



Ligand-based modeling followed by *in vitro* bioassay yielded new potent glucokinase activators



Mutasem O. Taha^{a,*}, Maha Habash^b, Ma'mon M. Hatmal^c,
Ahmed H. Abdelazeem^{d,e}, Amjad Qandil^{f,g}

^a Drug Discovery Unit, Department of Pharmaceutical Sciences, Faculty of Pharmacy, The University of Jordan, Amman, Jordan

^b Department of Pharmaceutical Chemistry and Pharmacognosy, Faculty of Pharmacy, Applied Science University, Amman, Jordan

^c Department of Medical Laboratory Sciences, Faculty of Allied Health Sciences, The Hashemite University, Zarqa, Jordan

^d Department of Medicinal Chemistry, Faculty of Pharmacy, Beni-Suef University, Beni-Suef, Egypt

^e Department of Pharmaceutical Chemistry, College of Pharmacy, Taif University, Taif, Saudi Arabia

^f Department of Medicinal Chemistry and Pharmacognosy, Faculty of Pharmacy, Jordan University of Science and Technology, Irbid, Jordan

^g Department of Basic Pharmaceutical Sciences, College of Pharmacy, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia

ARTICLE INFO

Article history:

Accepted 15 December 2014

Available online 24 December 2014

Keywords:

Glucokinase enzyme

Pharmacophore modeling

Quantitative structure–activity relationship

In-silico screening

In vitro assay

ABSTRACT

Glucokinase (GK) has received recent interest as a valid antidiabetic target. With this in mind, we applied a computational workflow based on combining pharmacophore modeling and QSAR analysis followed by *in silico* screening toward the discovery of novel GK activators. Virtual screening identified 10 promising bioactivators from the National Cancer Institute (NCI) list of compounds. The most potent NCI hit illustrated 6.3-fold GK activation at 10 μ M. These results demonstrated that our virtual screening protocol was able to identify novel GK activator leads for subsequent development into potential antidiabetic agents.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Glucokinase (GK), also referred to as hexokinase IV or D, is a member of the hexokinases family. It is predominantly expressed in the liver and pancreas. GK catalyses the phosphorylation of glucose to glucose-6-phosphate (G6P) via adenosine triphosphate (ATP) and Mg^{2+} . Furthermore, GK exerts high control in hepatic glucose metabolism. It acts as key player in the fed state by influencing glucose uptake, while in the fasted state it controls glucose production [1].

Several GK mutations have been linked to abnormalities in blood sugar levels due to either gain or loss of function in GK. Loss-of-function mutations in the GK gene is linked to type 2 diabetes of the young characterized by early onset of mild chronic fasting hyperglycemia [2]. On the other hand, rare activating mutations of GK in man cause hyperinsulinaemia with hypoglycaemia [3].

GK has a unique kinetic profile compared to other hexokinases. It has low affinity to glucose at low glucose concentrations; however, it becomes significantly more active at higher glucose levels.

This sigmoidal response to glucose concentration is referred to as 'positive kinetic cooperativity for glucose' and it seems to be related to the unique kinetic transition forms of GK [3].

GK has both open and closed crystal structures in the absence and presence of ligands (glucose and/or GK activators), respectively. It is postulated that in the presence of bound glucose the closed GK conformations are stabilized and GK becomes bioactive (switched on), while the open form is catalytically inactive and is the more stable form in the unbound state (switched off) [4].

The combination of positive kinetic cooperativity, low affinity to glucose at low glucose concentrations, and lack of end-product inhibition render GK activators of excellent potential as treatments of hyperglycemia and diabetes [5]. Activation of GK is expected to lead to better glycemic control through hepatic and pancreatic pathways. Additionally, the reduction in GK activity in response to low glucose levels reduces the possibility of hypoglycaemia during the treatment with GK activators [6].

Initial reports from Hoffmann-LaRoche Inc. about new GK activators (GKAs) (Fig. 1) prompted many pharmaceutical companies to initiate discovery projects to identify small-molecule GKAs as potential treatments for diabetes [5,7]. X-ray crystallographic images of GKAs co-crystallized within GK showed that these compounds bind to an allosteric pocket in the enzyme [3]. GKAs increase

* Corresponding author. Tel.: +962 6 5355000x2330; fax: +962 6 5339649.
E-mail address: mutasem@ju.edu.jo (M.O. Taha).

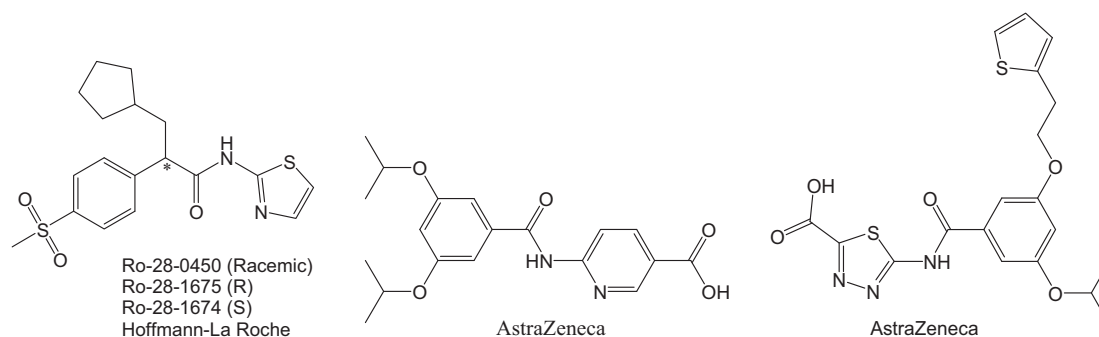


Fig. 1. Selected structures of GKA generated by Hoffmann-LaRoche and AstraZeneca.

the affinity of GK for glucose by stabilizing the closed conformations of this kinase, *i.e.*, in a similar manner to that of GK binding to glucose.

Unsurprisingly, recent discovery and optimization efforts for new GK activators relied heavily on structure-based ligand design [8]. Up to now, 11 X-ray complexes are found in the Protein Data Bank for human GKα (*e.g.*, PDB codes: 3ID8, 3IDH, 3FGU, 3H1V, 3IMX, 3AOI, 3GOI, 3FRO, 3F9M, 1V4S and 1V4T). However, crystallographic structures are restricted by limited resolution [9], crystallization-related artifacts of the ligand–protein complex [10] and negligence of protein anisotropic motion and conformational substrates [11].

The continuous interest in designing new GK activators, combined with problems of crystallographic structures and the induced-fit flexibility documented for GK [3,7,12] encouraged us to investigate the prospects of producing ligand-based pharmacophore(s) incorporated within quantitative structure–activity (QSAR) equation. This combination is independent of the structure of the binding site and thus should avoid the downsides of structure-based methodologies; furthermore, the resulting pharmacophore(s) can be used as search query(ies) for exploration of new GK activators.

We previously reported the use of this interesting methodology toward the discovery of new leads for glycogen synthase kinase 3β [13], hormone sensitive lipase [14], bacterial MurF [15], protein tyrosine phosphatase1B [16], influenza neuraminidase [17], β-secretase [18], CDK1 [19], cholesteryl ester transfer protein [20], and β-D-galactosidase [21].

Our computational workflow started by generating many reasonable pharmacophores for a list GK activators using CATALYST-HYPOGEN [22]. Subsequently, genetic algorithm (GA) coupled with multiple linear regression (MLR) were implemented to search for optimal quantitative structure–activity relationship (QSAR) that combine high-quality binding pharmacophore with other molecular descriptors that can explain bioactivity variation across the collected list of GK activators. The QSAR-selected pharmacophore was validated using receiver operating characteristic (ROC) curve analysis, and was subsequently employed to mine the national cancer institute's (NCI) compound database for new GK activators. Captured hits were evaluated *in vitro*.

2. Materials and methods

2.1. Molecular modeling

2.1.1. Software and hardware

Pharmacophore and QSAR modeling studies were performed using CATALYST (HYPOGEN module, version 4.11, from Accelrys, USA), CERIU2 (version 4.10, from Accelrys, USA) and Discovery Studio (version 2.5.5, from Accelrys, USA) software suites.

The chemical structures were drawn using ChemDraw Ultra 7.0 (Cambridge Soft Corp., USA).

2.1.2. Data set and conformational analysis

The structures of 30 GK activators (Table 1) were collected from the literature [23a,23b]. Their *in vitro* bioactivities were expressed as concentrations that activated GK by 50% (EC₅₀). Table 1 shows the collected structures and their corresponding EC₅₀ values. The logarithm of EC₅₀ (μM) values were used in modeling to correlate data linearly to the free energy change. However, in cases where EC₅₀ values were expressed as being >10 μM (*e.g.*, 2–6, 8 and 10) they were assumed to be 200 μM to maintain 4 log cycles difference from the most potent compound (28, EC₅₀ = 0.02 μM). This bioactivity spread is essential requirement for CATALYST pharmacophore modeling [22,24]. The logarithmic transformation of EC₅₀ values is expected to reduce any possible errors resulting from this supposition.

The chemical structures of the activators were drawn and saved as mol files. Then, they were converted into corresponding 3D structures and minimized to the closest local energy minima using the CHARMM force field within CATALYST. The resulting 3D conformers were utilized as starting points for conformational analysis.

