

QSAR Studies for the Pharmacological Inhibition of Glycogen Synthase Kinase-3

Pablo R. Duchowicz* and Eduardo A. Castro

Research Institute of Theoretical and Applied Physical Chemistry (INIFTA), Theoretical Chemistry Division, Faculty of Exact Sciences, Chemistry Department, La Plata National University, Suc.4, C.C. 16, La Plata 1900, Buenos Aires, Argentina

Abstract: The enzyme GSK-3 plays a central role in cells during the phosphorylation of various key regulatory proteins, and consequently pharmacological inhibitors of this enzyme potentially allow the treatment of diseases that include neurodegenerative and bipolar affective disorders, diabetes, and diseases caused by unicellular parasites. Today there is a huge number of reported empirical structure-activity relationships (SAR) that may guide a rational design of more potent and selective inhibitors. However, only a few studies based on Quantitative Structure-Activity Relationships (QSAR) are available for predicting the inhibitor potency against this specific kinase, and they involve mainly molecular modeling and 3D-QSAR. The present review deals with the recent search for a quantitative analysis of GSK-3 inhibition.

Key Words: QSAR Theory, Glycogen Synthase Kinase-3, Phosphorylation, CoMFA/CoMSIA, Partial Least Squares, Multivariable Linear Regression, Artificial Neural Network.

BIOLOGY OF THE ENZYME GSK-3

The reversible phosphorylation of serine, threonine and tyrosine residues by protein kinases is now considered to be one of the most common processes that regulates cellular metabolism and function, and abnormal regulation is associated with many diseases [1, 2]. There are about five hundred protein kinases encoded in the human genome, and they play essential roles in virtually all cellular processes and in almost all known diseases. Therefore, the screening for pharmacological inhibitors of kinases and phosphatases has become a promising strategy in the pharmaceutical industry during the last few decades.

Glycogen Synthase Kinase-3 (GSK-3) is a serine/threonine protein kinase that was discovered in the late 1970s for its phosphorylation activity toward glycogen synthase, the rate limiting enzyme of glycogen biosynthesis [3, 4]. It is known to be involved in various cellular and physiological events, such as the Wnt and Hedgehog signaling pathways, cell survival, response to DNA damage, HIV-1 Tat-mediated neurotoxicity, hyphosphorylation of protein tau (one of the diagnostic features of Alzheimer's disease), circadian rhythm, and others. These processes lead to different types of cancers, diabetes, neurodegenerative disorders (Alzheimer, manic depression), proliferation of protozoan parasites, and viral infections (HIV, cytomegalo virus, herpes virus).

Two human genes, that map to 19q13.2 and 3q13.3, encode two closely related GSK-3 isoforms, namely GSK-3 α (51 kDa) and GSK-3 β (47 kDa), respectively. They display 84% overall sequence identity (up to 98% within their cata-

lytic domains) with the main difference being an extra Gly-rich stretch in the N-terminal domain of GSK-3 α . However, they are not interchangeable functionally, as demonstrated by the embryonic-lethal phenotype observed when the gene that encodes GSK-3 β is knocked out. Recently, GSK-3 β 2 was identified, an alternative splicing variant of GSK-3 β that contains a 13-amino-acid insertion in the catalytic domain [5].

The overall shape of the crystallized GSK-3 β analyzed by X-ray is shared by all types of kinases [6, 7]. The primary structure of GSK-3 and Cyclin-Dependent Kinase (CDK) families are closely related, and this fact leads to interactions of similar nature with their inhibitors. It involves a small N-terminal lobe, consisting mostly of β -sheets and a large C-terminal lobe, formed essentially of α -helices [2]. The ATP-binding pocket of GSK-3 includes the amino acid residues Asp and Val, and is located between the two previously mentioned lobes. As the ATP-binding site is similar in both GSK-3 α and GSK-3 β , inhibitors that target this site are unlikely to differentiate between the two isoforms. Arg96, Arg180 and Lys205 form a small pocket where the phosphate group of the substrate binds. The enzyme was also co-crystallized with seven inhibitor compounds, providing a better understanding of their mechanism of interaction within the ATP-binding pocket.

GSK-3 has numerous physiological substrates *in vivo* that include transcription factors as well as enzymes involved in regulating metabolism, such as Glycogen Synthase, β -catenin, eukaryotic protein synthesis initiation factor-2B (eIF2B), cyclic AMP-response element-binding protein (CREB), etc. [8] The molecular basis for the unusual substrate specificity of GSK-3, such as it is for glycogen synthase or tau protein, is a consequence of the phosphorylation of proteins at serine (Ser) or threonine (Thr) residues that are normally located four amino acids (Xaa) N-terminal to another phosphoserine (pSer) or phosphothreonine (pThr) in the sequence Ser/Thr-

*Address correspondence to this author at the Research Institute of Theoretical and Applied Physical Chemistry (INIFTA), Theoretical Chemistry Division, Faculty of Exact Sciences, Chemistry Department, La Plata National University, Suc.4, C.C. 16, La Plata 1900, Buenos Aires, Argentina; E-mail: prduchowicz@yahoo.com.ar / duchow@inifta.unlp.edu.ar

Xaa-Xaa-Xaa-pSer/pThr, thus modifying the intrinsic activity of the substrate involved [9]. The C-terminal phosphoserine (termed the 'priming phosphorylation' side) is usually introduced by another protein kinase (see Fig. (1)).

The activation of GSK-3 seems to be an intramolecular autophosphorylation event of a single tyrosine residue (Tyr279 in the α -isoform and Tyr216 in the β -isoform) [10], while the inhibition of this enzyme results from the phosphorylation of a serine unit (Ser21 in the α -isoform and Ser9 in the β -isoform) [11]. For instance, during the signaling pathway by which insulin inhibits GSK-3 and glycogen synthase is activated, the serine phosphorylation transforms the N-terminal of GSK-3 into a pseudo-substrate, which occupies the 'priming phosphorylation' site of the substrate and blocks the access of other substrates to the catalytic side [6]. Unphosphorylated tyrosine units in the T-loop domain prevent access of substrates to the catalytic site, while phosphorylation releases this inhibition.

GSK-3 is also inhibited through interactions with many other proteins without the necessity of employing a priming phosphate [12]. Axin and Presenilin act as docking proteins that allow the substrates to make contact with the priming kinase (Casein Kinase (CK1) and Protein Kinase A (PKA)) and GSK-3. Docking proteins might, thus, specify different GSK-3 functions in the cell. An example of this are the secreted glycoproteins termed Wnts, that function in a crucial pathway for the specification of cell fates during embryonic development. Wnt signalling targets a particular subcellular pool of GSK-3 that is complexed to Axin, β -Catenin and the Adenomatous Polyposis Coli (APC) proteins, resulting in the displacement of Axin by other proteins, such as dishevelled and FRAT. This prevents GSK-3 from phosphorylating Axin and β -Catenin, but not other substrates, such as glycogen synthase and eIF2B [13]. The net effect of this process is the accumulation and translocation of β -Catenin to the nucleus, where it stimulates the transcription of genes required during embryonic development [14].

In summary, different inhibitors of GSK-3 exert their action by: (i) competing with either ATP or metal-binding sites involved directly in the catalytic process, (ii) inhibiting the protein-protein interactions necessary for binding of substrate (the primed phosphorylated serine binding area and the docking protein, Axin and Presenilin), (iii) modulating the

Tyr216 (GSK-3 β) and Tyr279 (GSK-3 α) activation sites and the Ser9 (GSK-3 β) and Ser21 (GSK-3 α) inhibition sites, and (iv) interfering with the intracellular targeting domain of GSK-3. Inhibition of the interaction between the docking protein and the priming kinase might change the substrate specificity of GSK-3.

GENERAL CHARACTERISTICS FOR GSK-3 INHIBITORS

The search for inhibitors of the enzyme GSK-3 is a very active field in both academic centers and pharmaceutical companies. This is usually accomplished by means of screening programs specifically aimed at finding GSK-3 inhibitory activity in other compounds previously reported with other biological properties. A typical challenge during the synthesis of new GSK-3 inhibitors relies on a high potency of the drug required, generally expressed as a low micromolar-range inhibitory activity (IC_{50} , concentration required for 50% inhibition of GSK-3 *in vitro*). The degree of selectivity towards different types of kinases is usually employed as a tool for demonstrating the participation of GSK-3 in a given cellular process, and this might only be achieved towards GSK-3 α and GSK-3 β by inhibitors that act at other sites on the kinases, by alternative molecular approaches, and when either the intracellular distribution or the expression is different for the two isoforms. The selectivity of most of the available GSK-3 inhibitors is not well known and based only on their evaluation with limited panels of kinases [15]. Apart from the two mentioned conditions, more important than all is the synthesis of pharmacologically acceptable drugs that have acceptable absorption, distribution, metabolism, excretion, and toxicological (ADMET) properties.

Over recent decades structurally diverse inhibitors have been reviewed [12, 16-19], many of which appear listed in Table 1 and whose molecular structures are presented in Fig. (2). These include lithium chloride, indirubins, paullones, maleimides, aloisines, hymenialdisine, etc. Lithium chloride was the first pharmacologically useful inhibitor of GSK-3 to be identified, being a drug widely employed despite the millimolar concentrations that are required to affect GSK-3 in living cells [20, 21]. Lithium ions are not selective for GSK-3 as they also inhibit, for example, inositol-phosphate phosphatases [22]. Paullones, especially kenpaullone, 6-bromo-substituted indirubins and the azaindoly maleimide com-

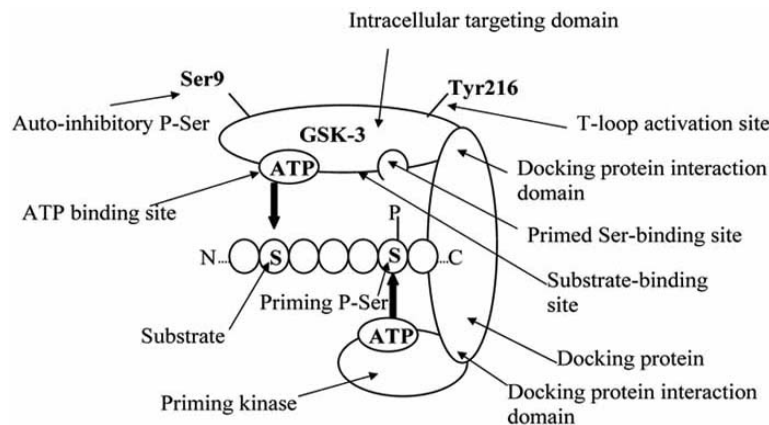


Fig. (1). Potential sites for the inhibition of GSK-3.

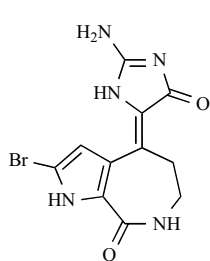
Table 1. Inhibitors of GSK-3

Inhibitor	Class	IC ₅₀ (μM)	Ref.
Hymenialdisine	Pyrroloazepine	0.010 (β)	[83]
Flavopiridol	Flavone	0.450	[84]
Kenpauillone	Benzazepinone	0.023 (β)	[84, 85]
Alsterpauillone	Benzazepinone	0.004 (α); 0.004 (β)	[84, 85]
Azakenpauillone	Benzazepinone	0.018 (β)	[86]
Indirubin-3'-oxime	Bis-Indole	0.022 (β)	[84]
6-Bromoindirubin-3'-oxime	Bis-Indole	0.005	[29, 49]
6-Bromoindirubin-3'-acetoxime	Bis-Indole	0.010	[29, 49]
Aloisine A	Pyrrolopyrazine	0.650	[87]
Aloisine B	Pyrrolopyrazine	0.750	[87]
TDZD8	Thiadiazolidinone	2.000 (β); 7.000 (α/β)	[16]
Compound 12	Pyridyloxadiazole	0.390 (β); 8.000 (α/β)	[88]
Pyrazolopyridine 18	Pyrazolopyridine	0.018 (α)	[60]
Pyrazolopyridine 9	Pyrazolopyridazine	0.022 (α)	[61]
Pyrazolopyridine 34	Pyrazolopyridine	0.007 (α)	[62, 63]
CHIR98014	Aminopyrimidine	0.00065 (α); 0.00058 (β)	[35]
CHIR99021 (CT99021)	Aminopyrimidine	0.010 (α); 0.007 (β)	[35]
CT20026	Aminopyridine	0.004 (α/β)	[19]
Compound 1	Pyrazoloquinoxaline	1.000	[89]
SU9516	Oxindole (indolinone)	0.330; 0.35 (α/β)	[90]
ARA014418	Thiazole	0.104 (β)	[24]
Staurosporine	Bisindolylmaleimide	0.015; 0.089	[91]
Compound 5a	Bisindolylmaleimide	0.018 (β)	[91]
Compound 29	Azaindolylmaleimide	0.034 (β)	[92]
Compound 46	Azaindolylmaleimide	0.036 (β)	[93]
GF109203x	Bisindolylmaleimide	0.190 (β)	[94]
Ro318220	Bisindolylmaleimide	0.003–0.038 (β)	[15, 94]
SB216763	Arylindolemaleimide	0.034 (α); 0.075 (α/β)	[34]
SB415286	Anilinomaleimide	0.078 (α); 0.13 (α/β)	[34, 58]
I5	Anilinoarylmaleimide	0.076 (α); 0.160 (β)	[23]
CGP60474	Phenylaminopyrimidine	0.010	[95]
Compound 8b	Triazole	0.280 (β)	[96]
TWS119	Pyrrolopyrimidine	0.030 (β)	[97]
Compound 1A	Pyrazolopyrimidine	0.016 (β)	[64]

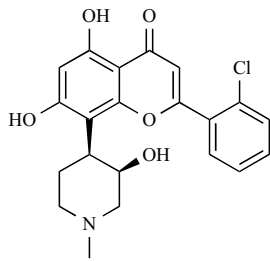
(Table 1. Contd....)

Inhibitor	Class	IC ₅₀ (μM)	Ref.
Compound 17	Chloromethyl thienyl ketone	1.00 (β)	[98]
Lithium	Atom (competition Mg ²⁺)	2000.0	[20, 21]
Beryllium	Atom (competition Mg ²⁺ and ATP)	6.00	[21]
Zinc	Atom (uncompetitive)	15.00	[99]

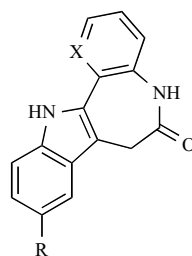
(α) or (β) correspond to the isoform tested. (α/β) represents a mixture of both isoforms, while the absence of specifications indicates that the study cited did not mention which isoform(s) was tested.



