

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Application of Computer Modeling to Drug Discovery: Case Study of PRK1 Kinase Inhibitors as Potential Drugs in Prostate Cancer Treatment

Abdulkarim Najjar, Fidele Ntie-Kang and
Wolfgang Sippl

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.68910>

Abstract

Computer modeling of natural products (NPs) and NP scaffolds is increasingly gaining importance in drug discovery, particularly in hit/lead discovery programs and at the lead optimization stage. Even though industry had lost interest in the implication of NPs in hit/lead searches, recent reports still show that computer modeling could be a useful assert for the identification of starting scaffolds from nature, which could be further exploited by synthetic modifications. In this chapter, the focus is on some useful tools for computer modeling aimed at the discovery of anticancer drugs from NP scaffolds. We also focus on some recent developments toward the identification of potential anticancer agents by the application of computer modeling. The chapter will lay emphasis on natural sources of anticancer compounds, present some useful databases and computational tools for anticancer drug discovery, and show some recent case studies of the application of computational modeling in anticancer drug discovery, as well as some success stories in virtual screening applications in anticancer drug discovery, highlighting some useful results on the application of on lead discovery (including promising NP scaffolds) against an interesting anticancer drug target, the protein kinase C-related kinase (PRK1).

Keywords: anticancer, molecular modeling, natural products, virtual screening, QSAR

1. Introduction

1.1. Cancer and natural products

Cancer is one of the most feared causes of death, as it represents several disease forms and treatments possibilities are still limited for late stages of the disease [1]. Among the known drugs for cancer treatment, camptothecin, vinblastine, vincristine, podophyllotoxin, and taxol are of natural origin [2]. Nature is known to be an immense repository of natural products (NPs), constituting the source of about half of the anticancer drugs currently in the market [3]. In spite of the drop in interest for NPs in drug discovery projects, from an industrial point of view, recent reports still show that NPs could constitute a useful asset for the identification of starting scaffolds for further discovery [4–6]. The quest for anticancer drugs of natural origin or with NP scaffolds has resulted in the development of NP databases, the most promising one being the naturally occurring plant-based anti-cancer compound activity-target database (NPACT), with ~1500 NPs, including experimentally verified *in vitro* and *in vivo* biological activities (in the form of IC_{50} s, ED_{50} s, EC_{50} s, GI_{50} s, etc.), along with physical, elemental, and topological properties of the compounds, the tested cancer types, cell lines, protein targets, commercial suppliers, and drug-likeness classification for each compound [7].

1.2. Prostate cancer

Among the many diverse cancer forms, prostate cancer is the second leading cause of cancer deaths in men worldwide [8]. Two different types of prostate cancer were identified: androgen-dependent prostate cancer and androgen-independent prostate cancer [9]. Androgen hormones (testosterone or dihydrotestosterone) are known to activate the androgen receptors located in the cell nucleus. The main function of these receptors is to modify gene expressions, thus controlling several biological activities in the cells, including cell growth and differentiation, development, and function of male reproductive and accessory sex tissues [10, 11]. It was found that the androgen receptors signaling pathway plays an important role in the progress and development of prostate cancer. In the first stage of androgen-dependent prostate cancer, the survival and growth of the cancer cells are mainly dependent on androgen hormones [11]. The initial treatment of androgen-dependent cancer, based on androgen ablation, is called hormone therapy. This procedure aims to stop the cell growth of cancer cells, which, in most cases, respond to this therapy. The recurring prostate cancer cells from hormone therapy would further not respond to androgen ablation, thus leading to the development of androgen-independent prostate cancer, which can further progress to metastasis [11]. The molecular mechanism of tumor recurrence is not completely clear. For the second stage of prostate cancer, there is no efficient therapy available.

1.3. Protein kinase C-related kinase

Protein kinase C-related kinase (PRK1, also known as PKN1) is a serine/threonine kinase known to play a role in controlling the activity of androgen receptors in prostate cancer.

It was shown that the activation of PRK1 stimulates the activity of androgen receptors and is involved in tumorigenesis [12]. In 2008, Schüle et al. showed that PRK1 phosphorylates histone H3 upon ligand-dependent recruitment to androgen receptor target genes [12]. The phosphorylation of histone H3 at threonine 11 (H3T11) increases demethylation of Lys-9 by Jumonji C (JmjC)-domain-containing protein (JMJD2C), which promotes androgen receptor-dependent gene expression and tumor cell proliferation [13]. Additionally, PRK1 may directly phosphorylate JMJD2C, thereby stimulating its activity [14]. Meanwhile, the role of PRK1 in androgen-independent prostate cancer is unknown. PRK1 can be activated by the Rho family GTPases, thus mediating several processes related to the migration and cancer cell invasion and consequently playing a major role in the formation of metastases [15, 16]. Thus, PRK1 is considered to be a promising therapeutic target, and the discovery of novel potent and selective inhibitors could supply a meaningful tool for the treatment of prostate cancer. On the other hand, the discovery of selective and potent inhibitors would provide a tool for understanding particular biological roles of PRK1. In spite of the importance of PRK1 in the targeted therapy of cancer, only a few known inhibitors have been identified. The known PKC inhibitors (**Figures 1 and 2**) include staurosporine (PubChem CID: 44259) and its analogue (Ro-318220; PubChem CID: 5083), bisindolylmaleimide I (BIM I; PubChem CID: 2396), and lestaurtinib (PubChem CID: 126565), as well as the nonselective Akt inhibitor GSK-690693 (PubChem CID: 16725726) and Pfizer's JAK nonselective inhibitor CP-690550 (tofacitinib; PubChem CID: 9926791) [17, 18]. In a previous work, several novel PRK1 inhibitors were identified containing different scaffolds, using a homology model, with varying potencies [19]. In the present work, the focus is to search for new small molecules and natural products, which could inhibit PRK1 by using the recently published crystal structures.

1.4. Structural analysis of PRK1

PRK1 kinase belongs to the protein kinase C (PKC) superfamily and was first identified in 1994 from a human hippocampal cDNA library [20]. Three isoforms are found in mammals (called PRK1, PRK2, and PRK3). They possess different enzymatic properties and are distributed among different tissues [21]. The PRK1 structure is divided into three conserved regions:

- an N-terminal lobe (which includes a regulatory region containing three homologous stretches and is rich in charged amino acids),
- an auto-inhibitory domain called C2-like region (which is sensitive to arachidonic acids), and
- a C-terminal lobe (which contains the catalytic domains or called kinase domain).

Both lobes are connected by a hinge region. The catalytic domain is located between both terminal lobes and shows high conservation and similarity to the PKC family kinase domain [21, 22]. Moreover, PRK1 and PKC are members of AGC kinase family [23]. The characteristic feature of these kinases is a C-terminal regulatory region (C-tail). The C-tails regulate the enzymatic activity and insert conserved phenylalanine residues into the ATP-binding site [23–25].

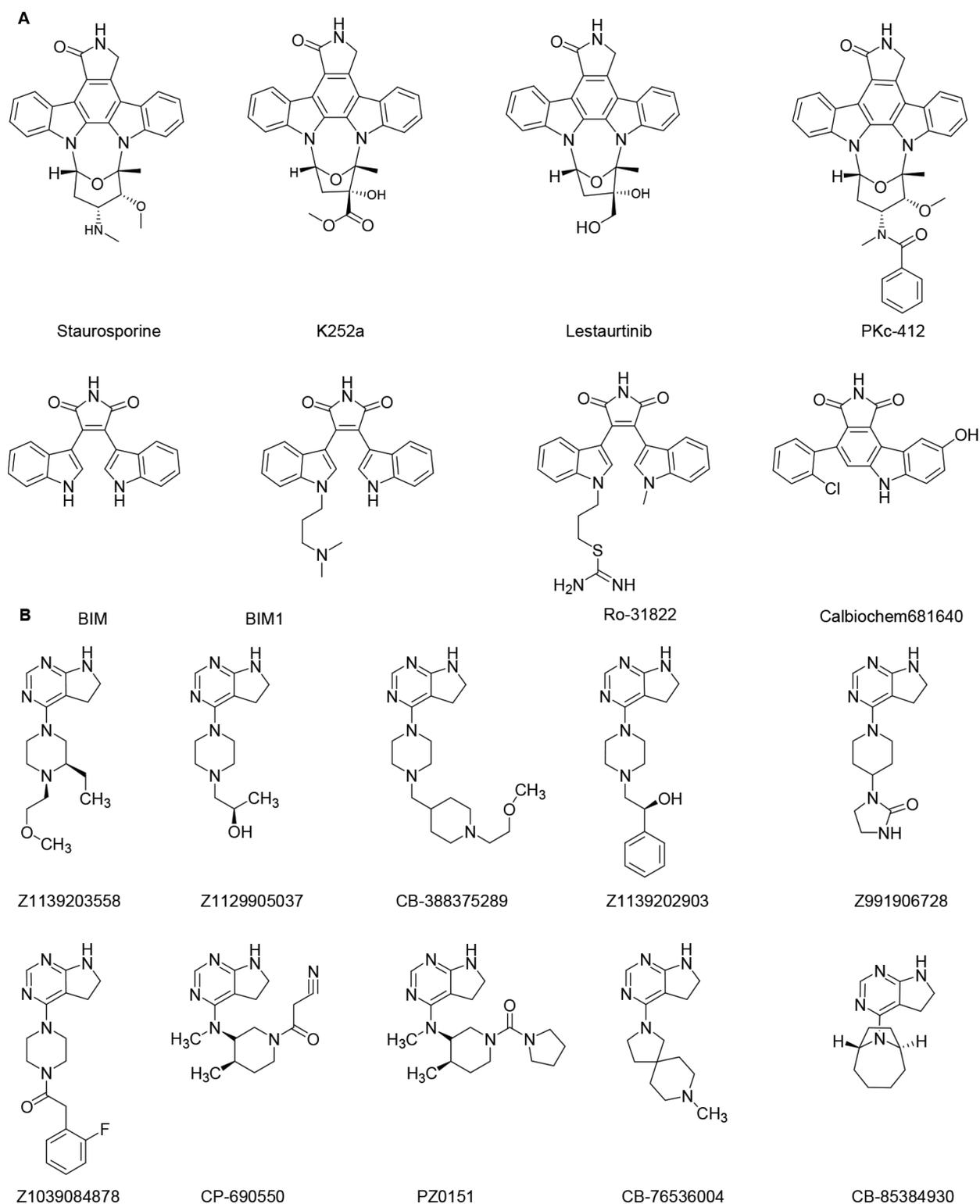


Figure 1. Chemical structures of PRK1 active compounds: (A) staurosporine and derivatives, (B) tofacitinib and other pyrrolopyrimidine derivatives [18].

1.5. Computational approaches in drug discovery

Virtual screening (VS) was first used at the end of the last century to refer to computational algorithms and techniques used to identify novel hit/lead compounds for biological target

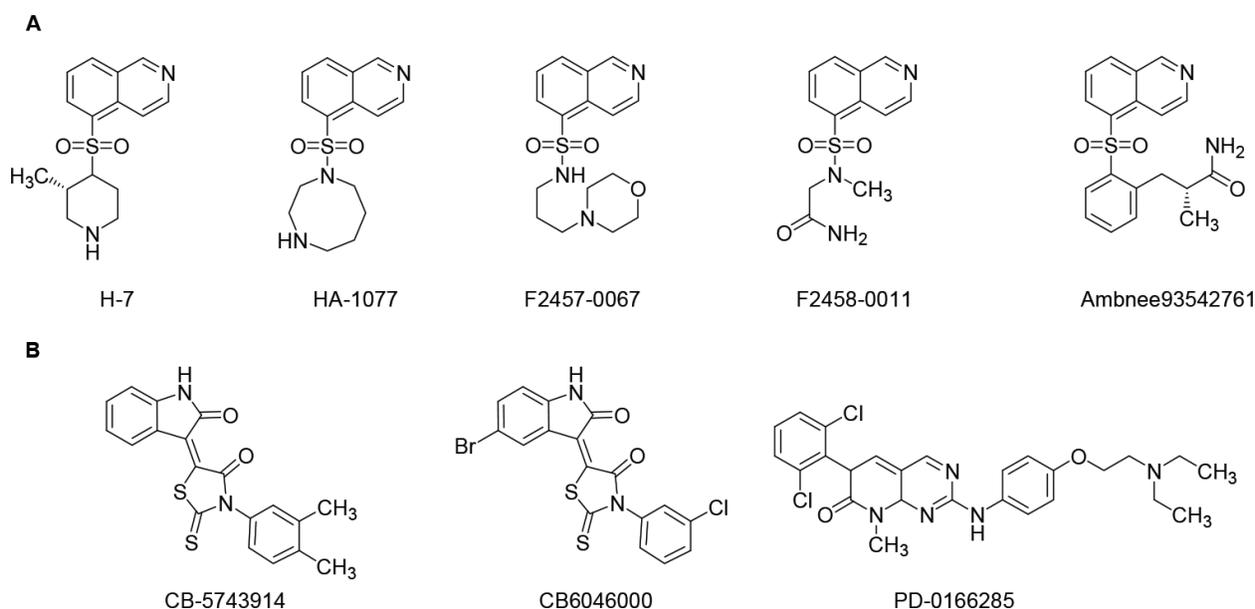


Figure 2. Chemical structures of more PRK1 active compounds. (A) Isoquinoline derivatives, (B) PD-0166285 and structurally diverse compounds identified by VS [18].

