European Journal of Medicinal Chemistry 44 (2009) 2361–2371

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

Fragment and knowledge-based design of selective GSK-3β inhibitors using virtual screening models

S. Vadivelan [a,b,*], Barij Nayan Sinha [b,**], Sunita Tajne [a], Sarma A.R.P. Jagarlapudi [a]

[a] GVK Biosciences Private Limited, No. 37, Sterling Road, Nungambakkam, Chennai 600 034, Tamil Nadu, India
[b] Department of Pharmaceutical Sciences, Birla Institute of Technology, Mesra 835215, Ranchi, India

ARTICLE INFO

Article history:
Received 8 May 2008
Received in revised form 26 August 2008
Accepted 29 August 2008
Available online 16 September 2008

Keywords:
Glycogen Synthase Kinase 3β
Pharmacophore
HypoGen
Docking
Virtual library screening
Comparative model

ABSTRACT

Glycogen Synthase Kinase 3β is one of the important targets in the treatment of type II diabetes and Alzheimer's disease. Currently this target is in pursuit for type II diabetes and a few GSK-3β inhibitors have been now advanced to Phases I and II of clinical trials. The best validated HypoGen model consists of four pharmacophore features; 1) two hydrogen bond acceptors, 2) one hydrogen bond donor and 3) one hydrophobic. This pharmacophore model correlates well with the docking model, one hydrogen bond acceptor is necessary for the H-bond interaction with VAL135, and second hydrogen bond acceptor is important for the H-bond interactions with ARG141 and the hydrophobic feature may be required for the weak H-bond interactions with ASP133. The comparative model was developed from analogue and structure-based models like Catalyst, Glide SP & XP, Gold Fitness & ChemScore and Ligand Fit using multiple linear regression analysis. A virtual library of 10,000 molecules was generated employing fragment and knowledge-based approach and the comparative model was used to predict the activities of these molecules. The H-bond with ARG141 appears to be unique to GSK-3β and explains the high GSK-3β selectivity observed for 1*H*-Quinazolin-4-ones and Benzo[*e*][1,3]oxazin-4-ones. This understanding of protein–ligand interactions and molecular recognition increases the rapid development of potent and selective inhibitors, and also helps to eliminate the increase in number of false positives and negatives.

1. Introduction

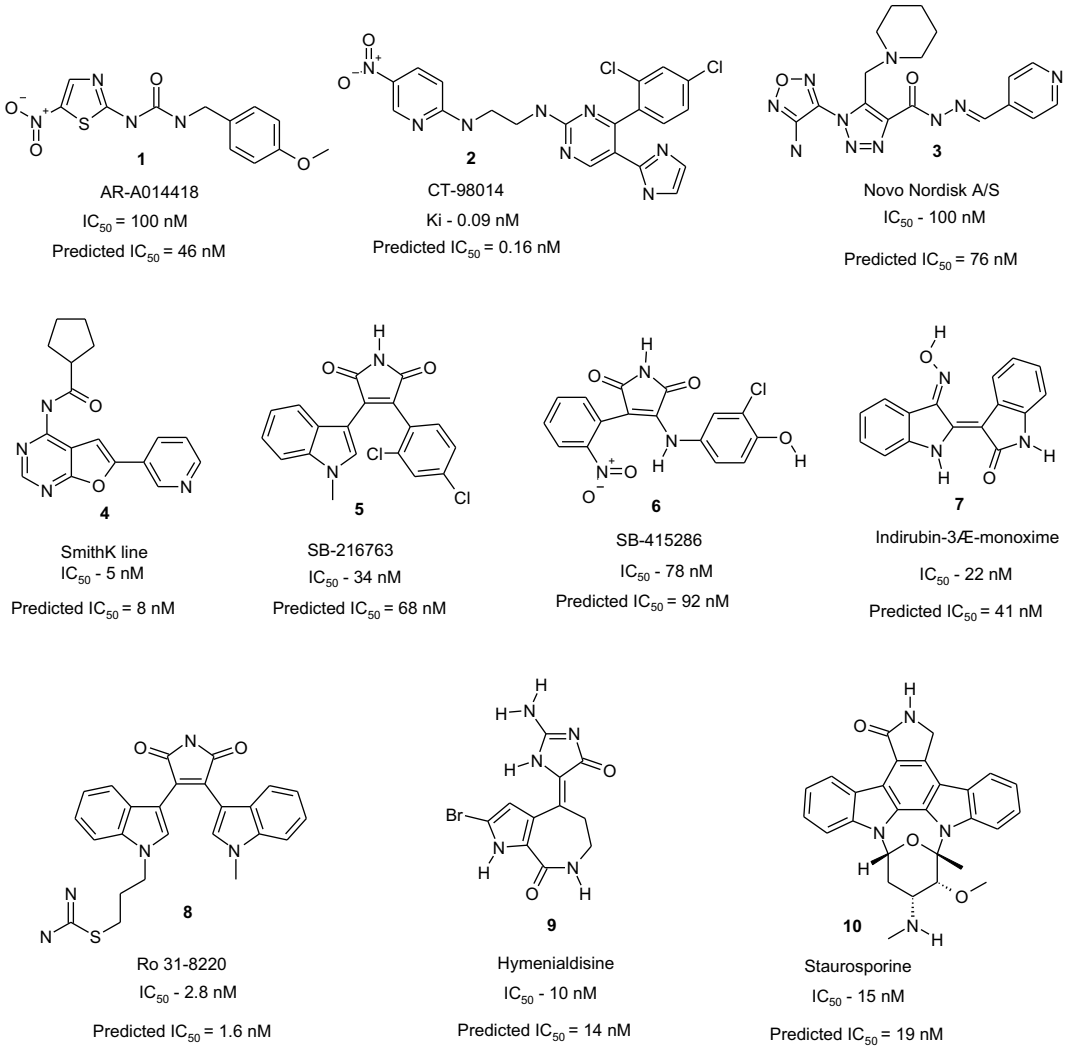
Glycogen synthase kinase 3 (GSK-3) is one of the first serine/threonine protein kinases to be identified and studied, for its function in the regulation of glycogen synthase. In humans, GSK-3 exists in two forms, GSK-3α (51 kDa) and GSK-3β (47 kDa). Both together display an 84% overall identity (98% within their catalytic domains) with the main difference being an extra Gly-rich stretch in the N-terminal domain of GSK-3α. Recently, GSK-3β2, an alternative splicing variant of GSK-3β that contains a 13-amino acid insertion in the catalytic domain, has been identified [1]. GSK-3 inhibitors have arisen as promising drugs for the pharmacotheraphy of several acute pathologies, such as diabetes, stroke, cancer, mood disorders, inflammation and Alzheimer's disease among others [2]. There are several small molecules as GSK-3β inhibitors (Scheme 1 and Table 1) presently, in the clinical trials for the treatment of type II diabetes [3]. The crystal structure of GSK-3 has become a powerful tool in improving the creativity of medicinal chemist for the rational design of specific inhibitors that target selective amino acids within the enzyme.

Purines were found to exhibit GSK-3 inhibitory activities and various heterocyclic skeletons such as pyrazines, pyrimidines, heterocyclic pyrimidones, aminopyrazoles, bisindolemaleimides, hymenialdisine, paullones, indirubines, and anilinomaleimides have been developed as ATP competitive GSK-3 inhibitors while thiadiazolidinones were reported to be the first ATP noncompetitive GSK-3 inhibitor [2,4–7]. However, except for the anilinomaleimides which were demonstrated as selective GSK-3 inhibitors relative to 24 protein kinases, the majority of the ATP competitive GSK-3 inhibitors showed significant activities at several other protein kinases [8,9]. To minimize the potential side effects, it is always desirable to have the clinical agent targeting the enzyme specifically [10,11].

Fragment-based drug design has become an important and powerful tool for the discovery and optimization of potent drug leads [12–14]. Thus, the design and optimization of the ultimate "lead" compound is carried out by identifying and optimizing the individual fragments, followed by synthetic linking or merging to produce the required high affinity for the target protein with appropriate lead-like properties [15,16]. The main attractions of fragment-based techniques lie in the synthesis and testing of

* Corresponding author. GVK Biosciences Pvt. Ltd., 37, Sterling Road, Nungambakkam, Chennai-600 034, Tamil Nadu, India. Tel.: +91 44 6629 3001; fax: +91 44-6629 3199.
** Corresponding author.
E-mail address: vadivelan@gvkbio.com (S. Vadivelan).

0223-5234/$ – see front matter © 2008 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2008.08.012



Scheme 1. GSK-3β inhibitors from clinical trials.

significantly fewer compounds, compared to HTS methods, structurally characterized binding modes and the potential high success rate of generating new chemical series with attractive lead-like properties [17]. Another key advantage of using molecular fragments for lead discovery is the significant amount of chemical space is sampled using a relatively small library of compounds [17]. Hence, different sets of molecular fragments are used to target a particular protein, based upon diversity or focused pharmacophore models. For example, family-specific fragment libraries can be assembled, such as focused kinase-fragment library. This diverse collection of synthetically tractable chemotypes is based on known ATP-site binders with acceptable lead-like properties. It is generally accepted that fragment leads need to be smaller (MW $< \sim 250$ Da) and less lipophillic (*c*log $P < \sim 3$) than conventional HTS leads to be successfully progressed into clinical candidates that possess good physicochemical properties [15].