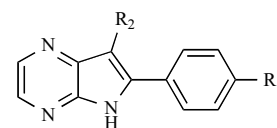
Hymenialdisine (1)



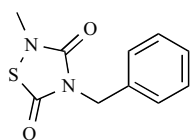
Flavopiridol (2)



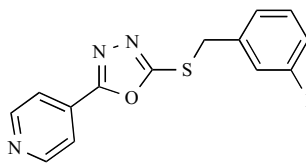
Paullone (R=H; X=C) (3)
 Kenpaullone (R=Br; X=C) (4)
 Alsterpaullone (R=NO₂; X=C) (5)
 Azakenpaullone (R=Br; X=N) (6)



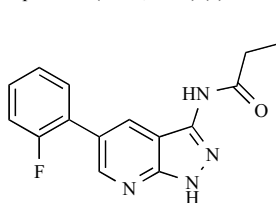
Aloisine A (R₁=OH; R₂=(CH₂)₃CH₃) (7)
 Aloisine B (R₁=Cl; R₂=CH(CH₃)₂) (8)



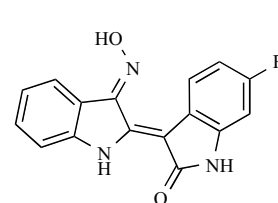
TDZD8 (9)



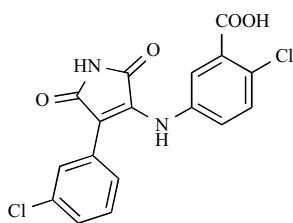
Compound 12 (10)



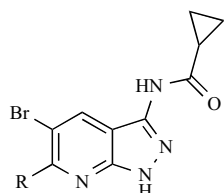
Pyrazolopyridine 18 (11)



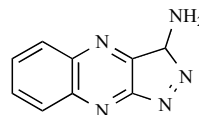
Indirubin 3' oxime (R=H) (12)
 BIO (R=Br) (13)



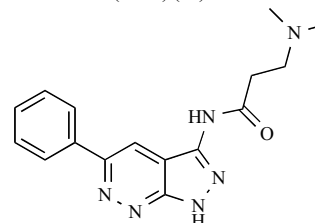
15 (14)



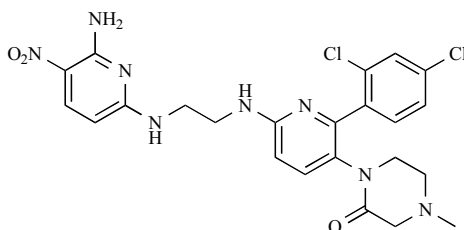
Pyrazolopyridine 34 (15)
 (R=2-furyl)



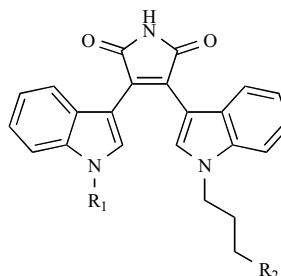
Compound 1 (16)



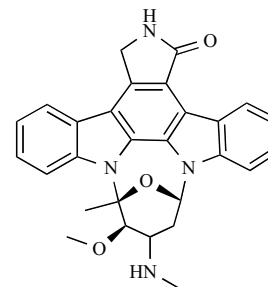
Pyrazolopyridine 9 (17)



CT20026 (18)

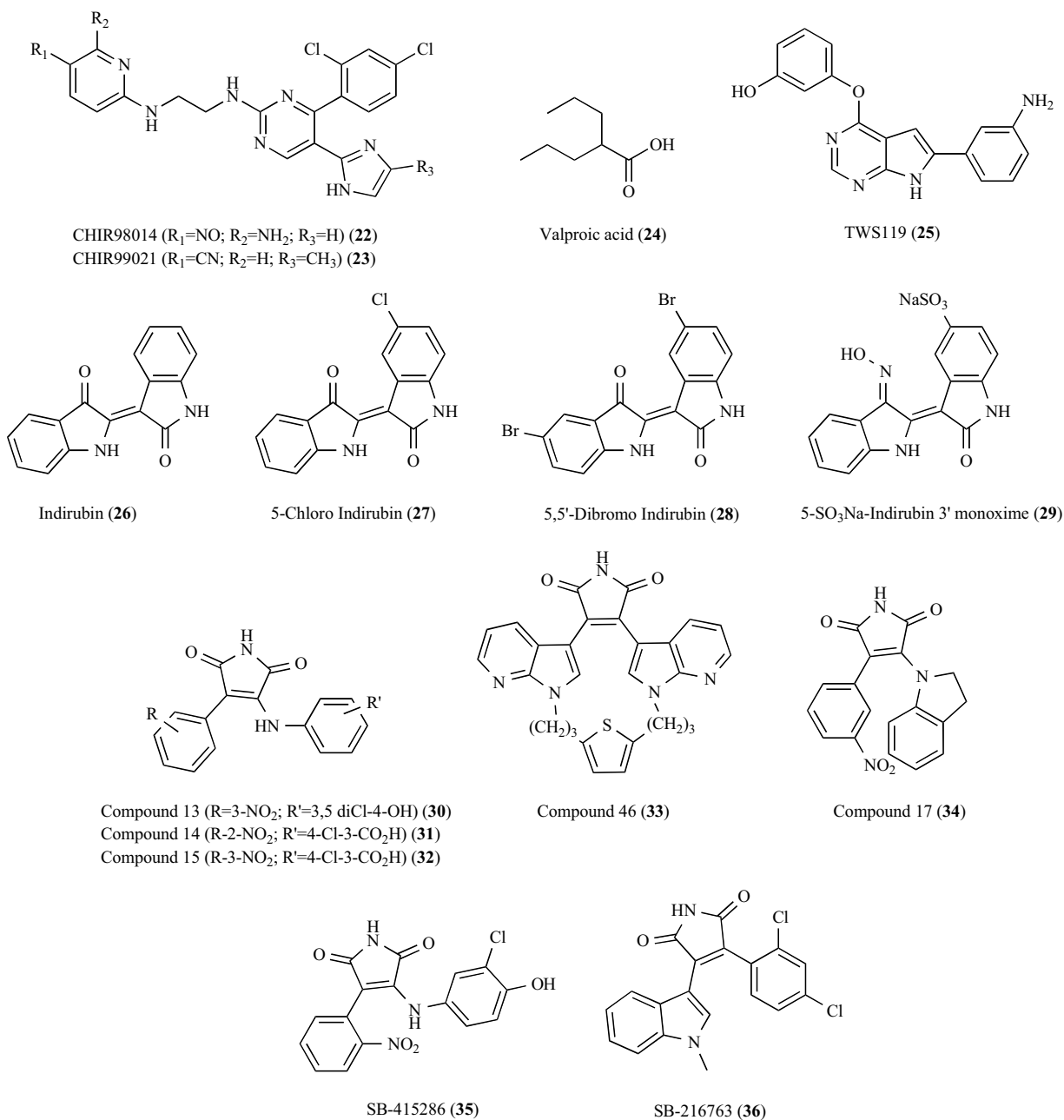


GF109203x (R₁=H; R₂=N(CH₃)₂) (19)
 Ro318220 (R₁=CH₃; R₂=S-C(NH)-NH₂) (20)



Staurosporine (21)

(Figure 2. Contd....)

**Fig. (2).** Structures of better known inhibitors.

pound 46 (Table 1), appear to be among the most selective inhibitors identified so far. However, the selectivity of these compounds against a large panel of other protein kinases has not yet been reported. The GSK-3 inhibitors elicit the following common characteristics: (a) small and flat hydrophobic heterocycles (molecular weights <600), since the kinase is usually embedded in multi-protein complexes; (b) kinetic experiments show that most of them act by competing reversibly with ATP for binding to the kinase, for exception of few like lithium or some thiadiazolidinones compounds (TDZD). (c) binding essentially through hydrophobic inter-

actions and 2–3 hydrogen bonds with the kinase; (d) the backbone carbonyl and amino side-chains of Val135 act as an H-bond acceptor and H-bond donor, respectively, to the inhibitors, whereas the backbone carbonyl of Asp133 often acts as an H-bond acceptor [23, 24].

SOME DISEASES ASSOCIATED WITH GSK-3

Nervous System Disorders

Lithium (Li^+) is the primary therapeutic agent that has long been used for the treatment of manic depression illness,

although the mechanism of biological action is yet unknown. Despite its valuable efficacy, this drug has side-effects and more potent and selective GSK-3 inhibitors possibly represent an alternative treatment for bipolar disorder [22]. Another mood-stabilizing drug, Valproate, also used in treating bipolar disorders and epilepsy, has been shown to inhibit GSK-3 *in vivo* [25].

Alzheimer's disease (AD) affects more than 18 million people worldwide. GSK-3 is related to different essential events that characterize AD [26], such as: (i) interaction and subsequent mutation of any one of the genes that encode presenilin-1, presenilin-2 and the β -amyloid precursor protein (β -APP), resulting in ~100% penetrance; (ii) the extracellular accumulation of amyloid- β (40–42 amino-acid peptides derived from proteolytic cleavage of β -APP by several enzyme complexes, including β - and γ -secretases); (iii) the intracellular aggregation of hyperphosphorylated forms of the microtubule-binding protein Tau leading to the formation of neurofibrillary tangles [5].

It is known that the activity of GSK-3 β increases in cells from familial AD patients that contain mutated presenilin-1/2 or β -APP, compared with control cells [27]. Furthermore, the activity of GSK-3 α is required for the production of amyloid- β [28] and amyloid- β toxicity is mediated by increased GSK-3 activity. In addition, GSK-3 is one of the major kinases that is involved in abnormal Tau phosphorylation on AD-specific sites. Presenilin-1 interacts directly with GSK-3, which favors the interaction of GSK-3 with substrates such as Tau. GSK-3 is involved in neuronal cell death and there are numerous examples of the neuroprotection provided by GSK-3 inhibitors following different insults [29]. One such example is provided by *Drosophila*, in which overexpression of shaggy, the GSK-3 β homolog, enhances neurodegeneration induced by the expression of human Tau, whereas expression of a loss-of-function mutant of shaggy prevents this neurodegeneration [30].

Parkinson's disease (PD) is a neurodegenerative disease that is characterized by the specific loss of dopamine containing neurons that project from the substantia nigra to the caudate-putamen (striatum). This event is facilitated by overexpression of GSK-3 β and attenuated by lithium [31]. Huntington's disease is another neurodegenerative disease caused by a polyglutamine extension of the huntingtin protein. The modified protein is toxic and forms intracellular aggregates. Decreased β -catenin levels and reduced transcription in cells that express mutant huntingtin indicates that GSK-3 modulates cellular toxicity [32]. Transmissible spongiform encephalopathies (TSEs, also known as prion diseases) are marked by the accumulation of an amyloidogenic, insoluble form of the prion protein. The pathological prion protein induces neuronal cell death, a process that is reduced significantly in culture by treatment with either lithium or insulin (which inhibits GSK-3) [33]. The activity of GSK-3 β enzyme is also present in the brain of patients with schizophrenia, and is implicated in forced alterations of the circadian rhythm.

Type 2 Diabetes

Diabetes mellitus is considered to be the most common disease of metabolism. Type 1 diabetics include pancreatic

β -cells that cannot produce insulin and require frequent injections of this hormone to survive, whereas type 2 diabetics (90% of cases) can make and secrete insulin, but are resistant to the hormone. The level of blood glucose is largely determined by the rate at which glucose is taken up by skeletal muscle and liver and converted into glycogen, and by the rate at which it is produced by the liver. Therefore, drugs that inhibit GSK-3 decrease the phosphorylation and increase the activity of glycogen synthase, overcoming the resistance to insulin [12, 34, 35]. Metformin (Glucophage; Merck-Santé) is the most commonly used drug introduced 50 years ago to treat type 2 diabetes. Although it results effective for normalizing the blood glucose levels, it lacks potency and its unpleasant side effects include stomach and intestinal problems.

Cancers

GSK-3 is a key component of the Wnt signaling pathway, which is a central process at many stages of development and is highly conserved between species. Many colorectal cancers have defects in elements of the Wnt pathway that result in the abnormal activation of Wnt signaling and accumulation of β -catenin [36], the co-transcription factor that regulates numerous genes involved in carcinogenesis. Resistance to TRAIL, an apoptosis-inducing ligand that is being evaluated as an anti-cancer agent, involves activation of GSK-3 β [37]. Inhibition of the kinase, by lithium, SB216673 and siRNA, dramatically enhances TRAIL-induced apoptosis in prostate cancer cell lines.

Unicellular Parasites

The design of new anti-parasitic drugs can be based on the evaluation of kinases from unicellular parasites. An initial study showed that *P. falciparum*, the most prevalent species causing a lethal form of Malaria, exports PfGSK-3 to the cytoplasm of host erythrocytes (which are devoid of GSK-3), where it co-localizes with parasite generated membrane structures known as Maurer's clefts [38]. The presence of PfGSK-3 supports the hypothesis that the kinase plays a role in regulating the strong circadian rhythm of the parasite, which is responsible for the circadian fevers that are characteristic of this infectious disease.

QSAR Studies

In a study of Kunick *et al.* [39], 3D-QSAR Comparative Molecular Similarity Indices Analysis (CoMSIA) calculations [40, 41] were developed for the inhibition of three serine/threonine kinases: CDK-1/cyclin B, CDK-5/p25 and GSK-3 β , by compounds from the paullone inhibitor family shown in Fig. 3. These models were based on the kinase inhibition data of 52 paullone entities reported by Leost *et al.*

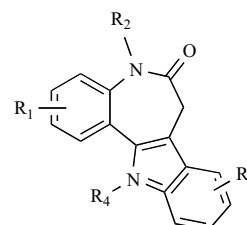


Fig. (3). General structure of paullone derivatives.

[42] The evaluation was carried out using a test set of 23 structural diverse paullone-related compounds, covering a broad range of kinase inhibitory activity. IC_{50} values for CDK-1, CDK-5, and GSK-3 were converted to pIC_{50} ($-\log IC_{50}$) values and used as dependent variables in the CoMSIA calculations. Although the selectivity toward only one kinase target is a desirable feature for a drug to avoid potential undesired side-effects, the simultaneous inhibition of especially CDK-5 and GSK-3 represents a therapeutical strategy of interest, as both kinases are involved in the hyperphosphorylation of the tau protein in AD.

The molecular geometries of all compounds analyzed were constructed using the SYBYL 6.8 Molecular Modeling Software (Tripos Associated Ltd.) and after that partial atomic charges were calculated using the semiempirical AM1 Hamiltonian [43]. Energy minimizations were performed using the Tripos Force Field [44] with a distance dependent dielectric parameter taking charge into account. The alignment of the compounds in the CoMSIA analysis is one of the key steps to obtain meaningful results, as the biologically active conformations of the structures should be aligned in a way representing a similar binding mode. For this reason the authors employed a docked alignment, and all paullone derivatives are manually docked into the ATP binding pocket of the CDK-1/cyclin B homology model from Gussio *et al.* [45], using a docked kenpaullone scaffold as template.

All inhibitor-protein complexes were minimized using the MAB Force Field [46] as implemented in the MOLOC program, keeping the protein fixed and allowing the inhibitor to move. This force field uses geometric hydrogen bonding

terms to model electrostatic interactions of polar groups. The alignment was then performed using the relative three-dimensional orientations of the compounds resulting from the MOLOC minimizations. A 3D-cubic lattice box with a grid spacing of 2 Å or 1 Å is created around the aligned molecules, leading to five CoMSIA similarity index fields available within SYBYL (steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor) [40, 41] and calculated using a C1+ probe of a 1 Å radius and the default value of 0.3 as attenuation factor.

Different 3D-QSAR models were established by means of Partial Least Squares (PLS) [47] regression analysis, where the CoMSIA fields are the independent variables. The optimal number of components was chosen by using the SAMPLS algorithm [48] and corresponds to the lowest *PRESS* (predictive sum of squares), calculated with the Leave-one-out Cross-Validation technique (loo) using a column filtering of 0.5, 1.0, and 2.0. Predictive R^2 from loo (q^2) is also calculated. The quality of these models in Table 2 is further indicated with the squared correlation coefficient (R^2), its standard deviation (S), and Fisher- F ratio. The best statistical values for the CoMSIA were obtained for the CDK-1-model: $R^2=0.929$ and $q^2=0.699$, which is clearly superior to the models for CDK-5: $R^2=0.874$ and $q^2=0.652$ and GSK-3: $R^2=0.871$ and $q^2=0.554$. The predictions achieved for each kinase are displayed in Fig. (4), revealing that not all the compounds from the training set are well predicted with the CDK-5 and GSK-3 models. There are two main reasons for this: (i) using a docked alignment with the 3D structure of CDK-1 as template, the side chain conformation

Table 2. Comparison of the Statistical Values and Field Distribution within the Selected CoMSIA Models of CDK-1, CDK-5, and GSK-3

	CDK-1	CDK-5	GSK-3
step size	2 Å	2 Å	1 Å
col. filt.	0.5	0.5	2.0
q^2	0.699	0.652	0.554
<i>PRESS</i>	0.663	0.846	0.853
R^2	0.929	0.874	0.871
S	0.321	0.509	0.458
F	120.916	81.422	62.292
opt. n°. compon.	5	4	5
Field Contribution			
steric	0.087	0.092	0.084
electrostatic	0.253	0.236	0.178
hydrophobic	0.181	0.227	0.232
donor	0.188	0.169	0.213
acceptor	0.29	0.277	0.293

col. filt.: column filtering; R : correlation coefficient; q^2 : predictive R^2 from leave-one-out cross validation technique; *PRESS*: predicted sum of squares from leave-one-out cross validation; S : standard deviation of the model; F : Fisher-test parameter; opt. no. compon.: optimal number of components.