from large chemical libraries, depending on the known structure or the drug target or active ligands [26–29]. The VS approach is considered as a complementary approach to experimental or physical screening or high-throughput screening (HTS). VS has been fully integrated to most modern day drug discovery projects [30]. However, VS does not require the physical existence of the compounds to be tested. It is only based on computer models or the three-dimensional (3D) structures of the drug target, also known as structure-based VS (SBVS) or on the chemical structure of known active ligand(s), also known as ligand-based VS (LBVS) [31]. The fundamentals of the SBVS approach depend on the availability of the 3D structure of the biological target and a database of small molecules. The SBVS approach often uses molecular docking to generate the protein-ligand complexes of small molecules from a database into the target active site. The aim is to identify the compounds, which interact favorably with the target binding site [32]. Meanwhile, LBVS methods utilize chemical similarity analysis of structurally diverse or known active ligands, with the view of identifying novel small molecules, which could show similar biological activities [33–35]. However, both approaches have practical limitations. Therefore, researchers often combine SBVS and LBVS, with the aim that this might improve the efficiency of the VS results [36, 37]. Computational approaches for the prediction of biological activity also include quantitative structure-activity relationships (QSAR) analysis, which is aimed at finding a correlation between predicted and experimental biological activities of small molecules. This approach incorporates the influence of the molecular descriptors. QSAR models can be used initially to predict the activity of untested hits from a virtual screening campaign [38]. On other hand, QSAR modeling is useful tool during lead optimization, which is aimed at the improvement of the biological activities of the identified hits [39]. Recently, many crystal structures of the PRK1 drug target have been published [24]. Due to the flexibility of the binding site of the PRK1 structure, which adapts to different conformations, each protein-ligand complex was individually analyzed to reveal the important intermolecular interactions responsible for ligand binding that are useful for the

identification new hits. The aim of this research project is to analyze ligand binding to PRK1 and to identify novel-specific inhibitors that could block the activity of PRK1, using structure-based VS. In a preliminary study, the applied docking methods were validated and used to investigate its ability to predict the experimental conformations of the co-crystallized ligands. Second, several QSAR models were generated in order to find the significant correlations between the computational and experimental biological activity data. The constructed QSAR models contain different scoring functions, including computed binding-free energy (BFE) values for the protein-ligand complexes, in addition to molecular descriptors of the isolated ligands. To investigate the predictive ability of the generated QSAR models, internal and external methods were performed. Moreover, an enrichment study was performed, in order to assess the ability of the scoring functions to identify known binders or actives in large databases containing actives and inactives. By focusing on kinase data sets, GSK data sets 1 and 2 were screened to search for novel lead compounds, employing virtual screening methods. A NP library including compounds isolated from African flora was also screened. This chapter begins by presenting database tools (small molecule libraries) useful in VS campaigns, with a focus on small molecule libraries and NPs with anticancer properties. This is followed by some recent success stories on VS for the identification of inhibitors and/or modulators of some anticancer drug targets. The end of the chapter shows the case study of VS for the identification of PRK1 inhibitors.

2. Databases of small molecule libraries and some recent success stories in anticancer drug discovery using computer-based methods

2.1. Small molecule databases for virtual screening

A summary of small molecule libraries utilizable in virtual screening experiments has been provided in **Table 1**. Known cancer drugs have been included in several databases, including ChEBI [43], ChEMBL [44], DrugBank [46], EpiDBase [48], NANPDB [51], NCI-DIS, NCI-DTP, and NCI-FDA [52], along with SANCDB [54], SuperDrug [55], p-ANAPL [58], PubChem [59], and ZINC15 [60]. However, those with tested and proven *in vitro* activities against known cancer cell lines and/or with known *in vivo* activities only include CancerDR [41], OAA [53], and SYFPEITHI [56], along with the NP libraries: AfroCancer [40], CancerHSP [42], CHMIS-C [43], InPACdb [49], MAPS [50], NPACT [7], and TIPdb [57].

2.2. Some recent success stories

Computational modeling has been applied to understand drug-target interactions in several validated anticancer drug targets, for the identification of novel inhibitors from small molecule databases and/or for the elucidation of modes of action. We here present some few recent cases. A typical example is the recent discovery of new inhibitors of CXC chemokine receptor 2 (CXCR2) by applying ligand-based pharmacophore models [61]. It should be mentioned that CXCR2 and its ligand, CXCL8, are known to be implicated in a number of inflammation-mediated diseases, including cancer [62–65]. In the study, Ha et al. generated a pharmacophore

Database	Description	Web accessibility	Reference
AfroCancer	A data set of natural anticancer products from African flora	http://www.african-compounds.org/about/afrocancer/	[40]
Cancer drug resistance database (CancerDR)	A database of 148 anticancer drugs and their effectiveness against around 1000 cancer cell lines	http://crdd.osdd.net/raghava/cancerdr/#thumb	[41]
CancerHSP	An anticancer herbs database of systems pharmacology	http://lsp.nwsuaf.edu.cn/CancerHSP.php/	[42]
ChEBI	A database for chemical entities of biological interest	http://www.ebi.ac.uk/chebi/	[43]
ChEMBL	An open large-scale bioactivity compound database	https://www.ebi.ac.uk/chembl/	[44]
CHMIS-C	A herbal medicines database for cancer	http://sw16.im.med.umich.edu/chmis-c/	[45]
DrugBank	A resource that combines detailed drug data with comprehensive drug target information	https://www.drugbank.ca/	[46]
DUDE datasets	Benchmarking data sets of actives and decoys for diverse targets, including proteins, GPCRs and ion channels, clustered ligands, etc. drawn from ChEMBL	http://dude.docking.org/	[47]
EpiDBase	A database for small molecule epigenetic modulators	http://www.epidbase.org/	[48]
InPACdb	Indian plant anticancer compounds database	http://www.inpacdb.org/	[49]
MAPS Database	A database of phytochemicals, including the data of >500 medicinal plants	http://www.mapsdatabase.com/	[50]
NANPDB	A natural products database for compounds of Northern African origin, with a significant number of bioactive metabolites exhibiting anticancer activity	http://www.african-compounds.org/nanpdb/	[51]
NCI Drug Information System (DIS)	A searchable database of 3D structures (mainly organic compounds) from the NCI Drug Information System (DIS)	https://cactus.nci.nih.gov/ncidb2.2/ http://dtp.nci.nih.gov/docs/3d_database/dis3d.html	[52]
NCI-DTP database	Tested compounds from the National Cancer Institute (NCI) Developmental Therapeutics Program (DTP)	https://dtp.cancer.gov/databases_tools/data_search.htm	[52]
NCI-FDA	Several sets of FDA-approved anticancer drugs to enable cancer research	https://wiki.nci.nih.gov/display/NCIDTPdata/Compound+Sets	[52]
NPACT	A database for plant-based anticancer compounds	http://crdd.osdd.net/raghava/npact/	[7]
Oral anticancer agents (OAA) from Singapore	A database of 39,772 oral anticancer agents prescribed to 8837 patients in Singapore, with 55 clinically significant drug-drug interactions for the evaluation of drug interaction facts	https://www.ncbi.nlm.nih.gov/pubmed/22795926/	[53]

Database	Description	Web accessibility	Reference
SANCDDB	A database of natural products from South Africa, containing also anticancer agents from the flora and fauna of this ecologically diverse country	https://sancdb.rubi.ru.ac.za/	[54]
SuperDrug	A database of 3D-structures of active ingredients of essential marketed drugs	http://bioinf.charite.de/superdrug/	[55]
SYFPEITHI	A database of major histocompatibility complex (MHC) ligands and peptide motifs	http://www.syfpeithi.de/	[56]
TIPdb	A database of anticancer, antiplatelet, and antituberculosis phytochemicals from indigenous plants in Taiwan	http://cwtung.kmu.edu.tw/tipdb/	[57]
p-ANAPL	A collection of samples of natural products from African sources	http://www.african-compounds.org/about/p-anapl/	[58]
PubChem	A public repository for information on chemical substances and their biological activities	https://pubchem.ncbi.nlm.nih.gov/	[59]
ZINC15	A free database of over 100 million purchasable compounds for virtual screening	http://zinc15.docking.org/	[60]

Table 1. Summary of some small molecule libraries currently available for anticancer drug discovery.

model based on known CXCR2 antagonists and used it to screen a database of 5 million commercially available compounds from different vendors. The authors were able to identify small molecule hits, which were further tested by *in vitro* screening in a cell-based CXCR2-mediated β -arrestin-2 recruitment assay, followed by several other cell-based assays. *In vivo* studies were conducted by lipopolysaccharide (LPS)-induced lung inflammation in mice. It was also shown that one of the compounds inhibits CXCR2 signaling through down regulation of surface CXCR2.

Moreover, the same compound was shown to inhibit CXCL8-mediated neutrophil migration and LPS-induced lung inflammation in mice significantly. The identified compounds were shown to also inhibit CXCR2/ β -arrestin-2 association, cell migration and proliferation, and acute inflammation in mouse models. Another study combined structure-based docking and pharmacophores to design novel indole and chromene analogues, which were cyclin-dependent kinase 2 (CDK2) inhibitors. The identified compounds proved to be active against MCF-7 and HeLa cell lines. The study was conducted by exploiting stereo-specific information obtained from crystal structures of CDK2, substituting the pharmacophores on their moiety and docking the target protein to calculate the binding affinities [66]. Other recent successful cases, where docking, QSAR, machine learning, and pharmacophore-based screening were combined to search for small molecules targeting cancer, are abundant in the literature [67–70].

3. Case study: structure-based virtual screening and QSAR studies on PRK1 inhibitors

3.1. Structure-based design studies (X-ray target structures and cross docking)

Four crystal structures of PRK1 are available in the protein data bank (PDB), **Table 2**. The PRK1 binding site is reported to exhibit either intrinsic or induced flexibility [17, 24]. Moreover, different bound inhibitors are known to induce different conformational changes in the binding site residues [71]. Furthermore, chemical characteristics of the co-crystallized ligands play an important role in the applicability of the X-ray structure for virtual screening. Therefore, a cross-docking procedure was carried out, as a retrospective study aimed at determining the appropriate structure for the virtual screening. The selected structure should be able to bind the nonnative ligands with low root mean square deviation (RMSD), with respect to the reference binding conformation. On the other hand, the performance of the docking power and scoring functions is varying for different targets [32]. The docking power measures the ability of docking algorithms to predict the correct conformation of the ligands [72].

The four X-ray structures of PRK1, including the apo-form, were used for a cross-docking study [24]. The co-crystallized ligands were docked toward the target binding sites, using the Glide cross-docking script, implemented in Schrödinger Suite (2014 version) with standard precision (SP) for flexible ligand docking. In order to optimize the docking solution, an option was selected to perform post-docking minimization and includes number of poses up to 5 per ligand. The structures were prepared using the default setting implemented in Protein Preparation Wizard (Schrödinger 2014). The co-crystallized water molecules were deleted. Hydrogen atoms and partial charges were assigned. Finally, the structure energy minimized applying the OPLS 2005 force field. The binding site was defined using Grid Generation of Schrödinger suite 2014 and sets the co-crystallized ligand as a center of the binding pocket. In the case of apo-structure 4OTD, the binding site was defined by applying centroid of selected residues Leu650 and Lys753 with box size 14 Å. A hydrogen bond constraint was defined at PRK1 hinge region residue Ser704. In addition, all ligands were prepared using LigPrep implemented in Schrödinger utility 2014 involving generation of ionization and tautomeric states at pH 7.4 with less than 10 low-energy ring conformations. The ligands were energy minimized using MMFFs force field.

Ligand	4OTD		4OTG		4OTH		4OTI	
	RMSD	SP	RMSD	SP	RMSD	SP	RMSD	SP
Lestaurtinib-4OTG	6.56	-4.29	0.24	-11.67	4.65	-7.82	7.56	-4.29
Ro-318220-4OTH	6.79	-7.01	2.98	-11.4	0.64	-11.76	6.47	-6.55
Tofacitinib-4OTI	3.14	-4.54	6.11	-8.09	2.7	-8.18	0.55	-8.8
Average RMSD	5.5		3.11		2.67		4.86	

Table 2. RMSD values (Å) for the top-ranked docking solutions (using Glide SP as scoring function).

The RMSD values were calculated, with respect to the respective reference ligand conformations. **Table 2** shows the RMSD values and Glide SP scores for the top-ranked poses in the corresponding structures. It was observed that none of the available structures had a binding site conformation, which allowed all the three different ligands to be docked with RMSD < 2.0 Å. Furthermore, the calculated average RMSD values for each structure were high. However, the applied docking method performed well by reproducing the binding mode of the co-crystallized inhibitor in self-docking. Moreover, the docking method correctly scored the co-crystallized inhibitors at the top of the list (**Table 2**), e.g., the X-ray structure of the ligand lestaurtinib was re-docked with an RMSD = 0.24 Å into its co-crystallized structure (PDB: 4OTG), being also ranked as the top scoring pose (Glide SP = -11.67 kcal/mol). Similar results were observed with the other inhibitors when X-ray structures were docked toward their respective co-crystallized target structures. Meanwhile, the binding pocket conformation for the apo-form of PRK1 was not found to be suitable for docking when using any of the current inhibitor structures. RMSD values were high and the docking scores were low when compared with those obtained in the rest PRK1 structures (**Table 2**). Therefore, an ensemble of PRK1 structures was proposed for the further virtual screening study.