The conformational surface of each activator (1–30, Table 1) was explored using the CHARMM force field implemented within CATALYST via the “best conformer generation” option. Conformational ensembles were generated for each training compound with energy threshold of 20 kcal/mol from the closest local minimum with a maximum limit of 250 conformers per molecule. The conformation search procedure implements a “poling algorithm” that penalizes closely related conformers to avoid entrapment in certain local minimum during conformational sampling [22], which endanger pharmacophore generation and subsequent *in silico* screening [25].

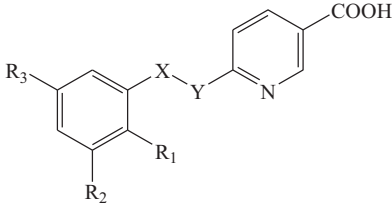
2.1.3. Pharmacophoric hypotheses generation

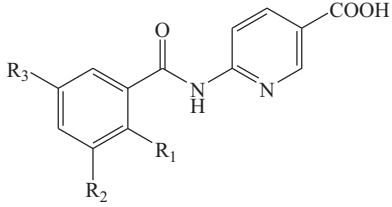
The training compounds (30 molecules) together with their associated conformational models were listed into a single spreadsheet with their EC₅₀ values combined with an “Uncertainty” of 3. This value assumes that the actual EC₅₀ value of any activator is situated somewhere in an interval ranging from one-third to three-times the reported EC₅₀ value [24b–d].

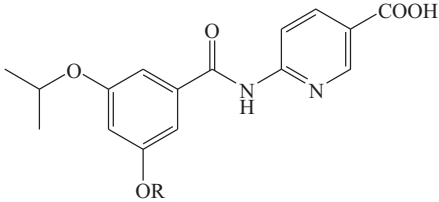
A structurally diverse training subset (Table 2) was selected for pharmacophore exploration through four modeling runs, as in Table 3. Different pharmacophores were produced by changing the interfeature spacing and the count of permissible features in the resulting models (Table 3). Section SM-1 under Supplementary Materials describes how CATALYST-HYPOGEN generates pharmacophoric models [24b–d]. Ultimately, our exploration efforts (4 automatic runs, Tables 2 and 3) yielded 40 binding models of variable qualities.

Table 1

The structures of GK activators utilized in modeling.

						
Compound	R ₁	R ₂	R ₃	X–Y	EC ₅₀ (μM) ^a	Reference
1	–OCH ₂ Ph	H	–SCH ₃	–CH=CH–	3.20	[23b]
2	H	–OCH ₂ Ph	–OCH ₂ Ph	–CH ₂ CH ₂ –	>10	<i>ibid</i>
3	H	–OCH ₃	–OCH ₃	–OCH ₂ –	>10	<i>ibid</i>
4	H	–OCH ₂ Ph	–OCH ₂ Ph	–CH ₂ O–	>10	<i>ibid</i>
5	H	–OCH ₂ Ph	–OCH ₂ Ph	–NHCO–	>10	<i>ibid</i>
6^b	–OCH ₂ Ph	H	–SCH ₃	–NHCO–	>10	<i>ibid</i>
7	–OCH ₂ Ph	H	–SCH ₃	–CONH–	4.22	<i>ibid</i>

					
Compound	R ₁	R ₂	R ₃	EC ₅₀ (μM) ^a	Reference
8	–OCH ₂ Ph	H	H	>10	<i>ibid</i>
9	H	–OCH ₂ - <i>o</i> -Cl-Ph	H	0.91	<i>ibid</i>
10^b	–OCH ₂ - <i>o</i> -Cl-Ph	–OCH ₂ - <i>o</i> -Cl-Ph	H	>10	<i>ibid</i>
11	H	–OCH(CH ₃) ₂	OCH ₂ CH(CH ₃) ₂	0.57	<i>ibid</i>
12	H	–OCH ₂ Ph	–OCH ₂ Ph	0.40	<i>ibid</i>
13	H	–OCH ₂ - <i>o</i> -F-Ph	–OCH ₂ - <i>o</i> -F-Ph	0.09	<i>ibid</i>

			
Compound	R	EC ₅₀ (μM) ^a	Reference
14	–CH ₂ CH ₂ -4-THP	1.33	<i>ibid</i>
15^b	–CH ₂ -CPent	0.65	<i>ibid</i>
16	–CH ₂ CH ₂ -CPent	0.17	<i>ibid</i>
17	–CH ₂ CH ₂ -3-pyridyl	1.26	<i>ibid</i>
18^a	–CH ₂ CH ₂ -4-pyridyl	1.78	<i>ibid</i>

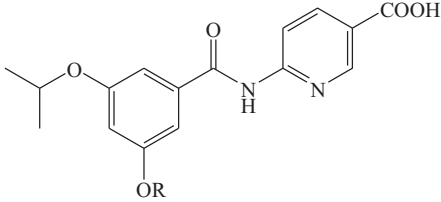
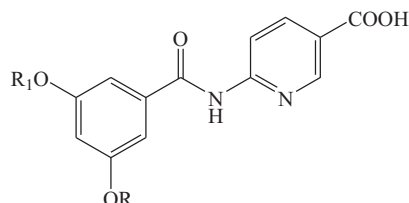
			
Compound	R	EC ₅₀ (μM) ^a	Reference
19	–CH ₂ CH ₂ Ph	0.13	<i>ibid</i>
20	–CH ₂ CH ₂ -3-thiophene	0.09	<i>ibid</i>
21^b	–CH ₂ Ph	0.29	<i>ibid</i>
22	–CH ₂ CH ₂ CH ₂ Ph	0.42	<i>ibid</i>
23	–CH ₂ - <i>o</i> -F-Ph	0.10	<i>ibid</i>

Table 1 (continued)

Compound	R	R ₁	EC ₅₀ (μM) ^a	Reference
24	(S)-CH(CH ₃)Ph	CH(CH ₃) ₂	0.11	[23a]
25	(R)-CH(CH ₃)Ph	CH(CH ₃) ₂	0.95	<i>ibid</i>
26	(S)-CH(CH ₃)CH ₂ OCH ₃	CH(CH ₃) ₂	0.61	<i>ibid</i>
27	(R)-CH(CH ₃)CH ₂ OCH ₃	CH(CH ₃) ₂	5.51	<i>ibid</i>
28	(S)-CH(CH ₃)CH ₂ Ph	CH(CH ₃) ₂	0.02	<i>ibid</i>
29^b	(R)-CH(CH ₃)CH ₂ Ph	CH(CH ₃) ₂	0.09	<i>ibid</i>
30	(S)-CH(CH ₃)CH ₂ Ph	(S)-CH(CH ₃)CH ₂ OCH ₃	0.03	<i>ibid</i>

^a EC₅₀ values were determined at 10 mmol/l glucose concentration.

^b These compounds were employed as external test subset in QSAR modeling.

Table 2

The training set employed in exploring the pharmacophoric space of GK activators, numbers correspond to compounds in Table 1. The compounds are classified according to their bioactivities.

Most active subset ^a	Intermediate subset	Least active subset ^b	No. of compounds
16, 19, 23, 24, 28, 29, 30	1, 7, 9, 11, 12, 14, 15, 17, 18, 21, 22, 25, 26, 27	2, 3, 4, 5, 6, 8	27

^a Potency categories as defined by Eq. (A) under Supplementary Materials.

^b Potency categories as defined by Eq. (B) under Supplementary Materials.

2.1.4. Assessment of the generated hypotheses

CATALYST.HYPOGEN minimizes certain cost function to select the best possible pharmacophore model. The cost function is comprised of 3 separate terms: Weight, Configuration and Error costs [22,24]. CATALYST also calculates the cost of the “null hypothesis” that assumes that experimental activities are normally distributed about their mean, i.e., no structure–activity relationship among training compounds. The larger the difference from the cost of the null hypothesis, the higher the probability that the particular pharmacophore is not produced by chance. In a successful CATALYST pharmacophore modeling trial, the generated models are ranked according to their total costs [22,24]. Detailed description of how CATALYST assesses pharmacophoric hypotheses during automatic pharmacophore modeling runs can be seen in the Supplementary Materials under section SM-2.

We also validated the generated pharmacophores using Fisher's permutation test implementing a 95% confidence level [26,36]. For more details about Fisher's randomization the reader is referred to section SM-5 under Supplementary Materials.

2.1.5. Clustering of the pharmacophore models

The successful models (40) were grouped into 7 clusters using the average linkage technique in CATALYST. Accordingly, similar pharmacophores were gathered in five-membered groups. Afterward, highest-ranking representatives (based on their

fit-to-bioactivity r^2 calculated against compounds **1–30**, Table 1) were chosen to represent their particular groups in consequent QSAR modeling (see Section 3, Table 4).

2.1.6. QSAR modeling

A group of 24 molecules taken from the collected list of activators (**1–30**, Table 1) was used as training list for QSAR statistical modeling. The remaining 6 compounds (20% of the collected dataset) were used as external list for validating optimal QSAR models (testing set). To select the testing list the collected activators (**1–30**, Table 1) were ranked based to their EC₅₀ values, subsequently every 5th molecule was chosen for the testing set beginning from the most potent compound. QSAR modeling was performed using calculated descriptors of various electronic and structural properties [27,31]. Furthermore, the training compounds were mapped against cluster-centers pharmacophores (7 models, Table 4 in Section 3), and their fit values were used as additional descriptors. The fit values were determined via equation (D) under Supplementary Materials [22]. For more experimental details about the QSAR modeling carried herein the reader is referred to section SM-4 under Supplementary Materials [22,27,31].