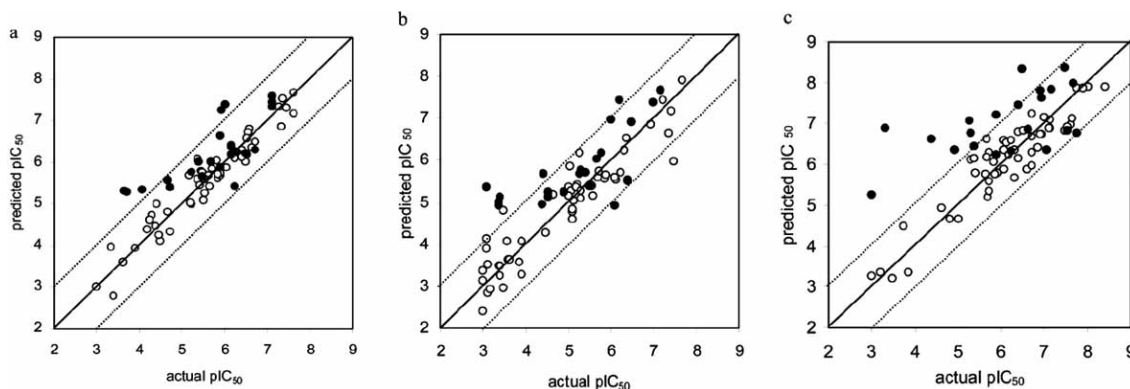


Fig. (4). Plot of observed versus predicted pIC_{50} for the training set (●) and test set (○) of CDK-1 model (a), CDK-5 model (b) and GSK-3 model (c).

is optimized for CDK-1 only but might be different in CDK-5 or GSK-3 and (ii) the uniqueness of the molecular structures in the training set.

The application of the three CoMSIA models to estimate pIC_{50} values in the test set lead to the conclusion that, in all the models, the 4-azapallones and 3-methoxy-substituted pallones are predicted with higher values than expected. This observation can be explained with the absence of similar compounds in the training set that could contribute required information to the model regarding the 4- or 3-position, respectively. Similarly, a thioimide is overpredicted. Although the training set comprised a related compound, that bears an unsubstituted indolo-nitrogen at position 12 and could be involved in hydrogen bonding, the thioimide from the test set is lacking this feature through methyl substitution. Consequently, it loses both hydrogen bond donor groups, a structural modification that could be responsible for a completely different alignment mode for this compound.

The CDK-1 and the CDK-5 model predict the activity of the thiophene analogues quite well, indicating that the absence of a comparable compound in the training set is not critical for these substances. In contrast, the GSK-3 model over-predicted these compounds. The latter observation might be due to structural differences between the CDKs and the GSK-3 that are not explained with the alignment that is used for the training set and test set of this study. With exception of the above-mentioned compounds, the inhibitory activities of the structures in the test set are predicted reasonably well. Present analysis provided with CoMSIA models for different kinases that enable new compounds with modified electron density throughout the heterocyclic ring system to be designed and evaluated *in silico* before synthetic approaches are undertaken.

In another study of the same year by Polychronopoulos *et al.* [49], molecular mechanics docking-scoring calculations were performed for a training set consisting on 23 indirubins with related chemical structure (Fig. (5)), in order to develop a model capable of predicting their binding affinity within the ATP-binding pocket of GSK-3 β , CDK-2 and CDK-5/p25.

All the ligands from the training set were energy-minimized (AMBER) using MACROMODEL software and re-

sorting to MOPAC for calculating partial atomic charges (AM1 Hamiltonian with EF minimizer and NOMM correction). These ligand molecules were then superimposed over the crystallographic position of 6-bromoindirubin 3'-oxime. The binding affinities were estimated using PrGen v.2.1 software [50], evaluating ligand-receptor interaction energies, ligand desolvation energies, and changes in both ligand internal energy and ligand internal entropy upon receptor binding according to the formula:

$$E_{\text{binding}} \approx E_{\text{ligand-receptor}} - T\Delta S_{\text{binding}} - \Delta G_{\text{solvation,ligand}} + \Delta E_{\text{internal,ligand}} \quad (1)$$

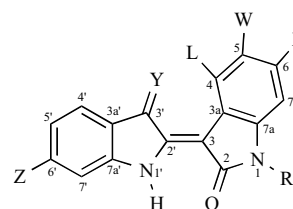


Fig. (5). Molecular structure of indirubin analogues.

This was accomplished through an iterative procedure implemented in PrGen referred to as the ligand equilibration protocol, involving two main steps: (a) correlation coupled receptor minimization and (b) unconstrained ligand relaxation/Monte Carlo search. The training set of molecules is first subjected to ligand equilibration combined with Monte Carlo search, leading to the model:

$$\Delta G_{\text{predicted}} = 0.961E_{\text{binding}} + 27.112 \quad (2)$$

producing $R=0.902$ and RMS of 0.93. Here, the experimental binding affinities are calculated as relative binding affinities using IC_{50} values by the following equations:

$$RBA = \frac{\text{indirubin } IC_{50}}{\text{ligand } IC_{50}} \cdot 100 \quad \Delta G \approx -\ln RBA \quad (3)$$

Relationship (2) was then used to predict the activity in a test set of 15 molecules that are introduced within the equilibrated receptor cavity and subjected to Monte Carlo mini-

zation. The resulting correlation between experimental and predicted binding affinities for both training and test sets is presented in Fig. (6). There is good agreement between calculated and experimental values with a *RMS* of 1.74: the largest deviation is obtained for 6-bromo-5-nitroindirubin.

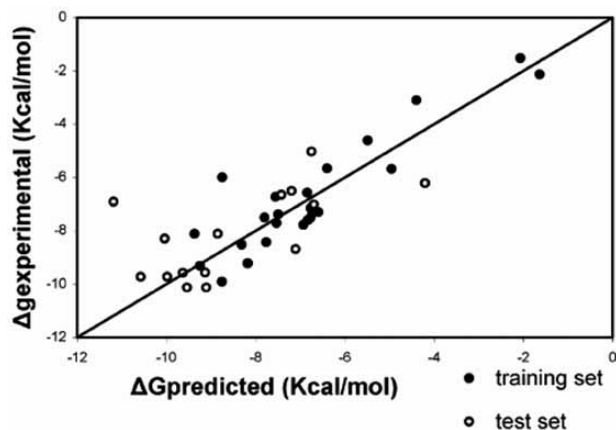


Fig. (6). Correlation between experimental and predicted GSK-3 β ΔG values for indirubins.

PrGen calculates three energy terms: van der Waals, electrostatics, and H-bonding. The affinity of indirubins for GSK3 depends mainly on the hydrophobic van der Waals energy term which accounts for 66% to 92% of the sum of these three energy terms. On the contrary, the contributions of the electrostatic and the H-bond term seem to be in most cases (except for substitution 6-F) responsible for the improved affinity/potency of the oxime-substituted indirubins when compared to their non-oxime analogues, suggesting that the =N-OH group stabilizing effect is of electrostatic character.

Crystallographic data of CDK-2 [51, 52], CDK-5 and GSK-3 β in complex with various indirubins provided valuable information on the active site differences along with specific interactions between those kinases and indirubin analogues. Substitution at position 6 turned out to be crucial for the selectivity while substitution at 3' was found to be important for the binding affinity. In example, for CDK-5 and CDK-2, the limits of the inner part of the binding cavity are defined by the side chain of Phe80 while in GSK-3 the corresponding residue is Leu132. The difference between the isobutyl side chain of Leu132 and the phenyl ring of Phe80 results in an increase of the pocket width in GSK-3 compared to that of CDK's and can obviously explain the selectivity of 6-bromoindirubin toward GSK-3 compared to CDKs. Regarding the potency, this was attributed to a possible indirect interaction of the group substituted in 3' with polar residue side chains through water molecules. Among the 6-substituted indirubins, the bromo substitution exhibits the highest activity against GSK-3. In the present study, calculations show that this is due mainly to van der Waals interactions which are optimal for both bromo and iodo derivatives, in agreement with biological test results. Combination of the substitutions on both positions 5 and 6 result in an additive effect on the activity. The *IC*₅₀ value is reduced about 2 times compared to corresponding single 5-substitution.

Several new 3'-oxime derivatives were synthesized in this analysis, exhibiting an increased inhibitory activity on all three kinases compared to their non-oxime counterparts. The best inhibitory activity on GSK-3 is observed for the 6-bromo- and the 5,6-disubstituted compounds. According to Eq. (2), this is mainly due to favorable contributions of the electrostatic and hydrogen bond energy, suggesting that the interactions of the oxime hydroxyl group should play an important role in the binding mode stabilization of these structures. Position 4 is found to be a disadvantageous site for substitution due to the size of the substituent that is a very restricting factor for the chemical synthesis. The modeling approach provided some insight into the molecular basis of indirubins' action and selectivity and allowed to forecast some improvements of this family of bis-indoles as kinase inhibitors. Predicted molecules, including 6-substituted and 5,6-disubstituted indirubins, were synthesized and evaluated as CDK and GSK-3 inhibitors.

In 2005, M. Zeng *et al.* [53] investigated the inhibition of GSK-3 by aloisine derivatives shown in Fig. (7) resorting to 3D-QSAR models. Thirty compounds constituted the training set, while another five conformed to the external test set.

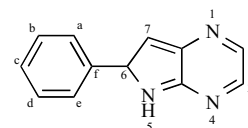


Fig. (7). General structure of aloisine derivatives.

Two alignment rules (template conformations) were employed; one extracted from the co-crystal structure of aloisine B (conformation 1 from Fig. (8)) with CDK-2, and the other from the lowest energy conformation (conformation 2 from Fig. (8)) which is found through a potential surface scan performed by Gaussian 98 Revision A.7. (Gaussian Inc.) Both alignments were used in order to check the influence of the spatial arrangement of the compounds in an approach that involves the CoMSIA and Comparative Molecular Field Analysis (CoMFA) techniques [39].

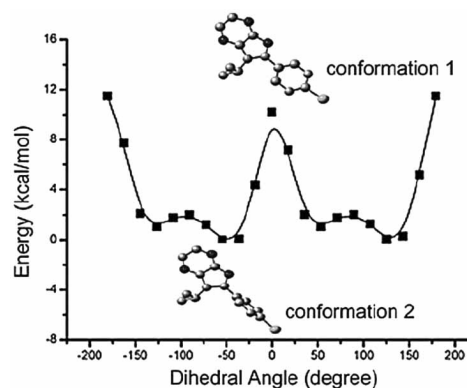


Fig. (8). Two template conformations of aloisine B, shown in the plot got from potential surface scan of dihedral angle ef67, performed by Gaussian 98 with B3LYP/6-31G* basis set.

Therefore, two sets of molecules were constructed and only the side chains were optimized, using AM1 method in BioMedCACHe 5.0. Gasteiger-Hückel charges were calculated with SYBYL 6.8 Molecular Modeling Software (Tripos Associated Ltd.), and 3D-structures were aligned based on template 1 and 2, respectively, generating two molecular aggregates. Since CoMFA model is highly sensitive to the space orientation of the molecular aggregate in the grid surrounding it, all-space searching (ASS) was employed to find the best CoMFA model, using Sybyl Pro-programming Language Script (SPL) written by Hou [54]. Grid step was 2.0 Å for the entire generated database. The standard fields available in Sybyl 6.8 were calculated, using sp^3 C-atom with +1 charge.

The Partial Least Square (PLS) method was applied to fit the 3D-structural features with their biological activities and check for consistence and predicting ability of the models. As it is appreciated from Table 3, CoMSIA provided the best QSAR model (highlighted in bold) for the higher energy conformation 1, with $R^2=0.938$ and $q^2=0.673$, involving steric, electrostatic and hydrophobic descriptors. This study demonstrated that hydrophobic interactions play an important role during inhibition, while the H-bond donor and acceptor fields do not possess such significance, manifested by the fact that all the other CoMSIA models in Table 3 that take into account these two fields do not generate good results. The predictions for both the training and test sets are plotted in Fig. (9).

The authors concluded that the biologically active conformation did not necessarily have the lowest energy because

when the inhibitor is complexed to the kinase, it cannot undergo conformation changes freely to lower its energy due to the confinement of kinase residues that interact with it. They supported this hypothesis by a superimposition of GSK-3 into the co-crystal structure of aloisine-CDK-2.

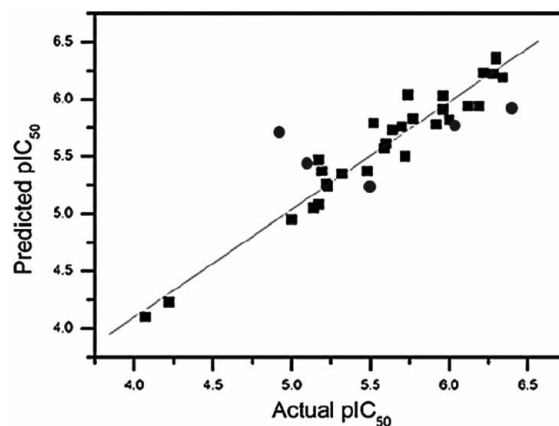


Fig. (9). Actual pIC_{50} versus predicted ones for the training set (squares) and the test set (circles) of CoMSIA 1 model.

In another publication by Lescot *et al.* [55], a 3D-QSAR CoMFA investigation for the inhibition of GSK-3 β on a training set consisting on 74 3-anilino-4-arylmaleimides is carried out (Fig. (10)). These studies allowed the analysis of a new GSK-3 β ligand as the thieno[2,3-*b*]pyrrolizinone de-

Table 3. Summary of CoMFA and CoMSIA Models for QSAR Study

	q^2	R^2	SEE	F	n	Field Contribution				
						S	E	H	D	A
CoMFA 1										
Before ASS	0.432	0.905	0.192	45.799	5	0.651	0.349			
After ASS	0.584	0.917	0.176	69.323	4	0.579	0.421			
CoMFA 2										
After ASS	0.344	0.894	0.200	48.765	5	0.696	0.304			
CoMSIA 1										
S+E	0.567	0.900	0.201	34.504	6	0.416	0.584			
S+E+H	0.673	0.938	0.158	58.372	6	0.248	0.398	0.355		
S+E+H+D	0.664	0.954	0.142	55.050	8	0.233	0.328	0.335	0.103	
S+E+H+A	0.665	0.956	0.139	57.391	8	0.233	0.327	0.356		0.084
S+E+H+D+A	0.650	0.962	0.134	55.513	9	0.224	0.280	0.344	0.073	0.080
CoMSIA 2										
S+E+H	0.289	0.964	0.124	83.967	7	0.273	0.336	0.392		

CoMFA 1 and CoMSIA 1 are based on aggregate 1; CoMFA 2 and CoMSIA 2 are based on aggregate 2.; S: steric field; E: electrostatic field; H: hydrophobic field; D: H-bond donor; A: H-bond acceptor; R^2 : squared correlation coefficient; q^2 : predictive R^2 from leave-one-out cross validation technique; SEE: standard error of estimation; F: Fisher-test parameter; n: number of descriptors of the model; ASS: all-space searching technique.

