Journal of Pharmaceutical Research International

24(3): 1-17, 2018; Article no.JPRI.44771
ISSN: 2456-9119
(Past name: British Journal of Pharmaceutical Research, Past ISSN: 2231-2919, NLM ID: 101631759)

Identifying Potential Dual Inhibitory Candidates for Non-Small Cell Lung Cancer through Molecular Docking, 3D-QSAR Pharmacophore-based Virtual Screening, Comparative Molecular Field and Similarity Indices Analysis Modeling

Xiao X. Zhu[1], Yan Li[2], Xuan R. Zhang[1*], Lun Yuan[2], Pei H. Luo[1], X. Gao[1,2] and Zhong S. Tan[2]

[1]School of Chemical Engineering, Sichuan University, Chengdu 610065, China.
[2]Department of Pharmaceutical Engineering, Sichuan University, Chengdu 610041, China.

Authors' contributions

This work was carried out in collaboration between all authors. Authors XXZ, YL, XRZ and XG designed the study, wrote the protocol and the first draft of the manuscript. Authors XXZ, YL, LY, PHL and XG performed the computational modelling and statistical analysis. Authors ZST and XXZ managed the analysis of the study and the literature searches. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JPRI/2018/44771
Editor(s):
(1) Dr. Vasudevan Mani, Universiti Teknologi MARA (UiTM), Selangor, Malaysia.
Reviewers:
(1) Vesna Rastija, Josip Juraj Strossmayer University of Osijek, Croatia.
(2) P. Saravana Kumara, Rathnavel Subramaniam College of Arts and Science, India.
(3) Alok Nahata, Ying Zhi Agricultural and Industries Sdn Bhd, Malaysia.
(4) Pall Emoke, University of Agricultural Sciences and Veterinary Medicine of Cluj-Napoca, Romania.
Complete Peer review History: http://www.sciencedomain.org/review-history/27250

Original Research Article

Received 14 August 2018
Accepted 06 November 2018
Published 16 November 2018

ABSTRACT

Aim: This work aims to understand potential inhibitory structural requirements and identify lead compounds for non-small cell lung cancer through 3D-QSAR pharmacophore-based virtual screening, molecular docking, CoMSIA and CoMFA QSAR modelling.
Materials and Methods: QSAR pharmacophore models were developed by HypoGen Module and validated by test data set, Fischer's randomization and Guner-Henry equation. The well-validated

**Corresponding author: E-mail: icddpharmscu@hotmail.com;*

pharmacophore model was employed to perform virtual screening to identify potent hits from ZINC database. The retrieved hits were subsequently subjected to filtering using ADMET descriptors and Lipinski's Rule of Five. CoMSIA and CoMFA were then utilized to produce QSAR models on phenylpyrimidine derivatives also.

Results: Validations on 3D-QSAR pharmacophore model indicate that the enrichment factor is 6.34, GH is 0.517 and a correlation coefficient is 0.83, implying its highly predictive ability. Top three hits: ZINC29356266, ZINC06589615, and ZINC03375633 were identified as promising potent inhibitory candidates with IC_{50} value of about 0.54 μM and fitness value of about 59.4. Interestingly, the top three hits indicate dual inhibitory activity targeting EGFR and PD-L1 from structure-based docking. Two developed QSAR models from CoMSIA and CoMFA modelling indicate a potential predictive ability (q^2=0.67, and 0.71 respectively). The designed compound **C** indicates a more potential (dual) inhibitory activity (pIC_{50}=7.39) targeting EGFR (fitness=59.78) and CTLA-4 (fitness =47.90).

Conclusion: Validations indicate that the developed 3D-QSAR pharmacophore model is highly predictive. Top three hits were identified as promising potent inhibitory candidates and indicated dual inhibitory activity targeting EGFR and PD-L1. The designed compound C indicates a more potential (dual) inhibitory activity targeting EGFR and CTLA-4. These important 3D-QSAR and molecular docking bioinformatics results achieved from this work should be valuable in designing more promising potent inhibitory candidates and developing novel lead compounds against advanced NSCLC in future.

Keywords: Bioinformatics; molecular docking; QSAR; pharmacophore; comparative molecular similarity indices analysis.

1. INTRODUCTION

The population of people developing lung cancer is gradually increasing every year and lung cancer remains the leading cause of carcinoma related deaths in the world. About 82% of lung carcinoma is caused by non-small cell lung cancer (NSCLC) and about 18% of lung carcinoma is caused by small cell lung cancer (SCLC). Standard conventional therapies for patients developed with metastatic or advanced lung cancer include surgery, chemotherapy and radiation therapy. Nevertheless, these treatments usually may not be able to improve more long-term survival of patients [1-9]. Therefore, much research work has been focused on investigating the mechanisms of lung cancer cell survival and proliferation as well as identifying their corresponding molecular targets [3,6,10,11].

Recently, experimental detection of NSCLC genomes has identified several important genomic mutational targets, such as epidermal growth factor receptor (EGFR), anaplastic lymphoma kinase (ALK), KIT, squamous cell carcinoma (SCC) and other genomic drivers [6,11]. As a result, many NSCLC and SCLC inhibitors harbouring these targets have been discovered. Genetic aberrations in EGFR domain have been predicted as one key driver of NSCLC proliferation. Harbouring the target, EGFR inhibitors, such as afatinib, gefitinib and erlotinib have been approved for treating patients with advanced NSCLC in clinic [11-15]. However, clinical researches indicate that about half of lung cancers with the chemotherapy treatments may lead to leptomeningeal, brain or central nervous system (CNS) metastases [15,16]. Moreover, it was worthy of mentioning that erlotinib, currently approved in USA for third-line treating patients with NSCLC, is only to improve anticancer drug efficacy as monotherapy treatment in patients with advanced NSCLC. Disappointing results were also found in gefitinib and erlotinib combined with the treatment of chemotherapy [16,17]. The factors causing the results are still unclear.

Anaplastic lymphoma kinase (ALK) gene, frequently occurring in NSCLC and other tumor cells, is also considered as a chemotherapy target in NSCLC. However, patients with NSCLC frequently relapse with the treatment of ALK inhibitors, such as crizotinib, due to drug resistance development which becomes the major barrier avoiding ALK inhibitors [18]. The mechanism of ALK inhibitory resistance in clinic is also still unknown.

More recently, molecular pathway- immune checkpoints have emerged and been employed to treat patients with advanced NSCLC, such as the programmed cell death protein-1 (PD-1), programmed cell death ligand-1 (PD-L1) and cytotoxic T-lymphocyte- associated antigen-4 (CTLA-4). PD-L1 and CTLA-4 are usually expressed on the membrane of T cells, antigen presenting cells and lymphocytes [19,20]. PD-L1

may play an important role in managing T cell responses in cancer immunity or other bacterial pathogenesis, which has been considered as one key therapeutic target. CTLA-4 is a receptor binding to CD86 and CD80 which induce impeding signals responding to the immune system. Atezolizumab and Pembrolizumab, recently developed inhibitors against PD-L1/PD-1, have shown improvements in the objective response rate (ORR) and overall survival (OS), and greatly changed the management of SCLC and NSCLC, indicating milder side effects. Early studies also suggested that combined chemotherapy with patients in NSCLC targeting PD-L1/PD-1 and CTLA-4 may further improve anticancer bioactivity. In one preclinical investigation on the combination of ipilimumab and nivolumab, suitable bioactivity was observed in NSCLC, but with extremely toxicity [21,22]. Furthermore, it has been noted that development of lung cancer inhibitors using PD-L1/PD-1 or CTLA-4 as targeting biomarkers may be limited by experimental testing possibility or feasibility, such as the acquisition of a suitable sample, kinetics or dynamics and hetero-genetic characteristics of PD-L1 expression, and uncertainty of determining pattern location of PD-L1 expression etc. Meanwhile, the molecular mechanism of targeting PD-L1 and interaction between PD-L1/PD-1 and ligands remain incompletely understood [23-30].

