diCl,4-OH	H	7.15	7.05	7.05
42	H	3-COOH	H	6.54	6.58	6.63
43 <sup>a</sup>	2-Cl	3-COOH	H	6.87	6.76	6.80
44	3-Cl	3-COOH	H	6.87	6.83	6.91
45	3-OCH <sub>3</sub>	3-COOH	H	6.71	6.74	6.74
46 <sup>a</sup>	3-NO <sub>2</sub>	3-COOH	H	7.10	7.18	7.20
47	4-Cl	3-COOH	H	6.73	6.61	6.59
48	4-OCH <sub>3</sub>	3-COOH	H	6.67	6.68	6.69
49 <sup>a</sup>	H	4-Cl,3-COOH	H	6.85	6.79	6.85
50	2-Cl	4-Cl,3-COOH	H	7.13	7.24	7.26
51	2-NO <sub>2</sub>	4-Cl,3-COOH	H	7.55	7.53	7.45
52	3-Cl	4-Cl,3-COOH	H	7.12	7.11	7.14
53	3-OCH <sub>3</sub>	4-Cl,3-COOH	H	7.07	7.08	7.08
54	3-NO <sub>2</sub>	4-Cl,3-COOH	H	7.59	7.57	7.56
55	4-Cl	4-Cl,3-COOH	H	6.96	6.99	6.98
56	H	4-SCH <sub>3</sub>	H	6.39	6.36	6.38
57 <sup>a</sup>	2-Cl	4-SCH <sub>3</sub>	H	6.79	6.50	6.53
58	2-OCH <sub>3</sub>	4-SCH <sub>3</sub>	H	6.96	6.82	6.72
59	3-Cl	4-SCH <sub>3</sub>	H	6.27	6.55	6.54
60	3-OCH <sub>3</sub>	4-SCH <sub>3</sub>	H	6.69	6.53	6.48
61	3-NO <sub>2</sub>	4-SCH <sub>3</sub>	H	6.82	6.90	6.94
62 <sup>a</sup>	4-Cl	4-SCH <sub>3</sub>	H	6.28	6.43	6.39
63	4-OCH <sub>3</sub>	4-SCH <sub>3</sub>	H	6.61	6.52	6.50
64	4-NO <sub>2</sub>	4-SCH <sub>3</sub>	H	6.41	6.56	6.51
65	H	H	CH <sub>3</sub>	5.58	5.55	5.61
66 <sup>a</sup>	3-NO <sub>2</sub>	H	CH <sub>3</sub>	5.85	6.06	6.16
67	4-Cl	H	CH <sub>3</sub>	5.64	5.62	5.60
68	2-Cl	–	–	6.47	6.49	6.49
69 <sup>a</sup>	2-OMe	–	–	6.73	6.64	6.64
70	2-NO <sub>2</sub>	–	–	6.88	6.88	6.77
71	3-Cl	–	–	6.34	6.35	6.36
72	3-NO <sub>2</sub>	–	–	6.79	6.70	6.77
73 <sup>a</sup>	4-Cl	–	–	5.85	6.21	6.20
74	4-OMe	–	–	6.16	6.26	6.31

<sup>a</sup> Compounds used as test set for 3D-QSAR analyses

### CoMFA and CoMSIA model development

CoMFA fields were generated using an sp<sup>3</sup> carbon as a probe atom with a formal charge of +1 to generate steric and electrostatic fields at each point in a 3D grid around the assembly of aligned compounds. The electrostatic field generated is based on a Coulomb potential whereas the steric field is based on a Lennard-Jones 6–12 potential. The cutoff value for the steric and electrostatic energy in



**Fig. 1** Structure of 3-anilino-4-phenylmaleimides, with 4-phenyl substituents labeled R and anilino substituents labeled R<sub>1</sub>. The fragment in red was used as a template for ligand database alignment

CoMFA was 30 kcal/mol. Scaling of variables was done using the CoMFA-STD method. For CoMSIA, we used hydrophobic, hydrogen bond donor, and hydrogen bond acceptor fields, in addition to the steric and electrostatic fields also used in CoMFA. An attenuation factor of 0.3 was used in the CoMSIA studies. Cross-validated SAMPLS (sample partial least squares) was carried out for an increasing number of components ( $N$ ) for each QSAR model. The optimal number of components for PLS analysis was determined by optimizing  $q^2$ : an additional component was allowed only if it increased  $q^2$  by more than 5%. For all models  $N < n/3$ , for  $n$ , the number of training set compounds, in accordance with the general rule of thumb. The optimal  $N$  was used in non-validated PLS analysis to obtain other statistical parameters: conventional or non cross-validated  $r^2$ ,  $SEE$  and  $F$ . Golbraikh and Tropsha emphasized that high internal predictivity ( $q^2$ ) is not sufficient for a robust 3D-QSAR analysis, but that high external predictivity ( $r^2_{pred}$ ) is also crucial [43]. Thus, we predicted the activity of the test set compounds using the derived CoMFA and CoMSIA models and calculated  $r^2_{pred} = 1 - \left[ \frac{\sum (Y_{obs} - Y_{pred})^2}{\sum (Y_{obs} - Y_{mean})^2} \right]$ , where  $Y_{pred}$ ,  $Y_{obs}$  and  $Y_{mean}$  are the predicted, observed and mean observed activities, respectively, of the test set molecules. It is common for either test or training set average to be used for  $Y_{mean}$ ; we chose to use the average of test set activities as implemented in the Sybyl software. The CoMFA and CoMSIA results were visualized and interpreted with the help of contour maps using STDEV\*COEFF field type.

### Homology model of GSK-3 $\alpha$

Sequence alignment of GSK-3 $\alpha$  and GSK-3 $\beta$  was performed with ClustalW [44] using default settings. This was facile because of their high percent identity. The GSK-3 $\alpha$  sequence was obtained from NCBI (Accession number: AAD11986). The homology model was built using Prime from Schrödinger [45]. The GSK-3 $\alpha$  sequence was aligned with the template GSK-3 $\beta$  (pdb accession number: 1Q4L, chain A) based on the sequence alignment protocol

available within Prime, which uses a combination of sequence and secondary structure information. The Prime sequence alignment was identical to that of ClustalW for the binding pocket region. For identical residues, the backbone atoms and conserved side chain coordinates for the GSK-3 $\alpha$  homology model were extracted from the GSK-3 $\beta$  template, whereas all the non-identical residues were predicted using Prime during model construction. The co-ordinates of the bound ligand, **52**, in the template were transferred to the built model. The final model was subjected to an all-atom minimization using the OPLS2000 force field in Prime.

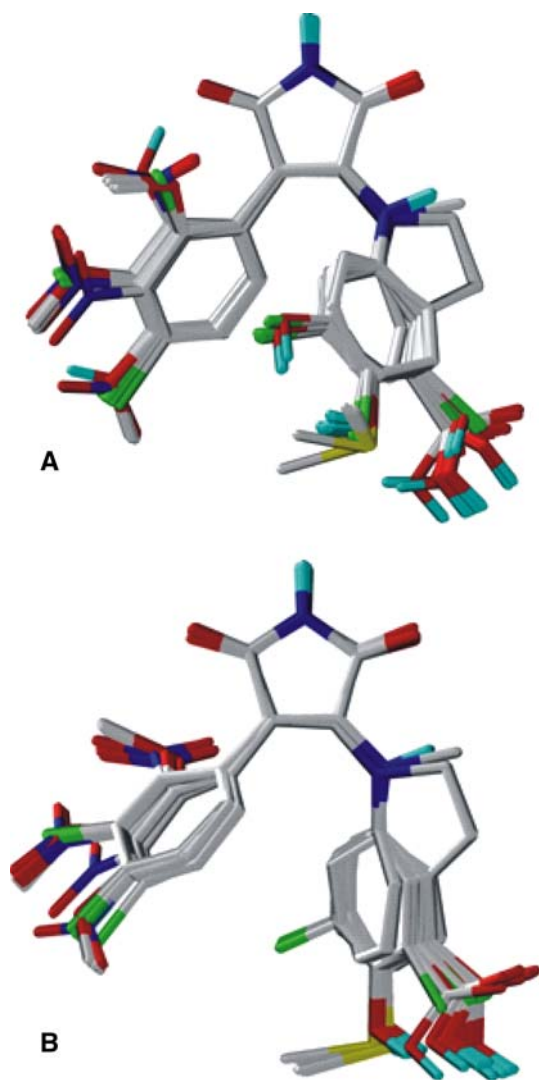
### Molecular docking

FlexX [46] implemented in Sybyl 7.0 was used for molecular docking studies. Structural information for the protein was obtained from the GSK-3 $\beta$  X-ray crystal structure (1Q4L) [24], which contains a co-crystallized ligand, **52**. The active site was defined as all complete amino acids containing any atom within 6.5 Å of any ligand atom, and a water molecule present in the crystal structure of the active site was also included. Flexible docking of each ligand into the active site was performed after removal of the co-crystallized ligand. The best out of 30 docking poses was found for each of a selection of compounds. We considered four scoring functions, D\_score, G\_score, ChemScore and PMF\_score, but found the FlexX total score gave the best correlation with activity.