Table 1
GSK-3β inhibitors in clinical trials

Clinical candidate	Company	Highest develop status	Indications
AR-A014418	AstraZeneca	Discovery	Alzheimer's disease
BI-5521	Boehringer Ingelheim Corp	Discovery	Type II diabetes
VX-608	Vertex Pharmaceuticals	No Development Reported	Type II diabetes
CT-98014	Chiron Corp	Discovery	Type II diabetes
CG-100179	Crystal Genomics Inc	Discovery	Diabetes mellitus
Cyclacel	Cyclacel Pharmaceuticals	Discovery	Type II diabetes
CP-70949	Pfizer Inc	Discovery	Diabetes mellitus
NP-12	Neuropharma SA	Phase 1	Alzheimer's disease
NNC-57-0511	Novo Nordisk A/S	No Development Reported	Type II diabetes
SB-216763	GlaxoSmithKline	Discovery	Type II diabetes
Ro- 31-8220	Roche Holding AG	Discovery	Osteoporosis

Knowledge-based approach is based on another popular technique applied in the drug discovery, known as 'scaffold-hopping' [18,19] where the goal is to 'jump' in chemistry space, i.e., to discover a new structure starting from a known active compound via the modification of the central core of this molecule [20].

In the present study, initial structure and analogue based design studies are done using Glide, Gold, Ligand Fit, Catalyst to develop a comparative model, then fragment-based and knowledge-based approaches are employed to design molecules of selective GSK-3β inhibitors and the activities of those molecules are predicted using the comparative model [21–24].

This article describes the fragment-based, knowledge-based design, identification and optimization of a new lead, 2,7-disubstituted 1*H*-Quinazolin-4-ones and 2,7-disubstituted

Benzo[*e*][1,3] oxazin-4-ones, as a novel series of potent and selective GSK-3β inhibitors using various virtual screening models.

2. Experimental techniques

Molecular modeling studies were carried out on a Silicon Graphics Octane R12000 computer running Irix 6.5.12 (SGI, 1600 Amphitheatre Parkway, Mountain View, CA 94043). Catalyst 4.11 software was used to generate pharmacophore models. Glide (Schrodinger, L.L.C., New York), Gold and Ligand Fit (Cerius^2) docking programs were used for the structure-based studies.

2.1. Database mining and training set

Kinase inhibitor database produced 26,913 GSK-3β inhibitors from 310 references (inclusive of Journals and Patents) while the modeling study was based on the diversity of both chemical structure and biological activity against human GSK-3β inhibitors [25]. The most critical aspect of pharmacophore hypothesis generation is the selection of the training set. There should not be any redundancy in information content in terms of both structural features and activity ranges. The 441 molecules are arranged in the decreasing order of their activities and were clustered into 100 bins. Based on atom–atom pair distance descriptors, similarity between most active molecules to all other molecules in 100 bins has been performed based on ISIS 960 keys and Tanimoto analysis. All these 441 molecules contain only IC_{50} values either in nanomolar or micrometers. All the training set compounds from the literature have similar GSK-3 β inhibitory assay. The molecules with K_i, ED_{50}, EC_{50} and other activity type values were ignored for modeling studies. The most diverse molecules from 25 bins have been selected as the training set and are given in Fig. 1 [6,26–40]. GSK-3β activities for the training set of 25 molecules covers five orders of magnitude ($0.00014\ \mu M \leq IC_{50} \leq 420\ \mu M$). This training set is used in HypoGen to generate pharmacophore models [24]. The mol files of all molecules from the database were exported and minimized using modified CHARMm force field in Catalyst package, installed on a SGI Octane 275 MHz MIPS R12000 processor, and conformational analysis was carried out using best method in Confirm module. The polling algorithm of Confirm module reduces

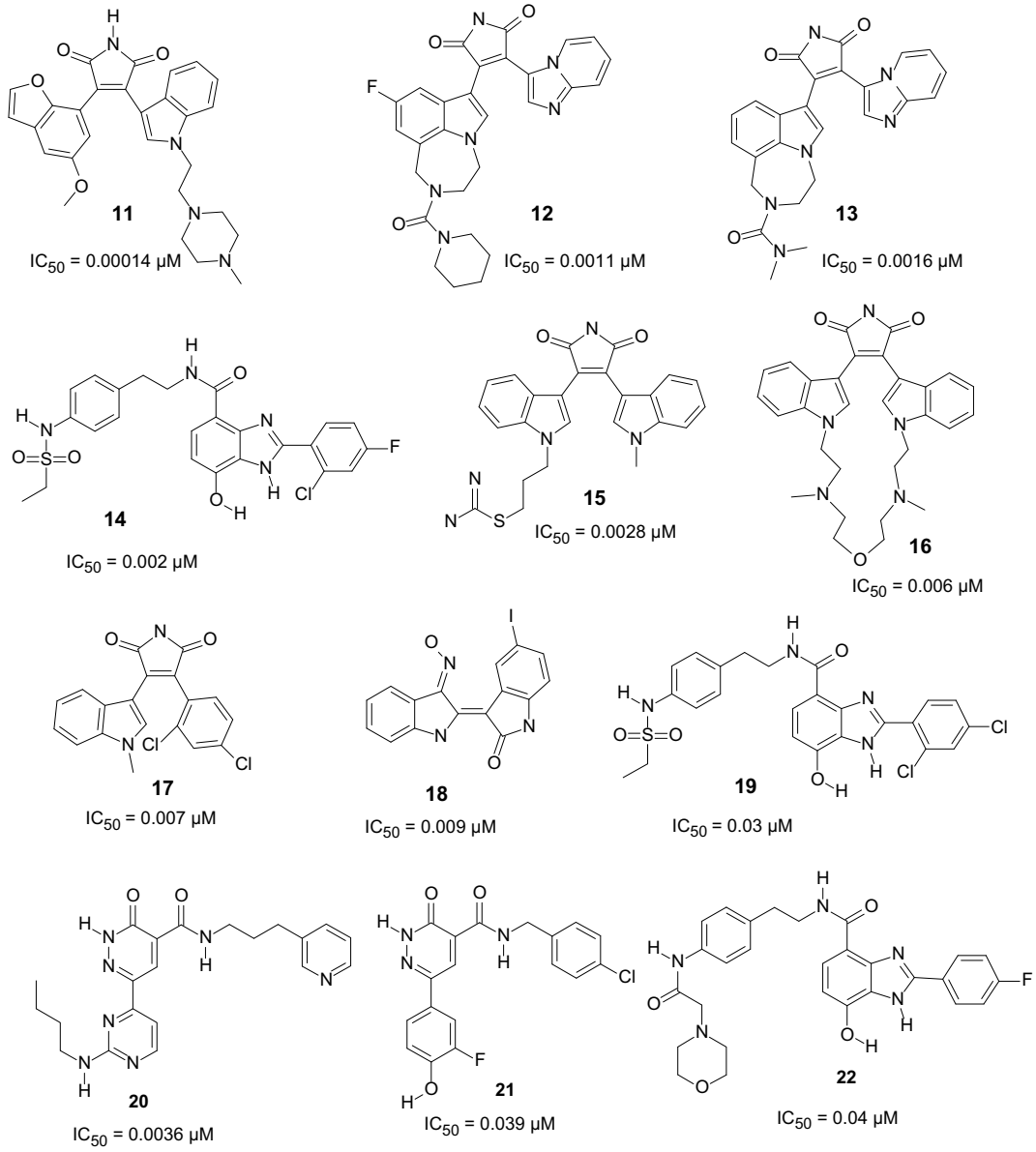


Fig. 1. Structures of 25 training set molecules with its experimental IC_{50} values.