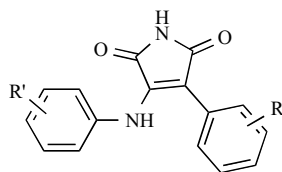


Fig. (10). General structure of maleimide derivatives.

rivative (shown in Fig. (11)). The comparison based on docking and simulation approaches confirm one preferential orientation of this particular ligand inside the active site, explaining the relationship with the reference 3-anilino-4-arylmaleimide derivatives and its biological affinity.

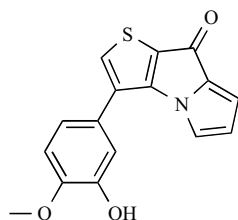


Fig. (11). Structure of the thieno[2,3-b]pyrrolizinone derivative.

Here, the 3D-structures of all compounds were obtained using the SYBYL 6.8 Molecular Modeling Software (Tripos Associated Ltd.). Partial atomic charges were calculated by the Gasteiger-Hückel method, and energy minimizations were performed using the Tripos Force Field [44] with a distance dependent dielectric and the Powell conjugate gradient algorithm (convergence criterion of 0.05 kcal.mol⁻¹). The conformation of a compound from Fig. (10) with R=3-Cl and R'=4-Cl, 3-COO co-crystallized with GSK-3 β was used as the basic structure for CoMFA alignment. Each molecule is initially aligned on the maleimide group, and then the dihedral angles were fixed by considering values of the reference compound. All substituents were oriented in the same way within the substitution series. 3D cubic lattice box with a grid spacing of 2 Å and 1 Å are created around the aligned molecules. Therefore, steric and electrostatic CoMFA fields were calculated using Csp³ probe atom (charge +1).

In this analysis, it was found that steric fields were highly correlated with the biological data, while the inclusion of electrostatic fields lead to a slight improvement of the PLS statistics. Electrostatic fields alone gave correct statistical result but clearly less than only steric fields. The best results were provided by a six component PLS model based only on steric fields and using a grid spacing of 2 Å: $R^2=0.891$, $q^2=0.805$, $S=0.146$. The steric contour plots from the CoMFA model reflected the unfavorable effect of methyl substitution onto the aniline NH linker and the favorable effect of *para* bulky groups on the 3-anilino ring. The results of the CoMFA analysis did not really correspond to the interactions recorded in the active site, but they characterized fundamental features (areas of the active site) of the interactions ligand-receptor.

Another study also based on the CoMFA methodology was developed by A. Martinez *et al.* [56], employing the


training set of 48 TDZD related compounds shown in Table 4. CoMFA models were calculated for each molecular field considered alone or in combination (see Table 5). The best PLS analysis was achieved by combining steric and electrostatic fields (model D), leading to a good correlation between the IC₅₀ values predicted from the principal components extracted from these two fields and the experimental data: $R^2=0.922$, $q^2=0.654$, $n=5$; see Fig. (12). In order to further validate model D, the activities of five representative compounds not included in the training set (compounds 4, 5, 7, 58, and 59) were predicted, being also in good agreement with experimental data (Table 6).

Since the relative contributions of steric and electrostatic molecular fields were 42% and 58%, respectively, it seemed that the inhibitory activity in TDZD and related compounds is modulated by a subtle balance of the two fields. In this study, the authors further performed mapping studies that enabled them to hypothesize two binding modes, which in turn might imply relevant differences in the mechanism that underlie the inhibitory activity of TDZDs.

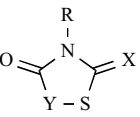
In a recent QSAR study of 2006 by Katritzky *et al.* [57], the Comprehensive Descriptors for Structural and Statistical Analysis (CODESSA-PRO) approach was applied. The authors reported QSAR models based on Multi-Linear Regression (MLR) and Artificial Neural Network (ANN) models, estimating the activity pIC₅₀ of 277 structurally diverse inhibitors of GSK-3 (GSK-3 α , GSK-3 β , and mixtures of both isoforms GSK-3 α/β) [58-65] by means of geometrical, topological, quantum mechanical and electronic type of descriptors. CODESSA-PRO enables the calculation of numerous quantitative descriptors solely on the basis of molecular structural information (Hansch-type approach) [66, 67]. Research using CODESSA PRO has successfully correlated and predicted various physical properties of interest [68-71].

The geometry of the inhibitors was first pre-optimized using the Molecular Mechanics Force Field (MM+) included in Hyperchem 7.5 (Hypercube). Final refined molecular geometries were obtained using the AM1 Semiempirical Method [72]. All the 960 molecular descriptors classified as (i) 38 constitutional, (ii) 38 topological, (iii) 14 geometrical, (iv) 367 charge-related, (v) 468 semiempirical, (vi) 35 thermodynamical, and were calculated using the CODESSA-PRO software. All the data employed were collected in Table 7, including the following information: (a) the CAS number for the 277 compounds (second column), and (b) experimental pIC₅₀ (third column), pIC₅₀ values calculated from the regression models (fourth column), and predicted pIC₅₀ values from a neural network model (fifth column). The experimental data points were divided by classes of compounds into four subsets as follows: *Class I*: $N=74$, 3-anilino-4-aryl-maleimides derivatives, (ii) *Class II*: $N=81$, 5-aryl-pyrazolo[3,4-b]pyridazines and N-phenyl-4-pyrazolo[1,5-b]pyridazin-3-yl-pyrimidin-2-amines derivatives, (iii) *Class III*: $N=62$, 5(6)-aryl-pyrazolo[3,4-b]pyridines and 6-hetero-aryl-pyrazolo[3,4-b]pyridines derivatives, and (iv) *Class IV*: $N=61$, [1-(1H-benzimidazol-7-yl)-1H-pyrazolo[3,4-d]pyrimidin-4-yl] arylhydrazones and [1-aryl-1H-pyrazolo[3,4-d]pyrimidin-4-yl] arylhydrazones derivatives. The templates for the four classes of GSK-3 inhibitors are shown in Fig. (13).

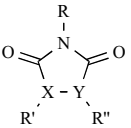
Table 4. GSK-3 Inhibition of TDZD Related Compounds

	X	Y	Z	A	B	R	IC ₅₀ (μM)
1	CH	CH	N	C=O	C=O	H	6
2	CH	CH	N	C=O	C=O	CH ₃	5
3	CH	CH	N	C=O	C=O	Bn	1
4	CH	CH	N	C=O	C=O	CH ₂ CO ₂ Et	3
5	CH	CH	N	C=O	C=O	CH ₂ Ph-4-OCH ₃	2.5
6	CH	CH	N	C=O	C=O	(CH ₂) ₂ Ph	2
7	CH	CH	N	C=O	C=O	(CH ₂) ₂ Ph-4-OCH ₃	3
8	CH	CH	N	C=O	C=O	(CH ₂) ₃ Ph	3
9	CH	CH	N	C=O	C=O	(CH ₂) ₃ Ph	3
10	CH	CH	CH ₂	C=O	C=O		12
11	CH	CH	O	C=O	C=O		>100
12	CH ₂	CH ₂	CH ₂	C=O	C=O		>100
13	CH ₂	CH ₂	N	C=O	C=O	H	>100
14	CH ₂	CH ₂	N	C=O	C=O	CH ₂ Ph	>50
15	CH ₂	CH ₂	N	C=O	CH ₂	CH ₂ Ph	>100
16	CH	CH	N	CH ₂	CH ₂	CH ₂ Ph	>100

Et: ethyl, Ph: phenyl, Bn: benzyl

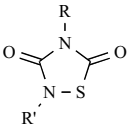
	n°.	X	Y	R	IC ₅₀ (μM)
	17	S	CH ₂	H	100
	18	S	CH ₂	CH ₃	100
	19	S	CH ₂	CH ₂ COOH	100
	22	S	CH ₂	Bn	25
	23	S	CH ₂	(CH ₂) ₂ Ph	35
	24	S	CH ₂	CH ₂ COPh	50
	25	O	CH ₂	Bn	100

Bn: benzyl

	n°.	X	Y	R	R'	R''	IC ₅₀ (μM)
28	S	S	CH ₃	-	-	-	100
31	N	N	Ph	H	H	H	100
32	N	N	Ph	CH ₃	CH ₃	CH ₃	100
35	N	N	Ph	-	-	-	100

Ph: phenyl

(Table 4. Contd....)

	R	R'	IC ₅₀ (μM)	n°	R	R'	IC ₅₀ (μM)
38	Et	CH ₃	5	49	Ph-2-Br	CH ₃	6
39	Et	Pr	10	50	Ph-4-CF ₃	CH ₃	6
40	Et	i-Pr	35	51	Ph-4-Cl	CH ₃	4
41	Et	C ₆ H ₁₁	10	52	Ph-4-F	CH ₃	4
42	Et	Et	25	53	Ph-4-OCH ₃	CH ₃	2
43	Bn	CH ₃	2	54	Ph-4-CH ₃	CH ₃	5
44	Bn	Et	7	55	Ph-4-NO ₂	Et	8.5
45	Bn	Bn	10	56	1-Np	CH ₃	2
46	Ph	CH ₃	2	57	C ₆ H ₁₁	CH ₃	100
47	Ph-4-Br	CH ₃	1.1	58	Ph	Et	6
48	Ph-3-Br	CH ₃	4	59	CH ₂ CO ₂ Et	CH ₃	2

Et: ethyl, i-Pr: isopropyl, Ph: phenyl, Bn: benzyl, Np: naphtyl

Table 5. Statistical Results of CoMFA Analysis

Eq.	Field	q^2	n	R^2	S	F
A	steric	0.403	5	0.862	0.251	49.85
B	electrostatic	0.495	3	0.775	0.313	48.22
C	H bond	0.194	2	0.393	0.507	13.92
D	steric+electrostatic	0.654	5	0.922	0.189	94.32
E	steric+lipophilic	0.274	2	0.498	0.462	21.30
F	electrostatic+lipophilic	0.440	4	0.780	0.313	36.33
G	steric+electrostatic+lipophilic	0.556	5	0.859	0.253	48.84

R^2 : squared correlation coefficient; q^2 : predictive R^2 from leave-one-out cross validation technique; S : standard deviation of the model; F : Fisher-test parameter; n : number of descriptors of the model.

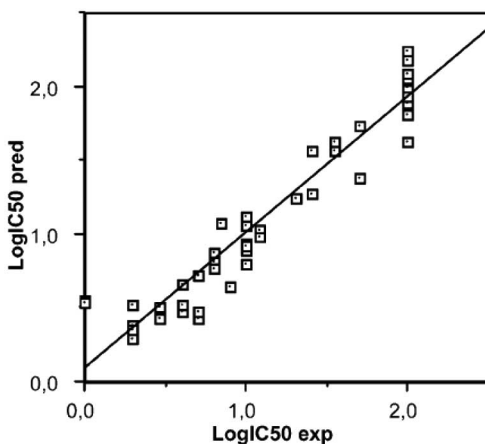
Fig. (12). Representation of experimental and predicted (model D) IC₅₀ for the training set.

Table 6. IC₅₀ Data (μM) for GSK-3 Inhibitors Included in the CoMFA Test Set

n°.	exp.	pred.	n°.	exp.	pred.
4	3	3.8	58	6	17
5	2.5	3.5	59	2	3.5
7	3	3			

n°.: number of compound; exp.: experimental IC₅₀; pred.: predicted IC₅₀ with the model.

Table 7. Experimental and Predicted pIC₅₀ Values^a for GSK-3

n°	CAS	exp.	MLR	ANN	n°	CAS	exp.	MLR	ANN
1	101291-07-0	6.277	6.133	6.422 ^b	140	681432-90-6	7.301	7.149	7.743
2	264207-26-3	6.521	6.357	6.628 ^b	141	681432-92-8	4.721	4.626	4.576
3	264214-83-7	6.152	6.029	6.352 ^b	142	681432-96-2	4.638	4.703	6.860 ^b
4	264214-79-1	6.827	6.821	7.153 ^b	143	748142-01-0	7.097	7.550	7.186
5	264214-82-6	6.536	6.707	6.669	144	548797-12-2	7.004	6.538	6.369
6	264214-81-5	6.845	6.707	6.669	145	405222-59-5	8.398	8.015	7.411 ^b
7	264206-86-2	6.394	6.417	6.150	146	439290-41-2	8.155	7.365	6.482
8	264208-87-9	5.583	5.579	5.595	147	557113-38-9	5.569	6.873	5.441
9	264213-20-9	6.666	6.673	6.590	148	557113-39-0	6.161	7.240	6.840
10	264210-58-4	6.710	6.764	6.812 ^b	149	405222-60-8	7.658	7.425	7.750
11	264214-76-8	6.427	6.372	6.828	150	405223-00-9	7.959	8.884	7.627
12	264216-52-6	6.818	6.851	6.865	151	405223-04-3	8.155	7.742	7.892
13	264214-72-4	7.032	7.208	6.826	152	405222-61-9	8.301	8.880	7.859
14	264214-75-7	6.866	6.864	6.842	153	405222-94-8	8.046	7.089	8.262
15	264214-74-6	7.131	7.139	7.469 ^b	154	405224-05-7	8.301	8.198	8.326 ^b
16	264210-60-8	6.793	6.598	6.623 ^b	155	405222-72-2	8.301	7.727	7.944
17	264222-25-5	6.472	6.569	6.182 ^b	156	583038-29-3	7.125	7.336	6.678
18	264213-24-3	6.666	6.683	6.633	157	583038-60-2	9.097	8.803	8.770 ^b
19	264210-49-3	6.943	6.759	6.777	158	583038-34-0	8.097	7.759	8.518
20	264213-50-5	6.587	6.378	6.749	159	583038-56-6	8.301	8.487	8.532
21	264213-63-0	6.857	6.936	6.635	160	583038-54-4	8.155	8.649	8.124 ^b
22	264213-72-1	7.086	7.174	6.837	161	583038-58-8	7.620	7.746	7.580
23	264208-22-2	6.959	6.956	7.765 ^b	162	583038-71-5	8.398	8.178	8.243 ^b
24	264222-23-3	6.728	6.574	6.834	163	583038-40-8	7.921	7.816	7.526
25	264214-12-2	6.983	7.050	6.751	164	548797-19-9	6.303	6.435	6.571 ^b
26	264218-33-9	6.600	6.701	6.838	165	548797-18-8	7.824	7.997	7.961 ^b
27	264218-23-7	6.983	7.108	7.254 ^b	166	548797-15-5	7.377	6.364	6.932
28	264216-07-1	7.284	7.416	7.230 ^b	167	548797-37-1	6.318	6.311	6.550 ^b
29	264217-67-6	7.553	7.558	7.455 ^b	168	548797-26-8	6.082	6.898	6.247

(Table 7. Contd....)