## Results

### Alignment template comparison

Figure 2 shows the ligand alignment to the ME and XR templates of both training and test sets. Several key geometrical parameters for the template maleimides **38** and **52** are shown in Table 2. In Fig. 3 the atom numbering for the parameters is shown. The bond lengths agree well between X-ray, force field and DFT, except for C3C4 and C6C7, which are most likely overestimated in the X-ray structure refinement of the 1Q4L:**52** complex. Another GSK-3 $\beta$  X-ray structure in the PDB containing a 3-indolyl-4-aryl-maleimide inhibitor, 1R0E, has the corresponding C3C4 = 1.36 Å and C6C7 = 1.40 Å, which matches the calculated bond lengths for **52**. The bond angles agree well except for two involving the aniline N, C3C4N5 and C4N5C6, which are overestimated and underestimated, respectively, in the calculations. The dihedral angles of **52** from X-ray and calculation agree qualitatively but there are discrepancies



**Fig. 2** Alignment of all the ligands: (a) To the ME template, based on the minimum energy conformation of **38**; (b) To the XR template, based on GSK-3 $\beta$  X-ray co-crystal conformation of **52**

of as much as 30°. These differences are likely due to crystal packing effects in the X-ray structure.

The main difference between the ME and XR template is in the relative orientation of the two side rings with respect to the central maleimide ring. The most useful thing to note from Table 2 is that while the bond lengths and bond angles of the ME and XR templates agree well, the dihedral angles involving atoms 1–7 (Fig. 3) are different, and approximately fit the pattern  $\theta$  in XR and  $360^\circ - \theta$  in ME, respectively. The lowest C4N5C6C7 dihedral angle for the 3-anilino ring is  $-14^\circ$  for the ME template but  $+13^\circ$  for the XR template. For the 4-phenyl ring, the lowest C1C2C3C4 dihedral angle is  $-44^\circ$  for the ME template but  $+43^\circ$  for the XR template. For the central maleimide with 3-anilino ring, the C3C4N5C6 dihedral angle is  $-43^\circ$  for the ME template and  $+46^\circ$  for the XR

template. The energetic difference between the two conformations is small and the one which is lowest in energy should depend on the particular functional groups attached on the rings, on their position of substitution and on the solvent environment.

## CoMFA

### Model statistics

CoMFA model statistics are summarized in Table 3. Consider first the models for the ME template. The single-field models are reasonably accurate, with the electrostatic (*E*) model substantially better than the steric (*S*) one. The *S* + *E* model is an improvement on either single-field model. For the *S* + *E* model, the *S* and *E* contributions are nearly the same. All the statistical measures show that the models for the XR template are significantly better than the corresponding ones for the ME template. For the XR template,  $r^2$ , *SEE*, and *F* are better for the *S* model than for the *E* only model, but  $q^2$  and  $r^2_{\text{pred}}$  are higher for the *E* model. The *S* + *E* model is the best CoMFA model, with  $q^2 = 0.844$ ,  $r^2 = 0.942$  and  $r^2_{\text{pred}}$  only slightly lower than for the *E* only model.

The importance of including the electrostatic field in the CoMFA model was overlooked in results published on the same data set by Lescot et al. [35] who stated that inclusion of electrostatic fields only led to a slight improvement in model statistical measures. Lescot et al. used an alignment based on an X-ray co-crystal structure very similar to our XR template alignment. For this data set,  $q^2$  for the *E* model is higher than for the *S* model. The statistical parameters for the *S* + *E* model are again better than those for the *S* or *E* only models. Thus it is essential to include the *E* field to get the best CoMFA model for this data set. Apart from the significant statistical improvement in the *S* + *E* models versus the *S*-only models, the value of having an *S* + *E* model is that we can use it to make predictions of which local electrostatic modifications will improve ligand-protein binding.

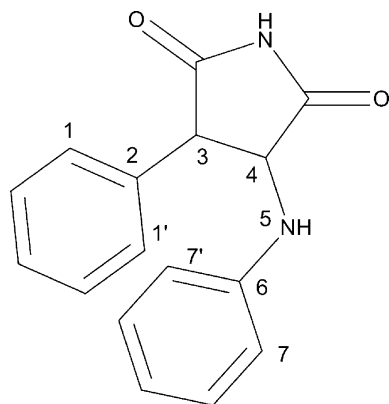
Note that Lescot et al. [35] used all 74 compounds in their model training set and did not have an external test set. Thus they failed to validate their models externally, which is a necessary criterion for a 3D-QSAR model [43]. Table 3 shows our models for the 74 compound training set. The statistical measures evaluating the models are all worse than for the 56/18 training/test set models. In addition our best model has the reliability verification provided by  $r^2_{\text{pred}} = \sim 0.8$ . Thus the *S* + *E* model with the XR template and 56/18 training/test set ratio is our best CoMFA model. The predicted activities of all the compounds using this best CoMFA model are given in Table 1.



**Table 2** Geometrical properties<sup>a</sup> for **52** (used for the XR template) and **38** (used for the ME template) (atoms 1–7 as numbered in Fig. 3)

	<b>52</b>				<b>38</b>		
	X-ray <sup>b</sup>		FF <sup>c</sup>	DFT <sup>d,e</sup>	FF <sup>c</sup>	HF <sup>e,f</sup>	DFT <sup>d,e</sup>
Bond lengths	Chain A	Chain B					
C1C2	1.431	1.433	1.404	1.407	1.403	1.394	1.405
C2C3	1.428	1.449	1.468	1.468	1.467	1.478	1.467
C3C4	1.437	1.444	1.346	1.371	1.346	1.345	1.374
C4N5	1.413	1.416	1.378	1.360	1.375	1.349	1.355
N5C6	1.399	1.385	1.410	1.405	1.411	1.416	1.408
C6C7	1.437	1.433	1.399	1.402	1.400	1.387	1.398
Bond angles							
C1C2C3	121.0	122.0	120.8	120.8	120.4	120.8	120.8
C2C3C4	130.6	131.6	128.5	131.8	128.4	131.6	132.0
C3C4N5	131.7	132.1	135.1	137.7	135.1	136.5	138.1
C4N5C6	134.5	134.7	127.9	131.1	128.5	127.2	131.2
N5C6C7	121.5	122.0	123.8	123.0	123.6	120.5	122.4
Torsions							
C1C2C3C4	41.4	40.6	43.4	31.9	315.9	315.0	329.4
C2C3C4N5	4.6	6.4	1.1	8.7	359.1	355.7	351.3
C3C4N5C6	18.4	13.6	45.5	23.0	317.1	341.4	340.1
C4N5C6C7	39.1	42.5	12.7	18.8	346.3	310.3	336.2

<sup>a</sup> Bond lengths in Å, angles in degrees; <sup>b</sup> 1Q4L with **52**: Chains A and B are both present in the crystal structure of 1Q4L; <sup>c</sup> MMFF94; <sup>d</sup> B3LYP; <sup>e</sup> 6-31G\*\* basis set; <sup>f</sup> Hartree-Fock

**Fig. 3** Numbering of atomic centers for the maleimide geometrical parameters listed in Table 2

### Steric field predictions

Figure 4 shows the steric contour maps obtained through the best CoMFA model around the most active compound of the series, **38**. The green regions show where steric bulk will increase GSK-3 $\alpha$  inhibitory activity whereas the yellow regions show where bulky groups will decrease the GSK-3 $\alpha$  inhibitory activity. Our *S* results generally agree with those of Lescot et al. [35] but some of the polyhedra are of different size. Two green polyhedra around the *ortho* and *meta* positions of the 4-phenyl ring suggest that bulky

substituents at these positions will increase GSK-3 $\alpha$  inhibitory activity. This confirms our Hansch 2D-QSAR findings that *ortho* steric descriptors  $\sigma$ ,  $f$ , and  $E^s$  for the 4-phenyl ring (where  $\sigma$  = Hammett electronic substituent constant,  $f$  = inductive parameter and  $E^s$  = steric descriptor), are positively correlated to activity [22]. The other big green region surrounding the *meta* and *para* positions of the 3-anilino ring shows that bulky groups in these positions enhance activity. In the model of Lescot et al. [35] this region is much smaller. The most active compounds of the series contain, on the 3-anilino ring, either 3,5 dichloro-; 4-hydroxy-; or 3-carboxyl-, 4-chloro-groups, rather than 3-chloro-, 4-hydroxy groups. The yellow polyhedron near the NH connecting the 3-anilino ring to the maleimide ring shows that bulky substitution at this position is detrimental to activity. The N-methyl anilino derivatives **65–67** and indoline derivatives **68–74** were less active than anilino derivatives.