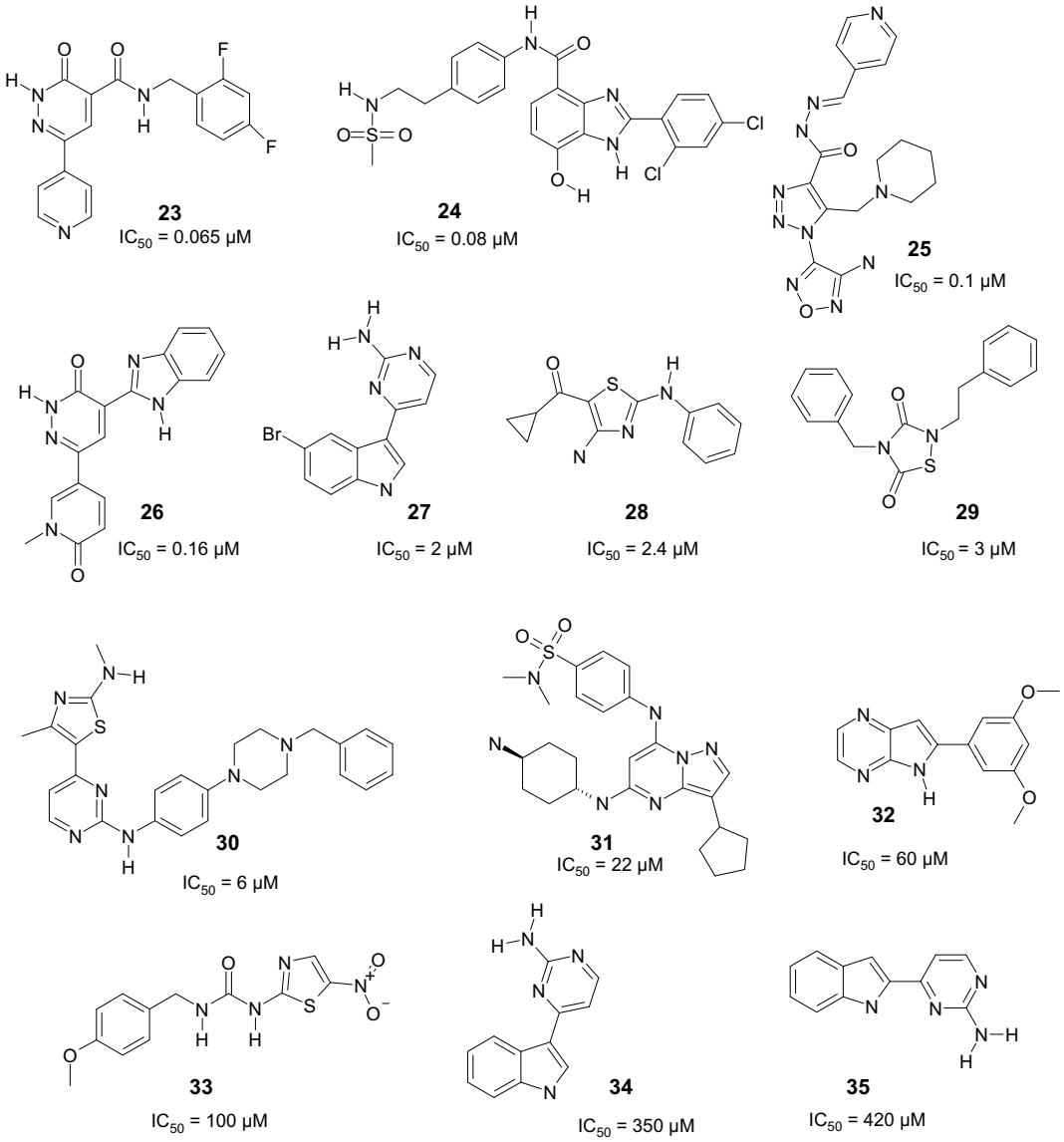


Fig. 1. (*continued*).

considerably the probability of reappearance of nearly similar conformers by usage of penalty function. For each molecule, a maximum of 250 conformers that lie within 10 kcal/mol from the observed global minimum was considered for the model generation. This method ensures an exhaustive conformational mapping even for most complex molecules.

2.2. HypoGen

The training set of 25 molecules defined earlier has been used in HypoGen to generate pharmacophore models, which could be used for quantitative estimation of activities while screening large virtual compound libraries. While generating HypoGen model, a minimum of 0 to a maximum of five features involving hydrogen bond acceptor, hydrogen bond donor and hydrophobic systems have been specified based on the common features and GSK-3β inhibitory activities. The molecules in the training set are broadly classified into three categories namely highly active (<0.02 μM), moderately active (>0.02–0.2 μM) and low active (>0.2 μM), while an uncertainty value of 3 was used. The quality of HypoGen models are best described in terms of Fixed Cost, Null Cost and total Cost and these terms are well defined by Debnath [41]. The cost for each hypothesis is the summation of the three cost components (error (*E*), weight (*W*), and configuration (*C*)) multiplied by a coefficient (default coefficient is 1.0 for each). The fixed cost represents the simplest model that fits the data perfectly. The null cost represents the cost of a hypothesis with no features that estimates every activity to be the average activity [24]. In simple terms there should be a large difference between fixed cost and null cost with a value of 40–60 bits for the unit of cost, which would imply a 75–90% probability for correlating the experimental and predicted activity data. For a good model, the total cost of any hypothesis should be close to the fixed cost. The best model Hypo-1 has been given in Fig. 2 while the parameters that describe the Hypo-1 are given in Table 2.

The best pharmacophore hypothesis was used initially to screen 416 GSK-3β inhibitors from the Kinase inhibitor Database. All queries were performed using the Best Flexible search databases/Spreadsheet method. To validate the model enrichment factors and goodness of hit has been calculated [42].

2.3. Docking studies

Crystal structure of GSK-3β (PDB code: 1q5k) was employed for the docking studies [43]. The 3D structure of protein was

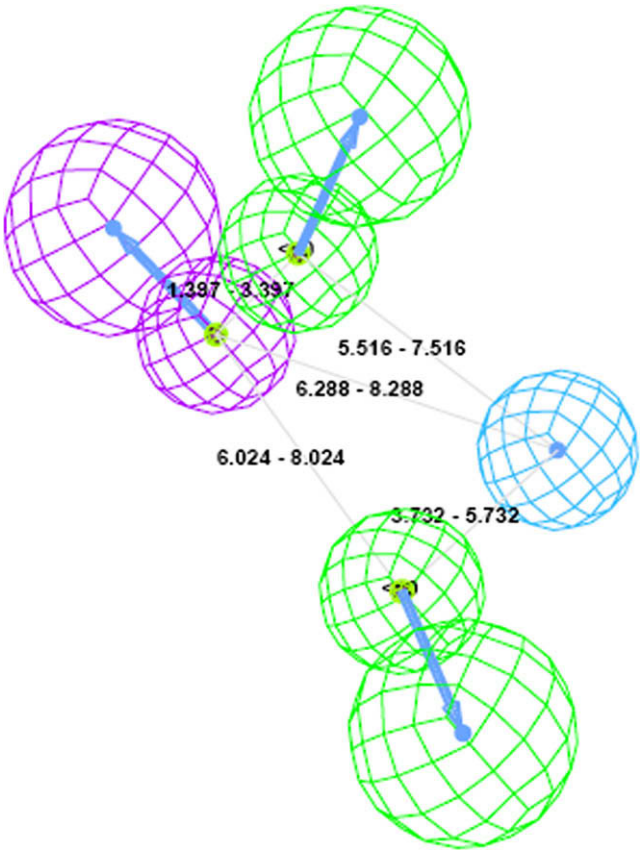


Fig. 2. Glycogen Synthase Kinase 3 β inhibitor pharmacophore model with its distance constraints, features are color coded with green: hydrogen bond acceptor, magenta: hydrogen bond donor, light blue: one hydrophobic aliphatic. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

downloaded from the protein databank (PDB) and loaded to the Maestro workbench [44]. Hydrogen adjust was made to calculate the bond orders for the protein and ligand. Protein preparation and refinement were done in glide and then used for all the three docking studies. Structure-based docking studies were carried out using Glide, GOLD and Ligand Fit on GSK-3β inhibitors to the 3D

Table 2
10 Pharmacophore models generated by the HypoGen algorithm

Hypo No.	Total cost	Cost difference[a]	Error Cost	RMS deviation	Training set (*r*)	Test set (*r*)	Features[b]
1	115.01	57.87	94.80	0.92	0.97	0.796	AADH
2	115.42	57.46	96.66	0.94	0.96	0.762	AADH
3	116.77	56.11	97.46	1.03	0.96	0.724	AADH
4	117.61	55.26	96.97	1.01	0.96	0.716	AADH
5	124.45	48.43	103.32	1.24	0.94	0.698	AADH
6	125.71	47.17	107.55	1.37	0.93	0.686	AADH
7	126.13	46.74	107.69	1.37	0.93	0.658	AADH
8	126.75	46.12	107.71	1.37	0.93	0.626	AADH
9	127.14	45.73	107.39	1.36	0.93	0.674	AADH
10	127.26	45.62	106.93	1.35	0.93	0.642	AADH

[a] (Null cost – Total cost); Null cost = 172.88, Fixed cost = 101.775. For the Hypo-1 Weight = 2.09, Configuration = 12.56, All cost units are in bits.
[b] A – Hydrogen Bond Acceptor, D – Hydrogen Bond Donor, H – Hydrophobic.

structure of GSK-3β and generated 50 best docking poses [45–47]. The best poses were selected based on the scoring functions and quality of pose orientation within the active site amino acids. In the GSK-3β protein, the binding site coordinates ($X = 24.0$, $Y = 22.8$ and $Z = 8.9$) were taken from the centroid of the ligand to define the active site region. Active site radius was taken as 12.0 Å so that the active site residues namely ASP133, VAL135, ARG141, ASP200, CYS199, LEU132, GLU185, GLU97, LYS85 and PHE67 could be included. The 441 GSK-3β inhibitors taken for the Pharmacophore studies were considered for docking studies to develop the comparative model. All ligand structures were built in Cerius^2 and minimized by OFF methods using the steepest descent algorithm with a gradient convergence value of 0.001 kcal/mol [23]. Open force field is an environment designed to maximize the effectiveness, flexibility, and ease-of use of forcefield-based atomistic simulation methods.

2.4. Comparative model

The comparative model was generated with various structure-based and pharmacophore studies like Glide, Gold, Ligand Fit and Catalyst using multiple linear regression analysis in Cerius^2. This type of comparative model is used to reduce the number of false positives and false negatives and for good prediction towards the enzyme.