Considering the limitations of available interaction mechanisms between NSCLC proteases and EGFR, PD-L1/PD-1 or CTLA□4 inhibitors as well as 3D-QSAR bioinformatics, and that some different factors interposing in antitumor cell proliferative bioactivity against NSCLC might not be observed *in vitro* or *in vivo*, in this paper we employed molecular docking and 3D QSAR pharmacophore modeling to discover potential lead compounds with novel scaffolds and investigate the mechanisms and modes of ligand-receptor binding against NSCLC as a supplement to the current inhibitory discovery. In addition, comparative molecular similarity analysis (CoMSIA) and comparative molecular field analysis (CoMFA) were also employed to design more promising inhibitory candidates against advanced NSCLC.

2. MATERIALS AND METHODS

2.1 Preparation of Ligand Data Set and Targeting Proteins

Sixty four known lung cancer inhibitors with diverse structures and experimental IC_{50} values were selected from the literatures [31-36]. Catalyst/Discovery studio as well as Tripos/SYBYL-X1.3-based standard guidelines were used for the selection of training and test set. Twenty one inhibitory molecules, of which IC_{50} values range from 0.5 to about 1000 μM and structures mainly contain indirubin, indoline, benzofuran, carbohydrazide and hydrazine functional groups, were utilized as training data set (TraDS) to build 3D-QSAR pharmacophore models (Fig. 1). Fifteen inhibitory molecules with similar structures and IC_{50} values were used as test data set (TesDS) to validate the built 3D-QSAR pharmacophore and an additional external test set of 5 compounds was chosen for validation on CoMFA models (Fig. 2). The remaining 23 compounds containing phenyl-pyrimidine functional groups were used to build CoMFA and CoMSIA models [37,38]. The crystallographic structures of targeting proteins such as EGFR, ALK, PD-L1 and CTLA‐4 were extracted from PDB (Protein Data Bank) deposited under accession codes 5FED, 2XP2, 4Z18, 5IUS and 3OSK, respectively, which were prepared for docking as described in previous work. ZINC database of about two million molecules was employed for 3D-pharmacophore-based virtual screening.

2.2 3D-pharmacophore Modeling, Validation and Virtual Screening

3D-QSAR pharmacophore models were built using DS 2.5. Feature mapping was initiated to determine important pharmacophore functionalities of those compounds in TraDS by an algorithm of molecular super-imposition. Conformations of 255 for each molecule were generated by Poling algorithm with the energy cutoff values of 83.6 kJ/mol to cover the best geometrical spatial conformation. Uncertain value of 2.0 was utilised for all compounds and the other parameters remained as defaults. Top four pharmacophore features, such as aromatic_ring (A_R), hydrogen bond acceptor (HB_A), hydrophobic (HP), hydrogen bond acceptor lipid (HB_AL) were employed to build 3D pharmacophore models by Hypogen algorithm, which predicts bioactivities for each compound through regression calculation correlating the fit value of 3D geometrical configuration at variance with the logarithm value of molecular bioactivity. Thus total ten pharmacophore hypotheses were developed, of which the best pharmacophore model was identified by Debnath' evaluation. The significance of those hypotheses was indicated

by statistical parameters such as total cost, fixed cost, correlation coefficient and RMSD. A correlating probability of above 85% between experimental and predicted bioactivities may be indicated by 42-68 bits for the difference in those costs [38].

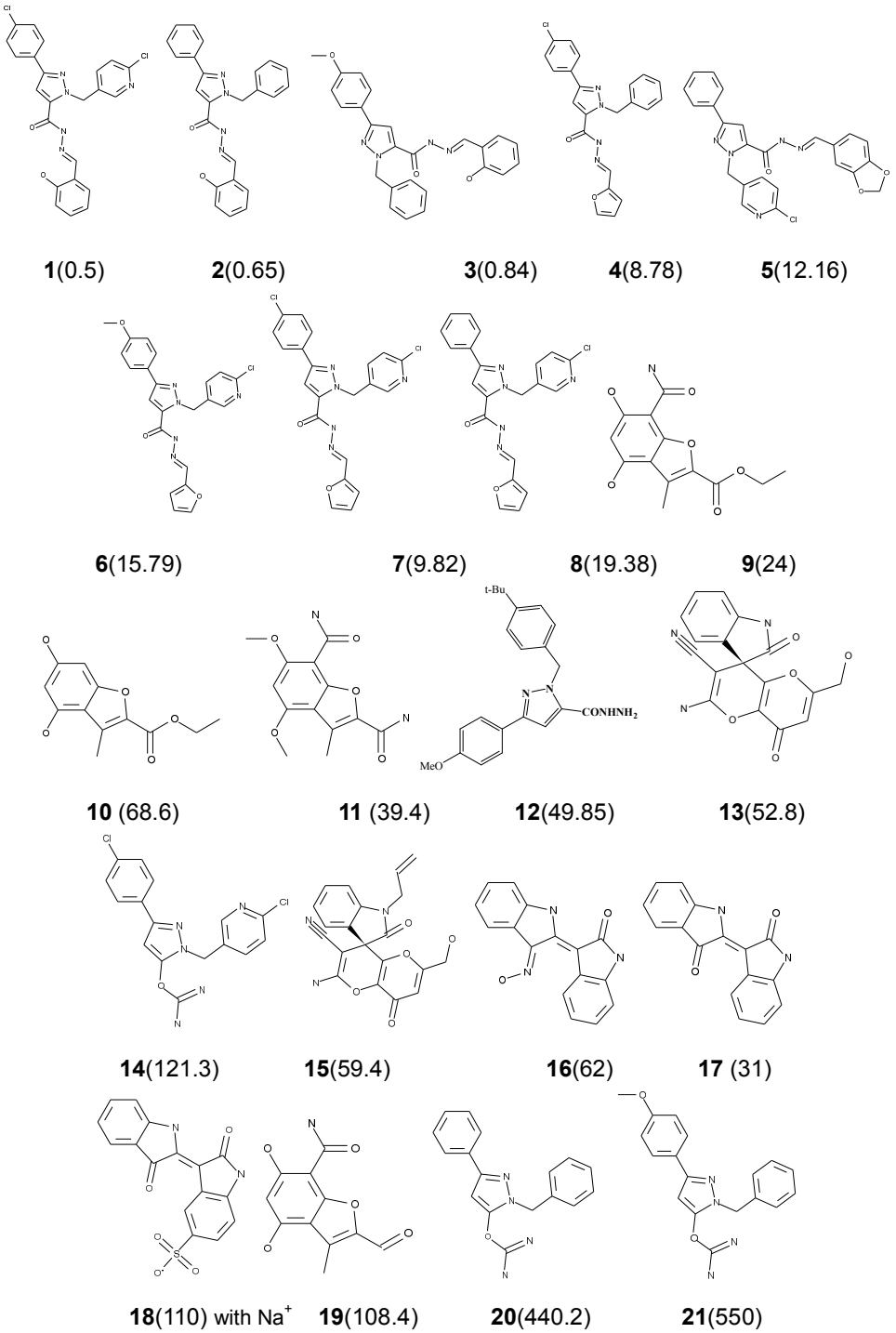


Fig. 1. 2D molecular configurations of 21 known lung cancer inhibitors used for developing QSAR pharmacophore models. Experimental IC_{50} (μM) values are given in parentheses, respectively