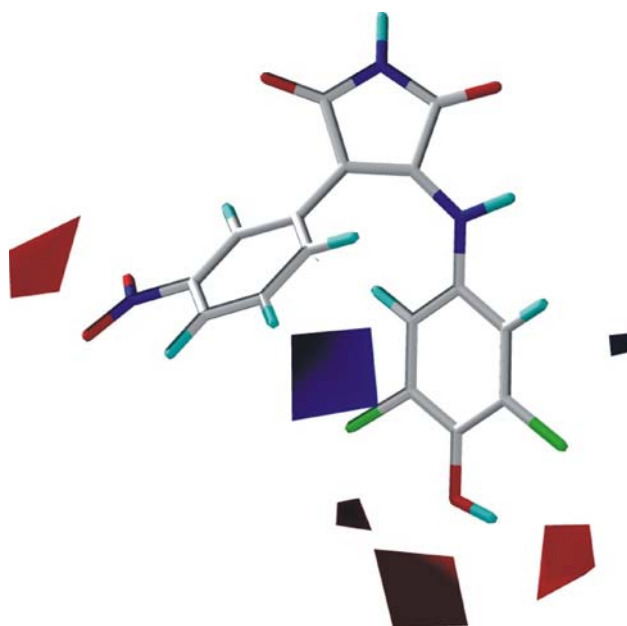
### Electrostatic field predictions

Figure 5 shows the electrostatic field contour maps. The red (blue) regions show where electronegative groups will increase (decrease) the GSK-3 $\alpha$  inhibitory activity. A red polyhedron around the *meta* position of the 4-phenyl ring indicates that increased negative charge at that position is favorable for activity. This is in good agreement with our

**Table 3** Statistical parameters of CoMFA models<sup>a</sup>

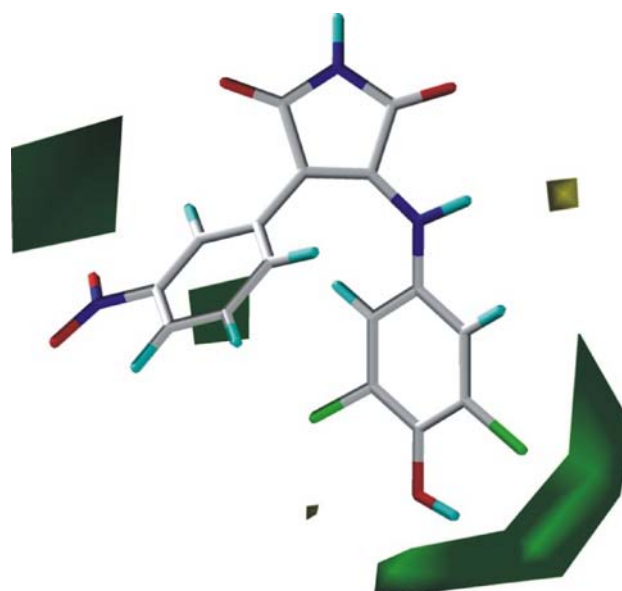
	$N$	$n$	$r^2$	$q^2$	$SEE$	$F$	$r^2_{\text{pred}}$	Field contribution	
								$S$	$E$
ME template									
$S$	4	56	0.753	0.635	0.212	38.9	0.322	1.0	–
$E$	4	56	0.879	0.764	0.148	92.9	0.739	–	1.0
$S + E$	6	56	0.926	0.779	0.118	101.7	0.716	0.514	0.486
XR template									
$S$	6	56	0.934	0.794	0.111	116.4	0.770	1.0	–
$E$	6	56	0.916	0.806	0.126	89.2	0.804	–	1.0
$S + E$	<b>5</b>	<b>56</b>	<b>0.942</b>	<b>0.844</b>	<b>0.104</b>	<b>162.5</b>	<b>0.779</b>	<b>0.512</b>	<b>0.488</b>
$S$	6	74	0.909	0.792	0.132	111.5	–	1.0	–
$E$	6	74	0.891	0.805	0.145	91.1	–	–	1.0
$S + E$	5	74	0.917	0.828	0.125	150.6	–	0.510	0.490

<sup>a</sup> The best CoMFA model used for the activity prediction in Table 1 is given in bold. *n*, number of compounds; *N*, number of components; *r*<sup>2</sup> and *q*<sup>2</sup> are conventional and leave-one-out cross validated correlation coefficient, respectively; *SEE* and *F* are the standard error of estimate and statistical significance of non-validated models; *r*<sup>2</sup><sub>pred</sub> = correlation coefficient for test set



**Fig. 5** CoMFA electrostatic contour map based on the 56 compound training set: the red (blue) regions show where electronegative groups increase (decrease) the GSK-3 $\alpha$  inhibitory activity. The most active member of the series, **38**, is also shown

2D-QSAR model showing the positive contribution of the electronic parameter,  $\sigma_{\text{metaR}}$ , suggesting that electron withdrawing groups at the *meta* position of R are crucial for GSK-3 $\alpha$  inhibitory activity [22]. Also, compounds with *meta* chloro groups (which are weakly electron-withdrawing due to inductive effects) are comparatively less potent than their corresponding nitro derivatives (**38** vs. **36**, pIC<sub>50</sub> = 7.70 vs. 7.24; **54** vs. **52**, pIC<sub>50</sub> = 7.59 vs. 7.12, respectively). The three small red polyhedra around the



**Fig. 4** CoMFA steric contour map based on the 56 compound training set: the green (yellow) regions show where steric bulk increases (decreases) the GSK-3 $\alpha$  inhibitory activity. The most active member of the series, **38**, is also shown

3-anilino ring show the importance of electron-withdrawing groups at the *meta* and *para* positions. This is in good agreement with the SAR data by Smith et al. [26] showing that 3-carboxyl; 3-carboxyl with 4-chloro; or 3-chloro or 3,5-dichloro with 4-hydroxy groups on the anilino ring are conducive to activity. The big blue polyhedron is actually around the NH linker of the aniline ring (Fig. 5) and indicates a favorable region for electropositive groups in the ligand to interact with complementary electronegative (electron rich) groups in amino acids of the active site.

## CoMSIA

## Model selection

CoMSIA models with both ME and XR template are listed in Table 4. The statistical measures of accuracy show that the CoMSIA models based on the ME template are almost all worse than those based on the XR template, as was found for CoMFA. Similarly, for the XR template, models with  $n = 74$  (no test set) have worse statistical measures than the  $n = 56$  models. Considering the XR template with

$n = 56$ , the best CoMSIA model with individual fields is for the  $E$  field alone. The 2-field model  $S + E$  is an improvement over the  $E$  model. The CoMSIA  $S + E$  model was consistent with but slightly less robust than the CoMFA  $S + E$  model in all statistical parameters except external prediction. The hydrophobicity ( $H$ ) plus hydrogen bond acceptor ( $A$ ) model has the highest  $r^2_{\text{pred}}$ , though its  $r^2$  and  $q^2$  are lower. The best 3-field model included the hydrogen bond donor field ( $D$ ):  $S + E + D$ . It was the best CoMSIA model with respect to  $q^2$ ,  $r^2$ ,  $SE$  and  $F$  and had a suitably high  $r^2_{\text{pred}} = 0.803$ . In fact it was better than the