3.2. Docking of active inhibitors

The data set DS1 includes 28 active (**Figures 1 and 2** [18]) and 300 inactive compounds. First, the active compounds were docked toward the ensemble of the three PRK1 structures using the previously described docking methods. The docking solutions were first inspected, with the aim of comparing them with the experimental conformations (the co-crystallized ligands in other kinases). The active compounds could be divided into four subsets, the first set being staurosporine derivatives (10 actives, including the co-crystallized inhibitors lestaurtinib and Ro-318220). The second subset is made of the tofacitinib or pyrrolopyrimidine family (nine actives, in addition to the inhibitor tofacitinib in PDB ID: 4OTI). The third set was made of two isoquinoline derivatives, forming a group of five actives. The last subset contains the remaining inhibitors, with diverse scaffolds. The binding mode for each compound was analyzed and compared with the experimental data. As previously seen in the cross-docking for the co-crystallized inhibitors (**Table 2**), the binding modes for lestaurtinib and Ro-318220 were correctly reproduced. The binding modes for the further staurosporine derivatives were compared with these two analogues. Since the ensemble docking procedure was applied to dock the actives, eight compounds were docked toward the 4OTG structure target site, while the others (Ro-318220 and BIM1) were docked toward the 4OTH site. It was interesting to notice that the top-ranked active poses had a quite conserved binding mode, which shares two H bonds, interacting with the hinge region of PRK1. The obtained binding modes for all staurosporine derivatives were identical to the published structure (in the PRK1 X-ray structure 4OTG). Since a part of the C-terminal regulation region (C-tail) is not resolved, resulting in a more open and accessible ATP binding pocket, most of the staurosporine derivatives could be docked into it. Furthermore, all staurosporine derivatives were docked into both PRK1 structures, which are co-crystallized with one of staurosporine derivatives (4OTG or 4OTH). For pyrrolopyrimidine and tofacitinib derivatives, the inhibitors were docked into three X-ray structures of PRK1. Tofacitinib was co-crystallized in the structure 4OTI. The RMSD value of re-docking tofacitinib was 0.55 Å. The binding modes

for the other analogues were compared with the experimental data (4OTI). Additionally, several other analogues were co-crystallized with the Akt1 kinase (PDB IDs: 3MV5, 3MVJ, and 3OCB).

Each active compound forms 2 H bonds with the hinge region, besides additional H bonds with the surrounding residues in the binding pocket. There are also some isoquinoline derivatives that were co-crystallized with other kinases, e.g., cAMP-dependent kinase (PDB IDs: 1YDS and 1YDR) and Rho kinase (PDB IDs: 2GNI). This subset of actives interacts with the PRK1 hinge region by mediating 1 H bond. A former virtual screening campaign had identified several actives, including CB-6046000 and CB-5743914, which were docked using the same previously described docking method. The binding mode showed two H bonds between the dihydroindol-2-one and the backbone of the hinge region, mainly Ser704 and Glu702. Furthermore, the binding mode of PD-0166285 matched the experimental data observed for the LCK kinase (PDB ID: 3KMM).

3.3. Scoring power

The rescoring for the derived docking poses was performed by calculating BFEs and using different solvation models implemented in the AMBER12 package. IC_{50} values had been previously determined for 26 of the active compounds (**Table 3**), with only percentage inhibitions available for two others) [19]. Thus, the correlation coefficient (R^2) between the experimental pIC_{50} and the calculated enthalpy changes (ΔH) for protein-ligand binding was calculated (**Table 3** and **Figure 3**). The BFE calculation was carried out using one snapshot after two consecutive minimizations steps and applying the Generalized Born solvation model [73, 74]. First, the minimization was carried out for water and counterion molecules without the ligand-protein complex, which was restrained to their initial coordinates with a force constant of 500 kcal/mol/Å². In this step, there were 2000 iterations (beginning with 1000 steepest descent and followed by 1000 conjugate gradient). The second minimization step was applied for the whole system through 10,000 iterations (first 5000 steepest descent and then 5000 conjugate gradients). A significant correlation was found when using the Nguyen and Simmerling (igb = 8) version of the Generalized Born solvation model [73, 74] and applying the two minimizations steps. The cross-validated R^2 was found to be 0.60, with a root mean square error (RMSE) of 0.89. To understand the effect of the chemical modification on the main scaffold, the binding mode of the actives and the interactions in the binding pocket was analyzed. As mentioned previously, PRK1 actives can be divided into four subsets (**Table 3**). The docking score (Glide SP) and the MM/GBSA BFE values were employed to explore the differences in the inhibitory activity. However, there was a weak correlation between the pIC_{50} and Glide SP scores (R^2 of 0.43, see **Table 4**).

It was observed, from docking, that the isoquinoline derivatives are able to form an H bond between the N atom of the isoquinoline motif and the NH backbone of Ser704 located in the hinge region. The isoquinoline motif occupied the adenine binding site; meanwhile, the substituents were located in the sugar-binding site, which is surrounded by several hydrophilic residues (e.g., Asp708, Asp750). Rescoring, using MMGBSA, showed that H-7 (IC_{50} = 658 μ M and ΔH = -34.51 kcal/mol) had a higher score than HA-1077 (IC_{50} = 1.95 μ M and ΔH = -29.42 kcal/mol), **Table 3**. The more favorable value resulted from the additional

Compound	IC ₅₀ (nM)	Exp. pIC ₅₀	Target structure	Glide SP	MM/GBSA**	PEOE_PC-	Pred. pIC ₅₀ [†]
Ro-318220	78.3	7.11	4OTH	-11.36	-45.97	-2.71	7.16
BIM1	579	6.24	4OTH	-9.82	-46.10	-2.29	6.58
PKC-412	14.2	7.85	4OTG	-10.98	-55.52	-3.26	8.26
Lestaurtinib	8.6	8.07	4OTG	-11.43	-46.92	-2.77	7.30
K252a	3.2	8.49	4OTG	-11.38	-51.40	-2.95	7.80
Staurosporine	0.81	9.10	4OTG	-11.83	-60.86	-2.72	8.59
Calbiochem681640	1441	5.84	4OTG	-10.61	-47.90	-1.96	6.71
BIM	154.2	6.81	4OTG	-9.16	-38.70	-1.96	5.58
K252c			4OTG	-10.53	-40.12		
Quercetin			4OTG	-8.43	-48.23		
CP-690550 (tofacitinib)	129	6.89	4OTI	-8.81	-38.88	-2.01	5.55
Z1139203558	55730	4.25	4OTG	-8.52	-27.91	-1.84	4.41
Z1129905037	53380	4.27	4OTI	-9.61	-35.25	-1.82	5.29
CB-38374289	42560	4.37	4OTI	-10.49	-33.36	-2.11	5.49
CB-85384930	34020	4.47	4OTH	-9.04	-35.48	-1.48	4.98
Z1139202903	31470	4.50	4OTI	-9.37	-37.77	-2.09	5.62
Z991906728	26400	4.58	4OTG	-9.18	-28.92	-2.02	4.75
CB-76536004	13090	4.88	4OTG	-9.43	-33.02	-1.45	4.83
Z1039084878	6210	5.21	4OTG	-8.76	-29.13	-2.10	4.73
PZ0151	1060	5.97	4OTI	-8.54	-43.19	-2.16	5.97
H-7	658.5	6.18	4OTH	-8.68	-34.51	-1.43	4.78
HA-1077	1945	5.71	4OTI	-8.71	-29.42	-1.39	4.31
F2457-0067	70300	4.15	4OTH	-8.49	-36.31	-1.76	5.10
F2458-0011	53730	4.27	4OTI	-8.07	-29.18	-1.70	4.34
Ambnee93542761	29330	4.53	4OTH	-9.12	-30.33	-1.68	4.66
CB-6046000	5350	5.27	4OTG	-7.39	-47.01	-1.57	5.70
CB-5743914	2940	5.53	4OTG	-9.89	-41.07	-1.72	5.81
PD-0166285	5517	5.26	4OTG	-7.48	-39.17	-2.41	5.53

Comparison of experimental pIC₅₀ versus predicted pIC₅₀ values calculated by QSAR_3 for the 26 training set molecules [18].

The BFE values were calculated according to the docked structure (target structure), which are supplied with the PDB ID.

Table 3. Biological activity, calculated BFE and docking scores of the active compounds.

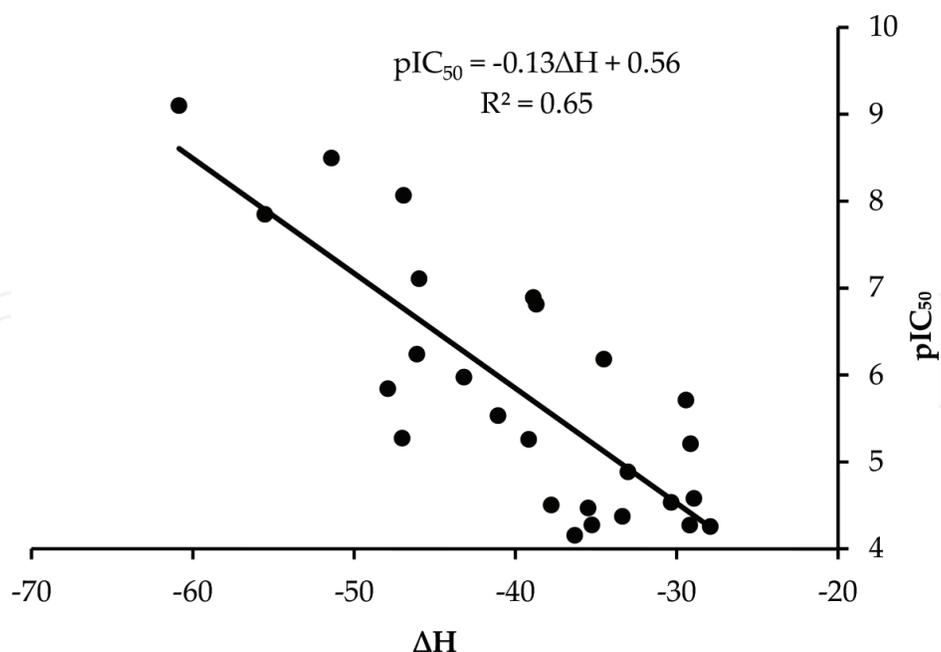


Figure 3. Correlation curve of pIC₅₀ versus ΔH scores of actives using MM/GBSA.

hydrophobic interactions between the front hydrophobic pocket residues and Phe910 (from the C-tail) and the methyl substituent on the piperazine ring. Meanwhile, the van der Waals interaction became weaker in the case of HA-1077. This could be explained by the shifting of the position of the substituent, since its seven-membered ring is bulkier. Both derivatives contain a positively charged ring nitrogen, which consequently forms an additional H bond in the ATP-sugar binding pocket (**Figure 4**). Meanwhile, the interaction between Asp708 and the piperazine ring enables the methyl group to get closer to Phe910 and its surrounding residues. Moreover, the sulfonyl moiety of both compounds is located under the P-loop, where it forms hydrophilic and hydrophobic interactions.

	GLIDE		ΔH
	SP	XP	GB8
R^2	0.43	0.27	0.65
EF1%	100	100	25
EF3%	30.3	33.33	9.1
AUC*	0.90	0.88	0.66

*AUC = area under the ROC curve.

Table 4. Enrichment study results using an ensemble of PRK1 structures, where the available scoring functions in Schrödinger and BFE methods were considered. Where **SP**: Standard Precision; **XP**: Extra Precision and **ΔH** : enthalpy score; **GB8**: Generalized Born theory model.

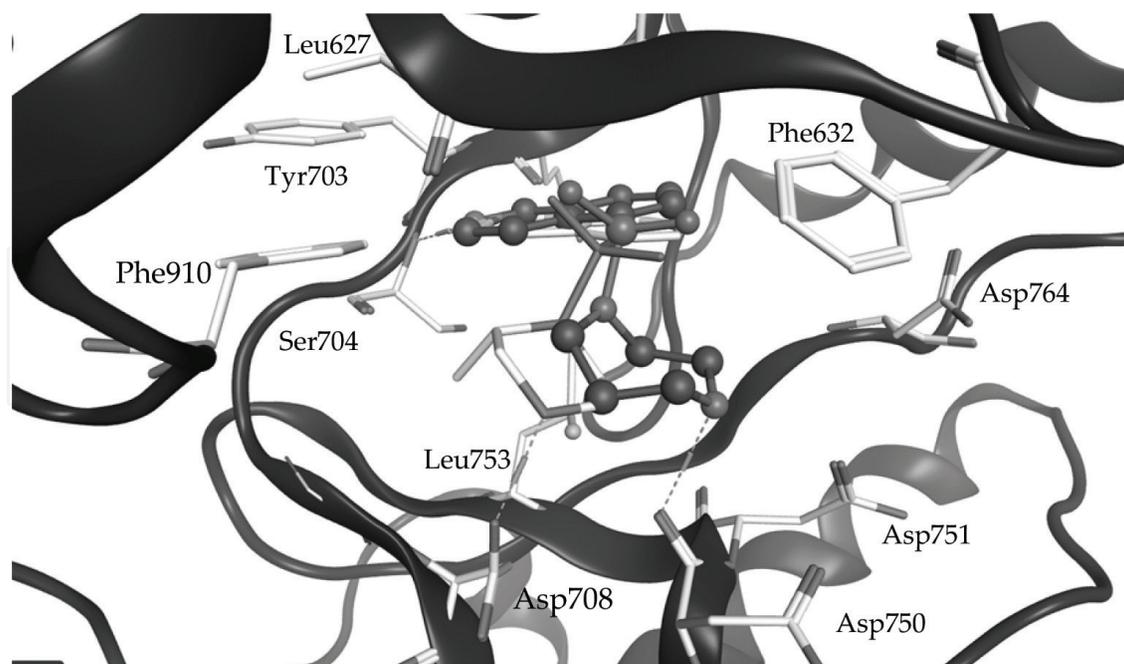


Figure 4. Comparison between the docking poses of two isoquinoline derivatives H-7 (magenta) and HA-1077 shown as green sticks (PRK1 PDB code: 4OTH). The figure was designed using MOE.