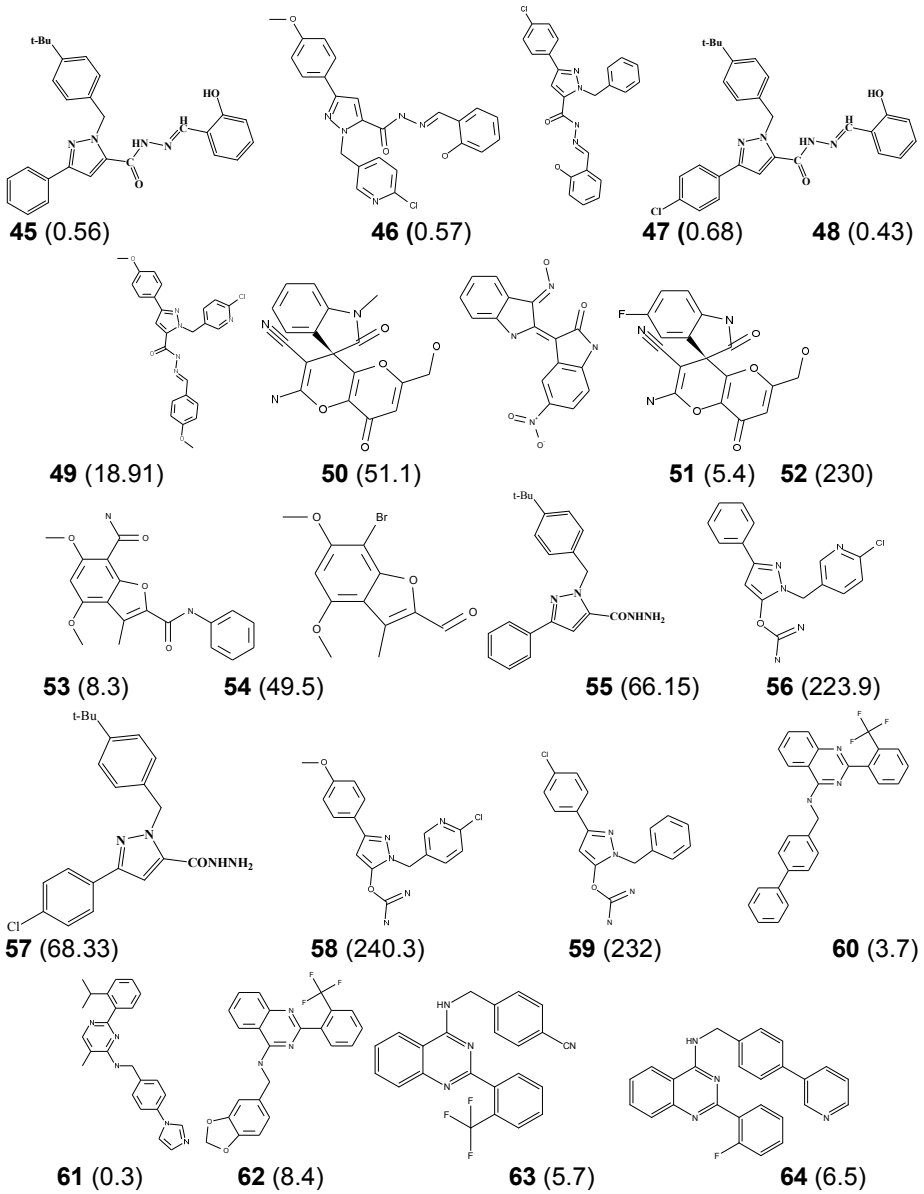


Fig. 2. 2D molecular configurations of 20 known lung cancer inhibitors used as test sets for validation on QSAR pharmacophore models (45-59) and on CoMFA models (60-64). Experimental IC_{50} (μM) values are given in parentheses, respectively

The best pharmacophore model developed (Hypo 1) was then validated using TesDS, Fischer's randomization and decoy test. Decoy test was conducted to evaluate the predictive ability of the developed model by following Guner-Henry (GH) scoring equation [39]:

$$E = \frac{H_a / H_t}{A / D}$$

$$GH = \frac{H_a}{4H_t \times A}\left(3A + H_t\right) \times \left(1 - \frac{H_t - H_a}{D - A}\right)$$

where GH is Guner-Henry factor, E is the enrichment factor. For the definition of H_a, H_t, A and D, readers may refer to ref. 39. In the decoy test, the database contains 128 chemicals (A), including 17 (D) known inhibitors with activity.

The final well-validated pharmacophore model was employed to perform virtual screening as a 3D configuration query with the flexible and best searching protocol to retrieve and identify potent hits with novel scaffolds from ZINC database. To obtain lead compounds and drug-like candidates, the retrieved hits were subsequently subjected to filtering using ADMET descriptors and Lipinski's Rule of Five.

2.3 Molecular Docking Modeling

Ligand binding affinity, binding mechanism and bioactivity against targeting proteins were predicted using GOLD program. Binding active site of each protein was defined by the coordinates of reference ligand bound to its protein and confined to the spatial region of 15 Å surrounding the reference ligand. A genetic algorithm was utilized to explore suitably 3D spatial configuration of ligands and partial flexibility of protein active sites. Binding affinity was computed from contributions of internal and external van der Waals (vdW), hydrophobic properties (HP) and hydrogen bond (HB). Initially, the cavities of binding active site were detected by a grid point spacing around 0.25 Å. At each grid point, ligand-protein interaction energies, such as HB, HP and vdW forces between the protein atoms and a hybridised C-sp^3 probe, were computed using Goldscore (empirical scoring) function, while total binding free energy difference was calculated using Chemscore function which was also utilised for rescoring. 10 space conformations were generated for each ligand and the best configuration was determined in terms of ligand-protein interaction energies during docking. Finally, binding affinity and bioactivity of ligands were evaluated by calculated fitness [39,40].

2.4 3D-QSAR CoMFA and CoMSIA Modeling

CoMSIA and CoMFA modelling were performed on 23 compounds containing phenyl-pyrimidine derivatives using SYBYL-X1.3. Prior to CoMFA modelling, all 23 molecules were docked into the active site of the protein (5L8E) to align the coordinates of compounds according to suitable active configurations by flexible docking. CoMFA coulombic electrostatic and Lennard-Jones (6-12) steric field descriptors were calculated using a 1.50 Å vdW radius and a probe atom of hybridized C-sp^3 with a +1 charge at a 3D spaced grid point of 2.0 Å. A filtering value of 7.52 kJ/mol was accepted to improve CoMFA simulating signal and a cutoff energy of 125.4 kJ/mol was set to truncate electrostatic and steric energies [41].

As a supplement to CoMFA modelling, hydrogen bond donor (HBD), hydrogen bond acceptor (HBA) and HP fields (descriptors) were calculated by CoMSIA modelling with a grid lattice. Gaussian function was utilised to compute CoMSIA similarity indices of five fields between the probe atom and a molecule with parameters such as hydrophobicity +1 and HB accepting or donating +1. Similar parameters employed in CoMFA modeling were also utilized in CoMSIA calculations.

3D QSAR models were then generated by partial least-squares (PLS) calculation following CoMSIA and CoMFA descriptor computations. PLS regression is suitable to correlate the bioactivity (dependent variables) with the descriptors (independent variables). Leave-one-out (LOO) was utilised to perform cross-validation and then evaluate the significance of built models. 4.18 kJ/mol column filtering was used to enhance 3D-QSAR modelling efficiency and improve signal. The predictive ability of built models was subsequently evaluated by a calculated statistical correlation coefficient q^2 of the cross-validation, an r^2-conventional correlation coefficient and the external test set [42].

Critical interpretation of the CoMFA and CoMSIA contour maps and docking results may be used for identifying key molecular structural features which might be investigated for designing new compounds with more potent inhibitory activity.

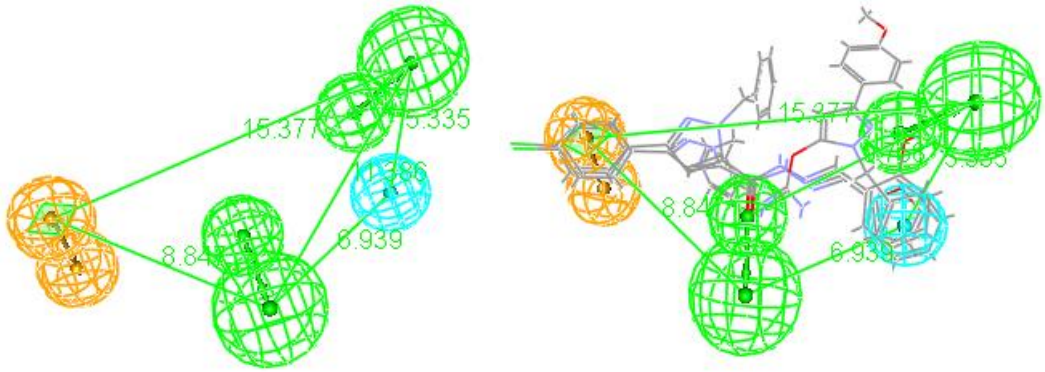
3. RESULTS AND DISCUSSION

3.1 3D-pharmacophore Modeling, Validation and Virtual Screening

In order to correlate reasonably bioactivity against lung cancer with 3D molecular configuration in TraDS, a number of common pharmacophore models (hypotheses) with the combination of several features were produced in DS. The top-ranked model (Hypo 1) includes four key features: A_R, HB_A, HP and HD_AL, as shown in Fig. 3(a). Computed statistical parameters for the model is given as followings: correlation coefficient is ~ 0.97, the configuration cost is 14.97 (between 12 and 17 to ensure that all possible spatial configurations have been included), the cost difference between the fixed

and null cost is ~79.84 (≥ 78 bits), and the difference between total and null cost is 69.27 (>62 bits for a >91% probability that a model may indicate a real correlation with experimental bioactivity). Furthermore, the bioactivities of all compounds were effectively predicted by their corresponding errors ranging from -1.0 to +2.2 as shown in Table 1, demonstrating that the developed model correctly identified the most key geometrical configuration elements involved in anticancer bioactivity of these compounds [38,40].