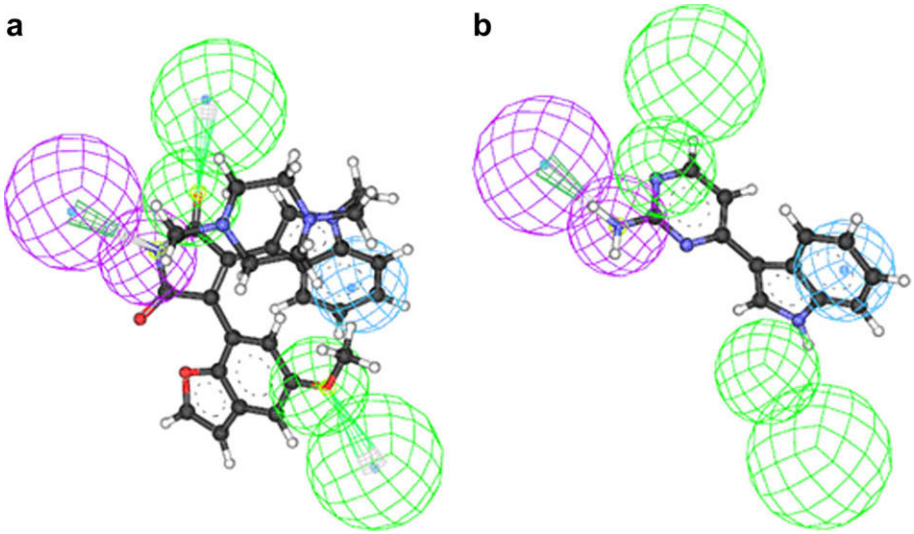


Fig. 3. Hypo-1 mapped to the most active molecule **11** ($IC_{50} = 0.00014$ μM, Fig. 3a) and also mapped to low active molecule **34** ($IC_{50} = 350$ μM, Fig. 3b) in the training set.

Table 3
Experimental and predicted IC_{50} data of 25 training set molecules

Molecule	Fit value[b]	Experimental IC_{50}, uM	Predicted IC_{50}, uM	Error[a]	Experimental scale[c]	Predicted scale[c]
11	11.68	0.00014	0.00018	1.29	+++	+++
12	9.3	0.0011	0.0043	3.91	+++	+++
13	9.48	0.0016	0.0028	1.75	+++	+++
14	9.25	0.002	0.0049	2.45	+++	+++
15	9.37	0.0028	0.0037	1.32	+++	+++
16	9.54	0.006	0.0025	−2.40	+++	+++
17	9.3	0.007	0.0043	−1.63	+++	+++
18	8.41	0.009	0.033	3.67	+++	++
19	8.38	0.03	0.036	1.20	++	++
20	8.23	0.036	0.051	1.42	++	++
21	8.71	0.039	0.02	−1.94	++	++
22	8.66	0.04	0.019	−2.11	++	++
23	7.71	0.065	0.17	2.62	++	++
24	8.22	0.08	0.052	−1.54	++	++
25	7.04	0.1	0.78	7.80	++	+
26	7.18	0.16	0.58	3.63	++	+
27	6.27	2	4.7	2.35	+	+
28	6.17	2.4	5.8	2.42	+	+
29	6.77	3	1.5	−2.00	+	+
30	6.41	6	3.4	−1.76	+	+
31	5.72	22	16	−1.38	+	+
32	5.02	60	83	1.38	+	+
33	5.96	100	9.4	−10.64	+	+
34	5.12	350	65	−5.38	+	+
35	4.81	420	130	−3.23	+	+

[a] + Indicates that the predicted IC_{50} is higher than the experimental IC_{50}; − indicates that the predicted IC_{50} is lower than the experimental IC_{50}; a value of 1 indicates that the predicted IC_{50} is equal to the experimental IC_{50}.
[b] Fit value[12] indicates how well the features in the pharmacophore overlap with the chemical features in the molecule. Fit = weight × [max(0,1 − SSE)] where $SSE = (D/T)^2$, D = displacement of the feature from the center of the location constraint and T = the radius of the location constraint sphere for the feature (tolerance).
[c] Activity scale – $IC_{50} < 0.02$ uM = +++ (highly active); $IC_{50} = 0.02$–0.2 uM = ++ (moderately active); $IC_{50} > 0.2$ uM = + (low active).

2.5. Construction of a large virtual scaffold library

The fragments were generated from kinase inhibitor databases using simple iterative disconnection algorithm (Fragment descriptors) developed In-house [48]. The fragmentation rules were coded using Java and applied on SDF file formats. Chemical fragments were classified as 1) Rings, 2) Linkers defined as acyclic graphs having more than one connecting atom to rings and 3) Substituents are defined as group connected to single ring systems. In general, the fragments in these sets have MW between 100 and 250, clog P of less than ~3 and are relatively simple with few functional groups, making them chemically tractable and suitable for rapid optimization.

3. Results and discussion

3.1. HypoGen studies

In the HypoGen studies, a couple of sets of 10 hypotheses were generated using the most diverse 25 molecules in the training set molecules. The best 10 hypotheses consist of 1) two hydrogen bond acceptors, 2) one hydrogen bond donor and 3) one hydrophobic. It is evident from the data summarized in Table 2, good cost values, correlation and low RMSD were observed for the generated hypotheses. The best hypothesis, Hypo-1, is characterized by the highest cost difference (58 bits), lowest RMS error (0.92) and with correlation 0.97. The fixed cost, pharmacophore (total) cost and null cost for hypo-1 are 102, 115 and 172 bits, respectively. It is evident that the error, weight and configuration component are very low and not deterministic to the model; the total pharmacophore cost is also low and close to the fixed cost. Also, as total cost is less than the null cost, this model appears to possess all the pharmacophore features and has a good predictability power. Fig. 2 shows the Hypo-1 pharmacophore model with their geometric parameters between the features. Fig. 3a and b represents the Hypo-1 aligned with the most active and low active molecules (**11** and **34**; IC_{50} values are 0.00014 and 350 μM, respectively) and shows a nice fit with all the features of Hypo-1. In case of 11, the two hydrogen bond acceptors seems to be mapped to the methoxyl group in benzofuran and carbonyl group in pyrrole, the hydrogen bond donor mapped to secondary amino group of pyrrole ring and the hydrophobic group fits well with the aromatic group in indole. On the other hand, for molecule **34**, the two features of hydrogen bond acceptor did not map at all which is the desired feature important for binding to Hinge region amino acids, whereas hydrogen bond donor and hydrophobic group fit well (Fig. 3b). Table 3 shows experimental and predicted IC_{50} values for 25 training set molecules, based on the Hypo-1 model, along with other details such as, error values and Fitness scores. In the training set, out of eight highly active molecules, seven were correctly predicted as highly active and one as moderately active, and in the eight medium active molecules, six were predicted well while two molecules were indicated as less active and all nine low active molecules were also predicted as low active. As presented in Table 2, it is noted that

Table 4
Statistical parameters from screening test set molecules

S. No.	Parameter	Molecules
1	Total molecules in database (D)	416
2	Total number of actives in database (A)	176
3	Total hits (H_t)	185
4	Active hits (H_a)	167
5	% Yield of actives $[(H_a/H_t) \times 100]$	90.27
6	% Ratio of actives $[(H_a/A) \times 100]$	94.89
7	Enrichment factor (E) $[(H_a \times D)/(H_t \times A)]$	2.13
8	False negatives $[A - H_a]$	9
9	False positives $[H_t - H_a]$	18
10	Goodness of Hit Score[a]	0.85

[a] $[(H_a/4H_tA)(3A + H_t)) \times (1 - ((H_t - H_a)/(D - A))]$; GH Score of 0.85 indicates a very good model.

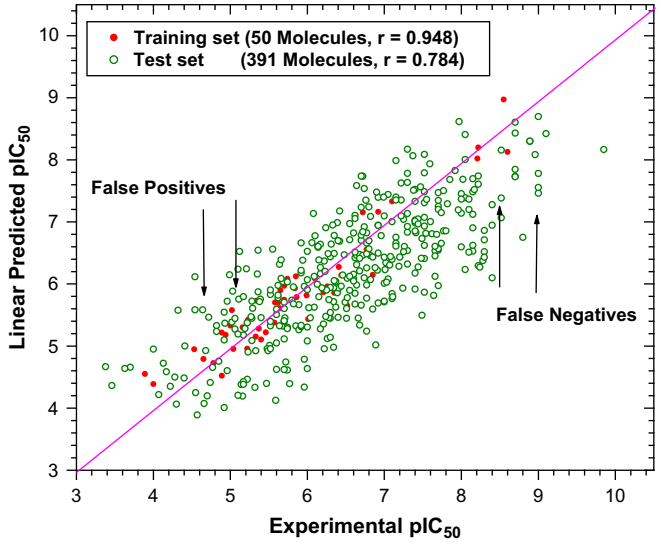


Fig. 4. Graph showing the correlation (r) between experimental and predicted activities for the 441 GSK-3β inhibitors (50 molecules in training set and 391 Molecules in test set) against comparative model (Multiple linear regression analysis) MLR equation: "0.754183 + 0.3594" × "Pharmacophore_Predicted_pIC50" + 0.017709 × "Gold Fitness" + 0.0068673 × "ChemScore" − 0.231128 × "Glide_SP" − 0.059958 × "Glide_XP" + 0.0017235 × "Dock Score".