The remaining compounds share common interactions (as previously mentioned), but the biological activities of these compounds could not be predicted using the derived BFE model. However, the lower activity of compound ambnee93542761 could be explained by the loss of the interaction between of the sulfonyl moiety and weak hydrophobic interactions with Phe910. Nonetheless, it forms a hydrogen bond with the catalytic Asp764 ($IC_{50} = 29.33 \mu\text{M}$ and $\Delta H = -30.32 \text{ kcal/mol}$). The results of this subset show the importance of the hydrophobic interactions with the Phe910 of the C-tail and its surrounding residues. However, the BFE values and docking scores failed to predict the affinity of F2457-0067, which shows a favorable BFE value but a low IC_{50} value (**Table 3**). This could probably result from the long linker between the isoquinoline motif and the positively charged morpholine ring. The generated conformations show additional van der Waals interactions. On the other hand, the morpholine ring shows more polar properties compared to the other compounds.

The second subset of actives is made of staurosporine derivatives. Most of these compounds form two H bonds between the pyrrole ring and the hinge backbone (NH of Ser704 and CO of Glu702). The docking results for staurosporine derivatives were compared with two PRK1 crystal structures (4OTG and 4OTH). The favorable BFE values of staurosporine derivatives could be attributed to the strong hydrophobic interactions with residues from the P-loop, e.g., Val635 and Leu627, in addition to the interactions with Leu753. The only observed deviation was with parts of the docked ligand interacting with the sugar pocket, e.g., staurosporine (the most active compound) showed a BFE value of $\Delta H = -60.85 \text{ kcal/mol}$ and an $IC_{50} = 0.0008 \mu\text{M}$ (**Table 3**). The favorable BFE value and activity of staurosporine could be estimated through the targeting of both hydrophobic pocket residues (behind the gatekeeper Met701 and near Phe910); on the other hand, several polar groups occupied the sugar polar

pocket. By comparison with lestaurtinib ($IC_{50} = 0.0086 \mu\text{M}$ and $\Delta H = -46.92 \text{ kcal/mol}$), the differences in the activity and BFE values could be attributed to the smaller ring/shorter linker in lestaurtinib for targeting Asp708 and Asp750 (**Figure 5**).

The third subset of PRK1 inhibitors is made of tofacitinib analogues. Ten compounds belong to this subset (**Table 3**). All derivatives interact with the hinge region by forming two H bonds through the pyrrolopyrimidine scaffold (**Figure 6**). The major substituents, which influence notably the BFE values and the biological activity of tofacitinib, are the N-methyl of the pyridine ring. Any chemical modification changing these interactions will consequently change the BFE value and the activity. A comparison of the calculated BFE values of tofacitinib with those of the other derivatives could clarify the effects of these interactions. As previously mentioned, the loss of the interactions made with the P-loop residues clearly influences the calculated BFE values and, therefore, affects the biological activities. Consequently, Z1129905037 ($IC_{50} = 55.73 \mu\text{M}$ and $\Delta H = -27.92 \text{ kcal/mol}$) possesses a BFE value lower than tofacitinib ($IC_{50} = 0.129 \mu\text{M}$ and $\Delta H = -38.88 \text{ kcal/mol}$) (**Table 3**). Furthermore, contrary to tofacitinib, the binding mode of Z1129905037 does not show hydrophobic interactions with the residue located at the bottom of the binding pocket. The ATP-sugar binding site is occupied by non-polar substituents in the case of Z1129905037, which is unfavorable and consequently reduces the activity. The other tofacitinib analogues, which do not contain any tertiary amine as the linker, show higher BFE values and obviously lower biological activities. One exception is PZ0151 ($IC_{50} = 1.06 \mu\text{M}$ and $\Delta H = -43.19 \text{ kcal/mol}$) which displays the same binding mode with tofacitinib. The pyrrolidine

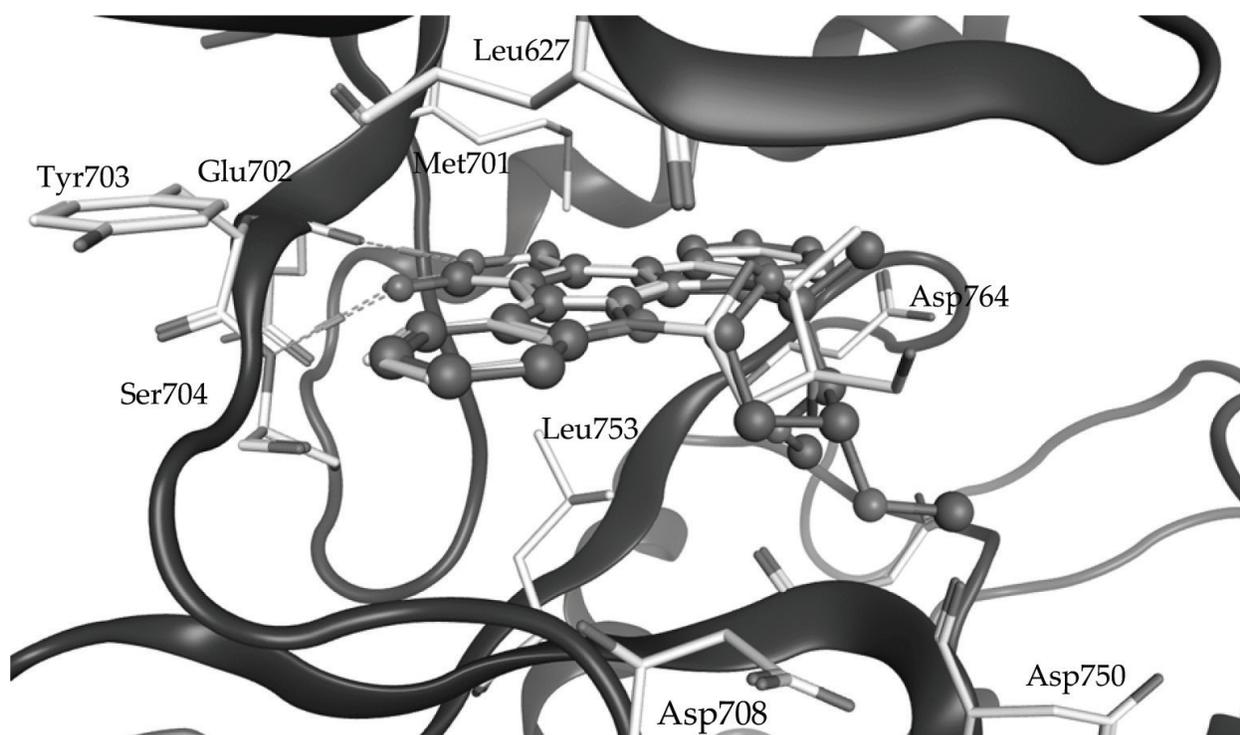


Figure 5. Comparison between the docking pose of staurosporine (seen in cyan sticks) and the binding mode of lestaurtinib (shown in white brown sticks) at PRK1 structure 4OTG. The figure was designed using MOE.

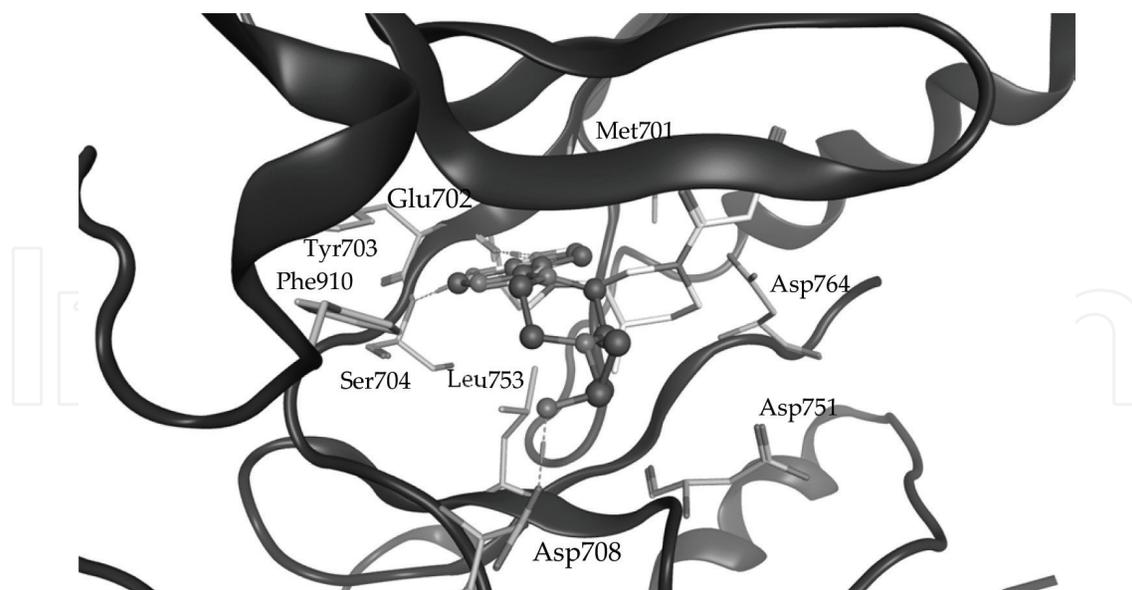


Figure 6. Comparison of the docking pose of Z1129905037 (shown as cyan sticks) and the binding mode of tofacitinib seen in magenta, at the binding site of PRK1 structure 4OTI. The figure was designed using MOE.

ring in PZ0151 takes the position of the CN group of tofacitinib but rather shows a lower activity. This could be as a result of the small size of the binding site under the P-loop residues, which is not suitable for large substituents. In general, it is difficult to explain the changes of the BFE values caused by chemical modification of all compounds under study. The observed correlation coefficient ($R^2 = 0.65$) could only be used to explain major differences in biological activities. However, it could be used as an indicator to identify the interactions, which play major roles in the determination of biological activity.

3.4. Ranking power

An enrichment study was performed in order to measure the ability to identify known binders or actives in large databases of inactives. The previously mentioned data set (named DS1), containing 28 actives and 300 inactives, was used to perform the enrichment study. The ligands were docked using the same ensemble and rescored depending on the previous findings. The results are presented in a BOX-PLOT and receiver operating characteristic (ROC) curves (**Figures 7 and 8**). **Table 4** shows enrichment study results, using an ensemble of PRK1 structures, where the available scoring functions in Schrödinger and BFE methods were considered (ΔH , Glide SP and Glide XP). As seen in **Figure 7**, Glide SP was able to discriminate between actives and inactives better than the BFE scoring (ΔH).

The median values of the pair active/inactive were $-9.17/7.23$ and $-38.79/34.02$ for Glide SP and ΔH , respectively. The area under the curve (AUC) and enrichment factors (EF) at two different percentages were calculated. **Table 4** shows the enrichment study results. It is interesting to note that at EF1%, all screened hits were actives when using both Schrödinger's scoring functions (Glide SP and XP). Interesting, it was also observed that for the EF3% (for Glide SP and XP), the rate of true positives was 100% (**Table 4**). The main factor to measure the ranking

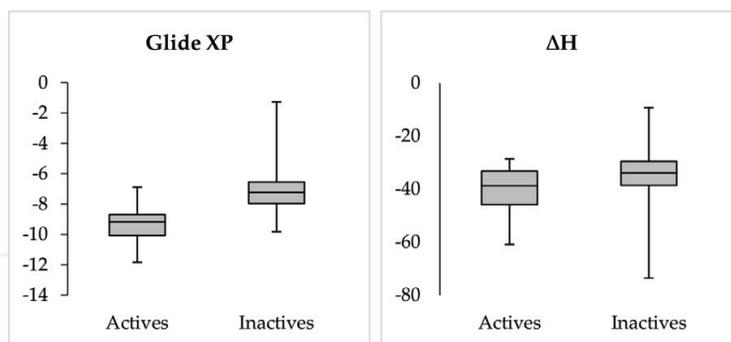


Figure 7. Box plots for the active and inactive compounds using the respective score values.

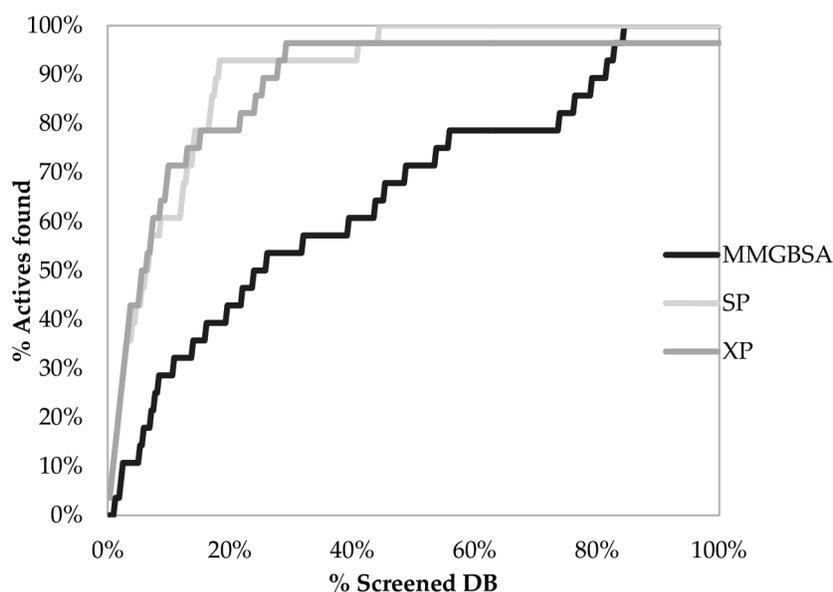


Figure 8. Enrichment plot showing comparison of the % actives found at a given percentage of the ranked database for DS1, using GLIDE SP, XP, and ΔH (MMGBSA).

power of the applied docking methods was the AUC value (Table 4). It is clear that the Glide scoring functions performed better to discriminate between the actives of inactives, having AUC = 0.90 and 0.88 for SP and XP, respectively.