Validation on the developed model using TesDS leads to a good correlation coefficient of 0.83 (experimental log IC_{50} versus predicted log IC_{50}) and 19 randomised runs from Fischer validation all show relatively higher cost values and lower correlation coefficients compared with the developed hypothesis, representing a 95%



(a) (b)

Fig. 3. (a) The developed pharmacophore model. A_R is represented by spheres in orange, HP is represented by a sphere in cyan, HB_A, and HB_AL are represented by spheres in green. The 3D spatial distances among the features are given in Å. (b) The mapping of inhibitor 1 and 21 in TraDS with the model

Table 1. Predicted activity and statistical parameters obtained from the developed QSAR pharmacophore modeling for 21 compounds in training data set

Compound	Active*/μM	Predicted(IC_{50})/μM	Fitted value	Error
1	0.50	0.4344	7.0534	-1.1510
2	0.65	0.4681	7.0209	-1.3884
3	0.84	0.8267	6.7739	-1.0161
4	8.78	13.0331	5.5762	1.4844
5	12.16	14.5777	5.5276	1.1988
6	15.79	15.2522	5.5079	-1.0352
7	9.82	17.0891	5.4585	1.7402
8	19.38	20.1956	5.3860	1.0420
9	24.00	30.2769	5.2101	1.2615
10	68.60	31.6754	5.1905	-2.1657
11	39.40	44.2336	5.0455	1.1226
12	49.85	45.6539	5.0318	-1.0919
13	52.80	50.6493	4.9867	-1.0424
14	121.30	60.0617	4.9127	-2.0195
15	59.40	61.3863	4.9032	1.0334
16	62.00	104.5561	4.6719	1.6863
17	31.00	110.5061	4.6479	3.5647
18	110.00	119.4970	4.6139	1.0863
19	108.40	134.5440	4.5624	1.2411
20	440.20	199.0670	4.3923	-2.2113
21	550.00	263.1890	4.2710	-2.0897

**experimental IC_{50} values*

confidence level (Fig. 4). Computation on the decoy test shows that the enrichment factor (E) is 6.34 and GH is 0.517. Fig. 3 (b) shows that the lowest active molecule **21** in TreDS was only mapped only with two key features, while the highest active molecule **1** was completely mapped with all four key features. All these calculations above supported the significance of the developed model, demonstrating its high predictive ability [38,39].

The developed 3D QSAR pharmacophore model was subsequently employed as a query to screen the hits for novel and potential lead compounds from ZINC database. An active cutoff value of 0.9 μM (IC_{50}) was set to select chemicals. Seventy eight hits with predicted active values (IC_{50}) ranging from 0.4 to 0.9 μM were obtained by virtual screening of about 2.5 million chemicals. ADMET descriptors and Lipinski's rule of five were then applied to filter these hits to evaluate their bioavailability. Seventeen hits finally met key parameters for drug-like lead compounds, such as toxicity, percent human oral absorption, blood brain barrier (BBB), penetration, etc.

3.2 Molecular Docking

In order to further evaluate anticancer bioactivity of 17 hits, these hits were docked into domain region of active sites of EGFR (PDB code: 5FED), PD-L1 (PDB: 4Z18), ALK (PDB: 2XP2), respectively. Docking scores (fitness values) determined by molecular docking simulations for 17 hits range from 32 to 70. Top three hits with novel structures, i. e. ZINC29356266 (IC_{50} value is 0.8655 μM, fitness value is 63.04 for EGFR and 40.50 for PD-L1), ZINC06589615 (IC_{50} value is 0.7140 μM, fitness value is 56.79 for EGFR and 41.93 for PD-L1) and ZINC03375633 (IC_{50} value is 0.6374μM, fitness value is 51.94 for EGFR and 42.39 for ALK), were identified as lead compounds or promising potential inhibitory candidates against advanced NSCLC in terms of higher fitness value and bioactivity as shown in Fig. 5.

The three compounds show strong interaction with NSCLC proteases. Compound ZINC29356266 forms four hydrogen bonds (HB) with SER7-ER72, MET7-ET79 GLN7-LN79 and THR7-HR79 residues of EGFR, and also three HBs with SER1-ER19, THR-HR17 and VAL1-AL16 residues of PD-L1, respectively, as shown in Fig. 6 (a) and (b). In the active site of EGFR, oxygen atom on furan ring interacts with amide group of SER7-ER72 to form a HB with O→H distance of 2.198 Å, while oxygen atom and amide group on phenyl-morpholine double rings make the other three HBs with amide and hydroxyl groups of MET7-ET79, GLN7-LN79 and THR7-HR79 respectively. Van der waals bonds and hydrophobic interaction were observed between triazole and phenyl ring of the compound and hydroxyl group of THR790 and carbonyl group of PRO794 respectively, which may further stabilize the binding to EGFR, leading to relatively higher binding affinity.

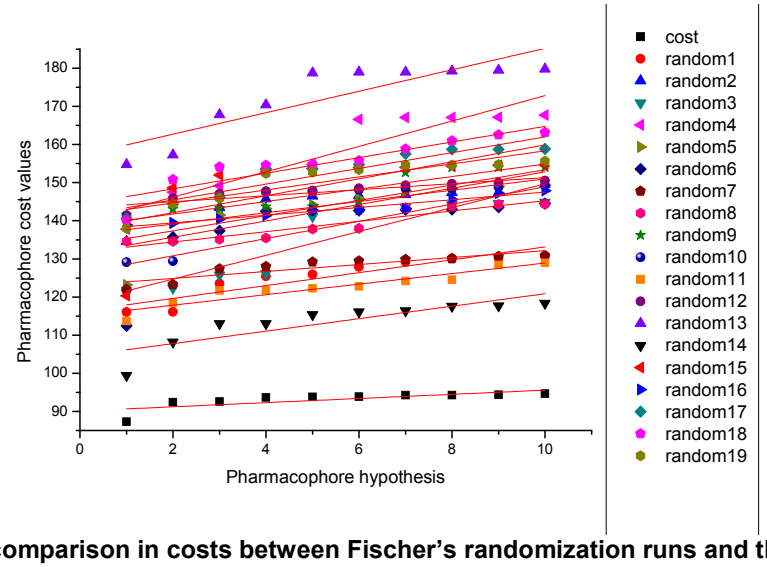
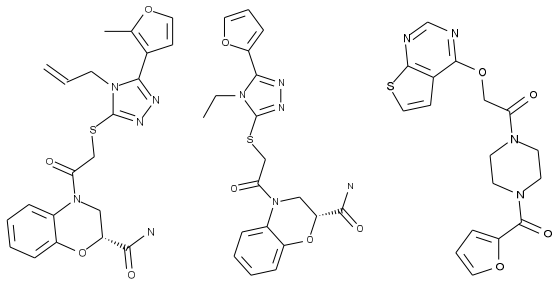


Fig. 4. The comparison in costs between Fischer's randomization runs and the developed QSAR pharmacophore model