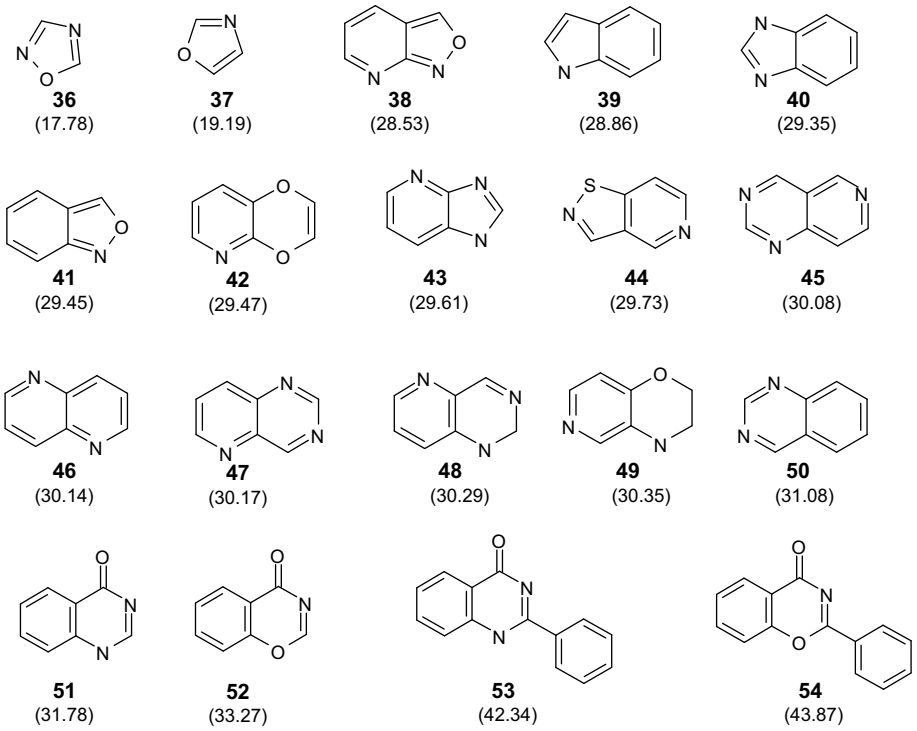


Fig. 5. Some of the fragments used for the knowledge-based design; values in parentheses indicate the Gold Fitness scores.

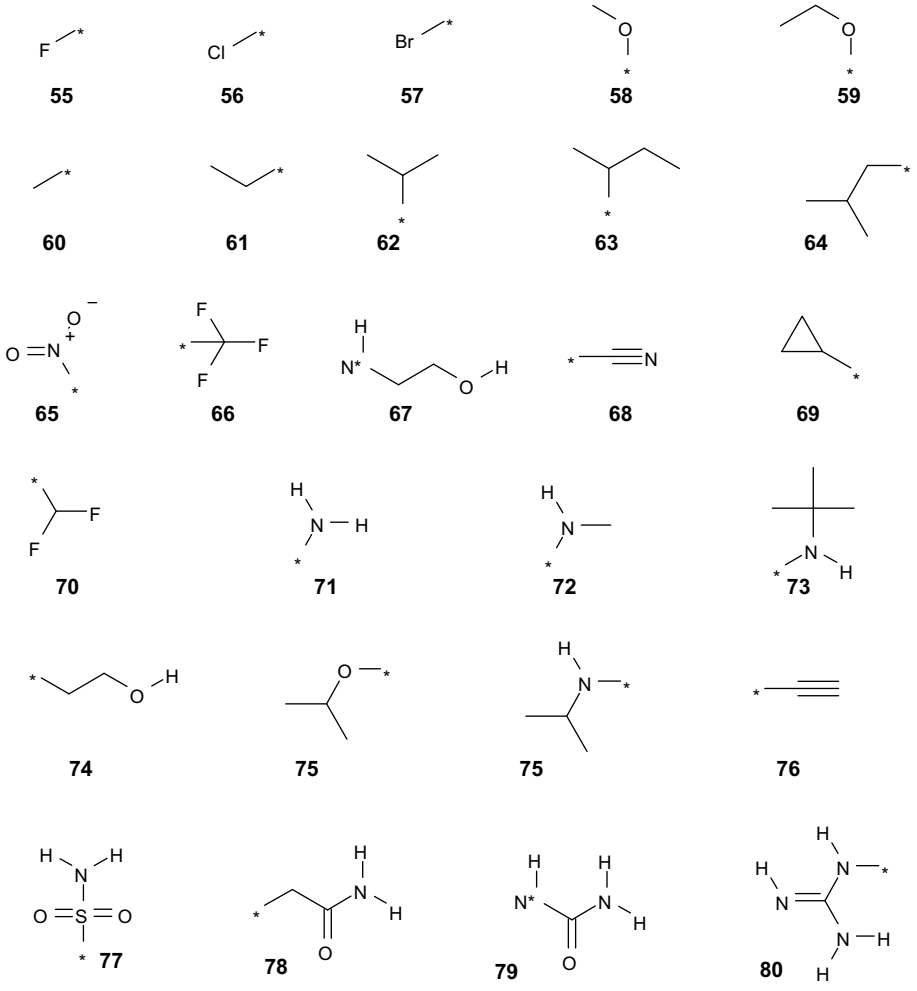


Fig. 6. Some of the substituents used in the ring systems, '*' indicates point of attachment.

molecules not only retrieved high, medium and low active molecules from database but also the difference between the experimental and predicted activities are also minimal. Thus the error values are very low, and the fit value gives a good measure of how well the defined features in Hypo-1 fit well with the pharmacophore of each molecule.

3.2. Pharmacophore model validation using known GSK-3β inhibitors

The validity of any pharmacophore model needs to be ascertained by screening some known inhibitors (test set) that are retrieved from the Kinase inhibitor databases in order to check how many active molecules are picked in the screening process, how their predicted activities are correlated with the experimental activities and the efficiency in reducing the false positives or false negatives. Hypo-1 was used to screen 416 GSK-3β of known high, medium and low active inhibitors of the test set. Database mining was performed in Catalyst software using the BEST flexible searching technique. A number of parameters such as hit list (Ht), number of active percent of yields (%*Y*), percent ratio of actives in the hit list (%*A*), enrichment factor of 2.13 (*E*), False negatives, False positives and Goodness of hit score of 0.85 (GH) are calculated (Table 4) while carrying out the pharmacophore model and virtual screening of test set molecules [42]. While the False positives and false negatives, 18 and 9, respectively, are minimal, enrichment factor of 2.13 against a maximum value of 2.25 is a very good indication of the high efficiency of the screening. In 176 molecules predicted to be active, 167 molecules were correctly picked, thus missing only 9 false negatives with 18 false positives overall. GH score assessment of hit lists was used for optimizing the working pharmacophore model by screening databases with known biological activities. It is to be noted that the technique can also be used to focus a list of active molecules as a post-HTS processing, or to prioritize a virtual library as a pre-HTS screening.

3.3. Docking studies

GOLD being widely regarded as one of the best docking programs, our initial fragment docking studies were carried out with this software [46]. GSK-3β kinase inhibitors were docked using different docking programs like Gold, Glide and Ligand Fit. The performance of the different scoring functions in docking of 441 molecules in the active site is analyzed in terms of correlation.

1) Gold Fitness score, correlation value = 0.59
2) Gold ChemScore, correlation value = 0.59
3) Glide Standard precision, correlation value = 0.55
4) Glide Extra precision, correlation value = 0.50 and
5) Ligand fit dock score, correlation value = 0.46.

On the basis of correlation alone, it would appear that GOLD Fitness, ChemScore and Glide standard precision are comparable, or perhaps one could state that GOLD performs slightly better than Glide standard precision. Most of the potent GSK-3β inhibitors form H-bonds with GSK-3β when binding in its ATP-binding site. The first two H-bonds are with VAL135 and ASP133 in the Hinge region and also third H-bond with ARG141 in some of the molecules. In some of the molecules it also forms H-bonds with GLN185, LEU132, ASP200 and LYS88 amino acids of GSK-3β.

3.4. Comparative model

The results obtained this far are mixed in nature but have several favorable features. To get a better VS model, a multiple linear regression analysis was carried out using Pharmacophore model and five different docking scoring functions (Comparative model) [49,50]. Out of the 441 ligands, 50 were used in the training set and 391 in the test set. A combination of Pharmacophore model, GOLD Fitness score, Gold ChemScore, Glide SP, Glide XP and Ligand fit dock score gave a better model than any other (Fig. 4) combination of models.

Activity = "0.754183 + 0.3594 × "Pharmacophore_Predicted_pIC50" + 0.017709 × "Gold Fitness" + 0.0068673 × "ChemScore" − 0.231128 × "Glide_SP" − 0.059958 × "Glide_XP" + 0.0017235 × "Dock Score".

The coefficient of each term is a measure of the contribution of each scoring function towards the final model. The final linear regression model $r^2 = 0.843$, $r^2_{cv} = 0.812$, PRESS = 11.92, $r^2_{bs} = 0.844$, was validated by the test set of 391 molecules. The model was able to correctly predict the activity of 94% of the 50 molecules in training set and 78% in case of 391 test set molecules within an order of magnitude.