2.1.7. Receiver operating characteristic (ROC) curve analysis

QSAR-selected pharmacophore model (i.e., Hypo2/4) was validated by assessing its ability to selectively capture diverse GK

Table 3

Training sets and CATALYST run parameters employed in exploring GK pharmacophoric space.

Run number	Selected input features: types and ranges ^a	Other run parameters ^{b,c}
1	Hbic (0–3), HBA (0–3), HBD (0–3), RingArom (0–3), Neglon (0–1)	Min–Max: 4–5, spacing: 100
2	Hbic (0–3), HBA (0–3), HBD (0–3), RingArom (0–3), Neglon (0–1)	Min–Max: 5–5, spacing: 100
3	Hbic (0–3), HBA (0–3), HBD (0–3), RingArom (0–3), Neglon (0–1)	Min–Max: 4–5, spacing: 300
4	Hbic (0–3), HBA (0–3), HBD (0–3), RingArom (0–3), Neglon (0–1)	Min–Max: 5–5, spacing: 300

^a HBA: hydrogen bond acceptor, HBD: hydrogen bond donor, RingArom: aromatic ring, Hbic: hydrophobic feature, Neglon: negative ionizable feature. The allowed ranges of input features are in brackets.

^b Min–max: allowed minimum and maximum number of output features.

^c Unmentioned parameters were set to their default values.

Table 4

The success criteria of representative GK pharmacophore hypotheses.

Run ^a	Hypotheses ^b	Pharmacophoric features in generated hypotheses	Total cost	Config cost	Cost of null hypothesis	Residual Cost ^c	R ^d	R ₂ -statistic ^e	Cat-scramble (%) ^f
1	6	HBD, HBA, 3× Hbic	131.5	18.6	182.9	51.4	0.88	0.623	95
	10	2× HBA, HBD, Hbic	132.3	18.6	182.9	50.6	0.89	0.327	95
	4^g	HBD, 2× Hbic, RingArom, Neglon	128.4	16.9	182.9	54.5	0.89	0.599	95
2	8	HBD, 2× Hbic, RingArom, Neglon	130.1	16.9	182.9	52.8	0.89	0.459	95
	9	HBD, HBA, 2× Hbic, Neglon	130.3	16.9	182.9	52.6	0.88	0.433	95
3	5	HBA, 3× Hbic, Neglon	134.1	17.8	182.9	48.8	0.86	0.432	95
4	7	2× HBA, 2× Hbic, Neglon	128.5	15.9	182.9	54.4	0.90	0.317	95

^a Run number as in Table 3.^b Serial numbers given by CATALYST for individual hypotheses and reported in the log book of the corresponding automatic run.^c The difference between the total cost and the cost of the corresponding null hypotheses.^d Correlation coefficient for given by CATALYST for individual hypotheses against training list compounds only.^e Correlation coefficients between pharmacophore-based bioactivity estimates and bioactivities of all collected compounds.^f Fisher confidence level calculated employing the Cat. scramble methods.^g This pharmacophore appeared in the best QSAR equation (bolded).

activators from a large testing list of actives and decoys. The testing list was prepared as described by Verdonk and co-workers [28]. ROC analysis and preparation of decoy list are described in details in the Supplementary Materials under section SM-3 [28–30,32].

2.1.8. *In silico* screening for new GK activators

Hypo2/4 was used to mine the National Cancer Institute list (238,819 compounds) for new hits. Captured compounds were filtered by Veber's and Lipinski's rules [32b,32c]. Surviving hits were fitted against Hypo2/4. Subsequently, the resulting fit values together with other molecular descriptors were substituted in QSAR Eq. (1) to predict GK EC₅₀ values. The best-ranking hits were tested *in vitro*.

2.1.9. *In vitro* GK activation

2.1.9.1. Chemicals. All chemicals needed for bioassay were acquired from Sigma–Aldrich Company and were used without further purification.

2.1.9.2. *In vitro* assay. Bioassay is based on the phosphorylation of D-glucose by GK to yield D-glucose-6-phosphate, which is oxidized by the enzyme glucose-6-phosphate dehydrogenase (G6PD) in the presence of NADP into 6-phospho-D-gluconate and NADPH. The later has λ_{\max} of 340 nm. The rate at which NADPH is generated is directly related to the catalytic activity of GK.

The bioassay procedure was performed as reported previously [33]. Briefly, stock solutions of test samples were prepared in DMSO, and then serially diluted with deionized water to give the desired working concentrations. However, the reaction pH was set at 9.0 to help dissolve the tested hits (they generally have acidic moieties combined with significantly hydrophobic fragments). Bioassay was performed by adding 3 μ L of tested sample solution to a reaction mixture (90 μ L) composed of Tris–HCl buffer (75 mM, 24 mL, pH 9.0 at 30 °C); MgCl₂ (600 mM in deionized water, 1 mL, equivalent to 20.10 mM in the reaction mixture); ATP (120 mM in deionized water, 1 mL, equivalent to 4.02 mM in the reaction mixture); β -D-(+)-glucose (360 mM in deionized water, 1 mL, equivalent to 12.10 mM in the reaction mixture, which is close to the S0.5 reported for GK, i.e., 8 mM [4]) and NADP (27 mM in deionized water, 1 mL, equivalent to 0.90 mM in the reaction mixture). Subsequently, G6PD (1000 U/mL in cold deionized water, 3 μ L, equivalent to 0.031 U/ μ L in the reaction mixture) was added followed by human GK solution (0.05 units/mL, 3 μ L, equivalent to 1.56×10^{-6} U/ μ L in the reaction mixture) in cold tris buffer (pH

8.5, 4 °C) to initiate the reaction. The samples' concentrations were fixed at 10 μ M in the reaction well. The change in absorbance at λ 340 nm is measured. The rate of enzyme reaction was considered as the reference for activation process. Change in absorbance (rate) was determined at 5, 10 and 15 min for all tested compounds. Activation of human GK was calculated as percent activity of the unactivated enzyme control. DMSO concentrations were kept <1% in all experiments and controls. Some samples were prepared in duplicates (see Table 6).

We tested GK enzyme stability under the reaction conditions by monitoring the rate of NADPH production in the bioassay reaction mixture without adding any tested hits. We repeated these tests each time we evaluated the GK activator bioactivities of a group of compounds (overall six times). We also tested the reaction rate in the presence of half the amount of GK enzyme (0.78×10^{-6} GKU/ μ L) to check the correlation between reaction rates and the amount of GK in the reaction mixture (this test was done three times). Moreover, we evaluated the rate of NADPH generation in the absence of GK to check whether the reaction conditions generate any NADPH in the absence of GK enzymatic activity. These assessments illustrated stable and reproducible GK enzymatic activity under the reaction conditions (including the reaction pH of 9.0). Furthermore, our tests indicate that the rate of NADPH production increases significantly only in the presence of GK enzyme while it remains negligible in its absence. Figure A under Supplementary Materials shows the rates of NADPH production under bioassay conditions including two different GK concentrations.

3. Results and discussion

CATALYST models ligand–protein interaction using information extracted from structures of known bioactive ligands. HYPOGEN module of CATALYST identifies three-dimensional matrix of up to five binding features commonly encountered among active training compounds. This matrix should provide comparative arrangement for each input compound corresponding with their binding to a postulated common binding site. The considered binding features can be hydrogen bond donors and acceptors (HBD and HBA), positive and negative ionizable (Poslon and Neglon) groups, aliphatic and aromatic hydrophobes (Hbic), and aromatic planar rings (RingArom). The conformational flexibility of training molecules is simulated by generating multiple conformer representatives to cover certain energy range. CATALYST binding models

have been used as three-dimensional queries to mine databases for new active compounds [13–15,22,24,34].

In the current work, we produced different pharmacophoric hypotheses for a group of GK activators (Table 1). The selected training subset included activators of significant diversity (Table 2) with EC₅₀ values extending over ca. 3.5 logarithmic cycles. Subsequently, we implemented a workflow comprised of genetic algorithm coupled with multiple linear regression to select the best combination of pharmacophore and other descriptors able of explaining GK bioactivation variations across the collected activators.

3.1. Exploration of GK pharmacophoric space

We decided to evaluate the pharmacophoric surface of GK activators employing a carefully chosen training subset from the collected molecules, *i.e.*, of maximum 3D diversity and continuous bioactivity spread over 3.5 logarithmic cycles (Table 2, see Section 2.1.3). We used HYPOGEN to explore and identify as many possible binding modes (pharmacophores) assumed by these GK activators into GK binding site. The reader is advised to see Section 2.1.3 and section SM-1 under Supplementary Materials to fully understand how HYPOGEN generates and assesses pharmacophore binding models [16].

We limited HYPOGEN to evaluate pharmacophoric hypotheses incorporating from 0 to 3 binding features of any particular type (*i.e.*, hydrophobic, HBD, ring aromatic, HBA and negative ionizable), *i.e.*, instead of the default range of 0–5. Furthermore, we limited the investigated pharmacophoric space to 4 and 5-featured models only (Table 3). We believe 3- and 2-featured binding hypotheses are promiscuous as three-dimensional search queries and probably insufficient descriptions of ligand-GK binding as judged from the structural diversity of the training compounds.