**Table 4** Statistical parameters of CoMSIA models

	$N$	$n$	$r^2$	$q^2$	$SEE$	$F$	$r^2_{\text{pred}}$	Field contribution				
								$S$	$E$	$H$	$A$	$D$
ME template												
$S$	4	56	0.748	0.515	0.214	37.9	0.566	1.0	–	–	–	–
$E$	4	56	0.771	0.591	0.204	43.0	0.496	–	1.0	–	–	–
$S + E$	6	56	0.883	0.637	0.148	61.9	0.633	0.162	0.838	–	–	–
$S + E + H$	6	56	0.909	0.675	0.131	81.5	0.754	0.096	0.606	0.298	–	–
$S + E + H + A$	6	56	0.923	0.727	0.120	98.4	0.748	0.070	0.440	0.263	0.227	–
$S + E + H + D$	6	56	0.920	0.675	0.123	93.7	0.743	0.074	0.484	0.243	–	0.199
$S + E + H + A + D$	6	56	0.921	0.726	0.122	95.0	0.745	0.199	0.375	0.209	0.197	0.169
XR template												
$S$	6	56	0.857	0.735	0.164	48.9	0.673	1.0	–	–	–	–
$E$	6	56	0.915	0.786	0.126	88.3	0.761	–	1.0	–	–	–
$H$	6	56	0.842	0.692	0.172	43.6	0.459	–	–	1.0	–	–
$A$	6	56	0.802	0.660	0.193	33.2	0.539	–	–	–	1.0	–
$D$	5	56	0.730	0.601	0.224	27.0	0.761	–	–	–	–	1.0
$S + E$	6	56	0.931	0.831	0.114	110.7	0.789	0.204	0.796	–	–	–
$H + A$	5	56	0.847	0.727	0.170	75.2	0.845	–	–	0.638	0.362	–
$H + D$	6	56	0.910	0.826	0.130	85.6	0.785	–	–	0.588	–	0.412
$A + D$	6	56	0.858	0.734	0.164	49.4	0.798	–	–	–	0.434	0.566
$S + E + H$	5	56	0.915	0.782	0.127	88.0	0.748	0.090	0.534	0.376	–	–
$S + E + A$	6	56	0.923	0.804	0.121	97.3	0.778	0.132	0.657	–	0.211	–
$S + E + D^a$	<b>6</b>	<b>56</b>	<b>0.932</b>	<b>0.833</b>	<b>0.113</b>	<b>111.7</b>	<b>0.803</b>	<b>0.153</b>	<b>0.602</b>	–	–	<b>0.245</b>
$H + A + D$	6	56	0.909	0.823	0.131	81.3	0.798	–	–	0.462	0.210	0.328
$S + E + H + A$	6	56	0.916	0.786	0.126	89.4	0.731	0.072	0.455	0.331	0.143	–
$S + E + H + D$	6	56	0.928	0.822	0.117	105.2	0.803	0.071	0.391	0.314	–	0.224
$S + E + H + A + D$	6	56	0.925	0.822	0.119	101.4	0.798	0.187	0.344	0.292	0.110	0.197
$S$	6	74	0.805	0.706	0.193	46.2	–	1.0	–	–	–	–
$E$	6	74	0.894	0.777	0.143	93.8	–	–	1.0	–	–	–
$S + E$	6	74	0.906	0.817	0.134	108.2	–	0.213	0.787	–	–	–
$S + E + H$	6	74	0.896	0.781	0.141	96.3	–	0.095	0.468	0.436	–	–
$S + E + A$	6	74	0.904	0.800	0.136	105.0	–	0.138	0.588	–	0.274	–
$S + E + D$	5	74	0.901	0.819	0.137	123.2	–	0.153	0.589	–	–	0.258
$S + E + H + A$	6	74	0.896	0.774	0.141	96.5	–	0.073	0.388	0.362	0.177	–
$S + E + H + D$	6	74	0.903	0.810	0.136	103.6	–	0.072	0.344	0.368	–	0.217
$S + E + H + A + D$	6	74	0.903	0.806	0.138	100.6	–	0.057	0.297	0.327	0.134	0.186

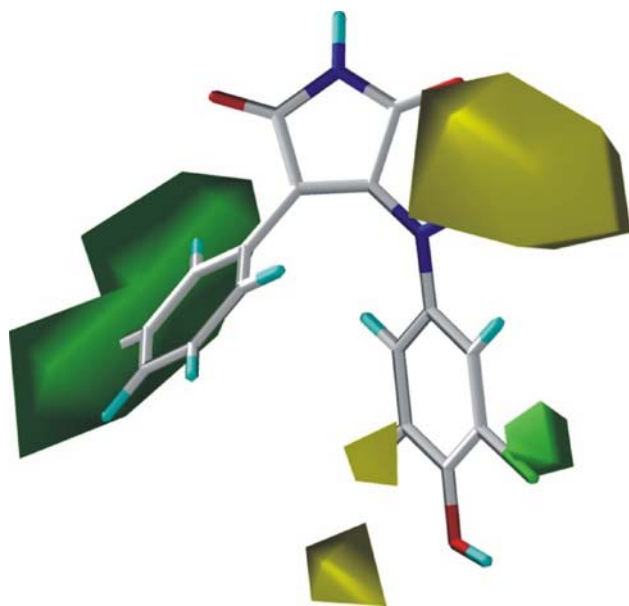
<sup>a</sup> The best CoMSIA model used for the activity prediction in Table 1 is given in bold



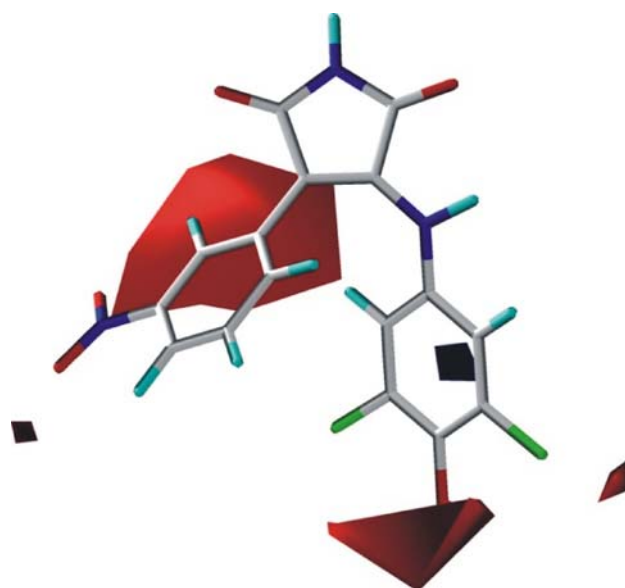
best 4- or 5-field models,  $S + E + H + D$  and  $S + E + H + A + D$ . The predicted activities of all the compounds using this best CoMSIA model are given in Table 1.

#### Steric and electrostatic field predictions

The steric and electrostatic contour maps obtained from the  $S + E$  CoMSIA model are given in Figs. 6 and 7. Both the steric and electrostatic contours from CoMSIA are similar to the CoMFA contours. Around the *ortho* and *meta* positions of the 4-phenyl ring, the two green polyhedra of the CoMFA steric contour are replaced by a single big polyhedron in the CoMSIA model. A big green polyhedron in the CoMFA contour around the *meta* and *para* positions of the 3-anilino ring is replaced by a small polyhedron around the *meta* position for CoMSIA. The CoMFA tiny yellow region accounting for the unfavorable effect of S-CH<sub>3</sub> substitution at the *para* position of the 3-anilino ring is replaced by two yellow polyhedra in the CoMSIA plot. The yellow contour around the NH position in the CoMFA contour is also found in the CoMSIA plot. All the blue and red regions from the CoMSIA electrostatic contour maps are the same as obtained with CoMFA except an additional red contour is seen around the *ortho* position of the 4-phenyl ring in CoMSIA. This indicates that electron rich groups are conducive to activity, in good agreement with our previous 2D-QSAR model showing the positive correlation of  $\sigma_{\text{orthoR}}$  with activity [22].



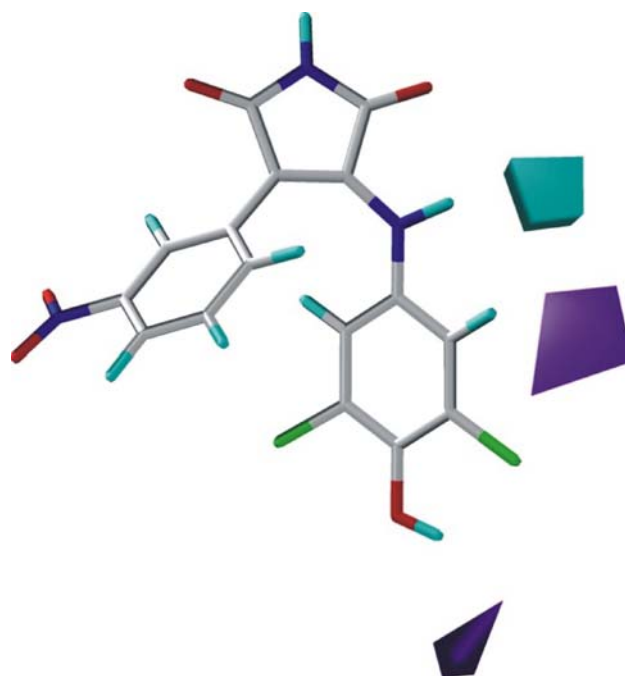
**Fig. 6** CoMSIA steric contour map based on the 56 compound training set: the green (yellow) regions show where steric bulk increases (decreases) the GSK-3 $\alpha$  inhibitory activity. The most active member of the series, **38**, is also shown



**Fig. 7** CoMSIA electrostatic contour map based on 56 compound training set: the red (blue) regions show where electronegative groups increase (decrease) the GSK-3 $\alpha$  inhibitory activity. The most active member of the series, **38**, is also shown

#### Hydrogen bond donor field predictions

The hydrogen bond donor field contour map obtained with the best  $S + E + D$  CoMSIA model is shown in Fig. 8.