3.5. QSAR model generation

Several QSAR models were generated to identify a possible correlation between experimentally obtained pIC_{50} values and calculated BFE or descriptor values (for the compounds on Figures 1 and 2). Calculated BFE (obtained by the MM/GBSA approach) from the previous study and 2D molecular descriptors, e.g., different descriptors referring to the partial charge and number of rotatable bonds, were considered. Since the Glide SP score showed a significant discrimination power between the active and inactive compounds, further scoring functions were also tested to optimize the QSAR models.

The best correlation coefficient in the previous study was found by applying Nguyen and Simmerling (igb = 8) Generalized Born solvation model [73, 74], after a minimization step ($R^2 = 0.65$, RMSE = 0.89 and cross-validated $q^2 = 0.60$, $n = 26$). In order to improve the correlation, partial charges of the ligands were calculated to compute several molecular descriptors for generating PLS models [75]. The inclusion of the PEOE_VSA-0 descriptor improved the correlation coefficient value only slightly (model QSAR_1, $R^2 = 0.68$, RMSE = 0.8, and $q^2 = 0.61$, $n = 26$). This descriptor indicates the partial equalization of electronegativities with approximated accessible van der Waals surface area (in Å²) of the molecule [76]. Furthermore, the inclusion of another descriptor based on ligand partial charges (PEOE_PC-) [76], which indicates to the total negative partial charge, was used instead of PEOE_VSA-0 to generate QSAR_2 ($R^2 = 0.70$, RMSE = 0.78 and $q^2 = 0.62$, $n = 26$; Eq. (1)). The model QSAR_3, containing three descriptors, showed a slightly better correlation coefficient ($R^2 = 0.71$, RMSE = 0.76 and $q^2 = 0.63$, $n = 26$; Eq. (2)). The relative importance of the individual descriptors was MM-GBSA: 1, GLIDE_SP: 0.417, and PEOE_PC-: 0.371. The other tested descriptors were not helpful and were not considered for the final models.

Further internal validation of the generated QSAR models was carried out by means of bootstrapping, in which the samples are randomly selected from the used inhibitors to form training and test sets [77, 78]. The respective models are generated, and their statistical parameters (R^2 and q^2) are calculated and compared with those of the original QSAR models (which had been generated from the whole data set). The random selection of the samples was performed within each cluster generated by hierarchical clustering, depending on the structural similarity or within activity distribution clusters of PRK1 inhibitors (low, moderate, and high active inhibitors) [77]. PRK1 inhibitors can be divided into four subsets depending on the chemical scaffolds (**Figures 1 and 2**). In the next step, hierarchical clustering using maximum common substructure (MACCS) keys and calculation of Tanimoto coefficients were carried out using the molecular operator environment (MOE) software tool (chemical computing group, Montreal, Canada, version 2014). Then, several samples within each cluster were taken, considering their activities. From 26 compounds, the sample subset contains only 20 inhibitors. The statistic parameters of the obtained QSAR models are presented in **Table 5**. The goal of the bootstrapping is to perturb the training set while not considering the statistical parameters of the test set. Since the obtained QSAR models were built from subset of the total data set, the values of R_{BT}^2 and q_{BT}^2 satisfy the minimum acceptable statistical parameters when

	Training set (20 inhibitors)			Test set (6 inhibitors)		
	R^2	RMSE	q^2	R_{BT}^2	RMSE _{pred}	q_{pred}^2
QSAR_BT_1	0.71	0.73	0.65	0.43	1.11	0.06
QSAR_BT_2	0.72	0.72	0.63	0.58	0.95	0.19
QSAR_BT_3	0.72	0.71	0.60	0.67	0.84	0.28

Table 5. Statistical parameters of obtained QSAR models in bootstrap validation.

compared with the three original QSAR models. These values also revolve around the values of the original QSAR models, generated from 26 PRK1 inhibitors (**Table 5**). It was observed that the obtained QSAR models possess acceptable statistic parameters ($q^2 > 0.5$) and were comparable with those found in the original QSAR models. However, the QSAR_BT_3 model showed poor correlation when compared with the results found in the whole data set ($q^2 = 0.60$, **Table 5**). The remaining six inhibitors were used as the external test set to verify the ability of the obtained QSAR models (QSAR_BTs) to predict the biological activities of the test set compounds. The values of R^2_{pred} between the calculated and measured pIC_{50} are 0.43, 0.58, and 0.67 for QSAR_BT_1, QSAR_BT_2, and QSAR_BT_3, respectively (**Table 5**). These results may confirm the ability of the model QSAR_3 and its derivatives (QSAR_BT_3) to predict the biological activities of novel compounds.

Interestingly, QSAR_3 have predicted pIC_{50} values higher or close to 7 for the most potent PRK1 inhibitors ($\text{IC}_{50} < 100$ nM). Moreover, QSAR_3 possessed the ability to distinguish between weak, moderate, and highly potent compounds (**Table 3**). Thus, the model could be used as a filter for the prioritization of specific compounds (hits) for synthesis and testing. **Figure 9** displays the graph of the correlation between the observed and predicted pIC_{50} values using QSAR_3 model.

3.6. Validation of the generated QSAR models and GSK databases screening

It has previously been demonstrated that considering only the statistic parameters (R^2 , RMSE, and q^2) to validate QSAR models is insufficient and sometimes misleading [78–82]. Therefore, it is important to validate QSAR models by testing the ability to accurately predict the biological activities of ligands not used for QSAR model generation. In the current study, two data

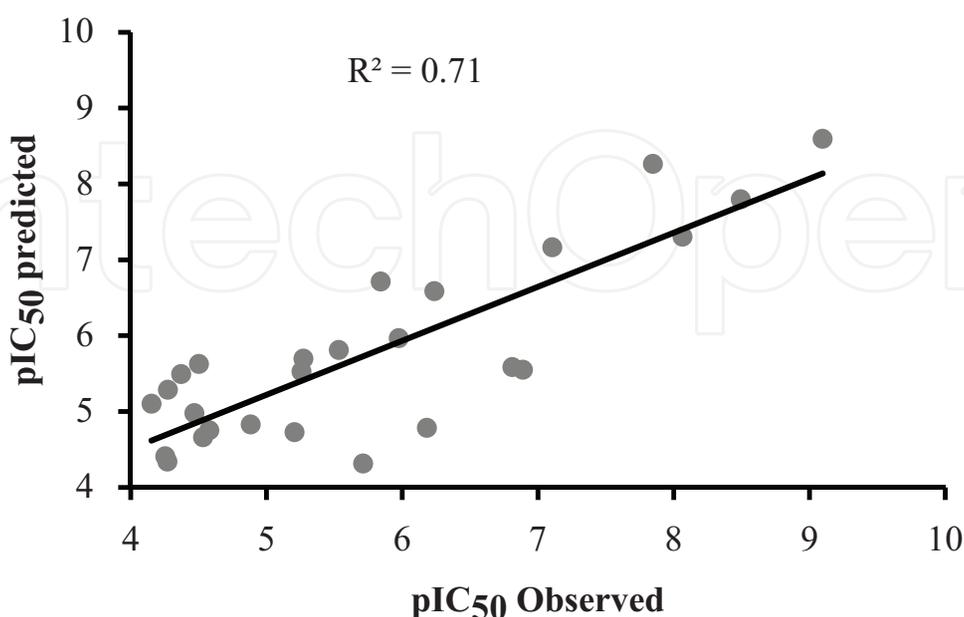


Figure 9. Correlation plot of the observed versus predicted pIC_{50} values of the training set ($n = 26$) using QSAR_3 model.

sets of compounds were available: the first set containing 26 PRK1 inhibitors with IC_{50} values was used as a training set for QSAR model generation (Figures 1 and 2). The second set was used as an external test set to validate the QSAR model. These compounds were identified from an *in vitro* screening (GSK PKIS1) and contain 35 compounds including 14 active compounds and 21 inactive compounds (Figures 10 and 11) [18]. Seven of the active compounds, for which IC_{50} values were measured (0.04–4.9 μ M, Table 6), were taken as external test set to further evaluate the predictive power of the QSAR models.

Table 5 provides a comparison of the calculated pIC_{50} versus experimental pIC_{50} values. The calculation of pIC_{50} values was performed by using the QSAR_1, 2, and 3 models. An important statistic parameter to test the predictive power of the QSAR models is the correlation coefficient R^2_{pred} between the predicted and observed pIC_{50} values for the test set [82]. The correlation coefficients between the experimental and predicted pIC_{50} values for the seven test set compounds were quite weak ($R^2_{pred} = 0.49$, $n = 7$) when using QSAR_3 model. The predictive power of the QSAR_1 and 2 models was investigated. The correlations of both models were within a similar range ($R^2_{pred} = 0.58$ and 0.37 , $n = 7$). A further analysis was performed to detect the outliers depending on Z-score values. The Z-score values were calculated using MOE, for QSAR_3. Golbraikh et al. mentioned that at least five compounds are required for a test set to validate a QSAR model [80]. After the analysis of the correlation and Z-scores, compound SB-750140 found to negatively affect the correlation coefficients. The correlation using QSAR_3 shows an improvement ($R^2_{pred} = 0.95$, $RMSD = 0.17$ and $q^2 = 0.89$, $n = 6$). A similar observation was made for the two other QSAR models (QSAR_1: $R^2_{pred} = 0.93$, $RMSD = 0.21$ and $q^2 = 0.82$, $n = 6$; QSAR_2: $R^2_{pred} = 0.98$, $RMSD = 0.12$ and $q^2 = 0.95$, $n = 6$). The outcomes confirm that the generated

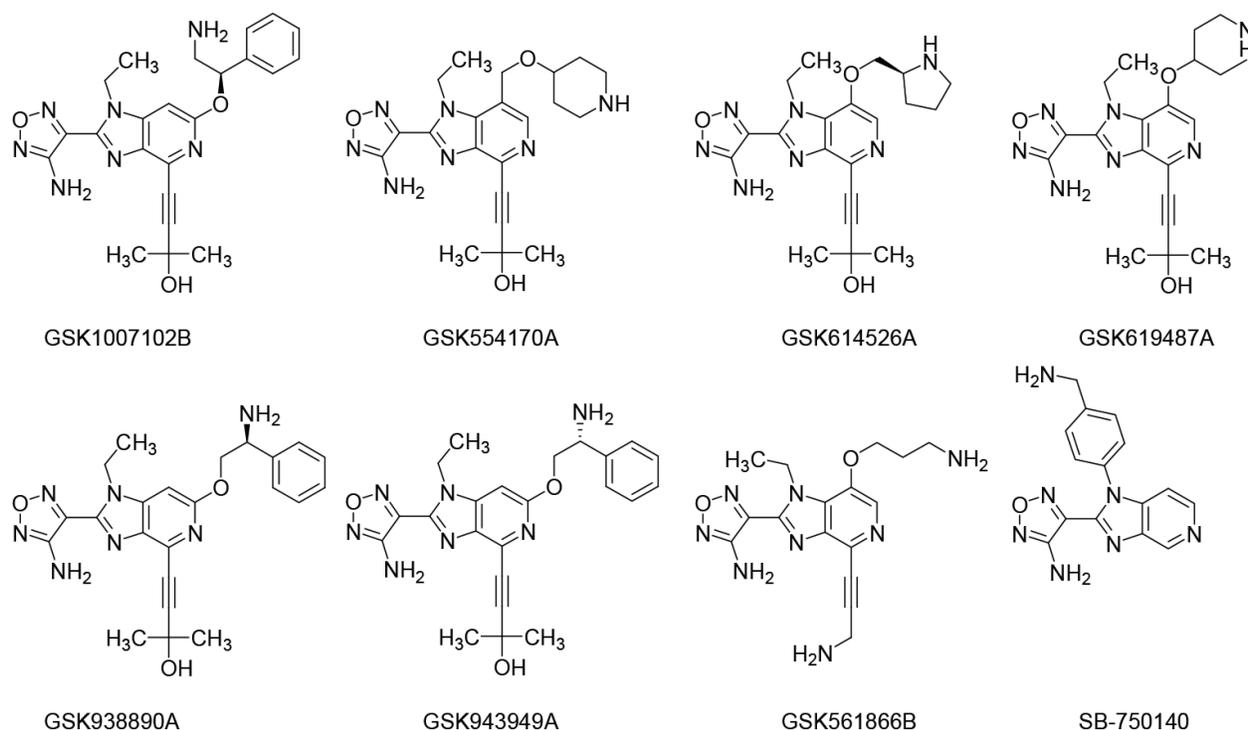


Figure 10. Chemical structures of PRK1 14 actives identified from screening the GSK PKIS1 [18].