The resulting pharmacophore models from each run were automatically ranked based on their “total cost”. Total cost is the sum of weight cost, configuration cost and error cost (see Section 2.1.4 and section SM-2 under Supplementary Materials). Error cost comprises most of the total cost and it is related to the 3D-QSAR capacity of the pharmacophore, *i.e.*, its ability to correlate the chemical structures to the respective biological responses. HYPOGEN determines also the cost of the null hypothesis, which assumes no structure–activity relationship in the data and that the observed bioactivities assume normal Gaussian distribution about their mean. The greater the separation from the null hypothesis cost (residual cost, Table 4), the more probably that the corresponding pharmacophore hypothesis reflects a real correlation (*i.e.*, not by chance) [17].

We further validated the resulting pharmacophores using Fisher's test [18]: The chemical structures and their corresponding biological data are randomized several times, and CATALYST is challenged to produce pharmacophoric models from the scrambled data. Confidence in the original pharmacophores (*i.e.*, produced from unscrambled SAR data) is reduced proportional to the number of trials CATALYST succeeds in producing pharmacophoric models from scrambled spreadsheets of superior total cost compared to the parent hypotheses (Section 2.1.4 provides more details).

Ultimately, 40 pharmacophore models resulted from 4 HYPOGEN runs, all illustrated Fisher confidence levels above 95%. The models were clustered into 7 subgroups and their finest representatives were used in QSAR analysis, as in Table 4.

3.2. QSAR modeling

The extrapolative value of pharmacophore models, *i.e.*, as 3D-QSARs, is usually limited by bioactivity-modifying electron-donating and electron-withdrawing groups as well as steric shielding [21]. This problem combined with the fact that

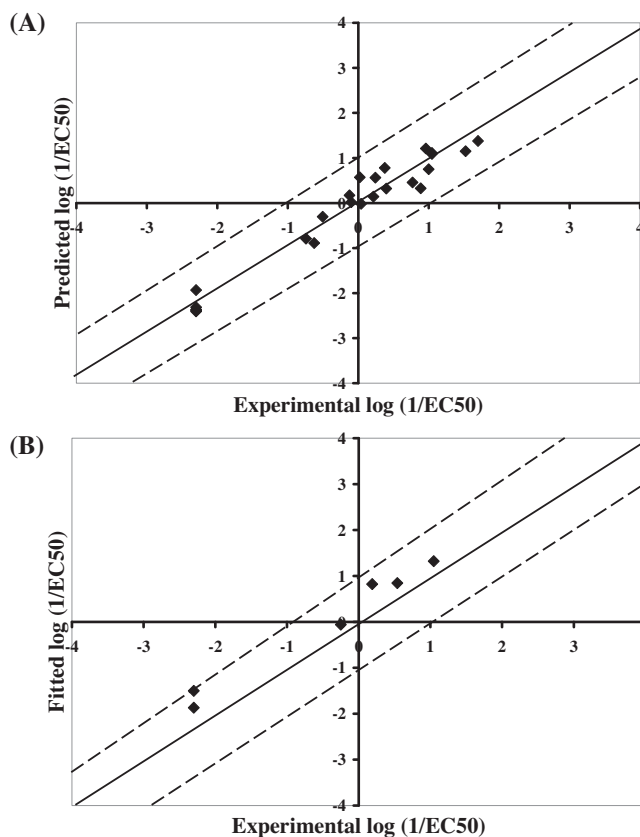


Fig. 2. Experimental versus fitted (A, 24 compounds, $r^2_{\text{LOO}} = 0.926$) and predicted (B, 6 compounds, $r^2_{\text{PRESS}} = 0.873$) bioactivities calculated from the best QSAR model (Eq. (1)). The solid lines are the regression lines for the fitted and predicted bioactivities of training and test compounds, respectively, whereas the dotted lines indicate the 1.0 log point error margins.

cluster representatives (*i.e.*, best pharmacophores produced during HYPOGEN exploration) had similar statistical properties (Table 4) impelled us to use Genetic algorithm/multiple linear regression-based QSAR (GFA-MLR-QSAR) modeling to select the most optimal combination of pharmacophore models and two-dimensional descriptors that can explain variation in GK bioactivation among the gathered activators (1–30, Table 1) [13,14,16,35].

The fit values obtained by fitting the representative pharmacophores (Table 4) against the entire list of collected GK activators (Table 1) were enrolled as descriptors in GFA-MLR-QSAR (see Section 2.1.6) [24c]. We randomly selected 6 molecules (labeled in Table 1) and used them as external testing molecules for validating the QSAR models (determination of r^2_{PRESS} , see Section 2.1.6). Moreover, QSAR equations were automatically validated using the leave-one-out technique in CERIU2 [22,24].

Eq. (1) represents the best QSAR we could achieve. Fig. 2 shows the related scatter plots of experimental versus estimated bioactivities for training and testing activators, respectively.

$$\text{Log} \left(\frac{1}{\text{EC}_{50}} \right) = -0.88 + 0.38 (\text{Hypo2/4}) + 1.12 (\text{AtypeS107}) \\ - 0.62 (\text{SssNH}) + 0.073 (\text{ShadowYZ}) - 0.29^0 \chi^v$$

$$r^2 = 0.96, \quad n = 24, \quad F = 75.8, \quad r^2_{\text{BS}} = 0.96, \quad r^2_{\text{LOO}} = 0.93, \\ r^2_{\text{PRESS}} = 0.87 \quad (1)$$

where r^2 is the correlation coefficient, F is Fisher statistical parameter, r^2_{BS} is the bootstrapping regression coefficient, r^2_{LOO} is the

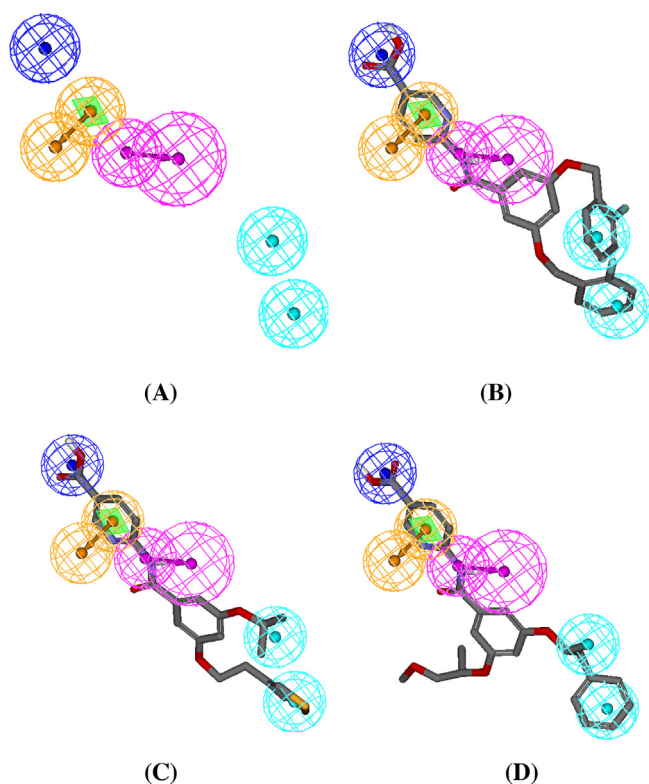


Fig. 3. (A) QSAR-selected Hypo2/4 fitted against training compounds: **13** ($EC_{50} = 0.09 \mu M$), **20** ($EC_{50} = 0.09 \mu M$) and **30** ($EC_{50} = 0.03 \mu M$) in (B), (C) and (D), respectively. Light blue spheres represent Hbic features, violet vectored spheres encode for HBDs, dark blue sphere represents Neglon feature and orange vectored spheres represent RingArom feature. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

leave-one-out correlation coefficient and r^2_{PRESS} is the predictive r^2 determined for 6 external test GK activators [36]. Hypo2/4 represents the fit values of the training compounds against the 4th pharmacophore from the 2nd modeling runs (Fig. 3, Tables 3 and 4) [24a]. AtypeS107 is part of the thermodynamic AlogP_atypes family of descriptors and it encodes for the hydrophobic contributions of sulfur atoms in LogP. SssNH is the electrotopological state index of

$\text{—}\overset{\text{H}}{\text{N}}\text{—}$ fragments. ShadowYZ is a Shadow descriptor related to the area of molecular shadow in the YZ plane calculated by aligning the molecules according to their principal moments of inertia in the X, Y and Z axes. $^0\chi^v$ is the zero order valence molecular connectivity index [13].

Emergence of one pharmacophore model in Eq. (1) suggests the existence of single mode of binding by which activators assume within GK's the binding pocket. Fig. 3 shows Hypo2/4 and how it maps active training compounds **13** ($EC_{50} = 0.09 \mu M$), **20** ($EC_{50} = 0.09 \mu M$) and **30** ($EC_{50} = 0.03 \mu M$). Table 5 shows the X, Y, Z coordinates of Hypo2/4.

Emergence of AtypeS107 in association with positive regression coefficient suggests that GK activation favors sulfur-containing hydrophobic ligands. The most probable explanation for this behavior can be either because ligand–receptor binding is mediated by critical hydrophobic interactions or because hydrophilic ligands favor hydration over binding with GK. In fact, we believe both factors contribute to this trend.

Emergence of the electrotopological state index SssNH combined with negative regression coefficient in Eq. (1) suggests certain role played by secondary N–H linker fragments in GK activation. The fact that N–H groups of training compounds (**1–30**, Table 1) are always amidic means SssNH in Eq. (1) is related to amidic linkers.