n°	CAS	exp.	MLR	ANN	n°	CAS	exp.	MLR	ANN
30	264222-45-9	6.883	6.764	6.656	169	548797-33-7	6.495	6.396	6.825
31	264217-40-5	5.830	6.254	6.825	170	548797-32-6	7.301	6.650	7.519
32	264217-28-9	7.027	6.939	7.152 ^b	171	548797-34-8	7.456	6.896	6.847
33	264215-79-4	7.237	7.159	6.886	172	548797-14-4	6.668	6.446	6.902
34	264215-80-7	6.873	6.797	7.221 ^b	173	548797-17-7	6.483	6.657	6.943
35	264217-24-5	7.119	7.060	6.858	174	583038-96-4	7.409	7.626	7.589
36	264210-27-7	6.274	6.455	6.461	175	583039-51-4	8.155	7.791	7.853
37	264222-36-8	6.337	6.135	6.434	176	583039-55-8	6.851	6.825	6.916
38	264210-48-2	6.590	6.475	6.719	177	583039-27-4	8.155	7.732	8.051
39	264215-20-5	6.326	6.156	6.791	178	583039-39-8	7.004	7.353	7.117
40	264215-16-9	6.848	6.833	6.820	179	583039-44-5	7.796	7.631	8.201 ^b
41	264215-19-2	6.710	6.722	6.890 ^b	180	583039-25-2	7.745	7.210	7.777
42	264215-18-1	7.071	6.963	7.175 ^b	181	583039-36-5	6.602	6.451	6.658
43	264207-93-4	6.693	6.666	6.791	182	107042-54-6	6.276	6.459	6.520
44	264213-31-2	6.851	6.864	6.439	183	405224-27-3	6.367	5.828	6.122 ^b
45	264209-34-9	7.155	6.908	6.634	184	405224-21-7	5.900	6.243	6.016 ^b
46	264209-39-4	6.627	6.604	6.949 ^b	185	439290-93-4	6.536	6.809	7.310 ^b
47	264213-38-9	6.910	6.890	6.848	186	405221-12-7	7.367	6.839	7.280 ^b
48	264213-25-4	7.229	7.138	6.860	187	405221-13-8	7.252	7.088	7.168 ^b
49	264211-21-4	7.699	7.441	7.485 ^b	188	405221-08-1	7.721	7.246	7.598 ^b
50	264213-05-0	7.102	7.161	7.372 ^b	189	405221-32-1	8.301	7.342	7.372 ^b
51	264214-59-7	7.585	7.455	6.866	190	405221-87-6	5.551	5.870	7.329 ^b
52	264209-30-5	6.818	6.794	6.859 ^b	191	557112-45-5	5.447	5.446	5.973 ^b
53	264208-99-3	5.854	5.838	5.621	192	557112-46-6	6.449	6.743	6.794
54	264222-21-1	6.793	6.641	6.698	193	557112-47-7	5.630	6.996	6.107
55	264207-11-6	6.289	6.309	6.401	194	557112-48-8	7.745	7.417	7.311 ^b
56	264207-01-4	6.350	6.565	6.533	195	405221-39-8	7.699	7.631	7.488 ^b
57	264208-84-6	6.390	6.208	6.814	196	405221-48-9	8.155	7.858	6.780
58	264209-73-6	6.499	6.492	6.832	197	405221-61-6	7.569	7.725	7.318 ^b
59	264216-61-7	6.762	6.729	6.839	198	405221-38-7	7.959	7.318	7.836
60	264209-67-8	7.041	7.029	7.409 ^b	199	405221-09-2	6.354	7.333	6.399
61	264211-44-1	6.730	6.853	6.729	200	405221-67-2	6.070	6.854	7.220 ^b
62	264211-48-5	6.963	7.133	6.846	201	405221-69-4	6.772	7.242	7.455 ^b
63	264207-08-1	6.277	6.449	6.486	202	557112-49-9	6.618	6.919	6.666
64	264208-93-7	5.641	5.848	6.473	203	405221-70-7	6.372	6.099	6.541
65	264222-01-7	5.850	6.179	6.164 ^b	204	583038-30-6	8.097	8.345	8.092
66	264206-99-7	6.409	6.409	6.603	205	583038-28-2	6.903	6.689	6.649

(Table 7. Contd....)

n°	CAS	exp.	MLR	ANN	n°	CAS	exp.	MLR	ANN
67	264207-22-9	6.807	6.812	6.967 ^b	206	583038-36-2	7.444	7.684	7.663
68	264215-15-8	6.318	6.245	6.783	207	583038-42-0	5.798	6.218	5.943
69	264215-12-5	7.081	6.962	6.771	208	583038-38-4	9.000	8.336	8.346
70	264215-14-7	6.670	6.912	6.674	209	583038-63-5	8.222	7.964	8.150
71	264206-97-5	6.614	6.706	6.782	210	583038-51-1	6.382	6.248	6.953 ^b
72	264222-11-9	6.159	6.280	6.358 ^b	211	583038-33-9	6.630	6.906	6.659
73	264215-97-6	7.149	7.273	6.887	212	583038-93-1	7.060	7.269	7.457
74	264211-18-9	6.407	6.539	6.543	213	583039-16-1	6.417	7.244	6.728
75	551919-61-0	7.721	7.567	7.468	214	583038-84-0	7.921	8.094	8.515
76	748142-08-7	5.222	7.148	6.005	215	583038-46-4	9.000	9.142	8.547
77	551920-54-8	8.000	7.480	7.925	216	583038-76-0	7.678	7.514	8.269 ^b
78	748142-06-5	7.509	6.857	7.016	217	583038-48-6	7.125	7.336	6.678
79	681432-33-7	6.903	7.012	7.166	218	319490-29-4	7.000	6.539	7.438
80	681432-36-0	7.301	6.821	7.019	219	710947-39-0	5.500	6.713	5.795
81	748142-09-8	7.921	7.995	8.024	220	710947-40-3	5.600	6.072	6.434 ^b
82	681432-47-3	8.000	7.882	7.763	221	650837-82-4	7.000	6.811	7.173 ^b
83	681432-38-2	8.000	7.618	7.530 ^b	222	710947-42-5	6.800	6.224	6.675
84	681432-32-6	7.921	7.211	7.796	223	710947-43-6	6.800	6.173	6.959
85	748142-07-6	7.796	8.321	7.960	224	650626-35-0	5.600	6.305	5.718
86	748141-84-6	7.301	7.151	7.431 ^b	225	650626-94-1	8.200	7.742	6.827
87	748141-85-7	6.701	6.062	6.664	226	650626-57-6	6.000	7.030	7.061
88	748141-88-0	6.100	6.177	6.294	227	650626-54-3	7.000	6.407	7.609
89	748142-10-1	7.000	7.295	7.769	228	650626-56-5	6.500	6.636	6.347
90	748142-11-2	7.301	7.360	7.448 ^b	229	650626-52-1	5.400	5.953	5.911 ^b
91	748142-12-3	7.301	7.565	7.383 ^b	230	650626-51-0	5.700	6.514	5.950
92	748142-13-4	7.102	7.151	7.653	231	710947-45-8	6.800	6.567	6.749
93	748142-14-5	6.903	6.948	6.372	232	710947-46-9	6.200	7.101	6.264 ^b
94	748142-15-6	6.701	6.692	6.549	233	650626-79-2	6.500	7.200	5.693
95	748142-16-7	5.500	6.028	6.779 ^b	234	650626-27-0	6.600	7.212	6.791
96	748142-17-8	6.201	5.565	6.426	235	650637-59-5	5.900	6.859	6.632 ^b
97	748142-18-9	7.409	7.747	7.420 ^b	236	710947-48-1	7.500	7.291	6.609 ^b
98	748142-19-0	7.301	7.912	6.853	237	650637-83-5	8.500	7.525	8.443
99	748142-20-3	8.000	6.964	7.836 ^b	238	650637-35-7	7.400	6.785	7.334
100	748142-21-4	8.000	8.437	7.803	239	650637-81-3	7.400	6.515	7.330
101	748141-95-9	6.801	6.356	6.675	240	650627-96-6	8.000	7.178	7.940
102	748142-22-5	6.903	6.640	7.051	241	650627-94-4	7.600	7.573	7.548 ^b
103	681432-49-5	8.000	7.438	6.686	242	650627-93-3	7.500	7.640	6.945 ^b

(Table 7. Contd....)

n°	CAS	exp.	MLR	ANN	n°	CAS	exp.	MLR	ANN
104	681432-52-0	7.921	7.348	6.722	243	650627-92-2	6.900	7.679	7.876
105	681432-56-4	8.000	7.659	6.727 ^b	244	650627-66-0	8.100	7.737	7.921
106	681432-55-3	7.959	7.566	7.879	245	650627-68-2	8.100	7.863	7.900
107	681432-53-1	7.699	7.476	6.598	246	650627-67-1	7.800	7.881	7.629
108	681432-57-5	7.699	6.932	7.802	247	650627-43-3	8.000	7.939	7.571
109	681432-61-1	7.699	7.646	7.425	248	650627-36-4	8.600	7.936	7.770
110	681432-63-3	7.602	6.963	7.597	249	650627-89-7	8.400	8.652	8.415
111	681432-62-2	7.398	6.661	7.458	250	650627-42-2	8.600	8.347	8.271 ^b
112	681432-60-0	8.000	7.086	7.749 ^b	251	650627-79-5	7.900	7.685	7.535
113	681432-02-0	5.300	5.856	5.463	252	650627-77-3	6.800	7.098	6.716
114	681432-05-3	4.500	5.267	4.539	253	650627-80-8	5.600	6.093	5.389
115	681432-06-4	5.100	5.335	4.818	254	650627-65-9	8.200	7.787	8.419
116	681432-10-0	4.509	5.230	4.705	255	650627-54-6	5.600	7.552	5.622
117	681432-07-5	5.300	5.874	5.037	256	650627-27-3	8.300	8.314	8.110 ^b
118	681432-65-5	6.801	6.794	6.605	257	710947-51-6	8.400	7.520	8.230
119	681432-70-2	6.500	6.429	6.174	258	710947-52-7	7.900	7.499	8.248 ^b
120	681432-67-7	7.301	6.820	6.689	259	710947-53-8	7.800	7.592	7.693
121	681432-69-9	6.000	6.293	6.020 ^b	260	710947-54-9	8.300	8.414	8.091 ^b
122	681432-20-2	5.903	7.141	6.618	261	650627-81-9	7.800	7.674	8.026 ^b
123	681432-23-5	5.600	6.522	6.546	262	706809-13-4	7.500	7.918	7.173
124	681432-24-6	6.000	6.519	6.046	263	706809-12-3	7.800	8.442	7.747
125	681432-26-8	5.400	6.142	5.922 ^b	264	710947-56-1	8.100	8.470	8.235
126	681432-25-7	6.099	6.826	6.349	265	710947-57-2	8.000	7.524	8.261 ^b
127	681432-71-3	7.301	8.152	7.449 ^b	266	710947-58-3	8.000	7.064	7.960 ^b
128	681432-75-7	7.398	7.093	7.248 ^b	267	650628-06-1	7.900	7.051	8.067 ^b
129	681432-77-9	7.509	6.661	7.122	268	650628-03-8	8.000	7.001	8.112 ^b
130	681432-76-8	7.000	7.065	7.093	269	650627-98-8	7.500	6.640	7.666
131	681432-74-6	7.495	7.904	7.684	270	710947-59-4	7.200	7.121	8.096 ^b
132	681432-78-0	6.100	6.184	6.733	271	710947-60-7	5.200	7.265	5.842
133	681432-83-7	5.801	5.981	5.881	272	650840-91-8	6.200	7.015	6.350 ^b
134	681432-82-6	6.201	6.135	6.239 ^b	273	706809-14-5	7.700	7.017	7.804
135	681432-84-8	6.801	6.888	7.215 ^b	274	650637-85-7	8.400	7.692	8.469
136	681432-85-9	7.602	7.182	7.599 ^b	275	650637-86-8	8.200	7.877	7.902 ^b
137	681432-89-3	7.602	6.900	8.329 ^b	276	706809-15-6	8.800	8.317	8.698
138	681432-91-7	7.796	7.737	8.220	277	650637-87-9	8.300	8.871	8.474 ^b
139	681432-88-2	7.796	7.605	7.380	278	706809-16-7	8.300	8.089	8.349 ^b

^a IC₅₀ (nM), taken from the original reference as follows: *Class I* from [58]; *Class II* from [59] (compounds 75-143) and [61] (compounds 144-155); *Class III* from [63] (compounds 156-181), [60] (compounds 182-203), and [62] (compounds 204-217); and *Class IV* from [65] (compounds 218-271) and [64] (compounds 272-278). ^b validation set. ^c n°.: number of compound; exp.: experimental pIC₅₀; MLR: predicted pIC₅₀ with Multi-Linear Regression Analysis; ANN: predicted pIC₅₀ with Artificial Neural Networks.

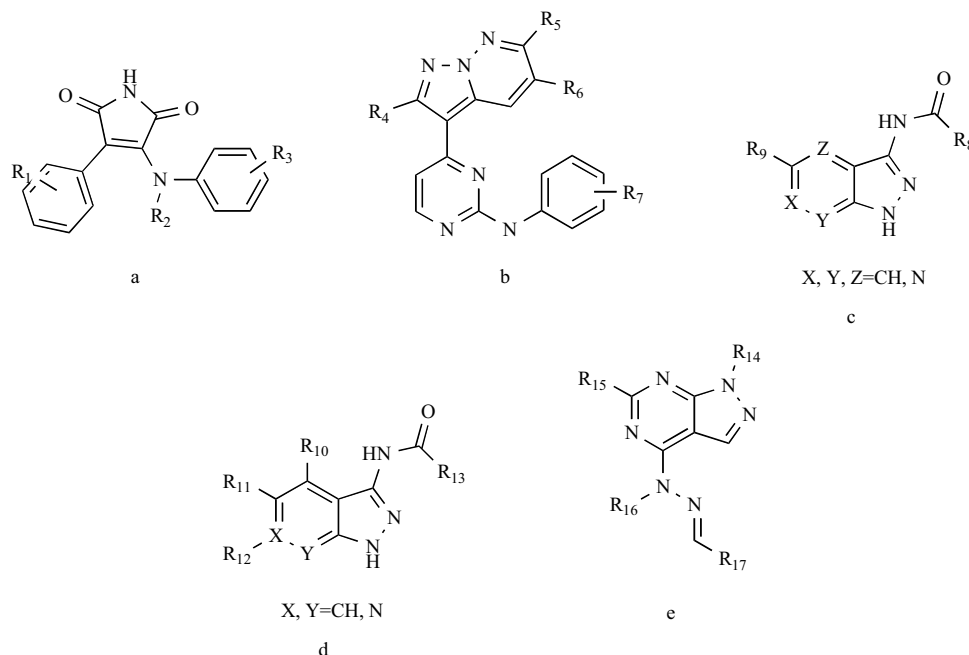


Fig. (13). The templates of GSK-3 inhibitors: (a) Class I; (b), (c) Class II; (d) Class III; (e) Class IV.

The Best Multilinear Regression (BMLR) [73] procedure was first applied on *Class I* to obtain the multilinear QSAR, selecting the optimal number of descriptors to be included in the model from the analysis of multi-parameter correlations containing up to 10 descriptors. Fig. (14) shows the relationships of R^2 with the number of variables. As can be seen, R^2 rises steeply as the number of parameters increases from 2 to 10 and a breakpoint (lower increase of R^2) occurs at the sixth descriptor. The R^2 for the six and seven parameter models are 0.896 and 0.904, respectively; $\Delta R^2=0.008<0.02$). Therefore, the best correlation equation corresponds to a model containing six descriptors. The QSAR equation for *Class I* is characterized by the following statistical parameters, shown in Table 8: $N=74$, $n=6$, $R^2=0.896$, $q^2=0.874$, $F=96.683$, $S^2=0.019$ where N is the number of data points; n is the number of descriptors and q^2 is the squared cross-validated correlation coefficient.

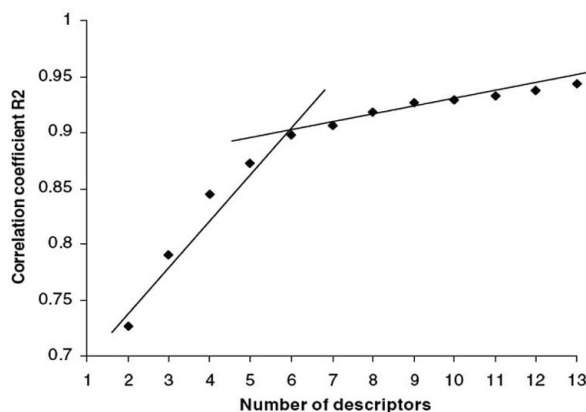


Fig. (14). Correlation coefficient versus the number of descriptors for Class I of compounds.