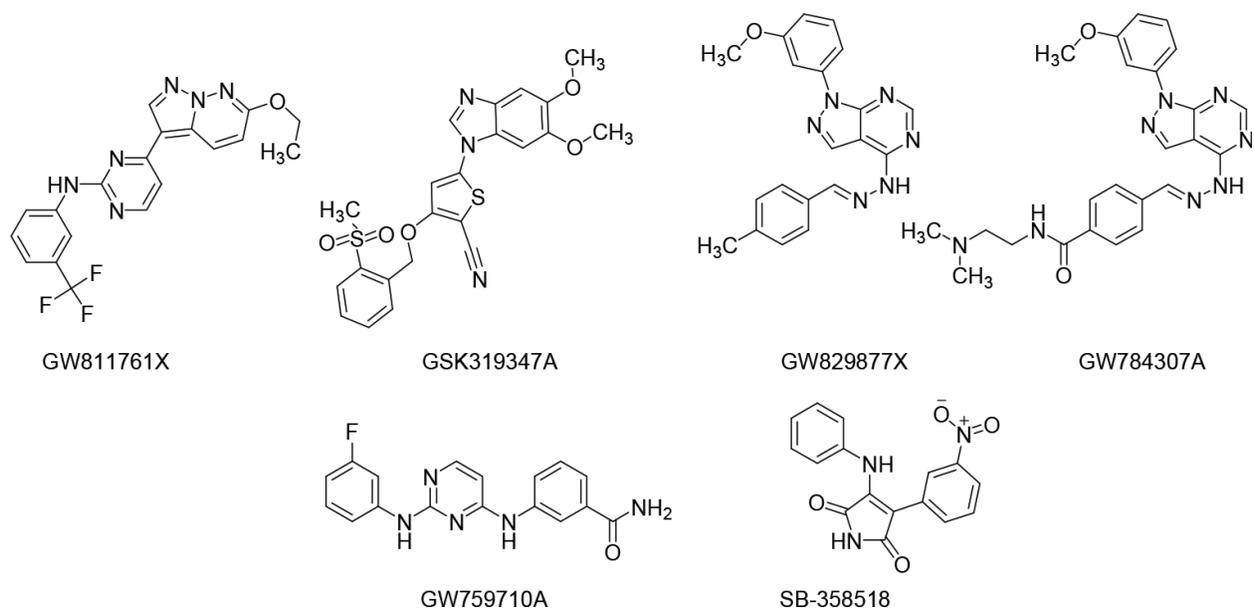


Figure 11. Chemical structures of PRK1 14 actives identified from screening the GSK PKIS1 (continued) [18].

Compound	Binding 1 μ M	IC ₅₀ nM	pIC ₅₀	Pred. pIC ₅₀ QSAR_3	Pred. pIC ₅₀ QSAR_1	Pred. pIC ₅₀ QSAR_2
GSK943949A	Yes	40	7.40	7.33	7.79	7.99
GSK614526A	Yes	68.1	7.17	7.09	7.68	7.69
SB-750140	Yes	120	6.92	5.70	6.06	5.88
GSK619487A	Yes	181.3	6.74	7.11	7.62	7.65
GSK554170A	Yes	710	6.15	6.34	6.76	7.02
GSK938890A	Yes	3600	5.44	5.67	5.77	6.49
SB-358518	Yes	4900	5.31	5.84	6.16	6.24
GW811761X	Yes	No IC ₅₀		5.72	5.97	6.05
GW784307A	Yes	No IC ₅₀		6.86	7.03	7.36
GSK1007102B	Yes	No IC ₅₀		6.34	6.72	7.21
GSK561866B	Yes	No IC ₅₀		5.82	6.10	6.45
GSK319347A	Weak	No IC ₅₀		6.77	6.26	6.96
GW759710A	Weak	No IC ₅₀		4.99	5.26	5.41
GW829877X	Weak	No IC ₅₀		5.26	5.52	5.53
GSK1030059A	No			7.66	8.08	7.98
GSK625137A	No			4.16	4.43	4.47
GW290597X	No			6.52	5.67	6.53
GW513184X	No			5.27	5.25	5.55

Compound	Binding 1 μ M	IC ₅₀ nM	pIC ₅₀	Pred. pIC ₅₀ QSAR_3	Pred. pIC ₅₀ QSAR_1	Pred. pIC ₅₀ QSAR_2
GW643971X	No			5.27	5.19	5.60
GW794607X	No			5.44	5.66	5.71
SB-409514	No			5.13	5.56	5.48
GSK994854A	No			5.93	5.90	6.50
GW874091X	No			5.26	5.65	5.56
GW693481X	No			3.25	3.80	3.92
GSK1000163A	No			6.21	7.02	6.91
GSK1220512A	No			6.03	5.51	6.58
GSK317314A	No			5.94	6.64	6.48
GSK579289A	No			5.14	4.89	5.60
GSK949675A	No			7.20	7.99	7.89
GW580496A	No			6.18	5.37	6.77
SB-476429-A	No			4.68	4.67	4.92
SB-744941	No			4.73	5.52	5.22
GSK571989A	No			6.26	6.42	6.75
GW779439X	No			4.83	4.76	5.35
GSK978744A	No			7.76	7.75	8.10

Table 6. Comparison of the experimental pIC₅₀ versus the calculated pIC₅₀ using the generated models QSAR_1, 2 and 3 [18].

QSAR models show satisfactory ability to predict the biological activities of the test set compounds. Thus, they could be accepted having an R^2_{pred} value > 0.5 [82]. Moreover, taking the last consideration of the threshold for the predicted pIC₅₀ > 7 found in case of actives to refer to potent compounds was clearly in the test set, since the most potent inhibitors in the test set (IC₅₀ < 100) were identified with predicted pIC₅₀ > 7 (Table 6). Meanwhile, the biological activities of the moderately active compounds were predictive with calculated pIC₅₀ in the range 5.5–7 (Table 6). Among the inactive compounds in the test set, three compounds were predicted to be highly potent compounds, but these compounds were inactive in the assay. The visualization of the false positives could not clarify the absence of activity of both compounds. However, GSK978744A and GSK1030059A, which mainly target the PLK kinase, possess the same chemical scaffold, and their binding modes were investigated and compared with the binding modes of analogues co-crystallized with other kinases (CDK2; PDB ID: 2I40, NEK2; PDB ID: 2XNN). However, it is not clear, from structural interactions, why this compound is not active. Additionally, it might be possible that the exclusion of entropy changes or solvation of the binding pocket could play a role in the wrong estimation of the biological activity.

Since the QSAR_2 model shows high value of R^2_{pred} in the test set (QSAR_2: $R^2_{\text{pred}} = 0.98$, RMSD = 0.12 and $q^2 = 0.95$, $n = 6$), the discrimination performance was therefore discussed. Similar to QSAR_3, the most potent inhibitors in the test set (IC₅₀ < 100) were identified with predicted pIC₅₀ > 7

(Table 6). Furthermore, between the remaining active compounds, two of them were predicted to be high potent inhibitors $pIC_{50} > 7$ (GW784307A: predicted $pIC_{50} = 7.36$; GSK1007102B: predicted $pIC_{50} = 7.21$). However, the IC_{50} values of these two compounds have not been measured yet, but it reported to be active since it binds to PRK1 at 1 μ M. Thus, the consideration of the threshold for the predicted $pIC_{50} > 7$ could also be applicable in the QSAR_2 model. The weak active compounds GSK319347A, GW759710A, and GW829877X (shown in Table 6) also predicted as moderate or low active compounds where predicted pIC_{50} values were less than 7. Among the inactive compounds, three compounds (the same compounds as in the case of the QSAR_3 model) were predicted to be highly potent compounds, but these compounds were inactive in the assay.

$$\begin{aligned} \text{QSAR_2: } pIC_{50} &= 0.15509 - 0.09778 \\ &*PEOE_PC - -0.09778 * MM\text{-}GBSA \\ (R^2 &= 0.70, RMSE = 0.78, q^2 = 0.62, n = 26) \end{aligned} \quad (1)$$

$$\begin{aligned} \text{QSAR_3: } pIC_{50} &= -1.37129 - 0.56938 * PEOE_PC - \\ &-0.08688 * MM\text{-}GBSA - 0.26312 * GLIDE_SP \\ (R^2 &= 0.71, RMSE = 0.76 \text{ and } q^2 = 0.63, n = 26) \end{aligned} \quad (2)$$

By comparing the three generated QSAR models (QSAR_1, 2, and 3) depending on the predicted pIC_{50} values among them to the measured ones, it was found the predicted pIC_{50} values when using QSAR_3 model were closer to the experimental values (Table 6), but QSAR_2 model showed the highest value of R^2_{pred} in the test set. Furthermore, in bootstrap validation, it was found that the QSAR_3 model had the ability to predict the biological activity of the test set compounds more accurately than the models QSAR_1 and 2 (Table 6). Thus, the QSAR_3 model could be used to predict the biological activities of the highly potent compounds, which should exhibit biological activities power in the nanomolar range and differentiate between the active compounds, according to their activity (low, moderate, and high).

3.7. Virtual screening using the p-ANAPL library

The above protocol was further used to virtually screen for possible PRK1 inhibitors from the p-ANAPL library [58]. As previously applied, ensemble docking was performed to dock the compounds to the X-ray structures of PRK1 stored in the PDB. Both data sets were prepared using Schrödinger 2014.U2, including the generation of several conformations per compound. The docking solutions were visualized using MOE to explore the conserved interactions with residues at the PRK1 hinge region. The selection of the hits was based first on the displayed binding mode, followed by their scoring, then predicting the pIC_{50} values using the generated QSAR models. Depending only on docking scores was insignificant for the selection of promising hit compounds. The docking solutions for the entire data set were visualized and several hits compounds were selected. In the next step, their predicted pIC_{50} values were calculated using QSAR_1, 2, and 3. Depending on the set threshold for the predicted pIC_{50} , several top scoring compounds could be selected as promising hits for the biological assays, e.g., those in Figure 12. Figure 13 displays the binding modes of the selected hits and the interaction with residues at the binding pocket of the corresponding X-ray structure of PRK1.

One of the promising hits, DBT-6b, is Bartericin B (MW = 408.49; PubChem CID: 12136210) from *Dorstenia barteri* [83]. The binding mode of DBT-6b is shown in **Figure 13A**. The compound is predicted to interact with hinge region residue Ser704 with one H bond. The phenol ring forms an additional H bond with the backbone of Asp764 from the DFG-motif. The interaction with the front polar pocket and the ribose binding pocket is mediated by two H bonds between Asp708 and propanol substituent in DBT-6b. The calculated pIC_{50} values were 7.85, 7.64, and 7.45 using the generated QSAR models QSAR_1, 2, and 3, respectively. The second hit (P87), vitexin or apigenin-8-C glucoside (MW = 432.38; PubChem CID: 5280441), is a flavonoid glycoside from diverse sources, e.g., *Hyparrhenia hirta* [84]. The binding mode of P87 (**Figure 13B**) shows that the compound might form one H bond with the hinge region residue Ser704. Additionally, the substituents on the

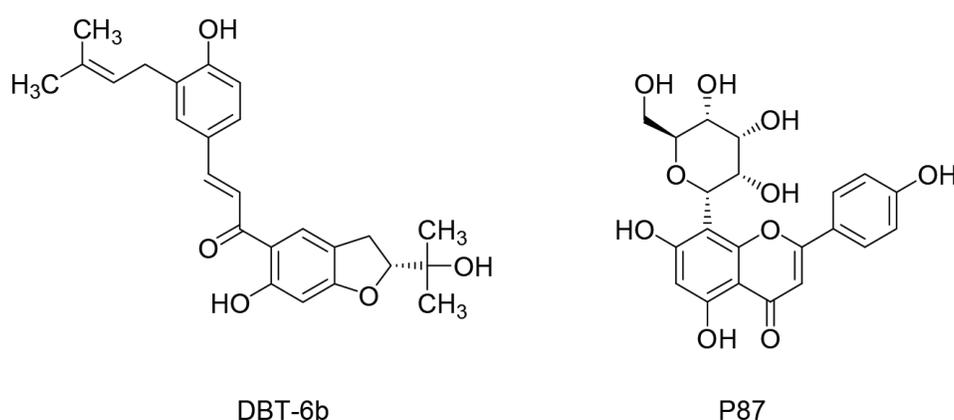


Figure 12. Chemical structures of the selected hits.

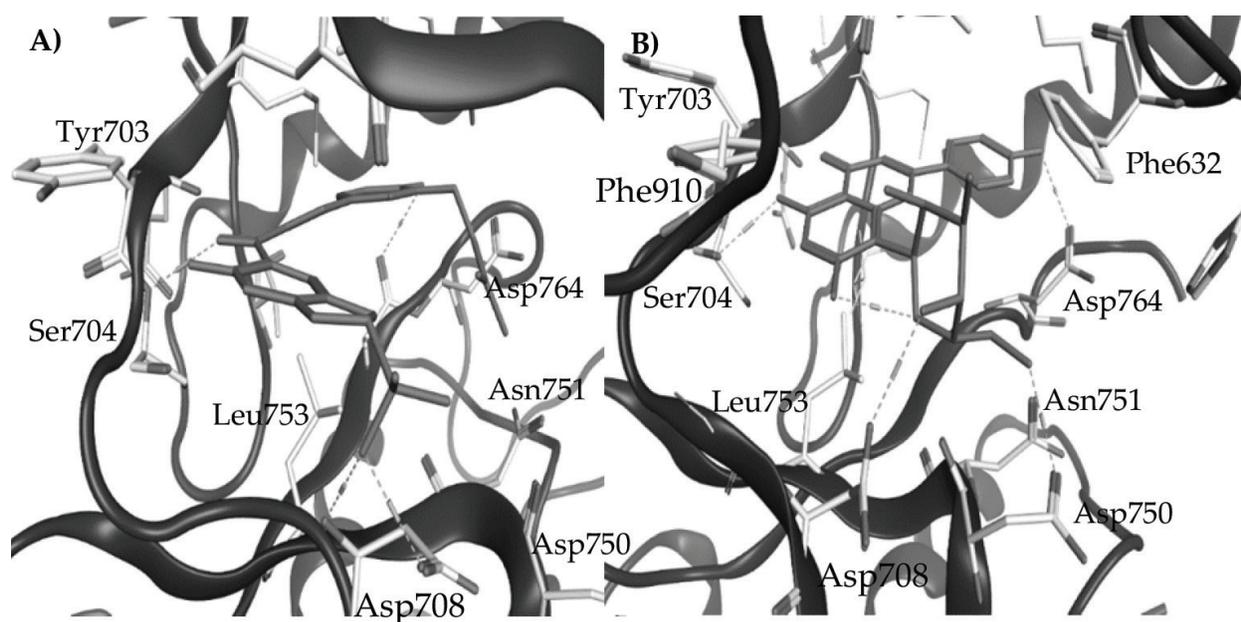


Figure 13. Docking poses of the selected hits in PRK1 binding pocket. (A) Docking poses of DBT-6b in the binding pocket of PRK1 structure 4OTG. (B) Docking pose of P87 in PRK1 structure 4OTH. The figure was designed using MOE.

oxane ring are targeting the front polar pocket and forming two H bonds with Asp708 and Asp750. The phenyl ring interacts with the back hydrophobic pocket and further forms an H bond with Asp764 in the DFG motif. P87 was predicted to be a highly potent compound, with a predicted pIC_{50} value greater than 7, using QSAR models ($pred_pIC_{50} = 8.02, 8.6, \text{ and } 9$, respectively). Bartericin B and vitexin are the only two examples of promising compounds from natural products which could target PRK1. Among the screened databases, other hits were selected as promising compounds, which could inhibit PRK1 activity in the nanomolar range. In further investigations, the selected compounds can be submitted for biological assay to figure out their ability to bind into PRK1.