3.5. Fragment-based and knowledge-based virtual screening

820 unique fragments were generated from Kinase inhibitor database of 113,868 molecules using the GVK in-house developed program fragment descriptors [48]. The structures of some of the fragments, substituents used for the incremental construction of molecules are shown in Figs. 5–7 [51,52]. These fragments were docked into the active site of GSK-3β (1q5k, 5 Å from the centroid of

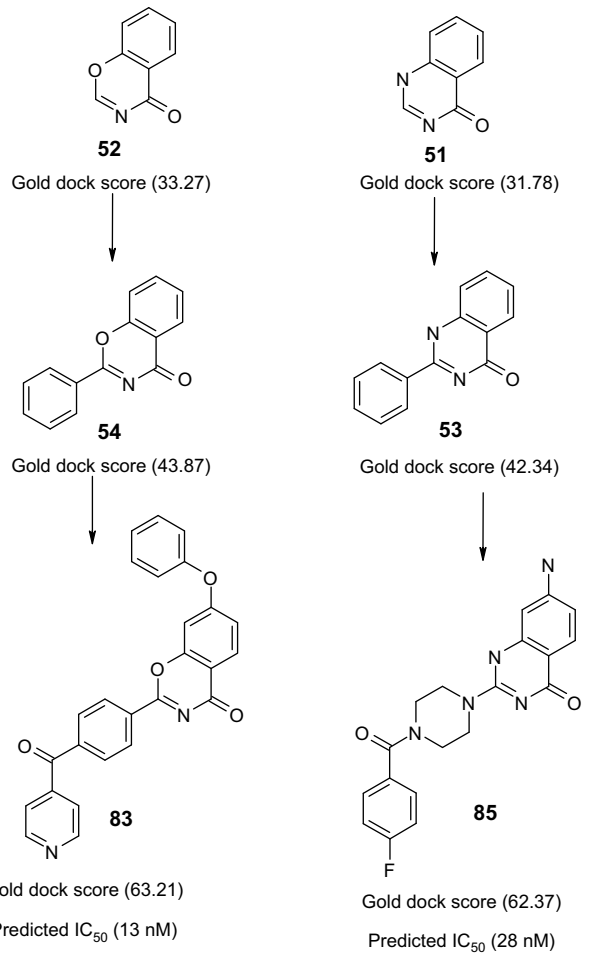


Fig. 7. Incremental construction of selective GSK-3β inhibitors using fragment-based and knowledge-based design.

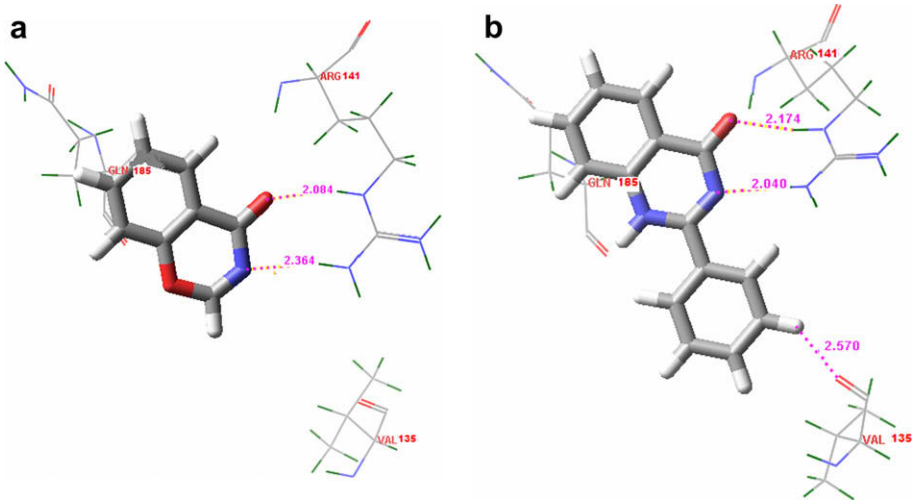


Fig. 8. Binding mode of fragments **52**, **53** in the ATP-binding site of GSK-3β, suggested by molecular docking studies. Fragments **52, 53** and its H-bonding patterns (VAL135 and ARG141) are represented in stick and line model. The H-bond between phenyl C–H and carbonyl is more likely due to van der Waals forces or desolvation effects. The key H-bonds between fragments and GSK-3β are indicated by the pink dotted lines, together with the bond lengths in angstroms. The atoms are colored as follows: H, green in protein and white in ligand; O, red; N, blue; C, gray. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ligand) and the fragments with high affinity and good score towards the hinge region of enzyme was selected for further virtual library generation by knowledge-based design using the structures of crystal ligand 1 with K_i of 38 nM (AstraZeneca molecule in Discovery Phase) in 1q5k PDB and Molecule **2** with IC_{50} of 0.09 nM. A total of 10,000 molecules were generated based on the knowledge of binding interaction of Ligand with the protein and also the common features necessary for the biological activity of molecule [44,45]. The binding mode derived from the docking pose of fragments **52** and **53** in the ATP-binding site of GSK-3β is illustrated in Fig. 8a and b, respectively. According to this docking model, Fragment **52** forms one H-bond and 53 forms two H-bonds with GSK-3β when binding in its ATP-binding site. Fragment **52** forms H-bond between the 3-aza and 4-one group of 1*H*-Quinazolin-4-one with side chain guanidine group of ARG141 with a distance of 2.364 and 2.084 Å. Fragment **53** forms first H-bonds between the 3-aza and 4-one group of 1*H*-Quinazolin-4-one with side chain guanidine group of ARG141 with a distance of 2.040 and 2.174 Å. The second H-bond is formed between Aromatic C–H and backbone carbonyl oxygen of VAL135 with a distance of 2.570 Å. The comparative

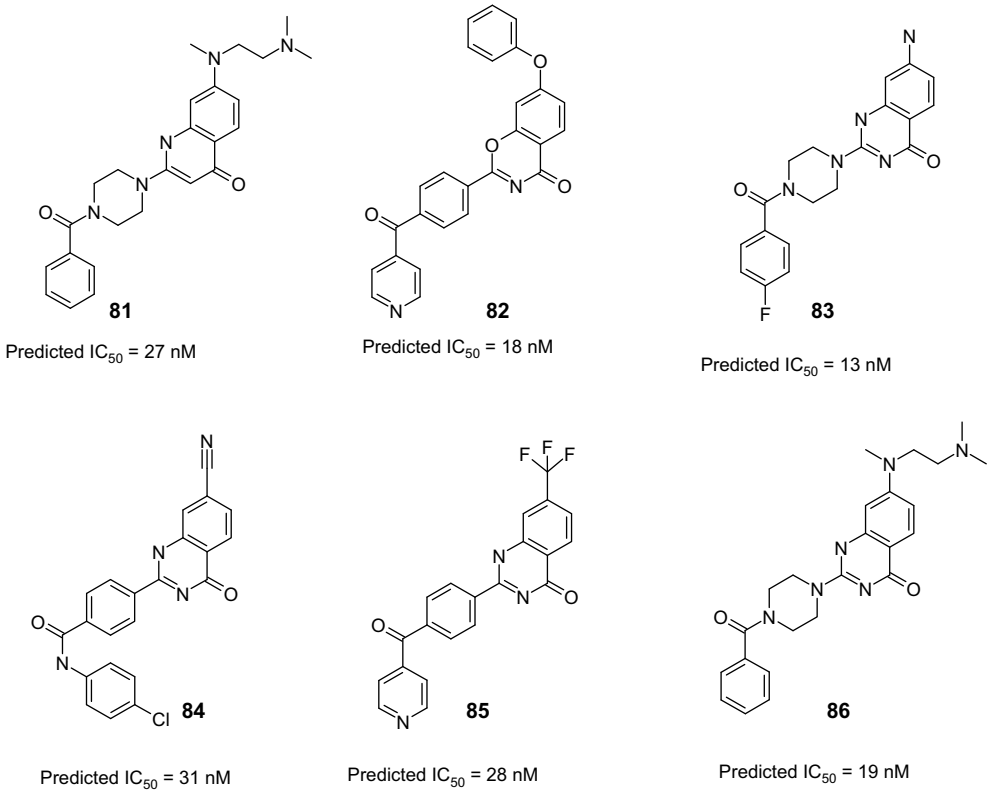


Fig. 9. Identified and optimized lead molecules through Virtual library screening.