ZINC29356266 ZINC06589615 ZINC03375633

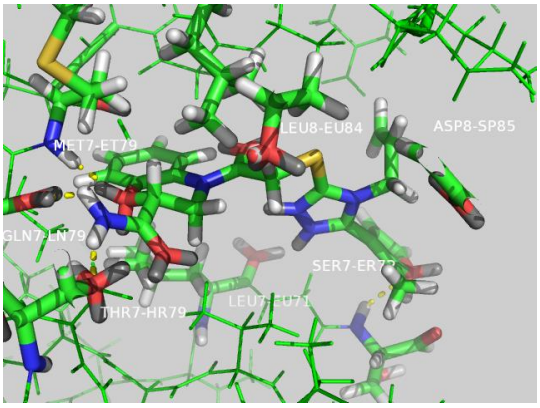
Fig. 5. 2D configurations of three lead compounds identified by the developed QSAR pharmacophore-based virtual screening and molecular docking. ZINC29356266 (IC_{50}=0.8655 μM, fitness value is 63.04 for EGFR and 40.50 for PD-L1); ZINC06589615 (IC_{50}=0.7140 μM, fitness value is 56.79 for EGFR and 41.93 for PD-L1) and ZINC03375633 (IC_{50}=0.6374μM, fitness value is 51.94 for EGFR and 42.39 for ALK)

In the active site of PD-L1, carbonyl group on the double rings of ZINC29356266 interacts with hydroxyl group of THR179 and VAL165 to form two HBs respectively, while amide group on the double rings makes the other HB with hydroxyl group of SER1 at N→H distance of 2.27 Å. $\pi-\pi$ stacking was observed between triazole ring and methyl group of GLU 164. Van der waals bonds were also formed between phenyl ring and carbonyl group of the compound and oxygen atom of GLU 164 and carbonyl group of GLY177 respectively. However, hydrophobic interaction was absent from triazole and the double rings, which may account for relatively lower binding affinity.

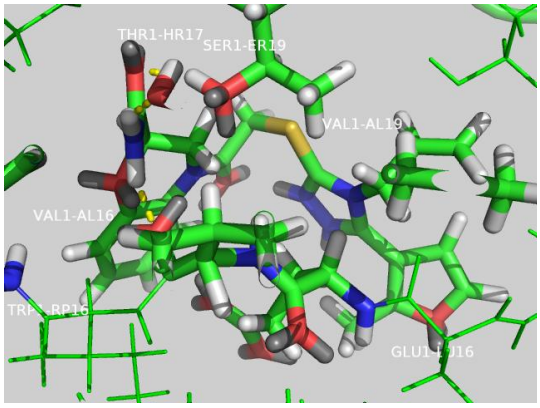
Compound ZINC06589615 forms four HBs with MET79, GLY724, THR79 and LEU718 residues of EGFR, and also three HBs with THR182,THR179 and GLY1-LY17 residues of PD-L1, respectively, as shown in Fig. 6 (c) and (d). In the active site of EGFR, N atom on triazole ring and carbonyl group on phenyl-morpholine double rings interact with hydroxyl group of THR79 and S atom of LEU718 to form two HBs. Furan ring makes the other two HBs with amide group of GLY 724 (Fig. 6 (c)). Similarly, Van der waals bonds and hydrophobic interaction were observed between the double rings and ethyl group of the compound and amide group of LYS85 and methyl group of LEU79 respectively. For the two similar scaffolds (ZINC29356266 and ZINC06589615), functional group triazole and the double rings make more strong interaction with THR 79 and MET 79 which may be identified as key binding residues in active site of EGFR.

In the active site of PD-L1, amide group connected to the double ring and triazole ring of ZINC06589615 interact with hydroxyl groups of THR181 and THR 179 to form HBs at N→H distance of ~2.74 Å respectively. Oxygen atom on furan ring forms the other HB with amide group of GLY17. $\pi-\pi$ stacking was absent from triazole ring and methyl group. The binding was also stabilised by hydrophobic interaction and Van der waals bonds formed between carbonyl group and furan ring of the compound and hydroxyl group of ALA16 and ethyl group of SER1-ER17. It is quite interesting that both ZINC29356266 and ZINC06589615 indicate dual potential inhibitory activity targeting EGFR and PD-L1 from docking simulations as well as 3D-QSAR pharmacophore mapping.

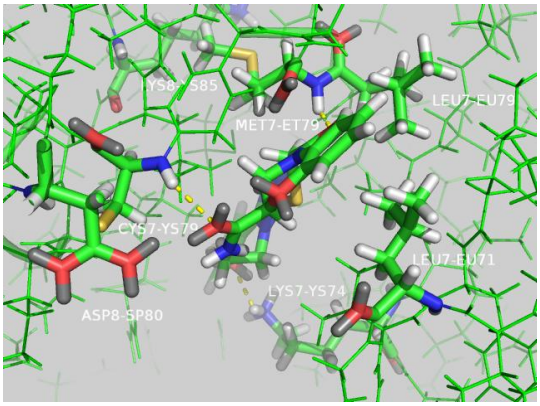
Compound ZINC03375633, of which structure is different from above two compounds, forms two HBs with LYS745 and MET793 residues of EGFR. N atoms on thoiphene- pyridazine double ring and carbonyl group interact with amide groups of LYS745 and MET793 to form two HBs respectively. More strong hydrophobic interaction and Van der waals bonds were observed between methyl group, S atom, furan ring, N atoms on the double rings of the compound and methyl group and amide group of residue LEU718 and LYS745 respectively. In the active site of ALK, the compound also makes two HBs with residue LYS1-YS11 and LEU1-EU12 as well as hydrophobic interaction and Van der waals bonds with residue ASP1-SP12, LEU1-EU12 and PHE1-HE12 respectively. ZINC03375633 indicates double potential inhibitory bioactivity targeting EGFR and ALK also.



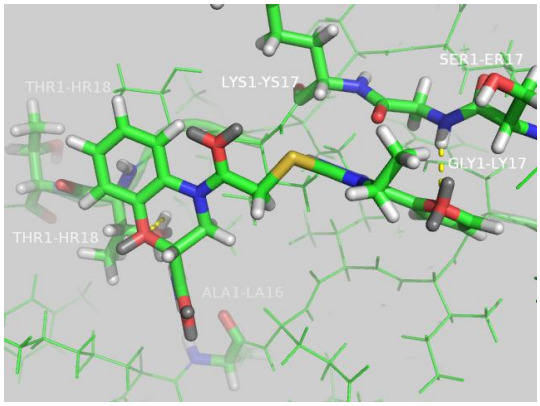
(a) Zinc29536266 binding to EGFR (5FED)



(b) Zinc29536266 binding to PD-L1 (4Z18)

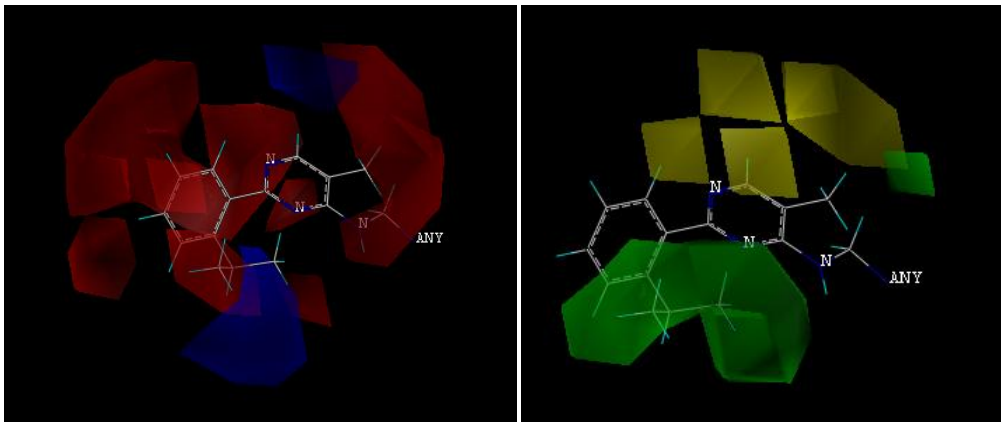


(c) Zinc06589615 binding to EFGR(5FED)