For this model, two compounds were registered as outliers: entries **31** and **65** from Table 7, displaying lower values for the predicted pIC_{50} values. Moreover, indoline (**65**) is supposed to have a different binding mode with respect to N-methyl-3-anilino-4-arylmaleimides and 3-anilino-4-arylmaleimides [58]. Table 9 includes the loo validation results for the QSAR models obtained for different classes. Furthermore, an internal validation was practiced that predicts the property values for each one third of the compounds with the model fitted for the remaining two third of the compounds.

The six types of descriptors involved in equation for *Class I* (Table 8) were: topological (D_1), molecular orbital related (D_2 , D_3), quantum chemical (D_6), electrostatic (D_4 , D_5). The t -test indicated the following order of significance for the descriptors included in the equation: $D_1>D_2>D_3>D_4>D_5>D_6$. The direct interpretation of the descriptors appearing in the model, nevertheless, is rather difficult, considering the complex nature of the GSK-3 inhibition processes. However, some indirect links between those descriptors and the physico-chemical phenomena behind the inhibitor-interactions with the GSK enzyme were suggested.

The most statistically significant descriptor is the Randic index (order 1) (D_1) that has a topological origin and is a measure of the compactness of the molecule [74]. In Eq. (4), D_i , D_j are the edges of the adjacent atoms i and j and the summation is performed over all pairs of edges of the molecule. The magnitude of the drug-receptor interaction is directly related to the degree of branching as well as to the molecular size. In Table 8, this descriptor has a large positive influence.

$$^m\chi = \sum_{\text{path}} (D_i D_j \dots D_k)^{-1/2} \quad (4)$$

Table 8. Selected Descriptors and Statistical Characteristics for the Best QSAR Models for Classes I-IV

Descriptor Name	ID	<i>b</i>	Δb	<i>t</i>	R^2	q^2	S^2
Class I							
Intercept	-	76.064	4.689	16.221	-	-	-
Randic index (order 1)	D ₁	0.539	0.034	15.839	0.440	0.410	0.099
Average bond order for atom H	D ₂	-89.158	6.410	-13.907	0.655	0.624	0.062
LUMO energy	D ₃	1.803	0.156	11.510	0.755	0.727	0.044
HACA-2 (MOPAC PC)	D ₄	-0.315	0.035	-8.863	0.822	0.795	0.032
Charged Surface Area (MOPAC PC) for atom N	D ₅	0.703	0.105	6.681	0.871	0.845	0.024
Max coulombic interaction for C-C	D ₆	1.755	0.436	4.024	0.896	0.874	0.019
Class II							
Intercept	-	44.352	23.914	1.854	-	-	-
Shadow plane ZX	D ₇	-0.055	0.006	-8.145	0.170	0.129	0.890
Max e-n attraction for H-C	D ₈	-1.906	0.254	-7.499	0.180	0.118	0.890
Max electroph. react. index for C	D ₉	197.519	27.334	7.225	0.374	0.311	0.689
Max SIGMA-PI bond order	D ₁₀	-20.261	3.243	-6.246	0.496	0.436	0.561
n°. H-donors sites (MOPAC PC)	D ₁₁	0.154	0.030	5.089	0.560	0.498	0.496
XY Shadow / XY Rectangle	D ₁₂	-8.689	2.370	-3.665	0.635	0.567	0.417
Max exchange energy for H-C	D ₁₃	18.463	5.996	3.078	0.677	0.593	0.374
Class III							
Intercept	-	95.451	27.757	3.438	-	-	-
HASA-1 (MOPAC PC) (all)	D ₁₄	0.013	0.002	6.334	0.254	0.204	0.612
Max partial charge (Zefirov) for atoms for atom H	D ₁₅	44.101	7.332	6.014	0.505	0.466	0.413
Min e-e repulsion for bond C-C	D ₁₆	-0.129	0.022	-5.700	0.622	0.570	0.321
Min 1-electron react. index for C	D ₁₇	-47.361	12.643	-3.745	0.707	0.660	0.253
Max atomic state energy for C	D ₁₈	-0.759	0.264	-2.867	0.745	0.703	0.224
Class IV							
Intercept	-	120.565	28.880	4.174	-	-	-
Max exchange energy for H-C	D ₁₃	-22.740	5.497	-4.136	0.126	0.070	0.867
Shadow plane YZ	D ₁₉	-0.089	0.024	-3.736	0.135	0.025	0.873
Structural Information content (order 2)	D ₂₀	0.105	0.034	-3.018	0.328	0.211	0.690
Tot hybridization comp. of the molecular dipole	D ₂₁	0.433	0.151	2.868	0.439	0.322	0.586
Highest normal mode vib transition dipole	D ₂₂	-0.109	0.047	-2.273	0.475	0.343	0.558
Max e-e repulsion for bond H-C	D ₂₃	0.216	0.117	1.846	0.507	0.370	0.535

b: regression coefficient of the linear model; Δb : standard error for each regression coefficient; *t*: *t*-test value for each coefficient *b*; ID: symbol for descriptor; R^2 : correlation coefficient; q^2 : predictive R^2 from leave-one-out cross validation technique; *S*: standard deviation of the model.

Table 9. Cross Validation Results for the Proposed Linear QSAR Models

Class	R^2	q^2	$\Delta R^2 - R^2 - q^2$
I	0.896	0.876	0.020
II	0.677	0.593	0.084
III	0.745	0.703	0.042
IV	0.507	0.370	0.137

R^2 : squared correlation coefficient; q^2 : predictive R^2 from leave-one-out cross validation technique

According to the *t*-test values, the second important descriptor is the average bond order for atom H. Bond orders can be related to the bonding nature σ/π . They are defined as the covariance of the electron population of two atoms and are related to the total atomic valences and atomic charges [75, 76]. The presence of this descriptor in the QSAR model (Table 8) suggests the hydrogen bonding ability of the inhibitor and emphasizes the electrostatic interactions GSK-3-inhibitor. LUMO energy is usually approximated with electron affinity, and characterizes the susceptibility of the molecule toward the attack by nucleophilic reactants. The maximum coulombic interaction for bond C-C, descriptor (D_6) accounts for the sum of the electronic repulsion energy, electron-nuclear attraction energy and nuclear repulsion energy between N and H atoms (Eq. (5)). In the equation found, descriptor D_6 has a small positive influence. The additive energy between two atoms A and B can be expressed as:

$$E(AB) = E_{ee}(AB) + E_{ne}(AB) + E_{nn}(AB) \quad (5)$$

where $E_{ee}(AB)$ is the electronic repulsion energy between two atoms, $E_{ne}(AB)$ is the electron-nuclear attraction energy between two atoms and $E_{nn}(AB)$ is the nuclear repulsion energy between two atoms.

Electrostatic descriptors HACA-2 (MOPAC PC) and charged surface area for atom N describe the electrostatic component of the inhibitor-GSK-3 interaction.

$$HACA = \sum_A \frac{q_A \sqrt{S_A}}{\sqrt{S_{tot}}} \quad (6)$$

where S_A represents the solvent accessible surface area of hydrogen-bonding acceptor atoms, selected by threshold charge, q_A is the partial charge on hydrogen-bonding acceptor atoms selected by threshold charge, and S_{tot} is the total solvent-accessible molecular surface area.

The resulting QSAR model (Table 8) obtained by CoDESSA PRO for Class II compounds has the following statistical characteristics: $N=81$, $n=7$, $R^2=0.677$, $q^2=0.593$, $F=21.900$, $S^2=0.374$, where the number of significant descriptors was again defined by using the break point rule. The best QSAR for Class III of compounds involved a five-parameter model, with $N=61$, $n=5$, $R^2=0.745$, $q^2=0.703$, $F=32.240$, $S^2=0.224$, while the QSAR for Class IV is characterized by: $N=61$, $n=6$, $R^2=0.507$, $q^2=0.370$, $F=9.245$, $S^2=0.535$. The model for Class IV is poorer according to its sta-

tistical criteria. Obviously, modeling the 61 [1-(1H-benzimidazol-7-yl)-1H-pyrazolo[3,4-d]pyrimidin-4-yl] arylhydrazones and [1-aryl-1H-pyrazolo[3,4-d]pyrimidin-4-yl] arylhydrazones derivatives was not an easy task.

The QSAR models obtained for classes II-IV involved descriptors that can be related to size and shape (shadow plane YZ (D_{19}), shadow plane ZX (D_7), XY shadow/XY rectangle(D_{12})), structural information content of 2nd order(D_{20})), electrostatic and hydrogen bonding (count of H donor sites (D_{11}), HASA-1 (MOPAC PC) all (D_{14}), maximum partial charge (Zefirov) for atom H (D_{15}), total hybridization component of the molecular dipole (D_{21})). Therefore, it was concluded that the interaction between inhibitor and GSK-3 occurs by hydrogen bonding. It is in accordance with 3D-QSAR studies on GSK-3 inhibition that reported the involvement of steric, hydrophobic and electrostatic factors into the calculated interaction energy, and emphasize the influence of these factors on the inhibitory activity [39, 49, 53]. Due to molecular flexibility, the compounds belonging to these classes include a large number of conformers. The lower quality of these QSAR could be explained taking into account the complexity of the modeled property and the errors related to the experimental data.

In this study, the authors employ the ANN methodology for predicting the activity values by partitioning the 277 experimental data for pIC₅₀ into training ($N=187$) and validation subsets ($N=90$). They discarded some descriptors from the large pool consisting on 961 descriptors resorting to the following procedures: (i) descriptors with both high intercorrelations ($R^2>0.6$) and at probability level $p<0.05$, thus 722 descriptors are excluded; (ii) descriptors with small variance ratio $\sigma/d_{\max} - d_{\min} < 10^{-6}$ are excluded (123); (iii) descriptors for which no values are available for all structures are excluded (83). Thus, the descriptor pool is reduced to 33 descriptors; (iv) from this reduced pool of variables, another 21 are rejected since they show random variations by exploring the scatter plots between the property and the corresponding descriptor. Therefore, the final descriptor pool is reduced to 12 descriptors for which sensitivity-stepwise analysis is performed by building the ANN models with simple 1-1-1 architecture for each relevant descriptor. Those descriptors that show the lowest prediction error at the ANN output were chosen for building the optimum ANN model, thus leading to six descriptors to be employed.

During the training stage the weights were adjusted according to the output prediction error by using the back-propagation ANN [77, 78]. The validation set error (and also

R^2) was monitored in order to avoid the over-training of the ANN and to stop the training process. It was found that a six descriptor model (6-6-6-1) was appropriate for the pIC_{50} property. The RMS for the training and validation data was 0.67 and 1.54, respectively. In addition, an exploration of the standard deviations of the neural networks models with different numbers of hidden units was performed. The seven descriptor models (7-6-6-1) did not give significant improvement over the six-descriptor models (RMS=1.11). The same result was found for the 5-6-6-1 models with increased hidden units (RMS=0.95). The predicted values of pIC_{50} obtained are given in Table 7. Graphical presentations as a linear fit for the training set and test set are given in Fig. (15).

The maximum squared correlation coefficient for the training set was $R^2=0.78$ for 187 experimental data points. The corresponding test set (which was not used to train the ANN model) had a maximum $R^2=0.67$ at which the training of the network was stopped. As can be seen from Table 7, the ANN model has superiority over the multilinear QSAR models. The ANN model included the following descriptors used as inputs: count of H-acceptor sites (Zefirov), HA dependent HDSA-1 (Zefirov), Kier and Hall index, Minimum partial charge for all atom types, RNCG Relative negative charged SA (SAMNEG/RNCG) (MOPAC PC), and Shadow plane ZX. Most of these descriptors are charge-related descriptors showing the importance of the electrostatic interactions between the inhibitors and the GSK enzyme. These studies gave an insight into the dominant role played by the electrostatic, bonding and steric interactions on the modulation of the inhibitory activity. As a result of the current investigations, the nature of GSK-3-inhibitor interaction was found to be electrostatic.

In a later study of 2006, Sivaprakasam *et al.* [79] reported on the inhibition of GSK-3 α by 3-anilino-4-phenylmaleimides (Fig. (16)) by means of two 2D-QSAR approaches: the Fujita-Ban [80] and the Hansch analysis [81]. Even though the GSK-3 α and GSK-3 β isoforms are 97% identical with respect to their kinase domain, studies performed by Phiel *et al.* [28] showed that selective reduction in concentration of the α isoform led to a decrease in the concentration of A β 40 and A β 42, primary constituents of amyloid plaques in Alzheimer's disease (AD), whereas decreased expression of the β isoform led to a slight increase in the

concentration of the two A β peptides. Thus, inhibition of GSK-3 α could potentially provide dual therapy against AD, preventing the buildup of amyloid plaques and of neurofibrillary tangles [28, 82].

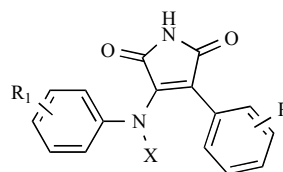


Fig. (16). Chemical structure of studied 3-anilino-4-phenylmaleimides.

Table 10 shows the two Fujita-Ban models established. Model0-a was obtained by fitting the 16 indicator variables to the set of 64 3-anilino derivatives, thus leading to $R^2=0.880$, $S=0.144$, $F=21.520$. Model0-b resulted from the application of the PLS technique with a significance level of 95%, thus leading to the selection of 9 variables only exhibiting $N=64$, $R^2=0.845$, $S=0.152$, $F=32.710$. It is evident that substituents 2-Cl, 3-Cl, 4-Cl, 2-OCH₃, 3-OCH₃, 4-OCH₃, 2-NO₂, 3-NO₂, and 4-NO₂ at the 4-phenyl ring (R) are positively correlated with GSK-3 α inhibitory activity in model0-a. Similarly, substituents such as 3-Cl, 4-Cl, 5-Cl, 3-COOH, 4-OH, and 4-SCH₃ at the 3-anilino ring (R_1) are positively and 3-OH is negatively correlated with GSK-3 α inhibitory activity. According to both models, 3-NO₂ and 2-NO₂ contribute the most to the activity.

The best Hansch-type models established by the authors resorting to the PLS technique for 64 3-anilino and 3 3-N-methyl anilino derivatives are shown in Table 11. Indicator variable I_{XCH_3} (which equals 1 for 3-N-methylanilino derivatives and 0 for 3-anilino derivatives) enabled to study both groups of compounds together. The negative contribution of I_{XCH_3} shows that the 3-N-methylanilino derivatives are unfavorable for GSK-3 α inhibitory activity in comparison to 3-anilino derivatives, as already noted by Smith *et al.* [58]

The indicator I_{4ClR1} represents the contribution of the structural feature 4-Cl at R_1 , which is a positive number as revealed by Table 11. This point to the enhancement of the activity at the anilino ring. The also positive contribution of

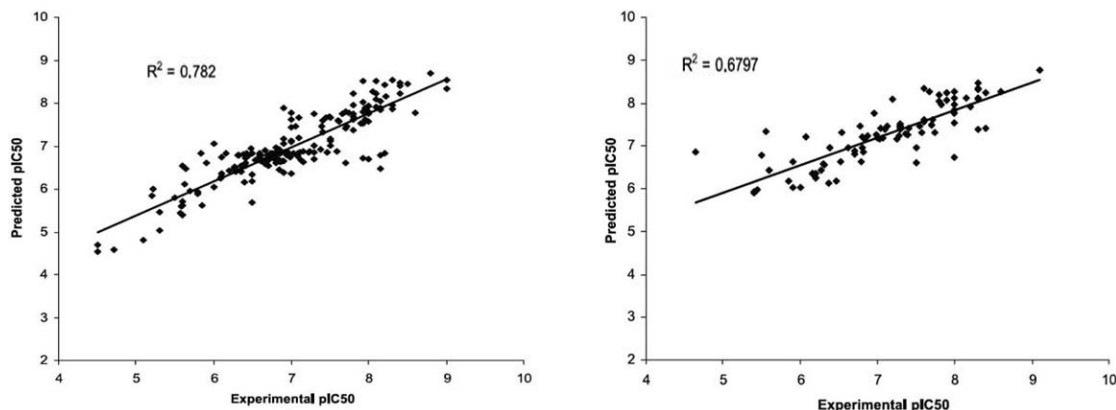


Fig. (15). Plot of predicted (with ANN) versus observed pIC_{50} values for the training set (left) and test set (right).