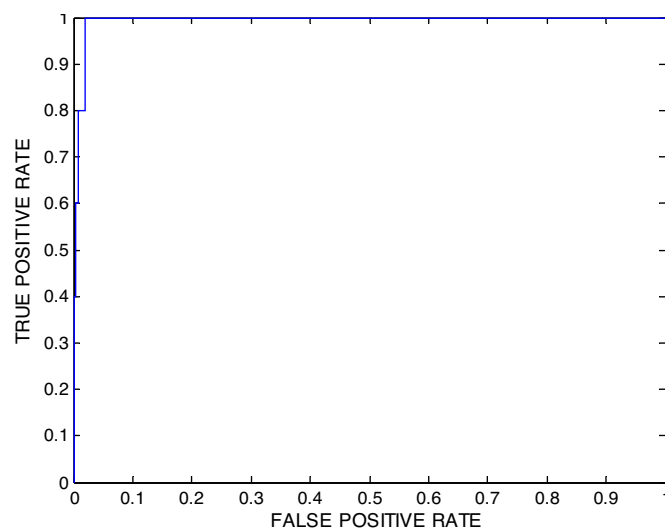


Fig. 4. Receiver operating characteristic (ROC) curves of QSAR-selected pharmacophore: Hypo2/4.

Because electrotopological state indices tend to increase in values with higher electronic densities [14] it can be concluded that the negative regression coefficient of SssNH in Eq. (1) indicates that electron-rich amidic NH fragments generally coincide with lower GK activation and vice versa. The most probable explanation for this trend is based on the hydrogen-bonding interaction that ties the amidic N–H in potent GK activators with the carbonyl of Arg63 in GK. This interaction is clearly seen in crystallographic structures of GK complexed with different potent ligands, e.g., IV4S, 3IMX, 3FRP, 3AOI, 3GOI, 3F9M, 3ID8 and 3H1V, as well as in docking experiments (see Section 3.4. Comparing pharmacophore model with docked GK–ligand complex).

The appearance of connectivity and shadow descriptors in Eq. (1) strongly suggests certain role played by ligand's topology in the binding process. However, in spite of their predictive value, the information content of such descriptors is rather vague.

3.3. Receiver operating characteristic (ROC) curve analysis

We subjected Hypo2/4 to receiver-operating characteristic (ROC) analysis to test its ability to correctly classify a list of compounds as actives or inactive (see Section 2.1.7 under and section SM-3 under Supplementary Materials for more details) [15]. Fig. 4 shows the ROC results of our QSAR-selected pharmacophore.

Clearly from the figure, Hypo2/4 illustrated exceptional overall performance. It achieved an AUC value of 99.4%, overall true positive rate of 80%, overall false negative rate of 1.5%, ACC of 98.6% and SPC of 98.5%. The excellent ROC behavior of Hypo2/4 provides additional validation for both pharmacophore and QSAR modeling.

3.4. Comparing Hypo2/4 with co-crystallized and docked GK–ligand complexes

Binding features obtained by pharmacophore/QSAR analysis can be matched up to GK binding site to identify residues important for binding and activation. Therefore, we compared the co-crystallized and docked poses of two potent GK activators with the way they fit Hypo2/4. Both compounds (compound **28** from Table 1, and ligand **OH6** from the protein databank, PDB Code: 3VF6) were fitted against our QSAR-selected pharmacophore, and the resulting fitted conformers were compared with the corresponding docked and co-crystallized poses of these compounds into GK binding site. The docking experiment was performed employing default

Table 5
Pharmacophoric features and corresponding weights, tolerances and 3D coordinates of Hypo2/4.

Definition	Chemical features							
	HBD		Hbic	Hbic	Neglonizable	Ring aromatic		
Weights		2.23	2.23	2.23	2.23		2.23	
Tolerances	1.60	2.20	1.60	1.60	1.60	1.60		1.60
Coordinates								
X	1.30	3.13	2.51	4.06	1.27	1.15		−0.89
Y	−4.06	−2.86	6.72	2.88	−9.82	−6.47		−6.70
Z	−0.23	−2.32	0.18	−2.16	1.44	0.40		−1.78

docking parameters of LIGANDFIT docking engine [16]. Fig. 5 shows both co-crystallized and docked poses and how they map Hypo2/4.

In the co-crystallized and docked poses (Fig. 5A and D), the carboxylate groups of the ligands were placed at close proximity to the guanidino of Arg63 (ca. 4.6 Å) suggesting mutual electrostatic attraction. This interaction corresponds to a Neglon feature in Hypo2/4 mapping the carboxylate group of both compounds (Fig. 5B and E). Similarly, both docked and co-crystallized poses suggest that the amide NH in **28** and **OH6** interact with carbonyl oxygen of Arg63 (Fig. 5A and D), which seems to agree with mapping the amide groups with HBD feature in Hypo2/4 (Fig. 5B and E). QSAR analysis clearly indicated the significance of this interaction (see Section 3.3). Similarly, projecting the pyridine rings of **28** and **OH6** perpendicular to the amidic linkage connecting Val455 with Ala456 suggests π -stacking interaction between the electron deficient pyridine and the electronically resonating amide. This interaction corresponds nicely to mapping the pyridine ring, in both compounds, against RingArom features in Hypo2/4. Finally, mapping the terminal phenylethyl moiety of **28** with two Hbic features in Hypo2/4 corresponds to hydrophobic interactions connecting this fragment with the hydrophobic side chains of Leu451

and Tyr215 (Fig. 5D and E). However, the co-crystallized pose of **OH6** misses one of these two Hbic features (i.e., Tyr215, Fig. 5A and B), which probably explains the weaker bioactivity of **OH6** (EC_{50} = 0.092 μ M) compared to compound **28** (EC_{50} = 0.02 μ M).

3.5. In-silico screening of databases for new GK activators followed by in vitro validation of captured hits

Based on its excellent ROC profile and excellent match with experimental crystallographic structure of GK co-crystallized with a GKA, we used Hypo2/4 to screen the national cancer institute (NCI) list of compounds (238,819 molecules) [17]. Hit compounds have their chemical moieties spatially overlap matching features in the pharmacophoric query. NCI hits were subsequently filtered based on Lipinski's [32b] and Veber's rules [32c], however, we allowed a single violation in Lipinski's filter.

Surviving hits were fitted against Hypo2/4 (fit values determined by Eq. D under Supplementary Materials) and their fit values, together with other 2D descriptors were substituted in QSAR Eq. (1) to determine their predicted bioactivities. However, to reduce the impact of any QSAR prediction errors on decisions regarding