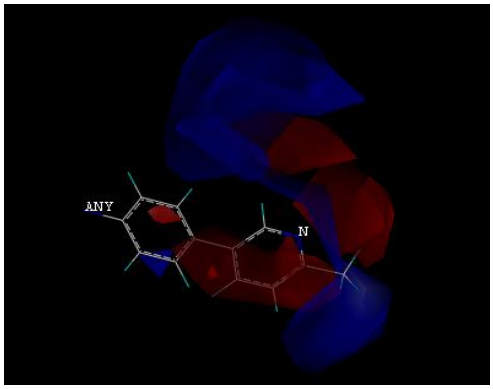


(d) Zinc06589615 binding to PD-L1(4Z18)

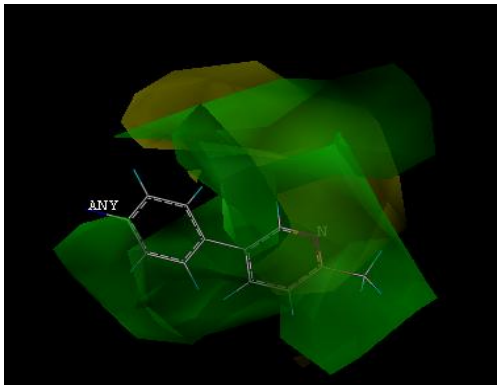
Fig. 6. The binding modes of Zinc29536266 and Zinc06589615 to EFGR and PD-L1, respectively. The hydrogen bonds are represented by dotted lines in yellow. Important residues are labeled and represented by sticks



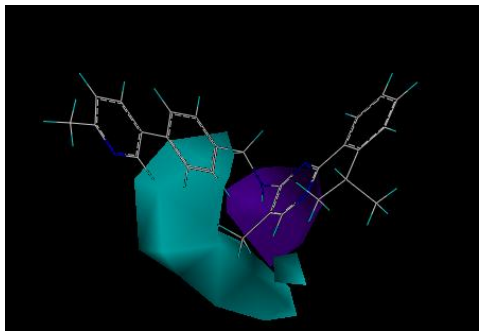
(a) R1 electrostatic contribution: blue 4.25 red 1.5; steric contribution: green 5.89.yellow 19.64



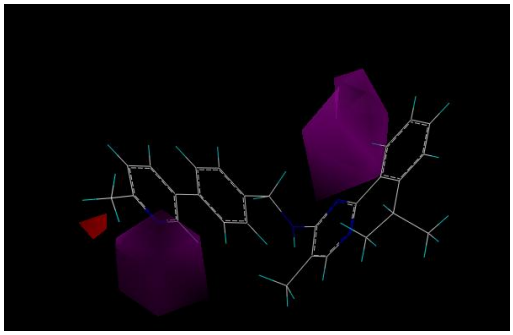
(b) R2 electrostatic contribution: blue 4.27 red 31.62;



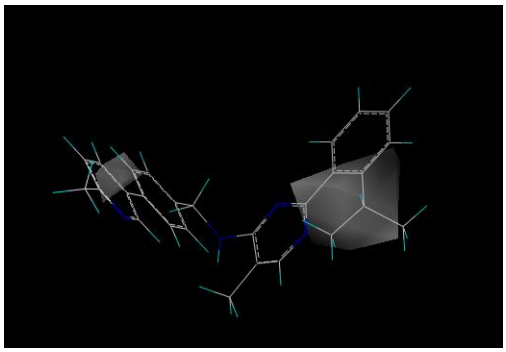
steric contribution: green 3.58 yellow 9.22



(c) HB donor field



(d) HB acceptor field



(e) Hydrophobic field

Fig. 7. Contour maps of developed CoMFA (a), (b) and CoMSIA (c), (d), (e) models for reference compound 37

3.3 CoMSIA and CoMFA Modeling

Considering the limitation of pharmacophore-based QSAR virtual screening and designing more potential inhibitory candidates, constructive CoMFA and CoMSIA modeling on 23 phenylpyrimidin derivatives (Table 2) were performed using their aligned configurations from flexible docking. It is worthy of mentioning that a prerequisite for developing 3D QSAR CoMSIA and CoMFA models is how to obtain suitable location-orientation and active configuration for each compound while determining proper configuration is a challenging assignment for some compounds. One known compound is usually utilised as a reference to align all other

compounds in a traditional CoMFA or CoMSIA alignment, leading to a fact that in some cases most aligned compounds might not fit into the corresponding active site of a receptor. Nevertheless, in docking-based alignment procedure used here, the feature of binding active site of the receptor is introduced into CoMFA or CoMSIA QSAR model, to ensure that all computed CoMFA descriptors are determined from real and suitable poses binding to the receptor and thus to guarantee that proper lead compounds may be found [41,42]. The statistical analysis results for both CoMSIA and CoMFA modelling are given in Table 3. The cross validation coefficient (q^2) and non-cross validation coefficient (R^2) of the 3D-QSAR model built by CoMFA are 0.72 and 0.99, respectively. The q^2 and R^2 values for CoMSIA model are 0.67 and 0.98, respectively. *SEE* values (standard errors of estimate) for CoMSIA and CoMFA are 0.13 and 0.07, respectively. CoMFA model appears to be slightly better on predictive ability than CoMSIA model. Nevertheless, the high *F* test value (1593 and 161.11 for CoMFA and CoMSIA model respectively) and bootstrapping r^2 values show the reliability of the two models constructed, while the cross-validation analyses further indicate that both CoMSIA and CoMFA are highly predictive as exhibited by the external test set (q^2=0.73) and those predicted values shown in Table 3.

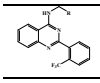
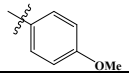
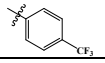
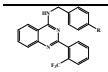
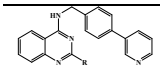
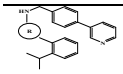
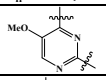
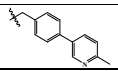
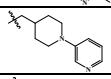
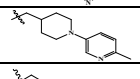
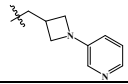
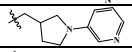
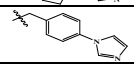
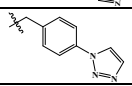
Contour maps of reference compound **37** (pIC_{50}=7.34 in Table 2) by CoMFA modelling are shown in Fig. 7 (a) and (b). Methyl group on pyridine ring occupies positive charge favoured blue area, and pyrimidine ring occupies the electronegative favoured red area. While pyridine ring, locating at positive charge favoured area in blue instead of methyl group in compound **41**, tends to make the region electronic density more negative, resulting in lower inhibitory activity (pIC_{50}=4.66). Three negative charge nitrogen atoms on triazole ring in compound **44** are situated at favoured electronegative area in red, exhibiting relatively high inhibitory bioactivity (pIC_{50}=7.10), implying the significance of negative charge functions at this location. In addition, compound **40** where substituent-methyl group at pyridine ring and the ring are situated at green area has relatively high bioactivity (pIC_{50}=7.09), implying the importance of sterically large groups in this location. Occupancy of sterically unfavourable region in yellow around the pyrimidine ring, for example, compounds **25** and **31** with substituents methyl and ethyl groups at the pyrimidine ring, may result in relatively lower bioactivity, pIC_{50}=5.54 and 6.82, respectively.