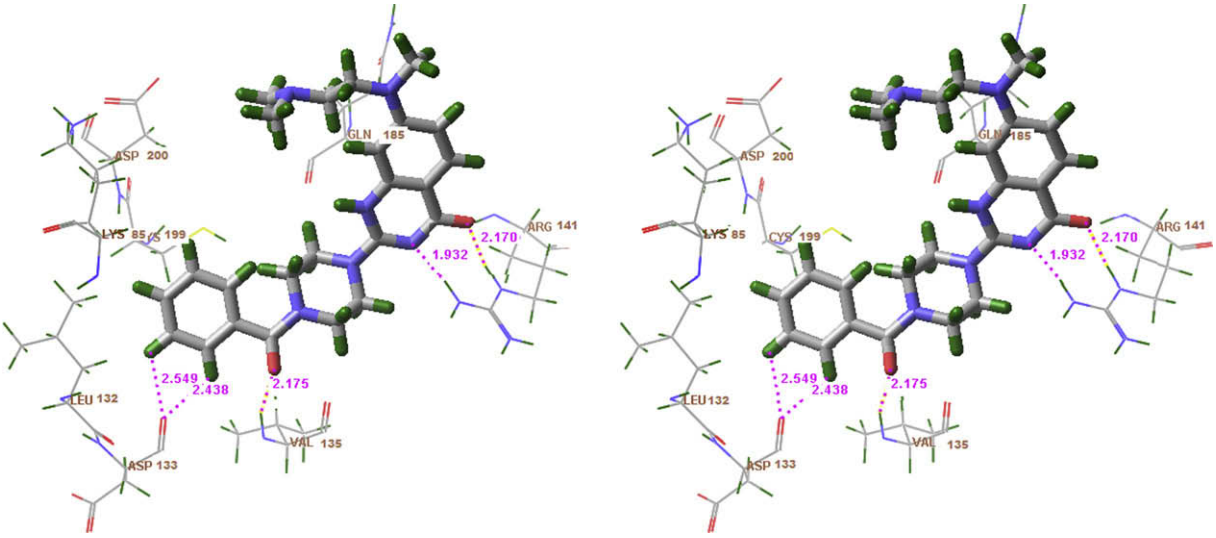


Fig. 10. Stereoview of the binding mode of molecule **86** in the ATP-binding site of GSK-3β, suggested by molecular docking studies. Molecule **86** and its H-bonding patterns (ASP133, VAL135 and ARG141) are represented in stick and line model. The H-bond between phenyl C–H and carbonyl is more likely due to van der Waals forces or desolvation effects. The key H-bonds between **81** and GSK-3β are indicated by the pink dotted lines, together with the bond lengths in angstroms. The atoms are colored as follows: H, green; O, red; N, blue; C, gray. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

model developed using Pharmacophore, Glide, Gold and Ligand Fit was used to screen the virtual library consisting of 10,000 molecules, which yielded hits comprising 1372 high, 5718 medium and 2910 low active molecules. The molecules were then broadly classified into three categories namely highly active (<0.02 μM), moderately active (>0.02–0.2 μM) and low active (>0.2 μM). Some of the potent lead molecules with their comparative predicted IC_{50} values are shown in Fig. 9.

Molecular docking studies were conducted for all the designed active molecules against the X ray structure of GSK-3β (1q5k.pdb) [10,11]. The binding mode derived from the docking pose of 81 in the ATP-binding site of GSK-3β is illustrated in Fig. 10. According to this docking model, Molecule **81** forms H-bonds with GSK-3 β when binding in its ATP-binding site. The first of the two H-bonds formed between the backbone carbonyl oxygen of ASP133 and the Aromatic C–H with a distance of 2.549 and 2.438 Å is more likely due to van der Waals interactions or desolvation effects. The second H-bond is formed between the backbone nitrogen of VAL135 with the carbonyl of benzoyl group with a distance of 2.175 Å. The third H-bond is formed between the side chain guanidine group of ARG141 and 3-aza and 4one group of 1*H*-Quinazolin-4-one with distances of 1.932 and 2.170 Å, respectively. The first two H-bonds that are set up with the backbone atoms are identical to those observed in Hinge region when staurosporine binds to various kinases. The third H-bond with ARG141 appears to be unique to GSK-3β and provides as interesting model to explain the high GSK-3β selectivity observed for 1*H*-Quinazolin-4-ones and Benzo[*e*][1,3]oxazin-4-ones. The positively charged ARG141 is quite unique to GSK-3β, with many of the other kinases having a negatively charged residue, will not only disrupt the H-bonding with the 3-aza and 4-one of 1*H*-Quinazolin-4-ones but also introduces an unfavorable repellent interaction with the lone-pair electrons on that atom. This unique binding feature explains how 1*H*-Quinazolin-4-ones and Benzo[*e*][1,3] oxazin-4-ones exhibit high selectivity at GSK-3 β versus other kinases. Most of the inhibitors had bulky substituents interacting with the hydrophobic pocket. In addition, it was observed that docking could limit the number of false positive hits compared to Hypo1. Interestingly, the pharmacophore and docking models showed complementary behavior in limiting the number of false positive and false negative hits. The overlay [53] of the pharmacophore model onto the docking pose of the enzyme is shown in Fig. 11. A full discussion for the consideration of several factors such as H-bonding, solvation/desolvation effects, including van der Waals forces and electrostatic interactions is required for the future computational studies. These studies are in progress and results will be published elsewhere.

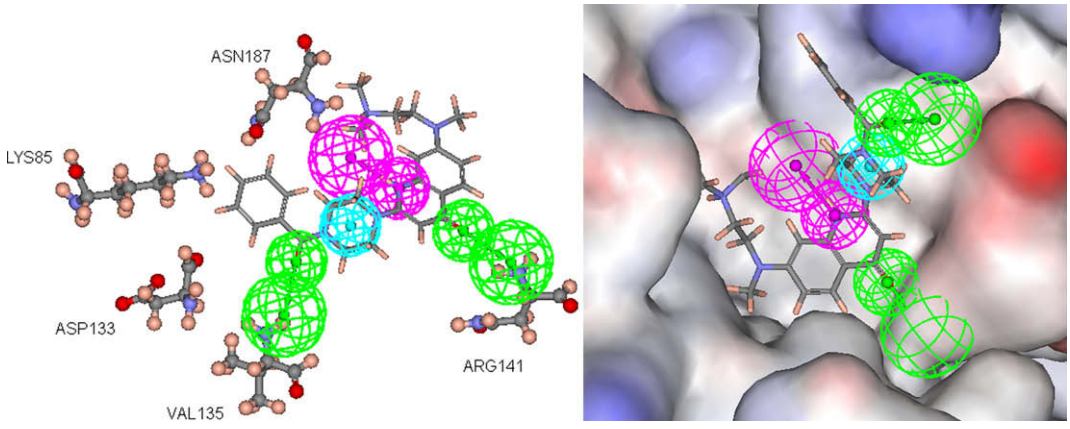


Fig. 11. Overlay of the pharmacophore model onto the docking pose of enzyme with molecule **81**.

4. Conclusions

GSK-3β enzymes have proven to be exciting and promising novel targets for the treatment of type II diabetes and Alzheimer's disease. The best quantitative pharmacophore model in terms of predictive value consisted of four features like two hydrogen bond acceptors, one hydrogen bond donor and one hydrophobic moiety, which is further validated by using a large set of 416 GSK-3β inhibitors that gives a r value of 0.796 and also Goodness of hit score of 0.85. The most active molecule **11** ($IC_{50} = 0.0001$ μM) in the training set fits very well with this top scoring pharmacophore hypothesis. Virtual screening produced some false positives and a few false negatives. It is believed that concurrent use of Docking and Multiple Linear Regression Analysis, which readily minimizes these errors, could be an added tool for Pharmacophore model based virtual screening in order to produce reliable true positives and negatives. The pharmacophore model correlates well with the docking model, one hydrogen bond acceptor is for H-bond interaction with VAL135 and the hydrophobic feature is for H-bond interactions with ASP133, one hydrogen bond donor and second hydrogen bond acceptor is necessary for the H-bond interactions with ARG141. The first two H-bonds with VAL135 and ASP133 are identical to those observed in Hinge region and the third H-bond with ARG141 appears to be unique to GSK-3β and portrayed as an interesting model to explain the high GSK-3β selectivity observed for 1*H*-Quinazolin-4-ones and Benzo[*e*][1,3]oxazin-4-ones. This pharmacophore model was further used to search the virtual library consisting of 10,000 structurally diversified molecules generated using fragment-based and knowledge-based design, which yielded 428 potential hits. The activities of those molecules were predicted using the developed pharmacophore and docking model and the highly active molecules are further refined using the Comparative model, which could be a valuable tool to produce reliable true positives and negatives to design more potent lead molecules against Glycogen Synthase Kinase 3β inhibitors. Fragment-based drug design continues to expand and evolve. As the understanding of protein–ligand interactions and molecular recognition increases, our ability to rapidly develop potent and selective inhibitors of therapeutically relevant targets could be dramatically enhanced. Their appropriate use in a drug discovery process should improve the ability to critically assess the identification and optimization of lead compounds that have high potential for generating new therapeutic agents.

Acknowledgements

Vadivelan would like to thank D.S. Brar, Chairman, G.V.Sanjay Reddy, MD and Manni Kantipudi, President of GVK Biosciences Pvt. Ltd., for their continuous support, S. Rama Devi, Saraswathi N. Balaji, Rambabu G, MNS Pavan Kumar and BVS Suneel Kumar for their technical support.