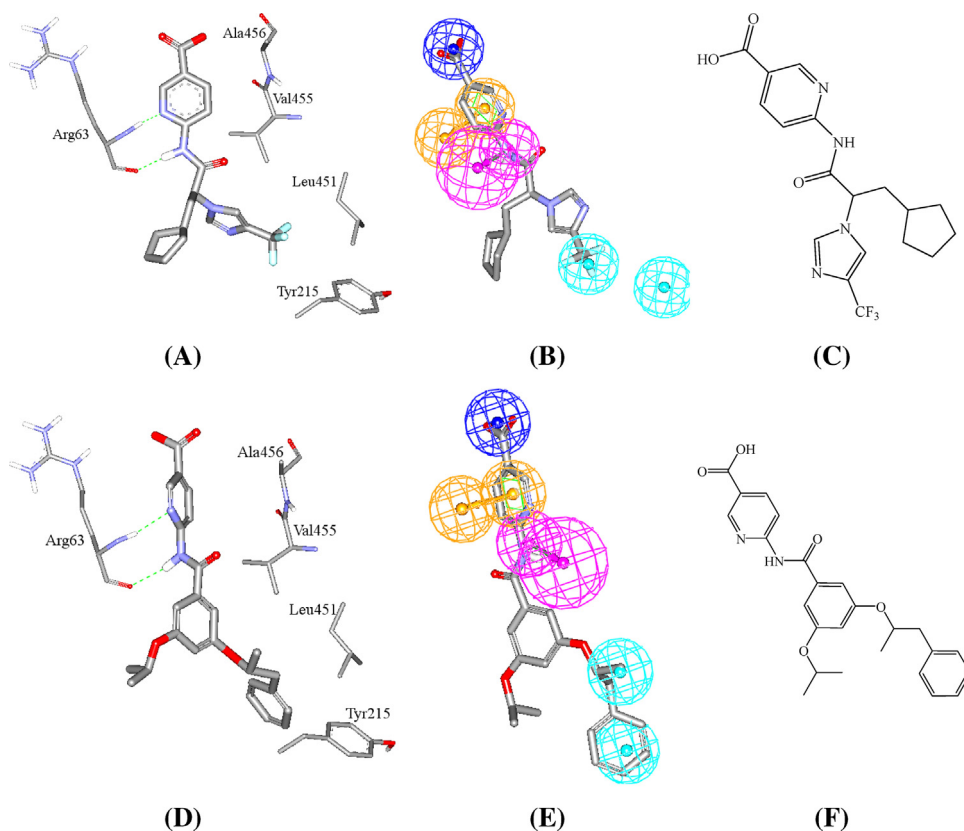
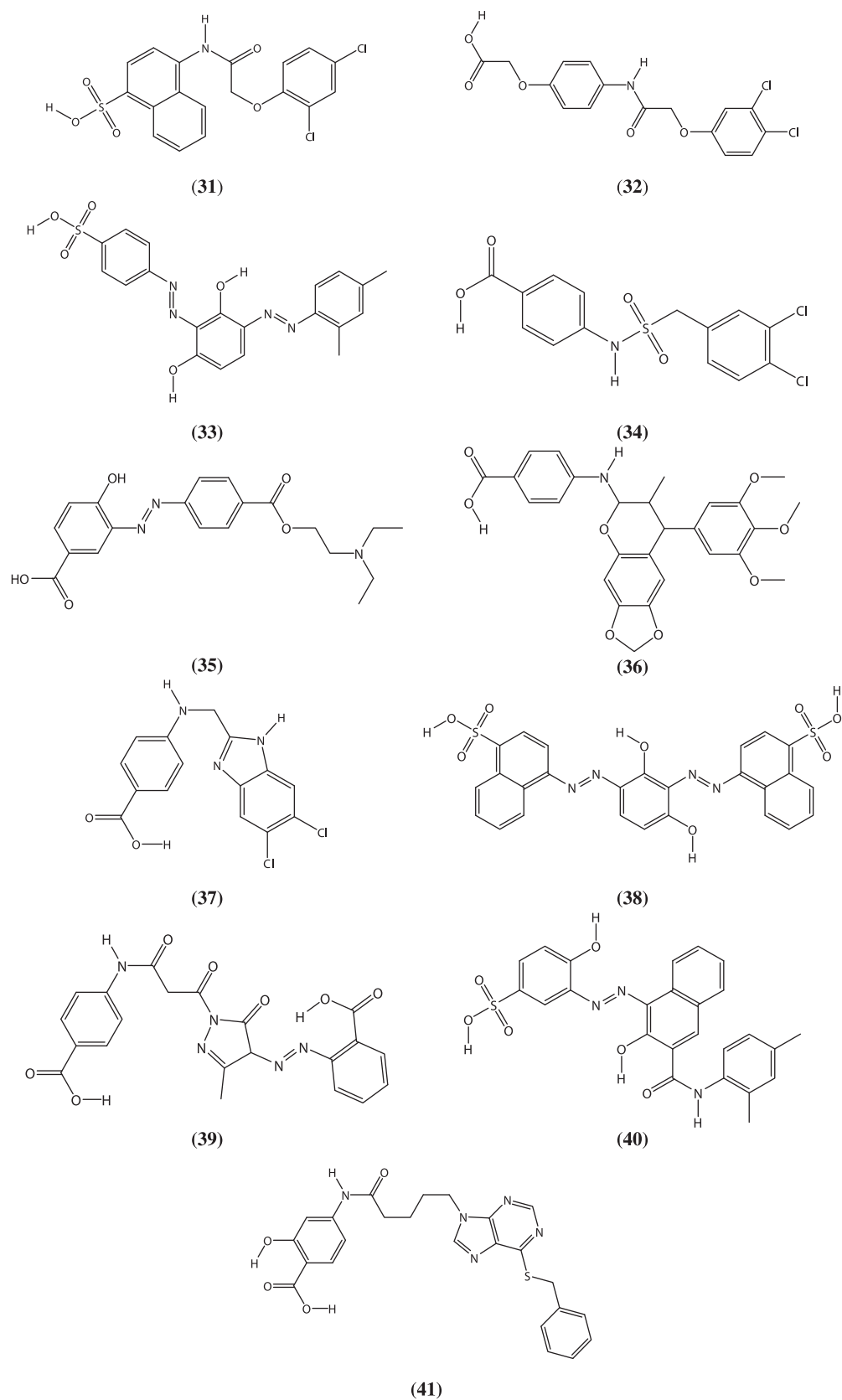


Fig. 5. (A) Co-crystallized pose of ligand **OH6** (EC_{50} = 0.092) [37] in the allosteric binding site of GK (PDB Code: 3VF6), (B) **OH6** mapped against Hypo2/4, (C) Chemical structure of **OH6**. (D) Docked pose of **28** (EC_{50} = 0.02 μ M, Table 1) in the allosteric binding site of GK (PDB Code: 1V4S), (E) **28** mapped against Hypo2/4, (F) chemical structure of **28**.

**Fig. 6.** Structures of the NCI hits captured by Hypo2/4.

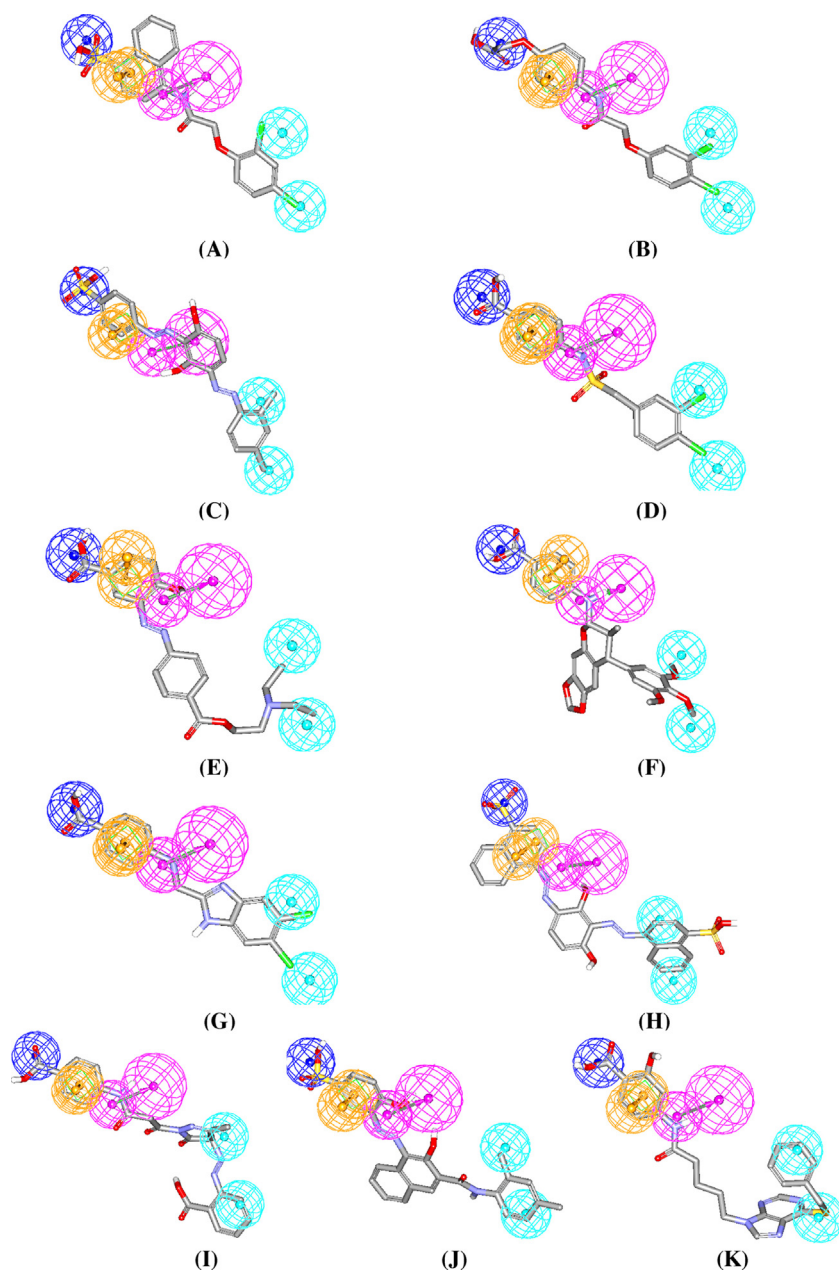


Fig. 7. Hypo2/4 fitted against captured active hits: (A) **31**, (B) **32**, (C) **33**, (D) **34**, (E) **35**, (F) **36**, (G) **37**, (H) **38**, (I) **39**, (J) **40**, and (K) **41**.

which hits merit subsequent *in vitro* testing, we only employed $\text{Log}(1/\text{EC}_{50})$ predictions to prioritize captured hits for *in vitro* testing [13–21]. The highest ranking hits were acquired for experimental validation.

We implemented a bioassay procedure that assesses GK activation upon binding to tested hits. The bioassay was performed via cascade enzymatic reactions that start with GK catalyzing the conversion of glucose into glucose-6-phosphate, which in turn acts as substrate for glucose-6-phosphate dehydrogenase (G6PD) to convert NADP to NADPH [27]. The later reaction causes measurable change in absorbance at λ 340 nm. The rate of enzyme reaction was considered as the reference for activation process, *i.e.*, activation of human GK was calculated as percent activity of the unactivated enzyme control. Incidentally, we used glucose concentration close to the $\text{S}_{0.5}$ value (8 mM) [4] to avoid masking the activating effects by saturating glucose concentrations.

Table 6 and Fig. 6 show the best predicted hits, as well as their experimental *in vitro* GK activations, while Fig. 7 shows how Hypo2/4 maps all captured active hits. Unsurprisingly, most tested hits illustrated significant GK activation with compounds **35** showing the highest activation of ≈ 6 folds at 10 μM .

However, because of the computational nature of the project we only measured the enzymatic bioactivation at a single ligand concentration, namely, 10 μM . Needless to say that higher bioactivation at 10 μM indicates more potent EC_{50} values (*i.e.*, determined by measuring bioactivation at several hit concentrations).