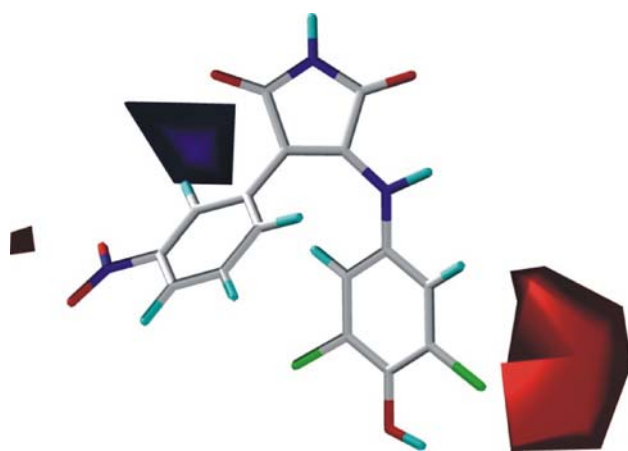


**Fig. 8** CoMSIA contour map based on the 56 compound training set illustrating the hydrogen bond donor field. The cyan (purple) regions show where hydrogen bond donor groups increase (decrease) the GSK-3 $\alpha$  inhibitory activity. The most active member of the series, **38**, is also shown

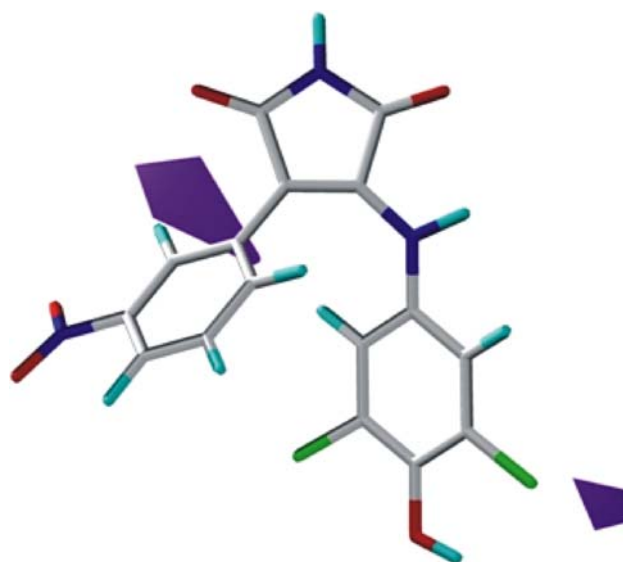
The cyan regions in Fig. 8 show where hydrogen bond donor groups will increase activity whereas the purple regions show where such groups will decrease the activity. This CoMSIA contour plot gives significant insights about the H-bond interactions between the ligand and the target structure. The small cyan polyhedron around the NH-region shown in Fig. 8 indicates a possibility of hydrogen bonding interaction between this group and the complementary hydrogen bond acceptor atom of the amino acid in the enzyme structure. The two purple polyhedra show the favorable effect of the acceptor atoms of 3-carboxylic acid groups on the 3-anilino ring. This is consistent with the fact that 3-carboxylic acid derivatives, namely **42–48**, are reasonably active ( $\text{pIC}_{50} > 6.50$ ), though not as active as the most active compound, **38** ( $\text{pIC}_{50} = 7.70$ ). The big purple polyhedron around the *meta* position of the 3-anilino ring shows that a hydrogen bond donor group at this position is not favorable for activity. Another cause of that purple region is the unfavorable contribution of the 3-OH donor group, least favored among all substitutions at this position in the original data set. A negative contribution of a variable indicating the presence of this group at the *meta* position was noted in our 2D-QSAR analysis of this data set [22].

#### Hydrogen bond acceptor and hydrophobic field predictions

In order to gain additional insights about the nature of interaction of these ligands with the GSK-3 $\alpha$  form, we visualized and analyzed the hydrogen bond acceptor and hydrophobicity fields from the final model which was developed with all the available fields present in CoMSIA:  $S + E + H + A + D$ , a model which was nearly as statistically robust as the  $S + E + D$  model. The red regions in Fig. 9 show where hydrogen bond acceptor groups will



**Fig. 9** CoMSIA contour map based on the 56 compound training set illustrating the hydrogen bond acceptor field. The red (blue) regions show where hydrogen bond acceptor groups increase (decrease) the GSK-3 $\alpha$  inhibitory activity. The most active member of the series, **38**, is also shown



**Fig. 10** CoMSIA contour map based on the 56 compound training set illustrating the hydrophobic field. The purple regions show where hydrophobic groups increase the GSK-3 $\alpha$  inhibitory activity. The most active member of the series, **38**, is also shown

increase the activity whereas the blue regions show where such groups will decrease the activity. The red polyhedron around the *meta* position of the 3-anilino ring shows that a donor OH group at this position is unfavorable or in other words that this position needs a good hydrogen bond acceptor atom for better binding. Also, the red polyhedron around the *meta* position of the 4-phenyl ring shows the importance of a nitro or another acceptor group there. The blue polyhedron around the *ortho* position of the 4-phenyl ring shows that acceptor atoms at this position are relatively less important than in the *meta* position. The purple regions in Fig. 10 show where hydrophobic groups will increase activity. A small purple contour around the *meta* position of the 3-anilino ring shows that the hydrophobic chlorine atom is favorable for activity. A purple polyhedron around the *ortho* position of the 4-phenyl ring indicates that 2-chloro substituents especially with either 3,5-dichloro, 4-OH substitution (cf. **33**) or 4-Cl, 3-COOH (cf. **50**) groups on the 3-anilino ring are favorable for the activity.

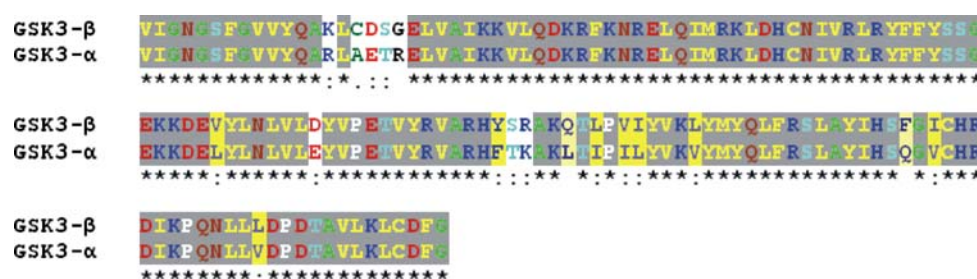
#### Homology model of GSK-3 $\alpha$

Witherington et al. [47–50] performed homology modeling for GSK-3 $\alpha$  based on the X-ray crystal structure of cyclin dependent kinase-2 (PDB code: 1HCL), but did not report any information about their model. We decided to prepare a new homology model. Sequence alignment of GSK-3 $\alpha$  and GSK-3 $\beta$  was facile because of the isoforms' high percent identity. Figure 11 shows that for a region encompassing all residues having atoms within 6.5 Å of

the co-crystallized **55** in GSK-3 $\beta$ , only 18 of 142 residues are not conserved. In particular, most of the key amino acids of GSK-3 $\beta$  having polar interactions with the ligands in this work, namely Val135, Thr138, Arg141, Gln185, are conserved in the  $\alpha$  isoform. However Asp133, which is involved in the hinge interaction [24] with the hydrogen atom of the maleimide nitrogen of these ligands, is replaced by the similar Glu196 in the  $\alpha$  isoform. If the side chain carboxylic acid were involved in the active site, than the one extra carbon in Glu196 of GSK-3 $\alpha$  could cause a significant change in the active site in comparison to Asp133 of GSK-3 $\beta$ , thus influencing the ligands' orientation and binding capabilities. However, the crucial hinge interaction in GSK-3 $\beta$  involves the backbone carbonyl oxygen of Asp133 acting as hydrogen bond acceptor for the complementary donor atom of the ATP competitive inhibitors. For the maleimide derivatives this hydrogen bond donor group is the NH of the central maleimide ring. In the X-ray crystal structure of GSK-3 $\beta$  (1q4l), the carboxyl group of this Asp133 side chain is pointed in the