Table 10. Fujita–Ban Contribution of Substituents Over 4-phenyl and 3-Anilino Ring

Substitution	Ring ^a	Freq. ^b	Model0-a	Model0-b
parent structure			6.255	6.436
2-Cl	<i>R</i>	8	0.286	0.167
2-OCH ₃	<i>R</i>	6	0.407	0.285
3-NO ₂	<i>R</i>	9	0.605	0.492
3-Cl	<i>R</i>	6	0.157	—
4-Cl	<i>R</i>	9	0.086	—
4-OCH ₃	<i>R</i>	6	0.212	—
2-NO ₂	<i>R</i>	5	0.485	0.371
3-OCH ₃	<i>R</i>	5	0.157	—
4-NO ₂	<i>R</i>	2	0.167	—
3-Cl	<i>R_I</i>	24	0.227	0.214
3-OH	<i>R_I</i>	9	-0.159	-0.220
4-OH	<i>R_I</i>	18	0.121	—
5-Cl	<i>R_I</i>	10	0.269	0.346
3-COOH	<i>R_I</i>	14	0.315	0.254
4-Cl	<i>R_I</i>	7	0.358	0.344
4-SCH ₃	<i>R_I</i>	9	0.096	—

^a The ring to which the substituent is attached, *R* for 4-phenyl and *R_I* for 3-anilino ring, respectively. ^b frequency of occurrence of substituent at *R* or *R_I*.

the hydrophobic constant for the substituents at the meta position of ring *R_I* (π_{metaR1}) shows that more-hydrophobic substituents at this location are favorable for enhanced GSK-3 α inhibitory activity. It has to be noticed that the most active compound among the whole series involves 3,5-dichloro substituents at *R_I*. Apparently, the inherent hydrophobicity of the chlorine atom makes it a suitable meta substituent for better hydrophobic interactions with the active site of GSK-3 α . The positive contribution of the electronic parameter σ_{metaR} suggests that electron-withdrawing groups at meta position of *R* are crucial for explaining GSK-3 α inhibitory activity. Also, the positive contribution of the descriptor HA_R indicates that hydrogen bonding interactions between acceptor probes such as methoxy or nitro groups of *R* and the complementary donor groups of the amino acids at the binding site of the enzyme are important. The inclusion of the steric descriptor E^s for reflecting the influence of ortho substituents at *R* (E_{orthoR}^s) lead to a negative contribution, sug-

gesting that the bulkier the ortho substituent at the 4-phenyl ring, the higher is the GSK-3 α inhibitory activity.

The correlation between the observed and predicted GSK-3 α inhibitory activity obtained from the best QSAR model 14 is shown in Fig. (17), with a single compound as outlier of the model. The reason for this behavior is not immediately apparent, although this compound is the least active among the 3-anilino derivatives. The statistical quality of model 14a is better than that of the corresponding models containing the outlier (see Table 11).

Fig. (18) summarizes the 2D-QSAR findings suggesting where functional groups having particular physicochemical properties may best improve GSK-3 α activity. As it was discussed above, activity was enhanced by hydrophobic substituents on the anilino ring. There is a preference for N–H over an N-methyl linker for the anilino ring. Of particular importance is that electron-withdrawing groups are favored

Table 11. Statistics for 2D-Hansch Models of 3-anilino-4-phenylmaleimides

Model	<i>N</i> ^b	<i>R</i> ²	<i>q</i> ²	<i>HA_R</i>	<i>I</i> _{4CIR1}	σ_{metaR}	<i>E</i> _{orthoR} ^s	<i>I</i> _{xCH3}	π_{metaR1}	Const
14	67	0.829	0.789	0.148	0.630	0.527	-0.265	-0.903	0.310	6.421
14a ^a	66	0.850	0.813	0.122	0.602	0.573	-0.262	-0.930	0.290	6.447

^a obtained without outlier compound. b: number of compounds; *R*² correlation coefficient; *q*²: predictive *R*² from leave-one-out cross validation technique; const: regression constant.

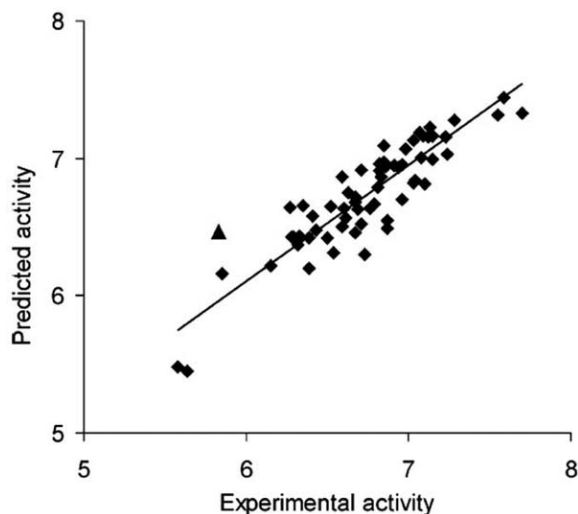


Fig. (17). Correlation between the observed and predicted GSK-3 α inhibitory activity showing the outlier (triangle).

at ortho and meta positions of the 4-phenyl ring. This is in good agreement with the most active compounds of the series which possess a 3-nitro or 2-nitro group at *R*. The nitro group is strongly electron-withdrawing due to available resonance structures, whereas the chloro group is weakly electron-withdrawing due to inductive effects. If we compare to derivatives with chloro substitution at these positions, we see that compounds with nitro groups are more potent.

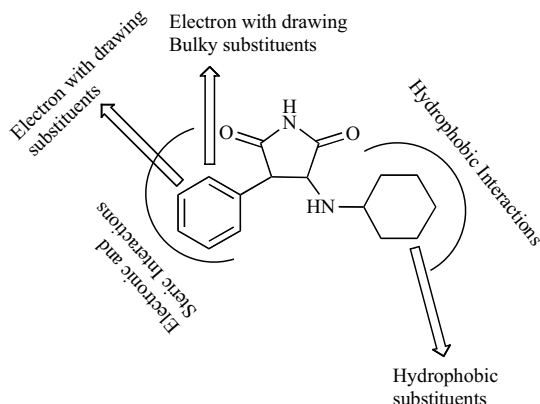


Fig. (18). Proposed model based on 2D-QSAR analyses showing the nature of interactions and substitution requirements for effective binding of 3-anilino-4-phenylmaleimides with the GSK-3 α isoform.

Present study agreed with that of Smith *et al.* [58] that on the 4-phenyl ring: 3-COOH; 4-Cl; and 4-OH in combination with 3-Cl or 3,5-di-Cl increase the activity, though in the analysis of Sivaprakasam *et al.* each of these groups helped to increase the activity. Also for the 4-phenyl ring Smith *et al.* noted that either electron-withdrawing or electron-donating groups are found on highly potent compounds, so they concluded that either electronic effects are relatively unimportant or that if they are important some of the compounds must adopt different binding modes. Present study

gave more specific analysis and information about the electronic effects on this ring, showing that ortho and meta electron-withdrawing groups were important for enhanced activity.

In contrast to the work of Lescot *et al.* [55], that also employed the same data set and where only steric factors were considered for modeling the GSK-3 β inhibition activity, present study also found electronic, hydrophobic, hydrogen bond acceptor, and some structural properties as important for activity enhancement. The 2D-findings were consistent with their steric contour plots from CoMFA, including the unfavorable effect of methyl substitution onto the aniline NH linker and the favorable effect of para bulky groups on the 3-anilino ring.

A comparison with the QSAR model of Katritzky *et al.* [57] also revealed that the same outlier compound appeared in the two reported models. Furthermore, Katritzky *et al.* interpreted the importance of the Randic topological index in their model by noting that the degree of branching and the bulk of the ligands are crucial for the activity. The importance that the 2D-QSAR analysis places on steric factors agreed with this. The average M \ddot{u} lliken bond order for any bond involving H is suggested to represent the hydrogen-bonding capability of the ligands, being in agreement with present analysis.

CONCLUDING REMARKS

Present review has discussed different QSAR studies devoted to GSK-3 inhibition, a topic that has attracted the attention of researchers from several disciplines especially in last decade. It is concluded that more experimental data on inhibitory potencies measured in similar biochemical conditions are required to achieve reliable QSAR models. Although a huge number of essays have been developed up today, these usually involve GSK-3 extracted from different tested animal species, or simply differ in the final reported expression of the inhibitory activity. In addition, a possible direction for further work in the QSAR field is the application of alternative modeling strategies for assisting the evaluation of new promising drug candidates.

ACKNOWLEDGEMENTS

We thank to the National Council of Scientific and Technological Research (CONICET) for the financial support of this work. We also would like to thank to the referee for the valuable comments that helped to enhance the quality of this work.

ABBREVIATIONS

AD	= Alzheimer Disease
ADME	= Absorption, Distribution, Metabolism and Excretion
AM1	= Austin Model-1 semiempirical method
ANN	= Artificial Neural Network
ASS	= All Space Searching technique
ATP	= Adenosine Triphosphate
BMLR	= Best Multilinear Regression procedure

CDK	= Cyclin-Dependent Kinase
CODESSA-PRO	= Comprehensive Descriptors for Structural and Statistical Analysis
CoMFA	= Comparative Molecular Field Analysis
CoMSIA	= Comparative Molecular Similarity Indices Analysis
F	= Fisher-test parameter
GSK-3	= Glycogen Synthase Kinase-3
IC ₅₀	= Concentration required for 50% inhibition of GSK-3 <i>in vitro</i>
loo	= Leave-one-out Cross Validation technique
MLR	= Multivariable Linear Regression
n	= Dimension of the model
N	= Number of compounds
PLS	= Partial Least Squares
PRESS	= Predicted sum of squares from loo
q ²	= Predictive R ² from loo
QSAR	= Quantitative Structure-Activity Relationships
R	= Correlation coefficient
RMS	= Root mean square deviation
R _{val}	= R for the validation set
S	= Standard deviation of the model
SAR	= Structure-Activity Relationships
S _{val}	= S for the validation set
TDZD	= Thiadiazolidinones

REFERENCES

- [1] Cohen, P. *Nat. Rev. Drug Discov.*, **2002**, *1*, 309.
- [2] Noble, M.E.; Endicott, J.A.; Johnson, L.N. *Science*, **2004**, *303*, 1800.
- [3] Cohen, P. *Biochem. Soc. Trans.*, **1979**, *7*, 459.
- [4] Embi, N.; Rylatt, D.B.; Cohen, P. *Eur. J. Biochem.*, **1980**, *107*, 519.
- [5] Meijer, L.; Flajolet, M.; Greengard, P. *Trends Pharm. Sci.*, **2004**, *25*, 471.
- [6] Dajani, R.; Fraser, E.; Roe, S.M.; Young, N.; Good, V.; Dale, T. C.; Pearl, L.H. *Cell*, **2001**, *105*, 721.
- [7] Ter Haar, E.; Coll, J.T.; Austen, D.A.; Hsiao, H.M.; Swenson, L.; Jain, J. *Nat. Struct. Biol.*, **2001**, *8*, 593.
- [8] Frame, S.; Cohen, P. *Biochem. J.*, **2001**, *359*, 1.
- [9] Fiol, C.J.; Mahrenholz, A.M.; Wang, Y.; Roeske, R.W.; Roach, P.J. *J. Biol. Chem.*, **1987**, *262*, 14042.
- [10] Hughes, K.; Mikolaki, E.; Plyte, S.E.; Totty, N.F.; Woodgett, J.R. *EMBO J.*, **1993**, *12*, 803.
- [11] Cohen, P. *Phil. Trans. R. Soc. Lond. B*, **1999**, *354*, 485.
- [12] Cohen, P.; Goedert, M. *Nat. Rev. Drug Discov.*, **2004**, *3*, 479.
- [13] Kim, L.; Kimmel, A.R. *Curr. Opin. Genet. Dev.*, **2000**, *10*, 508.
- [14] Seidensticker, M.J.; Behrens, J. *Biochim. Biophys. Acta*, **2000**, *1495*, 168.
- [15] Davies, S.P.; Reddy, H.; Caivano, M.; Cohen, P. *Biochem. J.*, **2000**, *351*, 95.
- [16] Martinez, A.; Castro, A.; Dorronsoro, I.; Alonso, M. *Med. Res. Rev.*, **2002**, *22*, 373.
- [17] Eldar-Finkelman, H.; Llouz, R. *Expert. Opin. Investig. Drugs*, **2003**, *12*, 1.
- [18] Van Wauwe, J.; Haefner, B. *Drugs News Perspect.*, **2003**, *16*, 557.
- [19] Wagman, A.S.; Johnson, K.W.; Bussiere, D.E. *Curr. Pharmacol. Des.*, **2004**, *10*, 1105.
- [20] Klein, P.S.; Melton, D.A. *Proc. Natl. Acad. Sci. USA*, **1996**, *93*, 8455.
- [21] Phiel, C.J.; Klein, P.S. *Annu. Rev. Pharmacol. Toxicol.*, **2001**, *41*, 789.
- [22] Gould, T.D.; Zarate, C.A.; Manji, H.K. *J. Clin. Psychiatry*, **2004**, *65*, 10.
- [23] Bertrand, J.A.; Thieffine, S.; Vulpetti, A.; Cristiani, C.; Valsasina, B.; Knapp, S.; Kalisz, H.M.; Flocco, M. *J. Mol. Biol.*, **2003**, *333*, 393.
- [24] Bhat, R.; Xue, Y.; Berg, S.; Hellberg, S.; Ormo, M.; Nilsson, Y.; Radesater, A.C.; Jerning, E.; Markgren, P.O.; Borgegard, T.; Nylof, M.; Gimenez-Cassina, A.; Hernandez, F.; Lucas, J.J.; Diaz-Nido, J.; Avila, J. *J. Biol. Chem.*, **2003**, *278*, 45937.
- [25] Chen, G.; Huang, L.D.; Jiang, Y.M.; Manji, H.K. *J. Neurochem.*, **1999**, *72*, 1327.
- [26] De Strooper, B.; Woodgett, J. *Nature*, **2003**, *423*, 392.
- [27] Ryder, J.; Su, Y.; Ni, B. *Cell. Signal.*, **2004**, *16*, 187.
- [28] Phiel, C.J.; Wilson, C.A.; Lee, V. M.-Y.; Klein, P.S. *Nature*, **2003**, *423*, 435.
- [29] Meijer, L.; Skaltsounis, A.L.; Magiatis, P.; Polychronopoulos, P.; Knockaert, M.; Leost, M.; Ryan, X.P.; Vonica, C.A.; Brivanlou, A.; Dajani, R.; Crovace, C.; Tarricone, C.; Musacchio, A.; Roe, S.M.; Pearl, L.; Greengard, P. *Chem. Biol.*, **2003**, *10*, 1255.
- [30] Jackson, G.R.; Wiedau-Pazos, M.; Sang, T.-K.; Wagle, N.; Brown, C.A.; Massachi, S.; Geschwind, D.H. *Neuron*, **2002**, *34*, 509.
- [31] King, T.D.; Bijur, G.N.; Jope, R.S. *Brain Res.*, **2001**, *919*, 106.
- [32] Carmichael, J.; Sugars, K.L.; Bao, Y.P.; Rubinstein, D.C. *J. Biol. Chem.*, **2002**, *277*, 33791.
- [33] Pérez, M.; Rojo, A.I.; Wandosell, F.; Díaz-Nido, J.; Ávila, J. *Biochem. J.*, **2003**, *372*, 129.
- [34] Coghlan, M.P.; Culbert, A.A.; Cross, D.A.; Corcoran, S.L.; Yates, J.W.; Pearce, N.J.; Rausch, O.L.; Murphy, G.J.; Carter, P.S.; Roxbee, C.L.; Mills, D.; Brown, M.J.; Haigh, D.; Ward, R.W.; Smith, D.G.; Murray, K.J.; Reith, A.D.; Holder, J.C. *Chem. Biol.*, **2000**, *7*, 793.
- [35] Ring, D.B.; Johnson, K.W.; Henriksen, E.J.; Nuss, J.M.; Go, D.; Kinnick, T.R.; Ma, S.T.; Reeder, J.W.; Samuels, I.; Slabiak, T.; Wagman, A.S.; Hammond, M.E.; Harrison, S.D. *Diabetes*, **2003**, *52*, 588.
- [36] Lustig, B.; Behrens, J. *J. Cancer Res. Clin. Oncol.*, **2003**, *129*, 199.
- [37] Liao, X.; Zhang, L.; Thrasher, J.B.; Du, J.; Li, B. *Mol. Cancer Ther.*, **2003**, *2*, 1215.
- [38] Droucheau, E.; Primot, A.; Thomas, V.; Mattei, D.; Knockaert, M.; Richardson, C.; Salicandro, P.; Alano, P.; Jafarshad, A.; Baratte, B.; Kunick, C.; Parzy, D.; Pearl, L.; Doerig, C.; Meijer, L. *Biochim. Biophys. Acta*, **2004**, *1697*, 181.
- [39] Kunick, C.; Lauenroth, K.; Wiekling, K.; Xie, X.; Schultz, C.; Gussio, R.; Zaharevitz, D.; Leost, M.; Meijer, L.; Weber, A.; Jorgensen, F.S.; Lemcke, T. *J. Med. Chem.*, **2004**, *47*, 22.
- [40] Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.*, **1994**, *37*, 4130.
- [41] Böhm, M.; Stürzebecher, J.; Klebe, G. *J. Med. Chem.*, **1999**, *42*, 458.
- [42] Leost, M.; Schultz, C.; Link, A.; Wu, Y.Z.; Biernat, J.; Mandelkow, E.M.; Bibb, J.A.; Snyder, G.L.; Greengard, P.; Zaharevitz, D.W.; Gussio, R.; Senderowicz, A.M.; Sausville, E.A.; Kunick, C.; Meijer, L. *Eur. J. Biochem.*, **2000**, *267*, 5983.
- [43] Stewart, J.J. *J. Comput. Aided Mol. Des.*, **1990**, *4*, 1.
- [44] Clark, M.; Cramer, R.D.; Van Obdenbosch, N. *J. Comput. Chem.*, **1989**, *10*, 982.
- [45] Gussio, R.; Zaharevitz, D.W.; McGrath, C.F.; Pattabiraman, N.; Kellogg, G.E.; Schultz, C.; Link, A.; Kunick, C.; Leost, M.; Meijer, L.; Sausville, E.A. *Anti-Cancer Drug Des.*, **2000**, *15*, 53.
- [46] Gerber, P.R.; Müller, K. *J. Comput.-Aided Mol. Des.*, **1995**, *9*, 251.
- [47] Wold, S.; Albano, C.W.J.; Dunn, I.; Ellund, U.; Esbensen, K.; Geladi, P.; Hellberg, S.; Johansson, E.; Lindberg, W.; Sjostrom, M. *Chemometrics: Mathematics and Statistics in Chemistry*, Reidel: Dordrecht, Netherlands, **1984**.
- [48] Bush, B.L.; Nachbar, R.B. *J. Comput.-Aided Mol. Des.*, **1993**, *7*, 587.