Interestingly, our captured hits exhibited significantly different scaffolds compared to the training compounds (see Table 1) and other known glucokinase activators (*e.g.*, see Fig. 1). For instance, published glucokinase activators were reported to require amidic linker cores to achieve potent bioactivities. Similarly, training compounds (published) were based on chemical scaffold comprised

Table 6

The captured hit molecules with their fit values, and corresponding *in vitro* bioactivities expressed as number of folds increase in enzymatic activity at 10 μ M.

Hits ^a	Name ^b	Best fit values against Hypo2/4 ^c	Experimental Activation Folds at 10 μ M
31	NSC190542	6.93	3.1
32	NSC211904	7.02	2.1
33	NSC10465	8.16	5.0 ^d
34	NSC18785	6.73	5.5
35	NSC32458	8.53	6.3 ^d
36	NSC675367	10.12	3.0
37	NSC60044	7.53	5.7 ^d
38	NSC75909	7.18	4.7 ^d
39	NSC337837	10.01	4.6
40	NSC97318	7.80	0.2
41	NSC159706	10.07	2.9

^a Chemical structures shown in Fig. 6.

^b NCI numbers.

^c Fit values calculated against Hypo2/4 using Eq. (D) under Supplementary Materials.

^d These values represent average results of duplicate measurements.

of two aromatic rings connected via amidic. In contrast, our hits showed diverse core linkers ranging from amides (hits **31**, **32**, **39** and **41**) to sulfonamides (hit **34**), diazo linkers (hits **33**, **35**, **38** and **40**) or hemi-aminal core (hit **36**). In one case, the captured hit exhibits imidazole linker moiety instead of the amidic core (i.e., hit **37**). Furthermore, most reported glucokinase activators, including our collected list of training compounds, exhibit terminal carboxylic acid moieties (however, Ro-28-0450 is a significant exception, Fig. 1). This feature was replaced with terminal sulfonic acid substituent in four of our active hits (compounds **31**, **33**, **38** and **40**).

Intriguingly, one of the tested hits inhibited GK (i.e., hit **40** in Table 6 and Fig. 6), which is not surprising as although we performed pharmacophore/QSAR modeling employing bioactivation EC₅₀ values, the fact that ligand-based bioactivation requires initial binding means each EC₅₀ value includes implicit affinity component. Therefore, the models (pharmacophores and QSAR) should, at least partially, reflect ligand-GK affinity explaining the emergence of inhibitory ligands among captured hits. Inhibitors have affinity to bind to GK, however, they seem to elicit non-specific conformational modifications in the enzyme leading to inhibition.

4. Conclusion

This work includes elaborate pharmacophore exploration of GK activators utilizing CATALYST-HYPOGEN. QSAR analysis was employed to select the best combination of molecular descriptors and pharmacophore models capable of explaining bioactivity variation across an informative list of training compounds. Single novel GK activation pharmacophore model appeared in the optimal QSAR equation. The resulting pharmacophore model yielded excellent ROC curve upon validation and was therefore used as 3D search query to screen the NCI for new GK activators. From the highest-ranking hits, five were found to be promising GK bioactivator leads for subsequent optimization.

Conflict of interest

The authors have no conflict of interest with any person or body.

Acknowledgments

The authors thank the Deanship of Scientific Research and Hamdi-Mango Center for Scientific Research at the University of Jordan for their generous funds.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jmglm.2014.12.003>.

References

- [1] K.J. Brocklehurst, V.A. Payne, R.A. Davies, D. Carroll, H.L. Vertigan, H.J. Wightman, S. Aiston, I.D. Waddell, B. Leighton, M.P. Coghlan, L. Agius, Stimulation of hepatocyte glucose metabolism by novel small molecule glucokinase activators, *Diabetes* 53 (3) (2004) 535–541.
- [2] J. Grimsby, R. Sarabu, W.L. Corbett, N.E. Haynes, F.T. Bizzarro, J.W. Coffey, K.R. Guertin, D.W. Hilliard, R.F. Kester, P.E. Mahaney, L. Marcus, L. Qi, C.L. Spence, J. Teng, M.A. Magnuson, C.A. Chu, M.T. Dvorozniak, F.M. Matschinsky, J.F. Grippo, Allosteric activators of glucokinase: potential role in diabetes therapy, *Science* 301 (5631) (2003) 370–373.
- [3] M. Larion, R.K. Salinas, L. Bruschweiler-Li, B.G. Miller, R. Bruschweiler, Order-disorder transitions govern kinetic cooperativity and allosterism of monomeric human glucokinase, *PLoS Biol.* 10 (12) (2012) 1001452.
- [4] S. Liu, M.J. Ammirati, X. Song, J.D. Knafels, J. Zhang, S.E. Greasley, J.A. Pfefferkorn, X. Qiu, Insights into mechanism of glucokinase activation: observation of multiple distinct protein conformations, *J. Biol. Chem.* 287 (2012) 13598–13610.
- [5] R. Sarabu, R. Taub, J. Grimsby, Glucokinase activation: a strategy for T2D therapy: recent developments, *Drug Discov. Today: Ther. Strateg.* 4 (2) (2007) 111–115.
- [6] S. Heuser, D.G. Barrett, M. Berg, B. Bonnier, A. Kahl, M.L. De La Puente, N. Oram, R. Riedl, U. Roettig, G.S. Gil, E. Seger, D.J. Steggles, J. Wanner, A.G. Weichert, Synthesis of novel cyclopropyl sulfones and sulfonamides acting as glucokinase activators, *Tetrahedron Lett.* 47 (16) (2006) 2675–2678.
- [7] M. Pal, Recent advances in glucokinase activators for the treatment of type 2 diabetes, *Drug Discov. Today* 14 (15–16) (2009) 784–792.
- [8] (a) M. Ishikawa, K. Nonoshita, Y. Ogino, Y. Nagae, D. Tsukahara, H. Hosaka, H. Maruki, S. Ohyama, R. Yoshimoto, K. Sasaki, Y. Nagata, Y. Eiki, T. Nishimura, Discovery of novel 2-(pyridine-2-yl)-1H-benzimidazole derivatives as potent glucokinase activators, *Bioorg. Med. Chem. Lett.* 19 (15) (2009) 4450–4454; (b) T. Nishimura, T. Iino, M. Mitsuya, M. Bamba, H. Watanabe, D. Tsukahara, K. Kamata, K. Sasaki, S. Ohyama, H. Hosaka, M. Futamura, Y. Nagata, Y. Eiki, Identification of novel and potent 2-amino benzamide derivatives as allosteric glucokinase activators, *Bioorg. Med. Chem. Lett.* 19 (5) (2009) 1357–1360; (c) L. Zhang, H. Li, Q. Zhu, J. Liu, L. Chen, Y. Leng, H. Jiang, H. Liu, Benzamide derivatives as dual-action hypoglycemic agents that inhibit glycogen phosphorylase and activate glucokinase, *Bioorg. Med. Chem.* 17 (20) (2009) 7301–7312.
- [9] N.R.A. Beeley, C. Sage, GPCRS: an update on structural approaches to drug discovery, *Targets* 2 (1) (2003) 19–25.
- [10] (a) G. Klebe, Virtual ligand screening: strategies, perspectives and limitations, *Drug Discov. Today* 11 (13–14) (2006) 580–594; (b) H. Steuber, M. Zentgraf, C. Gerlach, C.A. Sottriffer, A. Heine, G. Klebe, Expect the unexpected or caveat for drug designers: multiple structure determinations using aldose reductase crystals treated under varying soaking and co-crystallisation conditions, *J. Mol. Biol.* 363 (1) (2006) 174–187; (c) M.T. Stubbs, S. Reyda, F. Dullweber, M. Moller, G. Klebe, D. Dorsch, W.W. Mederski, H. Wurziger, pH-dependent binding modes observed in trypsin crystals: lessons for structure-based drug design, *Chembiochem: Eur. J. Chem. Biol.* 3 (2–3) (2002) 246–249.
- [11] M.A. DePristo, P.I. de Bakker, T.L. Blundell, Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography, *Structure* 12 (5) (2004) 831–838.
- [12] F.M. Matschinsky, Assessing the potential of glucokinase activators in diabetes therapy, *Nat. Rev. Drug Discov.* 8 (5) (2009) 399–416.
- [13] M.O. Taha, Y. Bustanji, M.A. Al-Ghoussein, M. Mohammad, H. Zalloum, I.M. Al-Masri, N. Atallah, Pharmacophore modeling, quantitative structure-activity relationship analysis, and in silico screening reveal potent glycogen synthase kinase-3beta inhibitory activities for cimetidine, hydroxychloroquine, and gemifloxacin, *J. Med. Chem.* 51 (7) (2008) 2062–2077.
- [14] M.O. Taha, L.A. Dahabiyeh, Y. Bustanji, H. Zalloum, S. Saleh, Combining ligand-based pharmacophore modeling, quantitative structure-activity relationship analysis and in silico screening for the discovery of new potent hormone sensitive lipase inhibitors, *J. Med. Chem.* 51 (20) (2008) 6478–6494.
- [15] M.O. Taha, N. Atallah, A.G. Al-Bakri, C. Paradis-Bleau, H. Zalloum, K.S. Younis, R.C. Levesque, Discovery of new MurF inhibitors via pharmacophore modeling and QSAR analysis followed by in-silico screening, *Bioorg. Med. Chem.* 16 (3) (2008) 1218–1235.
- [16] M.O. Taha, Y. Bustanji, A.G. Al-Bakri, A.M. Yousef, W.A. Zalloum, I.M. Al-Masri, N. Atallah, Discovery of new potent human protein tyrosine phosphatase inhibitors via pharmacophore and QSAR analysis followed by in silico screening, *J. Mol. Graphics Model.* 25 (6) (2007) 870–884.
- [17] A.M. Abu Hammad, M.O. Taha, Pharmacophore modeling, quantitative structure-activity relationship analysis, and shape-complemented in silico screening allow access to novel influenza neuraminidase inhibitors, *J. Chem. Inf. Model.* 49 (4) (2009) 978–996.
- [18] A. Al-Nadaf, G. Abu Sheikha, M.O. Taha, Elaborate ligand-based pharmacophore exploration and QSAR analysis guide the synthesis of novel pyridinium-based