References

[1] L. Meijer, M. Flajolet, P. Greengard, Trends Pharmacol. Sci. 25 (2004) 471–480.
[2] L. Dorronsoro, A. Castro, A. Martinez, Expert Opin. Ther. Patents 12 (2002) 1527–1536.
[3] Clinical Candidate Database, GVK Biosciences Private Limited, Hyderabad, India, 2006.
[4] A.M. Bray, P. Schultz, D.B. Ring, S.D. Harrison, Patent WO9816528, 1998.
[5] L. Naerum, L. Norskov-Lauritsen, P.H. Olesen, Bioorg. Med. Chem. Lett. 12 (2002) 1525–1528.
[6] A. Martinez, A. Castro, I. Dorronsoro, M. Alonso, Med. Res. Rev. 22 (2002) 373–384.
[7] A. Martinez, M. Alonso, A. Castro, C. Perez, F.J. Moreno, J. Med. Chem. 45 (2002) 1292–1299.
[8] D.G. Smith, M. Buffet, A.E. Fenwick, D. Haigh, R.J. Ife, M. Saunders, B.P. Slingsby, R. Stacey, R.W. Ward, Bioorg. Med. Chem. Lett. 11 (2001) 635–639.
[9] M.P. Coghlan, A.A. Culbert, D.A.E. Cross, S.L. Corcoran, J.W. Yates, N.J. Pearce, O.L. Rausch, G.J. Murphy, P.S. Carter, L.R. Cox, D. Mills, M.J. Brown, D. Haigh, R.W. Ward, D.G. Smith, K.J. Murray, A.D. Reith, J.C. Holder, Chem. Biol. 7 (2000) 793–803.
[10] G.H. Kuo, C. Prouty, A. DeAngelis, L. Shen, D.J. O'Neill, C. Shah, P.J. Connolly, W.V. Murray, B.R. Conway, P. Cheung, L. Westover, J.Z. Xu, R.A. Look, K.T. Demarest, S. Emanuel, S.A. Middleton, L. Jolliffe, M.P. Beavers, X.J. Chen, J. Med. Chem. 46 (2003) 4021–4031.
[11] F.X. Tavares, J.A. Boucheron, S.H. Dickerson, R.J. Griffin, F. Preugschat, S.A. Thomson, T.Y. Wang, H.Q. Zhou, J. Med. Chem. 47 (2004) 4716–4730.
[12] D.C. Rees, M. Congreve, C.W. Murray, R. Carr, Nat. Rev. Drug Disc. 3 (2004) 660–672.
[13] D.A. Erlanson, R.S. McDowell, T. O'Brien, J. Med. Chem. 47 (2004) 3463–3482.
[14] E.R. Zartler, M.J. Shapiro, Curr. Opin. Chem. Biol. 9 (2005) 366–370.
[15] J.H. Philip, J. Med. Chem. 49 (2006) 6972–6976.
[16] T.I. Oprea, J. Comput. Aided Mol. Des. 6 (2002) 325–334.
[17] A. Gill, A. Cleasby, H. Jhoti, ChemBioChem 6 (2005) 506–512.
[18] H.J. Bohm, A. Flohr, M. Stahl, Drug Discov. Today. Technol. 1 (2004) 217–224.
[19] H. Zhao, Drug Discov. Today 12 (2007) 149–155.
[20] P. Ertl, S. Jelfs, J. Muhlbacher, A. Schuffenhauer, P. Selzer, J. Med. Chem. 49 (2006) 4568–4573.
[21] Glide, Version 3.5, Schrodinger, L.L.C., New York, 2006.
[22] G. Jones, P. Willett, R.C. Glen, J. Mol. Biol. 245 (1995) 43–53.
[23] Cerius^2, Version 4.11, Accelrys Inc, San Diego, CA, USA, 2007.
[24] Catalyst, Version 4.11, Accelrys Inc, San Diego, CA, USA, 2007.
[25] Kinase Inhibitor Database, GVK Biosciences Private Limited, Hyderabad, India, 2005.
[26] H.C. Zhang, K.B. White, H. Ye, D.F. McComsey, C.K. Derian, M.F. Addo, P. Andrade-Gordon, A.J. Eckardt, B.R. Conway, L. Westover, J.Z. Xu, R. Look, K.T. Demarest, S. Emanuel, B.E. Maryanoff, Bioorg. Med. Chem. Lett. 13 (2003) 3049–3053.
[27] C. Kunick, K. Lauenroth, M. Leost, L. Meijer, T. Lemcke, Bioorg. Med. Chem. Lett. 14 (2004) 413–416.
[28] M. Gompel, M. Leost, E.B. De Kier Joffe, L. Puricelli, L.H. Franco, J. Palermo, L. Meijer, Bioorg. Med. Chem. Lett. 14 (2004) 1703–1704.
[29] D.S. Williamson, M.J. Parratt, J.F. Bower, J.D. Moore, C.M. Richardson, P. Dokurno, A.D. Cansfield, G.L. Francis, R.J. Hebdon, R. Howes, P.S. Jackson, A.M. Lockie, J.B. Murray, C.L. Nunns, J. Powles, A. Robertson, A.E. Surgenor, C.J. Torrance, Bioorg. Med. Chem. Lett. 15 (2005) 863–867.
[30] W.H. Bullock, S.R. Magnuson, S. Choi, D.E. Gunn, J. Rudolph, Curr. Top. Med. Chem. 2 (2002) 915–938.
[31] P.H. Olesen, A.R. Sorensen, B. Urso, P. Kurtzhals, A.N. Bowler, U. Ehrbar, B.F. Hansen, J. Med. Chem. 46 (2003) 3333–3341.
[32] Y. Mettey, M. Gompel, V. Thomas, M. Garnier, M. Leost, I. Ceballos-Picot, M. Noble, J. Endicott, J.M. Vierfond, L. Meijer, J. Med. Chem. 46 (2003) 222–236.
[33] T.A. Engler, J.R. Henry, S. Malhotra, B. Cunningham, K. Furness, J. Brozinick, T.P. Burkholder, M.P. Clay, J. Clayton, C. Diefenbacher, E. Hawkins, P.W. Iversen, Y. Li, T.D. Lindstrom, A.L. Marquart, J. McLean, D. Mendel, E. Misener, D. Briere, J.C. O'Toole, W.J. Porter, S. Queener, J.K. Reel, R.A. Owens, R.A. Brier, T.E. Eessalu, J.R. Wagner, R.M. Campbell, R. Vaughn, J. Med. Chem. 47 (2004) 3934–3937.
[34] S. Holder, T. Naumann, K. Schonafinger, D.W. Will, H. Matter, G. Muller, S.D. Le, B. Baudoin, T. Rooney, F. Halley, G. Tiraboschi, Patent US0176377A1, 2004.
[35] S. Wang, C. Meades, G. Wood, J. O'Boyle, C. McInnes, P. Fischer, Patent US0192300A1, 2005.
[36] S. Berg, S. Hellberg, Patent WO004478A1, 2003.
[37] P.A. Albaugh, J. Ammenn, T.P. Burkholder, J.R. Clayton, S.E. Conner, B.E. Cunningham, T.A. Engler, K.W. Furness, J.R. Henry, Y. Li, S. Malhotra, M.J. Tebbe, G. Zhu, Patent WO076398A2, 2003.
[38] J.M. Cho, S. Ro, G. Lee, K.J. Tae Lee, D. Shin, Y. Hyun, C. Lee, C. Seung J. Kim, Y.H. Jeon, Patent WO065370A1, 2004.
[39] K. Schoenafinger, S. Hoelder, Will, W. David, H. Matter, G. Mueller, M. Bossart, Patent WO085230A1, 2005.
[40] G.A. Martinez, D.I. Dorronsoro, C.M. Alonso, P.G. Panizo, del, H.A. Fuertes, P.M.J. Perez, P.M. Medina, Patent EP1586318A1, 2005.
[41] A.K. Debnath, J. Med. Chem. 45 (2002) 41–53.
[42] O.F. Guner, D.R. Henry, in: Osman F. Guner (Ed.), Phamacophore Perception, Development, and Use in Drug Design, International University Line, La Jolla, California, 2000, pp. 193–210.
[43] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, Nucleic Acids Res. 28 (2000) 235–242.
[44] Maestro, Version 1.0.9113, Schrodinger, L.L.C., New York, 2006.
[45] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, P.S. Shenkin, J. Med. Chem. 47 (2004) 1739–1749.
[46] G. Jones, P. Wilett, R.C. Glen, A.R. Leach, R. Taylor, J. Mol. Biol. 267 (1997) 727–748.
[47] C.M. Venkatachalam, X. Jiang, T. Oldfield, M. Waldman, J. Mol. Graph. Model 21 (2003) 289–307.
[48] Fragments Descriptors, GVK Biosciences Pvt. Ltd., Hyderabad, India, 2006.
[49] P.S. Charifson, J.J. Corkery, M.A. Murcko, W.P. Walters, J. Med. Chem. 42 (1999) 5100–5109.
[50] V. Aparna, G. Rambabu, S.K. Panigrahi, J.A. Sarma, G.R. Desiraju, J. Chem. Inf. Model. 45 (2005) 725–738.
[51] H.B. Broughton, I.A. Watson, J. Mol. Graph. Model. 23 (2005) 51–58.
[52] M.J. Hartshorn, C.W. Murray, A. Cleasby, M. Frederickson, I.J. Tickle, H. Jhoti, J. Med. Chem. 48 (2005) 403–413.
[53] Accelrys Software Inc., Discovery Studio Modeling Environment, Release 1.7, Accelrys Software Inc., San Diego, 2008.