4. Conclusions

This chapter provides a brief summary of database resources for anticancer drug discovery and some recent inputs and success stories for the identification novel inhibitors of selected anticancer drug targets by the use of computer modeling. While the experimental validation of some of the natural product hits identified in the case study is ongoing, the question about the applicability domains of the derived QSAR models reported is still to be answered. Among the identified hits, 50 closely similar compounds to Bartericin B (with cut-off Tanimoto coefficient of 0.7), having reported biological activities in PubChem [59] and ChEMBL [44], with 239 biological activities and 152 published patent data. A similarity search of vitexin in PubChem gives 112 similar compounds, corresponding to 524 biological activities and 208 patents. Bartericin B is known to exhibit antimicrobial activities against *Trichomonas gallinarum*, with minimum lethal concentrations (MLCs) of 0.244 and 0.121 $\mu\text{g/mL}$ after 24 H and 48 H, respectively [83], while its analogue (Bartericin A from *Dorstenia angusticornis*) is known to be very active against some bacteria and yeasts associated with human pathologies [85]. Both chalcones are known to also have potential antiprotozoal activities [86], e.g., against *Plasmodium falciparum* [87]. Vitexin, on the other hand, is known for its antioxidative and [88] spasmolytic effects [89]. The compound is known to be abundant in plants of the genus *Vitex* (Verbenaceae), which is known to exert insect antifeeding activities, among others [90]. These compounds could now be tested in biological assays for potential PRK1 inhibition.

Acknowledgements

A.N. is currently a doctoral candidate financed by the German Academic Exchange Services (DAAD), Germany. F.N.K. acknowledges a Georg Forster fellowship from the Alexander von Humboldt Foundation, Germany.

Competing interests

The authors declare that they have no competing interests.

List of abbreviations

AfroCancer	African Anticancer Natural Products Database
BFE	Binding free energy
EC ₅₀	Half maximal effective concentration, i.e., the concentration of a drug, antibody or toxicant which induces a response halfway between the baseline and maximum after a specified exposure time
ED ₅₀	The median effective dose, a dose that produces the desired effect in 50% of a population
GI ₅₀	The growth inhibition of 50%, drug concentration resulting in a 50% reduction in the net protein increase
IC ₅₀	The drug concentration causing 50% inhibition of the desired activity
NANPDB	Northern African Natural Products Database
NP	Natural product
NPACT	Naturally Occurring Plant-based Anti-cancer Compound Activity-Target Database
p-ANAPL	Pan-African natural products library
PDB	Protein databank
PLS	Partial least squares
PRK1	protein kinase C-related kinase
QSAR	Quantitative structure-activity relationship
RMSD	Root mean square deviation
RMSE	Root mean square error
SM	Secondary metabolite
VS	Virtual screening

Author details

Abdulkarim Najjar¹, Fidele Ntie-Kang^{1,2*} and Wolfgang Sippl¹

*Address all correspondence to: ntiekfidele@gmail.com

1 Department of Pharmaceutical Chemistry, Martin-Luther University of Halle-Wittenberg, Halle (Saale), Germany

2 Department of Chemistry, University of Buea, Buea, Cameroon and Department of Pharmaceutical Chemistry, Martin-Luther University of Halle-Wittenberg, Halle, Germany

References

- [1] Hanahan D. Rethinking the war on cancer. *Lancet*. 2014;**383**:558-563. DOI: 10.1016/S0140-6736(13)62226-6
- [2] Cragg GM, Grothaus PG, Newman DJ. Impact of natural products on developing new anti-cancer agents. *Chemical Reviews*. 2009;**109**:3012-3043. DOI: 10.1021/cr900019j
- [3] Newman DJ, Cragg GM. Natural products as sources of new drugs from 1981 to 2014. *Journal of Natural Products*. 2016;**79**:629-661. DOI: 10.1021/acs.jnatprod.5b01055

- [4] Harvey AL, Edrada-Ebel R, Quinn RJ. The re-emergence of natural products for drug discovery in the genomics era. *Natural Reviews Drug Discovery*. 2015;**14**:111-129. DOI: 10.1038/nrd4510
- [5] Ntie-Kang F, Lifongo LL, Judson PN, Sippl W, Efange SMN. How "drug-like" are naturally occurring anti-cancer compounds?. *Journal of Molecular Modeling*. 2014;**20**:2069. DOI: 10.1007/s00894-014-2069-z
- [6] Rodrigues T, Reker D, Schneider P, Schneider G. Counting on natural products for drug design. *Natural Chemistry*. 2016;**8**:531-541. DOI: 10.1038/nchem.2479
- [7] Mangal M, Sagar P, Singh H, Raghava GP, Agarwal SM. NPACT: Naturally occurring plant-based anti-cancer compound-activity-target database. *Nucleic Acids Research*. 2013;**41**:D1124-D1129. DOI: 10.1093/nar/gks1047
- [8] American Cancer Society, 2013. *Cancer Facts & Figures*. 2013. <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2013/cancer-facts-and-figures-2013.pdf> (Accessed 27-01-2017)
- [9] Jilg CA, Ketscher A, Metzger E, Hummel B, Willmann D, et al. PRK1/PKN1 controls migration and metastasis of androgen-independent prostate cancer cells. *Oncotarget*. 2014;**5**:12646-12664. DOI: 10.18632/oncotarget.2653
- [10] Heinlein CA, Chang C. Androgen receptor in prostate cancer. *Endocrine Reviews*. 2004;**25**:276-308. DOI: 10.1210/er.2002-0032
- [11] Lonergan PE, Tindall DJ. Androgen receptor signaling in prostate cancer development and progression. *Journal of Carcinogenesis*. 2011;**10**:20. DOI: 10.4103/1477-3163.83937
- [12] Metzger E, Yin N, Wissmann M, Kunowska N, Fischer K, et al. Phosphorylation of histone H3 at threonine 11 establishes a novel chromatin mark for transcriptional regulation. *Nature Cell Biology*. 2008;**10**:53-60. DOI: 10.1038/ncb1668
- [13] Wissmann M, Yin N, Müller JM, Greschik H, Fodor BD, et al. Cooperative demethylation by JMJD2C and LSD1 promotes androgen receptor-dependent gene expression. *Nature Cell Biology*. 2007;**9**:347-353. DOI: 10.1038/ncb1546
- [14] Di Croce L, Shiekhhattar R. Thrilling transcription through threonine phosphorylation. *Nature Cell Biology*. 2008;**10**:5-6. DOI: 10.1038/ncb0108-5
- [15] Lachmann S, Jevons A, De Rycker M, Casamassima A, Radtke S, et al. Regulatory domain selectivity in the cell-type specific PKN-dependence of cell migration. *PLoS One*. 2011;**6**:e21732. DOI: 10.1371/journal.pone.0021732
- [16] Turner EC, Kavanagh DJ, Mulvaney EP, McLean C, Wikström K, et al. Identification of an interaction between the TP α and TP β isoforms of the human thromboxane A2 receptor with protein kinase C-related kinase (PRK) 1: implications for prostate cancer. *Journal of Biological Chemistry*. 2011;**286**:15440-15457. DOI: 10.1074/jbc.M110.181180
- [17] Köhler J, Erenkamp G, Eberlin A, Rumpf T, Slynko I, et al. Lestaurtinib inhibits histone phosphorylation and androgen-dependent gene expression in prostate cancer cells. *PLoS One*. 2012;**7**:e34973. DOI: 10.1371/journal.pone.0034973

- [18] Slynko I, Schmidtkunz K, Rumpf T, Klaeger S, Heinzlmeir S, et al. Identification of highly potent protein kinase C-related kinase 1 inhibitors by virtual screening, binding free energy rescoring, and *in vitro* testing. *ChemMedChem*. 2016;**11**:2084-2094. DOI: 10.1002/cmdc.201600284
- [19] Slynko I, Scharfe M, Rumpf T, Eib J, Metzger E, et al. Virtual screening of PRK1 inhibitors: Ensemble docking, rescoring using binding free energy calculation and QSAR model development. *Journal of Chemical Information and Modeling*. 2014;**54**:138-150. DOI: 10.1021/ci400628q
- [20] Mukai H, Ono Y. A novel protein kinase with leucine zipper-like sequences: Its catalytic domain is highly homologous to that of protein kinase C. *Biochemical and Biophysical Research Communications*. 1994;**199**:897-904. DOI: 10.1006/bbrc.1994.1313
- [21] Mukai H. The structure and function of PKN, a protein kinase having a catalytic domain homologous to that of PKC. *Journal of Biochemistry*. 2003;**133**:17-27. PMID: 12761194
- [22] Lim WG, Tan BJ, Zhu Y, Zhou S, Armstrong JS, et al. The very C-terminus of PRK1/PKN is essential for its activation by RhoA and downstream signaling. *Cell Signal*. 2006 Sep;**18**(9):1473-1481. DOI: 10.1016/j.cellsig.2005.11.009
- [23] Kannan N, Haste N, Taylor SS, Neuwald AF. The hallmark of AGC kinase functional divergence is its C-terminal tail, a *cis*-acting regulatory module. *Proceedings of National Academy Science USA*. 2007;**104**:1272-1277. DOI: 10.1073/pnas.0610251104
- [24] Chamberlain P, Delker S, Pagarigan B, Mahmoudi A, Jackson P, et al. Crystal structures of PRK1 in complex with the clinical compounds lestaurtinib and tofacitinib reveal ligand induced conformational changes. *PLoS One*. 2014;**9**:e103638. DOI: 10.1371/journal.pone.0103638
- [25] Metzger E, Müller JM, Ferrari S, Buettner R, Schüle R. A novel inducible transactivation domain in the androgen receptor: implications for PRK in prostate cancer. *EMBO Journal*. 2003;**22**:270-280. DOI: 10.1093/emboj/cdg023
- [26] Rester U. From virtuality to reality - Virtual screening in lead discovery and lead optimization: a medicinal chemistry perspective. *Current Opinion in Drug Discovery and Developments*. 2008;**11**:559-568. PMID: 18600572
- [27] Rollinger JM, Stuppner H, Langer T. Virtual screening for the discovery of bioactive natural products. *Progress in Drug Research*. 2008;**65**:211,213-249. PMID: 18084917
- [28] Lavecchia A, Di Giovanni C. Virtual screening strategies in drug discovery: A critical review. *Current Medicinal Chemistry*. 2013;**20**:2839-2860. DOI: 10.2174/09298673113209990001
- [29] Heikamp K, Bajorath J. The future of virtual compound screening. *Chemical Biology and Drug Design*. 2013;**81**:33-40. DOI: 10.1111/cbdd.12054
- [30] Schneider G. Virtual screening: An endless staircase?. *Natural Reviews Drug Discovery*. 2010;**9**:273-276. DOI: 10.1038/nrd3139