- potent beta-secretase inhibitory leads, *Bioorg. Med. Chem.* 18 (9) (2010) 3088–3115.
- [19] M.A. Al-Sha'er, M.O. Taha, Discovery of novel CDK1 inhibitors by combining pharmacophore modeling, QSAR analysis and in silico screening followed by in vitro bioassay, *Eur. J. Med. Chem.* 45 (9) (2010) 4316–4330.
 - [20] R. Abu Khalaf, G. Abu Sheikha, Y. Bustanji, M.O. Taha, Discovery of new cholesteryl ester transfer protein inhibitors via ligand-based pharmacophore modeling and QSAR analysis followed by synthetic exploration, *Eur. J. Med. Chem.* 45 (4) (2010) 1598–1617.
 - [21] A.M. Abdula, R.A. Khalaf, M.S. Mubarak, M.O. Taha, Discovery of new beta-D-galactosidase inhibitors via pharmacophore modeling and QSAR analysis followed by in silico screening, *J. Comput. Chem.* 32 (3) (2011) 463–482.
 - [22] Catalyst, Version 4.11 (Software Package), Accelrys, Inc. (previously known as Molecular Simulations, Inc.), San Diego, 2005.
 - [23] (a) D. McKerrecher, J.V. Allen, P.W. Caulkett, C.S. Donald, M.L. Fenwick, E. Grange, K.M. Johnson, C. Johnstone, C.D. Jones, K.G. Pike, J.W. Rayner, R.P. Walker, Design of a potent, soluble glucokinase activator with excellent in vivo efficacy, *Bioorg. Med. Chem. Lett.* 16 (10) (2006) 2705–2709;
(b) D. McKerrecher, J.V. Allen, S.S. Bowker, S. Boyd, P.W.R. Caulkett, G.S. Currie, C.D. Davies, M.L. Fenwick, H. Gaskin, E. Grange, R.B. Hargreaves, B.R. Hayter, R. James, K.M. Johnson, C. Johnstone, C.D. Jones, S. Lackie, J.W. Rayner, R.P. Walker, Discovery, synthesis and biological evaluation of novel glucokinase activators, *Bioorg. Med. Chem. Lett.* 15 (2005) 2103–2106.
 - [24] (a) I.B. Bersuker, B.S.J.E. Boggs, *Pharmacophore Perception, Development, and Use in Drug Design*, International University Line, California, 2000;
(b) O. Guner, O. Clement, Y. Kurogi, *Pharmacophore modeling and three dimensional database searching for drug design using catalyst: recent advances*, *Curr. Med. Chem.* 11 (22) (2004) 2991–3005;
(c) J.S. Li, R. Hoffmann, *Pharmacophore Perception, Development, and Use in Drug Design* (Iul Biotechnology Series, 2), International University Line, La Jolla, CA, 2000;
(d) L.T. Poptodorov, T. Langer, R. Hoffmann, *Pharmacophores and Pharmacophore Searches*, Wiley-VCH, Weinheim, 2006;
(e) J. Sutter, G.O. Hoffmann, R.D.H. Li, Waldman M Effect of Variable Weights and Tolerances on Predictive Model Generation IUL Biotechnology Series, International University Line, La Jolla, CA, 2000.
 - [25] R.P. Sheridan, S.K. Kearsley, Why do we need so many chemical similarity search methods? *Drug Discov. Today* 7 (17) (2002) 903–911.
 - [26] E.M. Krovat, T. Langer, Non-peptide angiotensin II receptor antagonists: chemical feature based pharmacophore identification, *J. Med. Chem.* 46 (5) (2003) 716–726.
 - [27] CERIU2, Version 4.10 QSAR Users' Manual, Accelrys Inc., San Diego, CA, 2005, pp. 43–235.
 - [28] (a) J. Kirchmair, P. Markt, S. Distinto, G. Wolber, T. Langer, Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection – what can we learn from earlier mistakes? *J. Comput. Aided Mol. Des.* 22 (3–4) (2008) 213–228;
(b) M.L. Verdonk, V. Berdini, M.J. Hartshorn, W.T. Mooij, C.W. Murray, R.D. Taylor, P. Watson, Virtual screening using protein–ligand docking: avoiding artificial enrichment, *J. Chem. Inf. Comput. Sci.* 44 (3) (2004) 793–806.
 - [29] J.J. Irwin, B.K. Shoichet, ZINC – a free database of commercially available compounds for virtual screening, *J. Chem. Inf. Model.* 45 (1) (2005) 177–182.
 - [30] N. Triballeau, F. Acher, I. Brabet, J.P. Pin, H.O. Bertrand, Virtual screening workflow development guided by the “receiver operating characteristic” curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4, *J. Med. Chem.* 48 (7) (2005) 2534–2547.
 - [31] M. Jacobsson, P. Liden, E. Stjernschantz, H. Bostrom, U. Norinder, Improving structure-based virtual screening by multivariate analysis of scoring data, *J. Med. Chem.* 46 (26) (2003) 5781–5789.
 - [32] (a) A.P. Graves, R. Brenk, B.K. Shoichet, Decoys for docking, *J. Med. Chem.* 48 (11) (2005) 3714–3728;
(b) C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Adv. Drug Deliv. Rev.* 46 (1–3) (2001) 3–26;
(c) D.F. Veber, S.R. Johnson, H.Y. Cheng, B.R. Smith, K.W. Ward, K.D. Kopple, Molecular properties that influence the oral bioavailability of drug candidates, *J. Med. Chem.* 45 (12) (2002) 2615–2623.
 - [33] C.R. Goward, R. Hartwell, T. Atkinson, M.D. Scawen, The purification and characterization of glucokinase from the thermophile *Bacillus stearothermophilus*, *Biochem. J.* 237 (1986) 415–420.
 - [34] (a) D. Barnum, J. Greene, A. Smellie, P. Sprague, Identification of common functional configurations among molecules, *J. Chem. Inf. Comput. Sci.* 36 (3) (1996) 563–571;
(b) P.W. Sprague, R. Hoffmann, CATALYST pharmacophore models and their utility as queries for searching 3D databases, in: *Computer-Assisted Lead Finding and Optimization*, Verlag Helvetica Chimica Acta, 2007, pp. 223–240.
 - [35] (a) J. Singh, C.E. Chuaqui, P.A. Boriack-Sjodin, W.C. Lee, T. Pontz, M.J. Corbley, H.K. Cheung, R.M. Arduini, J.N. Mead, M.N. Newman, J.L. Papadatos, S. Bowes, S. Josiah, L.E. Ling, Successful shape-based virtual screening: the discovery of a potent inhibitor of the type I TGFbeta receptor kinase (TbetaRI), *Bioorg. Med. Chem. Lett.* 13 (24) (2003) 4355–4359;
(b) M.O. Taha, A.M. Qandil, D.D. Zaki, M.A. Aldamen, Ligand-based assessment of factor Xa binding site flexibility via elaborate pharmacophore exploration and genetic algorithm-based QSAR modeling, *Eur. J. Med. Chem.* 40 (7) (2005) 701–727.
 - [36] R. Fisher, *The Principle of Experimentation Illustrated by a Psycho-Physical Experiment*, 8th ed., Hafner Publishing Co., New York, 1966.
 - [37] J.A. Pfefferkorn, A. Guzman-Perez, J. Litchfield, R. Aiello, J.L. Treadway, J. Petersen, M.L. Minich, K.J. Filipinski, C.S. Jones, M. Tu, G. Aspnes, H. Risley, J. Bian, B.D. Stevens, P. Bourassa, T.D. Aquila, L. Baker, N. Barucci, A.S. Robertson, F. Bourbonais, D.R. Derksen, M. Macdougall, O. Cabrera, J. Chen, A.L. Lapworth, J.A. Landro, W.J. Zavadski, K. Atkinson, N. Haddish-Berhane, B. Tan, L. Yao, R.E. Kosa, M.V. Varma, B. Feng, D.B. Duignan, A. El-Kattan, S. Mardande, S. Liu, M. Ammirati, J. Knafels, P. Dasilva-Jardine, L. Sweet, S. Liras, T.P. Rolph, Discovery of (S)-6-(3-cyclopentyl-2-(4-(trifluoromethyl)-1Himidazol-1-yl)propanamido)nicotinic acid as a hepatoselective glucokinase activator clinical candidate for treating type 2 diabetes mellitus, *J. Med. Chem.* 55 (2012) 1318–1333.