Contour maps of reference compound **37** by CoMSIA modelling are shown in Fig. 7 (c)-(e). Methyl group at pyrimidine ring and phenyl ring are recognized as favoured HB donor region in cyan, and pyrimidine ring is marked as unfavourable HB donor area in purple in the compound (Fig. 7 (c)). Azedidine, pyridine and pyrrole rings occupy cyan favoured HB donor area instead of phenyl ring in compounds **41**, **39** and **40** respectively, hence relatively lower inhibitory activity is predicted for the three compounds respectively (pIC_{50}=4.66, 7.10, 7.09). While phenyl ring attached to pyrimidine ring, occupying purple region, also leads to the decrease in bioactivity in compounds **22**, **23**, **24**, demonstrating the importance of HB donor in this area. Pyrimidine and pyridine rings with substituent- methyl group are marked as HB acceptor favoured region in magenta (Fig. 7 (d)), but unfavourable HB acceptor region in red is likely not to be important in the reference **37**. Compound **34**, **35** with substituent-MeO and fluorine atom at pyrimidine ring occupying magenta favoured HB acceptor region have relatively higher bioactivity (pIC_{50}=7.15, 6.96 respectively), demonstrating the role of HB acceptor in this direction. The favored hydrophobic region in white grey is occupied by two phenyl rings at both sides of the compound **37** (Fig. 7 (e)). 2-trifluoride methyl and pyridine attached to the phenyl in compound **27** and an addition of morpholine ring instead of the phenyl at the position of amine group in compound **26**, occupying white grey favoured hydrophobic region respectively, may lead to the decrease in bioactivity (pIC_{50}=5.31, 4.98, respectively), suggesting the role of hydrophobic groups in the regions.

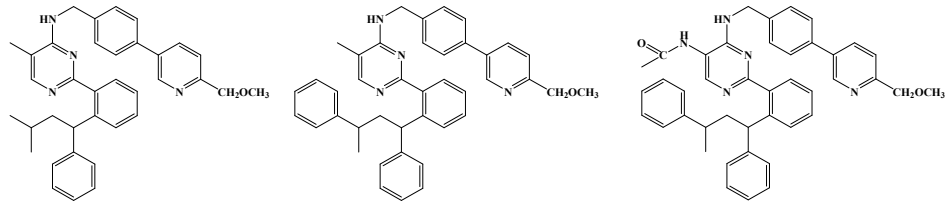
3.4 Design of Novel Compounds

Critical analysis of the CoMFA and CoMSIA contour maps results in identifying key molecular structural features which might be investigated for improving the inhibitory activity of the most potent inhibitor-reference molecule **37**. It may be deduced from CoMSIA and CoMFA QSAR modeling and molecular docking that the bioactivity and binding affinity against advanced NSCLC for phenylpyrimidine derivatives could be improved by substituents such as Ph-$CONH_2$, carboxyl-NH_2 and hydrophilic groups (-$CONH_2$, -CH_2OH) at the position of pyrimidine ring, and ethane, phenyl, bulk and hydrophilic groups at the position of pyridine ring.

Table 2. 2D molecular structures of compounds used for CoMFA and CoMSIA QSAR modeling and their bioactivities predicted from the two models

Compound	R	pIC_{50}*(μM)	Predicted CoMFA	Predicted CoMSIA
22		5.11	5.08	5.09
 **23**		5.21	5.18	5.23
24		5.05	5.13	5.07
25		5.52	5.54	5.50
26		4.98	4.95	4.89
 **27**		5.31	5.34	5.30
28		5.01	5.13	5.06
29		6.03	6.05	6.01
 **30**		5.67	5.76	5.71
31		6.82	6.84	_
 **32**		7.15	7.13	_
33		6.92	6.97	6.96
34		7.15	7.18	7.11
35		6.96	6.95	6.93
36		6.57	6.49	6.53
 **37**		7.30	7.34	7.31
38		6.85	6.78	6.84
39		7.10	7.03	7.07
40		7.09	7.04	7.08
41		4.66	4.63	4.68
42		6.09	6.03	6.07
43		6.52	6.58	6.60
44		7.12	7.10	7.11

**experimental IC_{50} values*



A (7.27) **B (7.30)** **C (7.39) (fitness value is 59.78 for EGFR and 47.90 for CTLA‐4**

Fig. 8. 2D molecular configurations of designed lead compounds and their corresponding activity predicted from CoMSIA and CoMFA modeling and molecular docking. pIC_{50} and fitness values are given in parentheses respectively

Table 3. Calculated statistics and cross-validation parameters from CoMSIA and CoMFA modeling

Model	Component	q^2	R^2	*F*	*SEE*	*SEE*
CoMFA	7	0.72	0.99	1593	0.04	0.04
CoMSIA	6	0.67	0.98	161.11	0.13	0.13

Considering the importance of red and blue contour maps shown in CoMFA contours in the corresponding pyrimidine and pyridine moiety, we might decide to increase electronegative favoured nature of the pyrimidine ring and positive charge favoured nature of pyridine ring as well as keep sterically large groups in the pyridine moiety making certain that the new moiety may form hydrophobic interactions with ARG 272, ASN 212 and LYS 190. Therefore, we introduced carbonyl and amide groups in the pyrimidine moiety, and methyl-O-methyl group in the pyridine moiety indicating an increase in inhibitory activity (compound **A**, **B** and **C** in Fig. 8).

Based on the green cyan, magenta and white grey contour maps shown in CoMSIA contours, we also might decide to increase HB donor nature of the pyrimidine ring and HB acceptor nature of phenyl ring as well as the steric bulk in the phenyl moiety. Hence, we also introduced methyl group connecting to carbonyl group in the pyrimidine moiety, and two additional phenyl rings in the phenyl moiety indicating an increase in inhibitory activity (compound **B** and **C**).

In order to further evaluate each bioactivity, ATMET filtering and molecular docking of the three compounds into 5L8E protein was performed. As shown in Fig. 8, their bioactivities are all relatively higher than those in TraDS. Designed compound **C** was then docked into EGFR and CTLA‐4 respectively, indicating more potent inhibitory activity (pIC_{50}=7.39) and binding affinity (fitness value is 59.78 for EGFR and 47.90 for CTLA‐4) as well as dual inhibitory bioactivity.

4. CONCLUSION

Validations on the satisfactory QSAR pharmacophore model using decoy test, TesDS and Fischer Fischer's randomisation indicate that the enrichment factor is 6.34, GH is 0.517 and a correlation coefficient is 0.83, supporting the significance of the developed model and implying its highly predictive ability. Top three hits: ZINC29356266, ZINC06589615, and ZINC03375633 were identified as promising potent inhibitory candidates with IC_{50} value of about 0.54 μM and fitness value of about 59.4. Interestingly, ZINC29356266 and ZINC06589615 indicate dual inhibitory activity targeting EGFR (fitness value =63.04, 56.79, respectively) and PD-L1 (fitness=40.50, 41.93, respectively), and ZINC03375633 also shows dual inhibitory activity targeting EGFR (fitness=51.94) and ALK (fitness=42.39) from structure-based docking.

CoMSIA and CoMFA modelling on phenylpyrimidine derivatives, developed using flexible docking configurations, was then utilized to produce satisfactory QSAR models also. The two developed QSAR models indicate a potential predictive ability (q^2=0.67, and 0.71 respectively). Considering the structural requirements from CoMSIA and CoMFA QSAR modelling, three new compounds were designed and the designed compound **C** indicates a more potential (dual) inhibitory activity (pIC_{50}=7.39) targeting EGFR (fitness=59.78) and CTLA□4 (fitness =47.90).

These important 3D-QSAR and molecular docking bioinformatics results achieved from this work should be valuable in designing more promising potent inhibitory candidates and developing novel lead compounds against advanced NSCLC in future.

CONSENT

It is not applicable.