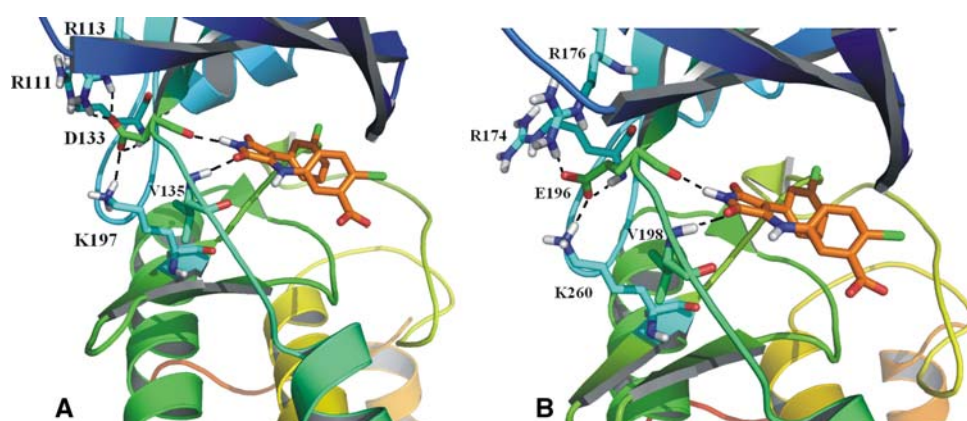
opposite direction away from the active site and forms salt bridges with Arg113, Lys197 and the backbone NH of Arg111 (see Fig. 12). Therefore, we predicted that the conservative change from Asp133 in GSK-3 $\beta$  to Glu196 in GSK-3 $\alpha$  would not affect the hinge interaction of the maleimides.

Since there is no X-ray crystal structure of the  $\alpha$  isoform we developed a homology model of GSK-3 $\alpha$  using 1q4l chain A as template (80% identity). The alignment of binding sites of the optimized model and the template structure showed a RMSD of 0.43 Å for the  $\alpha$  carbons of amino acids within 5 Å of the co-crystallized maleimide. The interactions exhibited by Glu196 with the maleimide **52** and the Glu side chain salt bridges with Arg176, Lys260 and the backbone NH of Arg174 found in our homology model (Fig. 12) suggest that the maleimide orientation and interactions are very similar in the two isoforms. Since the key amino acids in the  $\beta$  isoform that are involved in the interactions with the ligands in this work are the same as found in the  $\alpha$  isoform, these ligands would not be expected



**Fig. 11** Alignment of  $\beta$  and  $\alpha$  sequences to compare the similarity of the active site. Only the active site region within 6.5 Å around the crystallized ligand in 1Q4L for GSK-3 $\beta$  and the corresponding

regions in GSK-3 $\alpha$  are shown. The residues in grey background are conserved in both isoforms



**Fig. 12** (a) GSK-3 $\beta$  X-ray crystal structure (1q4l), with co-crystallized **52**. (b) GSK-3 $\alpha$  homology model, with coordinates of **52** extracted from the 1q4l structure. The backbone carbonyl of D133 in GSK-3 $\beta$  and E196 in GSK-3 $\alpha$  each play a key role in inhibitor binding into the active site but their sidechains face away from the inhibitor binding site. The key maleimide-Valine interaction is also

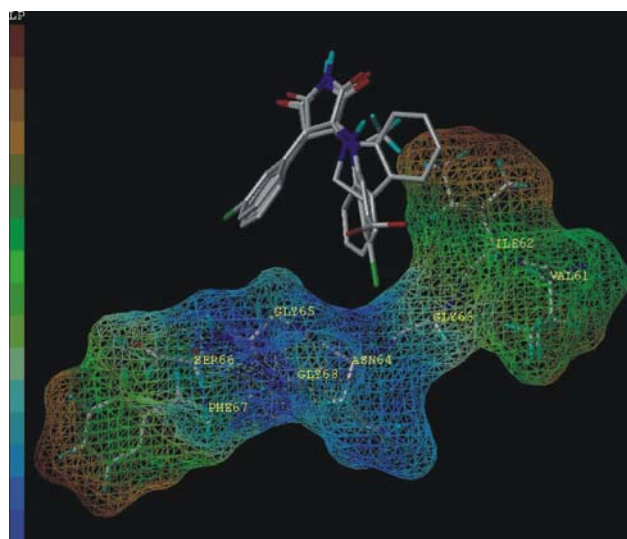
shown. One D133 carboxylic acid oxygen forms a salt bridge with R113; the other O forms a salt bridge with K197 and also interacts with the backbone NH of R111. One E196 carboxylic acid oxygen forms a salt bridge with R176; the other O forms a salt bridge with K260 and also interacts with the backbone NH of R174



to be selective against the  $\alpha$  isoform. Smith et al. [26] reported GSK-3 $\beta$  inhibitory activity data for a few of the ligands, showing that they are active against the  $\beta$  isoform as well (though IC<sub>50</sub> data was not reported). Since the active sites of the two isoforms are so similar, we used the X-ray crystal structure of the  $\beta$  isoform for the docking studies.

## Docking

We docked all the 74 compounds and obtained three slightly different poses for the three different categories of maleimides, anilino-, N-methylanilino- and indolino-. Representatives of these are shown in Fig. 13. FlexX correctly predicted that the aniline derivative **52** would bind better than those maleimides containing a tertiary amine linker to GSK-3 (Table 5). The key interactions between **52** and amino acid residues at the active site of GSK-3 $\beta$  are shown in Fig. 14. One of the oxygens of the carboxylic acid group of the 3-anilino ring forms a salt bridge with N <sup>$\eta$ 1</sup> and N <sup>$\eta$ 2</sup> of Arg141. The crystallographic water



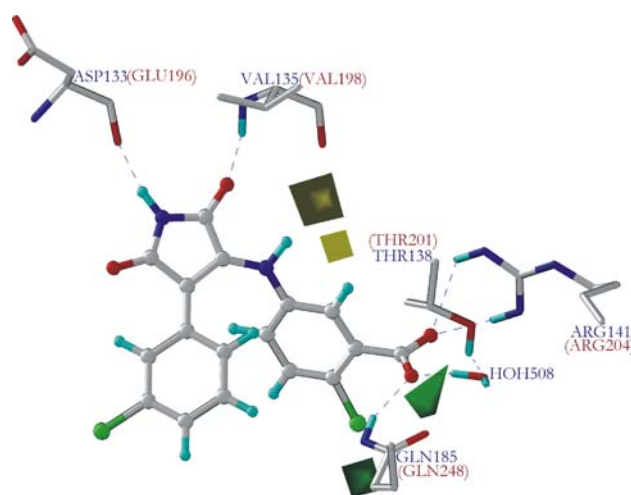
**Fig. 13** Binding mode of **52**, **65** and **71** within the nucleotide binding loop of active site of GSK-3 $\beta$ . Residues are rendered in Connolly molecular surface mapped with lipophilic potential. The color ramp shows the decreasing order of lipophilicity of the chosen residues, with blue least lipophilic

**Table 5** Scores obtained for the docking pose of some representative compounds

Compound	Total score	D_score	G_score	Chemscore	PMF_score
<b>52</b>	−33.94	−121.31	−138.64	−37.17	−37.00
<b>65</b>	−23.21	−95.24	−154.66	−30.51	−36.22
<b>71</b>	−23.71	−109.76	−171.86	−34.12	−47.07

molecule bridges the O <sup>$\gamma$</sup>  of Thr138 and the other carboxyl acid oxygen of the 3-anilino ring of these ligands. Also, N <sup>$\epsilon$ 2</sup> of Gln185 interacts with this carboxyl oxygen atom. These results are analogous to the findings of Bertrand et al. [24]. In general, the 3-anilino ring substituents, especially the carboxylic acid group, interact with the water molecule, so it is important to include the water for docking. The presence of a water molecule does not influence the docking results for the three N-methyl maleimides and seven indolino derivatives studied since they are unsubstituted at the 3-anilino ring. Lescot et al. [35] compared their CoMFA steric contour with the crystallographic structure of the enzyme-ligand complex and also considered electrostatic and hydrophobic interactions for docking poses of selected ligands. They did not include a water molecule for the docking.

We superimposed the CoMFA contours within the docked pose for a direct comparison of how well they match with the docking interactions observed in the active site of the enzyme (Fig. 14). The common docking pose for the maleimides had the maleimide NH forming a hydrogen bond with the carbonyl oxygen of Asp133 and the C=O group adjacent to this maleimide nitrogen pointing to the backbone NH of Val135, the hinge interaction of these ligands within the ATP-binding site of the  $\beta$  isoform [24]. The two yellow polyhedra around the NH of the anilino linker predict that any bulky substitution on the nitrogen of NH would probably lead to a steric clash with the carbonyl oxygen of Val135. Docking results show that this NH group indeed is close to and capable of interacting with the carbonyl group of Val135. The carbonyl oxygen is 3.29 Å away from the NH hydrogen in the docking pose for **52** (Fig. 14). The docking shows that one of the carboxylic



**Fig. 14** Docked pose of **52** showing key interactions with the GSK-3 $\beta$  isoform. CoMFA steric contour maps are also shown for comparison. The corresponding amino acids in the GSK-3 $\alpha$  isoform are shown in parentheses

acid oxygen atoms of **52** interacts with Arg141 while the other oxygen interacts with a water molecule and the amide of Gln185, which is consistent with the two purple polyhedra found in the CoMSIA hydrogen bond donor field mentioned earlier.