- [49] Polychronopoulos, P.; Magiatis, P.; Skaltsounis, A.L.; Myrianthopoulos, V.; Mikros, E.; Tarricone, A.; Musacchio, A.; Roe, S.M.; Pearl, L.; Leost, M.; Greengard, P.; Meijer, L. *J. Med. Chem.*, **2004**, *47*, 935.
- [50] Vedani, A.; Zbinden, P.; Snyder, J.; Greenidge, P. *J. Am. Chem. Soc.*, **1995**, *117*, 4987.
- [51] Hoessel, R.; Leclerc, S.; Endicott, J.; Noble, M.; Lawrie, A.; Tunnah, P.; Leost, M.; Damiens, E.; Marie, D.; Marko, D.; Niederberger, E.; Tang, W.; Eisenbrand, G.; Meijer, L. *Nat. Cell. Biol.*, **1999**, *1*, 60.
- [52] Davies, T.G.; Tunnah, P.; Meijer, L.; Marko, D.; Eisenbrand, G.; Endicott, J.A.; Noble, M.E.M. *Structure*, **2001**, *9*, 389.
- [53] Zeng, M.; Jiang, Y.; Zhang, B.; Kewen, Z.; Zhang, N.; Yu, Q. *Bioorg. Med. Chem. Lett.*, **2005**, *15*, 395.
- [54] Hou, T.J.; Xu, X.J. *Chemometr. Intell. Lab. Syst.*, **2001**, *56*, 123.
- [55] Lescot, E.; Bureau, R.; Santos, J.S.O.; Rochais, C.; Lisowski, V.; Lancelot, J.-C.; Rault, S. *J. Chem. Inf. Model.*, **2005**, *45*, 708.
- [56] Martinez, A.; Alonso, M.; Castro, A.; Dorronsoro, I.; Gelpi, J.L.; Luque, F.J.; Pérez, C.; Moreno, F.J. *J. Med. Chem.*, **2005**, *48*, 7103.
- [57] Katritzky, A.R.; Pacureanu, L.M.; Dobchev, D.A.; Fara, D.C.; Duchowicz, P.R.; Karelson, M. *Bioorg. Med. Chem.*, **2006**, *14*, 4987.
- [58] Smith, D.G.; Buffet, M.; Fenwick, A.E.; Haigh, D.; Ife, R.; Saunders, M.; Slingsby, B.P.; Stacey, R.; Ward, R.W. *Bioorg. Med. Chem. Lett.*, **2001**, *11*, 635.
- [59] Tavares, F.X.; Boucheron, J.A.; Dickerson, S.H.; Grin, R.J.; Preugschat, F.; Thomson, S.A.; Wang, T.Y.; Zhou, H.Q. *J. Med. Chem.*, **2004**, *47*, 4716.
- [60] Witherington, J.; Bordas, V.; Garland, S.L.; Hickey, D.M.B.; Ife, R.J.; Liddle, J.; Saunders, M.; Smith, D.G.; Ward, R.W. *Bioorg. Med. Chem. Lett.*, **2003**, *13*, 1577.
- [61] Witherington, J.; Bordas, V.; Haigh, D.; Hickey, D.M.B.; Ife, R.J.; Rawlings, A.D.; Slingsby, B.P.; Smith, D.G.; Ward, R.W. *Bioorg. Med. Chem. Lett.*, **2003**, *13*, 1581.
- [62] Witherington, J.; Bordas, V.; Gaiba, A.; Garton, N.S.; Naylor, A.; Rawlings, A.D.; Slingsby, B.P.; Smith, D.G.; Tackle, A.K.; Ward, R.W. *Bioorg. Med. Chem. Lett.*, **2003**, *13*, 3055.
- [63] Witherington, J.; Bordas, V.; Gaiba, A.; Naylor, A.; Rawlings, A.D.; Slingsby, B.P.; Smith, D.G.; Tackle, A.K.; Ward, R.W. *Bioorg. Med. Chem. Lett.*, **2003**, *13*, 3059.
- [64] Peat, A.; Garrido, D.; Boucheron, J.A.; Schweiker, S.L.; Dickerson, S.H.; Wilson, J.R.; Wang, T.Y.; Thomson, S.A. *Bioorg. Med. Chem. Lett.*, **2004**, *14*, 2127.
- [65] Peat, A.J.; Boucheron, J.A.; Dickerson, S.H.; Garrido, D.; Mills, W.; Peckham, J.; Preugschat, F.; Smalley, T.; Schweiker, S.L.; Wilson, J.R.; Wang, T.Y.; Zhou, H.Q.; Thomson, S.A. *Bioorg. Med. Chem. Lett.*, **2004**, *14*, 2121.
- [66] Katritzky, A.R.; Lobanov, V.S.; Karelson, M. *Chem. Soc. Rev.*, **1995**, *24*, 279.
- [67] Katritzky, A.R.; Karelson, M.; Lobanov, V. *Pure Appl. Chem.*, **1997**, *69*, 245.
- [68] Karelson, M.; Maran, U.; Wang, Y.; Katritzky, A.R. *Collect. Czech. Chem. Commun.*, **1999**, *64*, 1551.
- [69] Katritzky, A.R.; Fara, D.C.; Yang, H.; Karelson, M.; Suzuki, T.; Solovev, V.P.; Varnek, A. *J. Chem. Inf. Model.*, **2004**, *44*, 529.
- [70] Katritzky, A.R.; Dobchev, D.A.; Fara, D.C.; Karelson, M. *Bioorg. Med. Chem.*, **2005**, *13*, 1623.
- [71] Katritzky, A.; Dobchev, D.; Fara, D.; Karelson, M. *Bioorg. Med. Chem.*, **2005**, *13*, 6598.
- [72] Dewar, M.J.S.; Zebisch, E.G.; Healy, E.F.; Stewart, J.J.P. *J. Am. Chem. Soc.*, **1985**, *107*, 3902.
- [73] Katritzky, A.R.; Ignachenko, E.; Barcock, R.; Lobanov, V.; Karelson, M. *Anal. Chem.*, **1994**, *66*, 1799.
- [74] Randic, M. *J. Am. Chem. Soc.*, **1975**, *97*, 6609.
- [75] Julg, A.; Julg, P. *Int. J. Quantum Chem.*, **1978**, *13*, 483.
- [76] Ángyan, J.G.; Rosta, E.; Surján, P.R. *Chem. Phys. Lett.*, **1999**, *299*, 1.
- [77] Masters, T. *Practical Neural Network Recipes in C++*, Academic Press: Boston, **1993**.
- [78] Zupan, J.; Gasteiger, J. *Neural Networks in Chemistry and Drug Design*, Wiley-VCH: Weinheim, **1999**.
- [79] Sivaprakasam, P.; Xie, A.; Doerksema, R.J. *Bioorg. Med. Chem.*, **2006**, *14*, 8210.
- [80] Fujita, T.; Ban, T. *J. Med. Chem.*, **1971**, *14*, 148.
- [81] Hansch, C.; Fujita, T. *J. Am. Chem. Soc.*, **1964**, *86*, 1616.
- [82] Bhat, R.V.; Budd Haeberlein, S.L.; Avila, J. *J. Neurochem.*, **2004**, *89*, 1313.
- [83] Meijer, L.; Thunnissen, A.M.; White, A.W.; Garnier, M.; Nikolic, M.; Tsai, L.H.; Walter, J.; Cleverley, K.E.; Salinas, P.C.; Wu, Y.Z.; Biernat, J.; Mandelkow, E.M.; Kim, S.H.; Pettit, G.R. *Chem. Biol.*, **2000**, *7*, 51.
- [84] Leclerc, S.; Garnier, M.; Hoessel, R.; Marko, D.; Bibb, J.A.; Snyder, G.L.; Greengard, P.; Biernat, J.; Wu, Y.Z.; Mandelkow, E.M.; Eisenbrand, G.; Meijer, L. *J. Biol. Chem.*, **2001**, *276*, 251.
- [85] Bain, J. *Biochem. J.*, **2003**, *371*, 199.
- [86] Kunick, C.; Lauenroth, K.; Leost, M.; Meijer, L.; Lemcke, T. *Bioorg. Med. Chem. Lett.*, **2004**, *14*, 413.
- [87] Mettrey, Y.; Gompel, M.; Thomas, V.; Garnier, M.; Leost, M.; Ceballos-Picot, I.; Noble, M.; Endicott, J.; Vierfond, J.-M.; Meijer, L. *J. Med. Chem.*, **2003**, *46*, 222.
- [88] Naerum, L.; Nørskov-Lauritsen, L.; Olesen, P.H. *Bioorg. Med. Chem. Lett.*, **2002**, *12*, 1525.
- [89] Ortega, M.A.; Montoya, M.E.; Zarranz, B.; Jaso, A.; Aldana, I.; Leclerc, S.; Meijer, L.; Monge, A. *Bioorg. Med. Chem.*, **2002**, *10*, 2177.
- [90] Lane, M.E.; Yu, B.; Rice, A.; Lipson, K.E.; Liang, C.; Sun, L.; Tang, C.; McMahon, G.; Pestell, R.G.; Wadler, S. *Cancer Res.*, **2001**, *61*, 6170.
- [91] Zhang, H.-C.; White, K.B.; Ye, H.; McComsey, D.F.; Derian, C.K.; Addao, M.F.; Andrade-Gordon, P.; Eckardt, A.J.; Conway, B.R.; Westover, L.; Xu, J.Z.; Look, R.; Demarest, K.T.; Emanuel, S.; Maryanoff, B.E. *Bioorg. Med. Chem. Lett.*, **2003**, *13*, 3049.
- [92] Kuo, G.H.; Prouty, C.; DeAngelis, A.; Shen, L.; O'Neill, D.J.; Shah, C.; Connolly, P.J.; Murray, W.V.; Conway, B.R.; Cheung, P.; Westover, L.; Xu, J.Z.; Look, R.A.; Demarest, K.T.; Emanuel, S.; Middleton, S.A.; Jolliffe, L.; Beavers, M.P.; Chen, X. *J. Med. Chem.*, **2003**, *46*, 4021.
- [93] Shen, L.; Prouty, C.; Conway, B.R.; Westover, L.; Xu, J.Z.; Look, R.A.; Chen, X.; Beavers, M.P.; Roberts, J.; Murray, W.V.; Demarest, K.T.; Kuo, G.-H. *Bioorg. Med. Chem.*, **2004**, *12*, 1239.
- [94] Hers, I.; Tavaré, J.M.; Denton, R.M. *FEBS Lett.*, **1999**, *460*, 433.
- [95] Ruetz, S.; Fabbro, D.; Zimmermann, J.; Meyer, T.; Gray, N. *Curr. Med. Chem. Anti-Canc. Agents*, **2003**, *3*, 1.
- [96] Olesen, P.H.; Sørensen, A.R.; Urso, B.; Kurtzhals, P.; Bowler, A.N.; Ehrbar, U.; Hansen, B.F. *J. Med. Chem.*, **2003**, *46*, 3333.
- [97] Ding, S.; Wu, T.Y.; Brinker, A.; Peters, E.C.; Hur, W.; Gray, N.S.; Schultz, P.G. *Proc. Natl. Acad. Sci. USA*, **2003**, *100*, 7632.
- [98] Conde, S.; Perez, D.I.; Martinez, A.; Perez, C.; Moreno, F. *J. Med. Chem.*, **2003**, *46*, 4631.
- [99] Ilouz, R.; Kaidanovich, O.; Gurwitz, D.; Eldar-Finkelman, H. *Biochim. Biophys. Res. Commun.*, **2002**, *295*, 102.