- [31] McInnes C. Virtual screening strategies in drug discovery. *Current Opinion in Chemical Biology*. 2007;**11**:494-502. DOI: 10.1016/j.cbpa.2007.08.033
- [32] Cheng T, Li Q, Zhou Z, Wang Y, Bryant SH. Structure-based virtual screening for drug discovery: A problem-centric review. *AAPS Journal*. 2012;**14**:133-141. DOI: 10.1208/s12248-012-9322-0
- [33] Sun H. Pharmacophore-based virtual screening. *Current Medicinal Chemistry*. 2008;**15**:1018-1024. DOI: 10.2174/092986708784049630
- [34] Ripphausen P, Nisius B, Bajorath J. State-of-the-art in ligand-based virtual screening. *Drug Discovery Today*. 2011;**16**:372-376. DOI: 10.1016/j.drudis.2011.02.011
- [35] Anighoro A, Bajorath J. Three-dimensional similarity in molecular docking: prioritizing ligand poses on the basis of experimental binding modes. *Journal of Chemical Information and Modeling*. 2016;**56**:580-587. DOI: 10.1021/acs.jcim.5b00745
- [36] Ballester PJ, Westwood I, Laurieri N, Sim E, Richards WG. Prospective virtual screening with ultrafast shape recognition: The identification of novel inhibitors of arylamine N-acetyltransferases. *Journal of Royal Society Interface*. 2010;**7**:335-342. DOI: 10.1098/rsif.2009.0170
- [37] Li GB, Yang LL, Yuan Y, Zou J, Cao Y, et al. Virtual screening in small molecule discovery for epigenetic targets. *Methods*. 2015;**71**:158-166. DOI: 10.1016/j.ymeth.2014.11.010
- [38] Verma J, Khedkar VM, Coutinho EC. 3D-QSAR in drug design - A review. *Current Topics in Medicinal Chemistry*. 2010;**10**:95-115. DOI: 10.2174/156802610790232260
- [39] Winkler DA. The role of quantitative structure - activity relationships (QSAR) in biomolecular discovery. *Briefings in Bioinformatics*. 2002;**3**:73-86. PMID: 12002226
- [40] Ntie-Kang F, Nwodo JN, Ibezim A, Simoben CV, Karaman B, et al. Molecular modeling of potential anticancer agents from African medicinal plants. *Journal of Chemical Information and Modeling*. 2014;**54**:2433-2450. DOI: 10.1021/ci5003697
- [41] Kumar R, Chaudhary K, Gupta S, Singh H, Kumar S, et al. CancerDR: Cancer drug resistance database. *Scientific Reports*. 2013;**3**:1445. DOI: 10.1038/srep01445
- [42] Tao W, Li B, Gao S, Bai Y, Shar PA, et al. CancerHSP: Anticancer herbs database of systems pharmacology. *Scientific Reports*. 2015;**5**:11481. DOI: 10.1038/srep11481
- [43] Hastings J, Owen G, Dekker A, Ennis M, Kale N, et al. ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Research*. 2016;**44**:D1214-D1219. DOI: 10.1093/nar/gkv1031
- [44] Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, et al. The ChEMBL database in 2017. *Nucleic Acids Research*. 2017;**45**:D945-D954. DOI: 10.1093/nar/gkw1074
- [45] Fang X, Shao L, Zhang H, Wang S. CHMIS-C: A comprehensive herbal medicine information system for cancer. *Journal of Medicinal Chemistry*. 2005;**48**:1481-1488. DOI: 10.1021/jm049838d

- [46] Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J. DrugBank: A comprehensive resource for *in silico* drug discovery and exploration. *Nucleic Acids Research*. 2006;**34**:D668-D672. DOI: 10.1093/nar/gkj067
- [47] Mysinger MM, Carchia M, Irwin JJ, Shoichet BK. Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking. *Journal Medicinal Chemistry*. 2012;**55**:6582-6594. DOI: 10.1021/jm300687e
- [48] Loharch S, Bhutani I, Jain K, Gupta P, Sahoo DK, Parkesh R. EpiDBase: A manually curated database for small molecule modulators of epigenetic landscape. *Database (Oxford)*. 2015;**2015**.pii: bav013. DOI: 10.1093/database/bav013
- [49] Vetrivel U, Subramanian N, Pilla K. InPACdb - Indian plant anticancer compounds database. *Bioinformatics*. 2009;**4**:71-74. PMID: 20198172
- [50] Ashfaq UA, Mumtaz A, Qamar TU, Fatima T. MAPS database: Medicinal plant activities. *Phytochemical and structural database*. *Bioinformatics*. 2013;**9**:993-995. DOI: 10.6026/97320630009993
- [51] Ntie-Kang F, Telukunta KK, Döring K, Simoben CV, Moumbock, et al. The Northern African Natural Products Database (NANPDB); 2017. www.african-compounds.org/nanpdb/
- [52] Milne GW, Nicklaus MC, Driscoll JS, Wang S, Zaharevitz D. National cancer institute drug information system 3D database. *Journal of Chemical Information and Computer Science*. 1994;**34**:1219-1224. DOI: 10.1021/ci00021a032
- [53] Ko Y, Tan SL, Chan A, Wong YP, Yong WP, et al. Prevalence of the coprescription of clinically important interacting drug combinations involving oral anticancer agents in Singapore: A retrospective database study. *Clinical Therapy*. 2012;**34**:1696-1704. DOI: 10.1016/j.clinthera.2012.06.025
- [54] Hatherley R, Brown DK, Musyoka TM, Penkler DL, Faya N, et al. SANCDB: A South African natural compound database. *Journal of Cheminformatics*. 2015;**7**:29. DOI: 10.1186/s13321-015-0080-8
- [55] Goede A, Dunkel M, Mester N, Frommel C, Preissner R. SuperDrug: A conformational drug database. *Bioinformatics*. 2005;**21**:1751-1753. DOI: 10.1093/bioinformatics/bti295
- [56] Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanović S. SYFPEITHI: Database for MHC ligands and peptide motifs. *Immunogenetics*. 1999;**50**:213-219. DOI: 10.1007/s002510050595
- [57] Lin YC, Wang CC, Chen IS, Jheng JL, Li JH, et al. TIPdb: A database of anticancer, antiplatelet, and antituberculosis phytochemicals from indigenous plants in Taiwan. *Scientific World Journal*. 2013;**2013**:736386. DOI: 10.1155/2013/736386
- [58] Ntie-Kang F, Amoa Onguéné P, Fotso GW, Andrae-Marobela K, Bezabih M, et al. Virtualizing the p-ANAPL library: A step towards drug discovery from African medicinal plants. *PLoS One*. 2014;**9**:e90655. DOI: 10.1371/journal.pone.0090655

- [59] Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, et al. PubChem substance and compound databases. *Nucleic Acids Research*. 2016;**44**:D1202-D1213. DOI: 10.1093/nar/gkv951
- [60] Sterling T, Irwin JJ. ZINC 15 - ligand discovery for everyone. *Journal of Chemical Information and Modeling*. 2015;**55**(11):2324-2327. DOI: 10.1021/acs.jcim.5b00559
- [61] Ha H, Debnath B, Odde S, Bensman T, Ho H, et al. Discovery of novel CXCR2 inhibitors using ligand-based pharmacophore models. *Journal of Chemical Information and Modeling*. 2015;**55**:1720-1738. DOI: 10.1021/acs.jcim.5b00181
- [62] Zhu YM, Webster SJ, Flower D, Woll PJ. Interleukin-8/CXCL8 is a growth factor for human lung cancer cells. *British Journal of Cancer*. 2004;**91**:1970-1976. DOI: 10.1038/sj.bjc.6602227
- [63] Rubie C, Kollmar O, Frick VO, Wagner M, Brittner B, et al. Differential CXC receptor expression in colorectal carcinomas. *Scandinavian Journal of Immunology*. 2008;**68**:635-644. DOI: 10.1111/j.1365-3083.2008.02163.x
- [64] Murphy C, McGurk M, Pettigrew J, Santinelli A, Mazzucchelli R, et al. Nonapical and cytoplasmic expression of interleukin-8, CXCR1, and CXCR2 correlates with cell proliferation and microvessel density in prostate cancer. *Clinical Cancer Research*. 2005 ;**11**:4117-4127. DOI: 10.1158/1078-0432.CCR-04-1518
- [65] Yang G, Rosen DG, Liu G, Yang F, Guo X, et al. CXCR2 promotes ovarian cancer growth through dysregulated cell cycle, diminished apoptosis, and enhanced angiogenesis. *Clinical Cancer Research*. 2010;**16**:3875-3886. DOI: 10.1158/1078-0432.CCR-10-0483
- [66] Kumar AV, Mohan K, Riyaz S. Structure guided inhibitor designing of CDK2 and discovery of potential leads against cancer. *Journal of Molecular Modeling*. 2013;**19**:3581-3589. DOI: 10.1007/s00894-013-1887-8
- [67] Al-Sha'er MA, Khanfar MA, Taha MO. Discovery of novel urokinase plasminogen activator (uPA) inhibitors using ligand-based modeling and virtual screening followed by *in vitro* analysis. *Journal of Molecular Modeling*. 2014;**20**:2080. DOI: 10.1007/s00894-014-2080-4
- [68] Ren JX, Li LL, Zheng RL, Xie HZ, Cao ZX, et al. Discovery of novel Pim-1 kinase inhibitors by a hierarchical multistage virtual screening approach based on SVM model, pharmacophore, and molecular docking. *Journal of Chemical Information and Modeling*. 2011;**51**:1364-1375. DOI: 10.1021/ci100464b
- [69] Li L, Khanna M, Jo I, Wang F, Ashpole NM, et al. Target-specific support vector machine scoring in structure-based virtual screening: Computational validation, *in vitro* testing in kinases, and effects on lung cancer cell proliferation. *Journal of Chemical Information and Modeling*. 2011;**51**:755-759. DOI: 10.1021/ci100490w
- [70] Khanfar MA, Taha MO. Elaborate ligand-based modeling coupled with multiple linear regression and k nearest neighbor QSAR analyses unveiled new nanomolar mTOR

- inhibitors. *Journal of Chemical Information and Modeling*. 2013;**53**:2587-2612. DOI: 10.1021/ci4003798
- [71] B-Rao C, Subramanian J, Sharma SD. Managing protein flexibility in docking and its applications. *Drug Discovery Today*. 2009;**14**:394-400. DOI: 10.1016/j.drudis.2009.01.003
- [72] Bjerrum EJ. Machine learning optimization of cross docking accuracy. *Computational Biology and Chemistry*. 2016;**62**:133-144. DOI: 10.1016/j.compbiolchem.2016.04.005
- [73] Tan C, Yang L, Luo R. How well does Poisson-Boltzmann implicit solvent agree with explicit solvent? A quantitative analysis. *Journal of Physical Chemistry B*. 2006;**110**:18680-18687. DOI: 10.1021/jp063479b.
- [74] Wang J, Tan C, Chanco E, Luo R. Quantitative analysis of Poisson-Boltzmann implicit solvent in molecular dynamics. *Physical Chemistry Chemical Physics*. 2010;**12**:1194-1202. DOI: 10.1039/b917775b
- [75] Mateos-Aparicio G. Partial least squares (PLS) methods: Origins, evolution, and application to social sciences. *Communications in Statistics - Theory and Methods*. 2011;**40**:2305-2317. DOI: 10.1080/03610921003778225
- [76] Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity - A rapid access to atomic charges. *Tetrahedron*. 1980;**36**:3219-3228. DOI: 10.1016/0040-4020(80)80168-2
- [77] Kiralj R, Ferreira MMC. 2009. Basic validation procedures for regression models in QSAR and QSPR studies: Theory and application. *Journal of Brazilian Chemical Society*. 2009;**20**:770-787. DOI: 10.1590/S0103-50532009000400021
- [78] Veerasamy R, Rajak H, Jain A, Sivadasan S, Varghese CP, et al. Validation of QSAR models - Strategies and importance. *International Journal of Drug Design Discovery*. 2011;**2**:511-519
- [79] Golbraikh A, Tropsha A. Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. *Molecular Diversity*. 2002;**5**:231-243. PMID: 12549674
- [80] Golbraikh A, Tropsha A. Beware of q^2 ! *Journal of Molecular Graphics and Modeling*. 2002;**20**: 269-276. DOI: 10.1016/S1093-3263(01)00123-1
- [81] Nahum OE, Yosipof A, Senderowitz H. A multi-objective genetic algorithm for outlier removal. *Journal of Chemical Information Modeling*. 2015;**55**:2507-2518. DOI: 10.1021/acs.jcim.5b00515
- [82] Pratim Roy P, Paul S, Mitra I, Roy K. On two novel parameters for validation of predictive QSAR models. *Molecules*. 2009;**14**:1660-1701. DOI: 10.3390/molecules14051660
- [83] Omisore NOA, Adewunmi CO, Iwalewa EO, Ngadjui BT, Adenowo TK, et al. Antitrichomonal and antioxidant activities of *Dorstenia barteri* and *Dorstenia convexa*. *Brazilian Journal of Medical and Biological Research*. 2005;**38**:1087-1094. DOI: 10.1590/S0100-879X2005000700012

- [84] Bouaziz M, Simmonds MS, Grayer RJ, Kite GC, Damak M. Flavonoids from *Hyparrhenia hirta* Stapf (Poaceae) growing in Tunisia. *Biochemical Systematics and Ecology*. 2001;**29**:849-851. DOI: 10.1016/S0305-1978(01)00028-X
- [85] Kuete V, Konga Simo I, Ngameni B, Bigoga DJ, Watchueng J, et al. Antimicrobial activity of the methanolic extract, fractions and four flavonoids from the twigs of *Dorstenia angusticornis* Engl. (Moraceae). *Journal of Ethnopharmacology*. 2007;**112**:271-277. DOI:10.1016/j.jep.2007.12.017
- [86] Fotie J. The antiprotozoan potential of flavonoids. *Pharmacognosy Reviews*. 2008;**2**:6-19
- [87] Ngameni B, Watchueng J, Boyom FF, Keumedjio F, Ngadjui BT, et al. Antimalarial prenylated chalcones from the twigs of *Dorstenia barteri* var. *subtriangularis*. *Arkivoc*. 2007;**13**:116-123
- [88] Kim JH, Lee BC, Kim JH, Sim GS, Lee DH, et al. The isolation and antioxidative effects of vitexin from *Acer palmatum*. *Archives of Pharmacal Research*. 2005 ;**28**:195. DOI:10.1007/BF02977715
- [89] Ragone MI, Sella M, Conforti P, Volonté MG, Consolini AE. The spasmolytic effect of *Aloysia citriodora*, Palau (South American cedrón) is partially due to its vitexin but not isovitexin on rat duodenums. *Journal of Ethnopharmacology*. 2007;**113**:258-266. DOI: 10.1016/j.jep.2007.06.003
- [90] Hernández MM, Heraso C, Villarreal ML, Vargas-Arispuro I, Aranda E. Biological activities of crude plant extracts from *Vitex trifolia* L. (Verbenaceae). *Journal of Ethnopharmacology*. 1999;**67**:37-44. DOI: 10.1016/S0378-8741(99)00041-0