Smith et al. [26] suggested that the reason for the reduction of potency of N-methyl derivatives in comparison to NH-derivatives is that NH...protein interactions for these compounds are crucial. Since some of the indoline derivatives are equipotent to the NH-derivatives, they predicted that indoline derivatives adopt a different binding mode than N-methyl derivatives. We analyzed the binding mode of **52**, **65** and **71** as representatives of NH-, N-CH<sub>3</sub> and indoline derivatives, respectively. In Fig. 13, the Connolly molecular surface of selected residues (using a probe radius of 1.4 Å, resembling a water molecule as solvent) is shown with a lipophilicity map to illustrate the hydrophobic interaction of these residues with the 3-anilino ring systems. Figure 13 shows the best binding pose of these ligands within the nucleotide binding site of the  $\beta$  isoform. The nucleotide binding site, especially the segment Ile62-Gly63-Asn64, can shift to accommodate the substituents in the 3-anilino ring and the 3-anilino ring can rotate freely to maximize van der Waals interaction of its substituents with this nucleotide binding region. This is facilitated by the ease of rotation around the C–N bonds in **52**, whereas C–N bonds to the tertiary amine in N-CH<sub>3</sub> and indoline derivatives have restricted rotation and could be expected not to be able to adjust as well to fit into the GSK-3 binding pocket, explaining their lower activity. A possible explanation for some of the most conformationally restricted indoline ring derivatives being equipotent to NH-derivatives is that this ring system would have more hydrophobic interactions with the apolar residues of the nucleotide binding loop.

## Discussion

Our work has some overlap with the work of Lescot et al. [35] so it is important to summarize the major differences and novel findings of our research: We made use of a training and test set whereas they used only a training set, considered a faux pas in CoMFA [43]. The use of a training/test set combination allows us to demonstrate that our final models are predictive. Their CoMFA model included only a steric field, whereas we found that the best CoMFA model included both steric and electrostatic fields. The  $q^2$  of our CoMFA electrostatic model is significantly different than the corresponding model developed by Lescot et al. [35] 0.806 vs. 0.598. This could be because of differences in alignment, coordinate orientation or criterion for choosing the optimal number of PLS components. We used Sybyl's

automated database alignment procedure with common fragment that included three rings, not just the central maleimide ring. They used the cutoff criterion for choosing the optimal number of components as 'first minimum in the standard error', whereas we optimized based on maximizing  $q^2$ , a more typically used criterion. Our ME template *E* model also had a relatively high  $q^2$ . It is best to include *E* in the final model; even Lescot et al. in their Figs. 8 and 11 illustrated the importance of electrostatic interactions for this series of ligands. We used both CoMFA and CoMSIA methods whereas they used only CoMFA. Thus our models were able to give insight into the role not only of steric factors but also of electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor fields, in explaining the bioactivity of the maleimides.

Our CoMFA and CoMSIA models obtained from the alignment to the XR template were more highly predictive than the ones obtained from alignment to the ME template. This case is an excellent example to emphasize that the bioactive ligand conformation found in the receptor site is not necessarily the ligand's minimum energy conformation and that necessary care should be taken when choosing an alignment template for CoMFA analysis.

Bertrand et al. [24] commented about the X-ray co-crystal conformation of **52** with GSK-3 $\beta$  that the central maleimide ring is sandwiched between Ala83 and Leu188 with the 4-phenyl ring and 3-anilino ring both rotated out of the maleimide ring plane in order to maximize favorable interactions with other amino acids. This aspect favors the maleimides as GSK-3 inhibitors in comparison to rigid inhibitors such as staurosporine, whose rings are rigidly interconnected. Zeng et al. [39] found similarly for aloisines in GSK-3 $\beta$  that the ligands do not adopt the uncomplexed minimum energy conformation but rather undergo internal rotation to optimize interactions with the kinase.

## Conclusion

GSK-3 inhibition is an attractive target for developing novel inhibitors for the treatment of diseases like diabetes, Alzheimer's disease, stroke, malaria and bipolar disorder. In this work, CoMFA and CoMSIA investigations and docking studies on reported maleimides were performed to gain insight into their GSK-3 $\alpha$  enzyme inhibition. We developed the best, highly predictive CoMFA and CoMSIA models using an alignment based on the bioactive conformation derived from an X-ray co-crystal structure rather than from the minimum energy conformer of the most active member of the series. Our best CoMFA model contained steric and electrostatic fields and had  $n = 56$ ,  $q^2 = 0.844$ ,  $r^2 = 0.942$ ,  $SEE = 0.104$ ,  $F = 162.49$  and  $r^2_{\text{pred}} = 0.779$  for five components. CoMFA electrostatic



contours revealed that increased negative charge at the *meta* position of the 4-phenyl ring was favorable for the activity. We concluded that strongly electronegative groups at these positions are valuable sources to improve the activity of the existing derivatives. Also, we found that electron withdrawing groups at the *meta* and *para* positions around the anilino ring were important for enhancing activity. Electron-withdrawing bulky *ortho* substituents on the 4-phenyl ring were conducive to GSK-3 $\alpha$  inhibition. We derived a highly predictive CoMSIA model showing the importance of hydrogen bond donor groups on these ligands for enhanced activity. The best CoMSIA model ( $S + E + D$ ) had  $n = 56$ ,  $q^2 = 0.833$ ,  $r^2 = 0.932$ ,  $SEE = 0.113$ ,  $F = 111.67$  and  $r^2_{\text{pred}} = 0.803$  for six components. Comparatively, 3-N-methylanilino derivatives were less active than 3-anilino derivatives. Based on GSK-3 $\alpha$  and GSK-3 $\beta$  sequence alignment, for interaction with the maleimides, only Asp133, which is involved in the hinge interactions with the hydrogen atom from the imide nitrogen of these ligands, is replaced by Glu196 in the  $\alpha$  isoform. The X-ray crystal structure of maleimide **52** in GSK-3 $\beta$  (1q4l) has the Asp133 side chain pointing away from the active site. We prepared a homology model for GSK-3 $\alpha$ , based on the 1q4l structure, and confirmed that the change from Asp to Glu should not affect maleimide binding significantly. All other key amino-acids in the  $\beta$  isoform are conserved in the  $\alpha$  isoform, though other residues differ in the active site region we considered for docking studies.

We also studied the difference in the binding mode of NH, N-CH<sub>3</sub> and indoline derivatives in the active site of the  $\beta$  isoform. Docking results suggested that conformational restriction of rotation around the linker of the maleimide ring to the 3-anilino ring of these derivatives limits the interaction with the segment Ile62-Gly63-Asn64 in the nucleotide binding site. However, the bulky indoline derivatives could interact well with the bulky hydrophobic Ile62, yielding a potency roughly equal to that of NH derivatives. Results discussed herein were helpful for understanding the binding site of both  $\beta$  and  $\alpha$  isoforms at the molecular level. Our investigations may be useful for design of more potent and selective analogues from this class of compounds for treatment of a wide variety of diseases.

**Acknowledgements** Funding from University of Mississippi, including from its Faculty Research Program and from the Office of Research and Sponsored Programs; from the National Center for Zoonotic, Vector-borne, and Enteric Diseases (CK) of the Centers for Disease Control and Prevention (U01/CI000211); from National Science Foundation (EPS-0556308); and from National Institutes of Health's National Center for Research Resources (P20 RR021929); as well as Laboratory for Applied Drug Design and Synthesis and MCSR computing facilities are greatly appreciated. SP is a Natural Products Neuroscience Fellow. This investigation was conducted in a facility constructed with support from research facilities improvement

program C06 RR-14503-01 from the NIH National Center for Research Resources.

## References

- Hanks SK, Hunter T (1995) FASEB J 9:576
- Ali A, Hoefflich KP, Woodgett JR (2001) Chem Rev 101:2527. doi:10.1021/cr000110o
- Sarabu R, Tilley J (2004) Annu Rep Med Chem 39:41–56
- Wagman AS, Johnson KW, Bussiere DE (2004) Curr Pharm Des 10:1105. doi:10.2174/1381612043452668
- Kunnimalaiyaan M, Vaccaro AM, Ndiaye MA, Chen H (2007) Mol Cancer Ther 6:1151. doi:10.1158/1535-7163.MCT-06-0665
- Manoukian AS, Woodgett JR (2002) Adv Cancer Res 84:203–229. doi:10.1016/S0065-230X(02)84007-6
- Mulholland DJ, Dedhar S, Wu H, Nelson CC (2006) Oncogene 25:329. doi:10.1038/sj.onc.1209020
- Doerig C, Billker O, Pratt D, Endicott J (2005) Biochim Biophys Acta-Proteins Proteomics 1754:132. doi:10.1016/j.bbapap.2005.08.027
- Droucheau E, Primot A, Thomas V, Mattei D, Knockaert M, Richardson C et al (2004) Biochim Biophys Acta-Proteins Proteomics 1700:139. doi:10.1016/j.bbapap.2004.04.005
- Droucheau E, Primot A, Thomas V, Mattei D, Knockaert M, Richardson C et al (2004) Biochim Biophys Acta-Proteins Proteomics 1697:181. doi:10.1016/j.bbapap.2003.11.023
- Kumar R, Singh VP, Baker KM (2007) J Mol Cell Cardiol 42:1. doi:10.1016/j.yjmcc.2006.09.005
- Murphy E, Steenberg C (2005) Expert Opin Ther Targets 9:447. doi:10.1517/14728222.9.3.447
- Alvarez G, Munoz-Montano JR, Satrustegui J, Avila J, Bogonez E, Diaz-Nido J (2002) Bipolar Disord 4:153. doi:10.1034/j.1399-5618.2002.01150.x
- Bhat RV, Budd SL (2002) Neurosignals 11:251. doi:10.1159/000067423
- Bhat RV, Haerberlein SLB, Avila J (2004) J Neurochem 89:1313. doi:10.1111/j.1471-4159.2004.02422.x
- Huang HC, Klein PS (2006) Curr Drug Targets 7:1389
- Phiel CJ, Wilson CA, Lee VMY, Klein PS (2003) Nature 423:435. doi:10.1038/nature01640
- Gould TD, Zarate CA, Manji HK (2004) J Clin Psychiatry 65:10
- Jope RS, Yuskaitis CJ, Beurel E (2007) Neurochem Res 32:577. doi:10.1007/s11064-006-9128-5
- Aghdam SY, Barger SW (2007) Curr Alzheimer Res 4:21. doi:10.2174/15672050779939832
- Liang MH, Chuang DM (2007) J Biol Chem 282:3904. doi:10.1074/jbc.M605178200
- Sivaprakasam P, Xie AH, Doerksen RJ (2006) Bioorg Med Chem 14:8210. doi:10.1016/j.bmc.2006.09.021
- ter Haar E, Coll JT, Austen DA, Hsiao HM, Swenson L, Jain J (2001) Nat Struct Biol 8:593. doi:10.1038/89624
- Bertrand JA, Thieffine S, Vulpetti A, Cristiani C, Valsasina B, Knapp S et al (2003) J Mol Biol 333:393. doi:10.1016/j.jmb.2003.08.031
- Alonso M, Martinez A (2004) Curr Med Chem 11:755. doi:10.2174/0929867043455738
- Smith DG, Buffet M, Fenwick AE, Haigh D, Ife RJ, Saunders M et al (2001) Bioorg Med Chem Lett 11:635. doi:10.1016/S0960-894X(00)00721-6
- Coghlan MP, Culbert AA, Cross DAE, Corcoran SL, Yates JW, Pearce NJ et al (2000) Chem Biol 7:793. doi:10.1016/S1074-5521(00)00025-9
- Klebe G, Abraham U, Mietzner T (1994) J Med Chem 37:4130. doi:10.1021/jm00050a010

29. Hansch C (1964) *J Am Chem Soc* 86:1616. doi:[10.1021/ja01062a035](https://doi.org/10.1021/ja01062a035)
30. Duchowicz PR, Castro EA (2007) *Med Chem* 3:393. doi:[10.2174/157340607781024375](https://doi.org/10.2174/157340607781024375)
31. Dessalew N, Bharatam PV (2006) *Chem Biol Drug Des* 68:154. doi:[10.1111/j.1747-0285.2006.00430.x](https://doi.org/10.1111/j.1747-0285.2006.00430.x)
32. Dessalew N, Bharatam PV (2007) *Bioorg Med Chem* 15:3728. doi:[10.1016/j.bmc.2007.03.048](https://doi.org/10.1016/j.bmc.2007.03.048)
33. Dessalew N, Patel DS, Bharatam PV (2007) *J Mol Graph Model* 25:885. doi:[10.1016/j.jmgm.2006.08.009](https://doi.org/10.1016/j.jmgm.2006.08.009)
34. Katritzky AR, Pacureanu LM, Dobchev DA, Fara DC, Duchowicz PR, Karelson M (2006) *Bioorg Med Chem* 14:4987. doi:[10.1016/j.bmc.2006.03.009](https://doi.org/10.1016/j.bmc.2006.03.009)
35. Lescot E, Bureau R, Santos JSD, Rochais C, Lisowski V, Lancelot JC et al (2005) *J Chem Inf Model* 45:708. doi:[10.1021/ci050008y](https://doi.org/10.1021/ci050008y)
36. Martinez A, Alonso M, Castro A, Dorronsoro I, Gelpi JL, Luque FJ et al (2005) *J Med Chem* 48:7103. doi:[10.1021/jm040895g](https://doi.org/10.1021/jm040895g)
37. Patel DS, Bharatam PV (2006) *J Comput Aided Mol Des* 20:55. doi:[10.1007/s10822-006-9036-4](https://doi.org/10.1007/s10822-006-9036-4)
38. Xiao JF, Guo ZR, Guo YS, Chu FM, Sun PY (2006) *Protein Eng Des Sel* 19:47. doi:[10.1093/protein/gzi074](https://doi.org/10.1093/protein/gzi074)
39. Zeng M, Jiang YJ, Zhang B, Zheng KW, Zhang N, Yu QS (2005) *Bioorg Med Chem Lett* 15:395. doi:[10.1016/j.bmcl.2004.10.060](https://doi.org/10.1016/j.bmcl.2004.10.060)
40. Zhang N, Jiang YJ, Zou JW, Zhang B, Jin HX, Wang YH et al (2006) *Eur J Med Chem* 41:373. doi:[10.1016/j.ejmech.2005.10.018](https://doi.org/10.1016/j.ejmech.2005.10.018)
41. Cramer RD, Patterson DE, Bunce JD (1988) *J Am Chem Soc* 110:5959. doi:[10.1021/ja00226a005](https://doi.org/10.1021/ja00226a005)
42. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA, In, Gaussian, Inc., Wallingford CT, 2004
43. Golbraikh A, Tropsha A (2002) *J Mol Graph Model* 20:269. doi:[10.1016/S1093-3263\(01\)00123-1](https://doi.org/10.1016/S1093-3263(01)00123-1)
44. Larkin MA, Blackshields G, BN P, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) *Bioinformatics* 23:2947. doi:[10.1093/bioinformatics/btm404](https://doi.org/10.1093/bioinformatics/btm404)
45. Schrödinger (2008) LLC, New York
46. Rarey M, Kramer B, Lengauer T, Klebe G (1996) *J Mol Biol* 261:470. doi:[10.1006/jmbi.1996.0477](https://doi.org/10.1006/jmbi.1996.0477)
47. Witherington J, Bordas V, Gaiba A, Garton NS, Naylor A, Rawlings AD et al (2003) *Bioorg Med Chem Lett* 13:3055. doi:[10.1016/S0960-894X\(03\)00645-0](https://doi.org/10.1016/S0960-894X(03)00645-0)
48. Witherington J, Bordas V, Gaiba A, Naylor A, Rawlings AD, Slingsby BP et al (2003) *Bioorg Med Chem Lett* 13:3059. doi:[10.1016/S0960-894X\(03\)00646-2](https://doi.org/10.1016/S0960-894X(03)00646-2)
49. Witherington J, Bordas V, Garland SL, Hickey DMB, Ife RJ, Liddle J et al (2003) *Bioorg Med Chem Lett* 13:1577. doi:[10.1016/S0960-894X\(03\)00134-3](https://doi.org/10.1016/S0960-894X(03)00134-3)
50. Witherington J, Bordas V, Haigh D, Hickey DMB, Ife RJ, Rawlings AD et al (2003) *Bioorg Med Chem Lett* 13:1581. doi:[10.1016/S0960-894X\(03\)00135-5](https://doi.org/10.1016/S0960-894X(03)00135-5)