ETHICAL APPROVAL

It is not applicable.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Pitrodal SP, Stack ME, Liu GF, Song SS, Chen L, et al. JAK2 inhibitor SAR302503 abrogates PD-L1 expression and targets therapy-resistant non–small cell lung cancers. Mol Cancer Therap. 2018;17(4): 732-739.
2. Soria JC, Marabelle A, Brahmer JR, Gettinger S. Immune checkpoint modulation for non-small cell lung cancer. Clin Cancer Res. 2015;21(10):2256–62.
3. Van Allen EM, Golay HG, Liu Y, Koyama S, Wong K, et al. Long-term benefit of PD-L1 blockade in lung cancer associated with JAK. Cancer Immu Res. 2015;3(8):855-863.
4. Govindan R, Ding L, Griffith M, Subramanian J, Dees ND, Kanchi KL, et al. Genomic landscape of non-small cell lung cancer in smokers and neversmokers. Cell. 2012;150:1121-1134.
5. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science. 2015;348:124-128.
6. Tan PS, Aguiar Jr. P, Haaland B, Lopes G. Comparative effectiveness of immune-checkpoint inhibitors for previously treated advanced non-small cell lung cancer – A systematic review and network meta-analysis of 3024 participants. Lung Cancer. 2018;15:84-88.
7. Chen DS, Irving BA, Hodi FS. Molecular pathways: Next-generation immunotherapy? inhibiting programmed death-ligand 1 and programmed death-1. Clin Cancer Res. 2012;18:6580–6587.
8. Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus docetaxel in advanced nonsquamous non-Small-Cell lung cancer. New Eng J Med. 2015;373:1627–1639.
9. Rittmeyer A, Barlesi F, Waterkamp D, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): A phase 3, openlabel, multicentre randomised controlled trial. The Lancet. 2017;389:255–265.
10. Fehrenbacher L, Spira A, Ballinger M. Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): A multicentre, open-label, phase 2 randomised controlled trial. The Lancet. 2016;387:1837-1846.
11. Pick A, Wiese M. Tyrosine kinase inhibitors influence ABCG2 expression in EGFR-positive MDCK BCRP cells via the PI3K/AKT signaling pathway. Chem Med Chem. 2012;7(4):650-662.
12. Zhang YK, Zhang GN, Wang YJ, Patel BA, Talele TT, Yang DH, et al. Bafetinib (INNO-406) reverses multidrug resistance by inhibiting the efflux function of ABCB1 and ABCG2 transporters. Sci Rep. 2016;6: 25694.
13. Ramalingam SS, O'Byrne K, Boyer M, Mok T, Janne PA, Zhang H, et al. Dacomitinib versus erlotinib in patients with EGFR-mutated advanced nonsmall cell lung cancer (NSCLC): Pooled subset analyses from two randomized trials. Ann Onco. 2016;27(7):423-429.
14. Zhang GN, Zhang Y, Wang YJ, Gupta P, Ashby Jr. CR, Alqahtani S. Epidermal growth factor receptor (EGFR) inhibitor PD153035 reverses ABCG2-mediated multidrug resistance in non-small cell lung cancer: *In vitro* and *in vivo*. Cancer let. 2018;424:19-29.
15. Zeng Q, Wang J, Cheng Z, Chen K, Johnström P. et al. Discovery and evaluation of clinical candidate AZD3759, a potent, oral Active, central nervous system-penetrant, epidermal growth factor receptor tyrosine kinase inhibitor. J Med Chem. 2015;58:8200−8215.
16. Gril B, Evans L, Palmieri D, Steeg PS. Translational research in brain metastasis is identifying molecular pathways that may lead to the development of new therapeutic strategies. Euro J Cancer. 2016;46:1204−1210.

17. Herbst RS, Sandler AB. Overview of the status of human epidermal growth factor receptor inhibitors in lung cancer. Clin Lung Cancer. 2004;6:S7-S19.
18. Dardaei L, Wang HQ, Singh M, Fordjour P, Shaw KX. SHP2 inhibition restores sensitivity in ALK-rearranged non-small-cell lung cancer resistant to ALK inhibitors. Nat Med. 2018;24:512-517.
19. Bianco A, Malapelle U, Rocco D, Perrotta F, Mazzarella G. Targeting immune checkpoints in non small cell lung cancer. Curr Opinion in Pharm. 2018;40:46-50.
20. de Mello RA, Veloso AF, Catarina PE, Nadine S, Antoniou G. Potential role of immunotherapy in advanced non-small-cell lung cancer. Onco Therapy. 2017;10:21-30.
21. Tumeh PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJM, Robert J, et al. PD- 1 blockade induces responses by inhibiting adaptive immune resistance. Nature. 2014;568-571.
22. Antonia SJ, Gettinger SN, Chow LQM, et al. Nivolumab (anti‐PD‐1; BMS‐936558, ONO‐4538) and ipilimumab in first‐line NSCLC: Interim phase I results. J Clin Onco. 2014; 32(suppl):8023a.
23. Chan CWH, Tsui S, Law BHW, So WK, Tang FW, Wong CL. The utilization of the immune system in lung cancer treatment: Beyond chemotherapy. Inter J Mol Sci. 2016;17:286.
24. Hansen AR, Siu LL. PD-L1 testing in cancer: Challenges in companion diagnostic development. JAMA Onco. 2016;2:15- 16.
25. Morabito A. Second-line treatment for advanced NSCLC without actionable mutations: Is immunotherapy the 'panacea' for all patients? Morabito BMC Med. 2018; 16:24.
26. Reck M, Popat S, Reinmuth N, De Ruysscher D, Kerr KM, Peters S, on behalf of the ESMO Guidelines Working Group. Metastatic non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Anna Onco. 2014;25:27–39.
27. Masters GA, Temin S, Azzoli CG, Giaccone G, Baker S, Brahmer JR, et al. Systemic therapy for stage IV non-small-cell lung cancer: American Society of Clinical Oncology clinical practice guideline update. J Clin Onco. 2015;33:3488–515.
28. Passiglia F, Bronte G, Bazan V, Natoli C, Rizzo S, et al. PD-L1 expression as predictive biomarker in patients with NSCLC: A pooled analysis. Oncotarget. 2016;7(15):19738–19747.
29. Aguiar Jr PN, Santoro IL, Tadokoro H, de Lima LG, Filardi BA, et al. The role of PD-L1 expression as a predictive biomarker in advanced non-small cell lung cancer: A network meta-analysis. Immunotherapy. 2016; 8(4):479–88.
30. Griqq C, Rizvi NA. PD-L1 biomarker testing for non-small cell lung cancer: Truth or fiction? J Immun Cancer. 2016; 4:UNSP48.
31. Lee JW, Moon MJ, Min HY, Chung HJ, et al. Induction of apoptosis by a novel indirubin-5-nitro-30-monoxime, a CDK inhibitor, in human lung cancer cells. Bioorganic & Med Chem Let. 2005;15: 3948–3952.
32. Parthasarathy K, Praveen C, Balachandran C, Kumar PS, Ignacimuthu S, Perumal PT. Cu(OTf)2 catalyzed three component reaction: Efficient synthesis of spiro[indoline-3,40-pyrano[3,2-b]pyran derivatives and their anticancer potency towards A549 human lung cancer cell lines. Bioorganic Med Chem Let. 2013;23:2708–2713.
33. Bazin M, Bodero L, Tomasoni C, Rousseau B. Synthesis and antiproliferative activity of benzofuran-based analogs of cercosporamide against non-small cell lung cancer cell lines. Euro J Med Chem. 2013;69:823-832.
34. Xia Y, Dong Z, Zhao BX, Ge X, et al. Synthesis and structure–activity relationships of novel 1-arylmethyl-3-aryl-1H-pyrazole-5-carbohydrazide derivatives as potential agents against A549 lung cancer cells. Bioorganic Med Chem. 2007;15:6893–6899.
35. Xia Y, Fan CD, Zhao BX, Zhao J, Shin D, et al. Synthesis and structure-activity relationships of novel 1-arylmethyl-3-aryl - 1H-pyrazole-5- carbohydrazide hydrazonederivatives as potential agents against A549 lung cancer cells. Euro J Med Chem. 2008;43:2347-2353.
36. Dexheimer TS, Rosenthal AS, Luci DK, Liang Q, et al. Synthesis and structure−activity relationship studies of N‐Benzyl-2- phenylpyrimidin-4-amine derivatives as potent USP1/UAF1 deubiquitinase inhibitors with anticancer

activity against non-small cell lung cancer. J Med Chem. 2014;57:8099−8110.
37. Gu¨ner OF. Pharmacophore perception, development and use in drug design. International University Line: La Jolla, CA; 2000.
38. Thomas S, Andreas B, Gary T, et al. Recognizing pitfalls in virtual screening: A critical review. J. Chem. Inf. Model. 2012; 52(4):867-881.
39. Clement OO, Freeman CM, Hartmann RW, et al. Three dimensional pharmacophore modeling of human CYP17 inhibitors. Potential agents for prostate cancer therapy. J Med Chem. 2003;46:2345-2351.
40. Leach AR, Gillet VJ. An introduction to chemoinformatics. Kluwer Dordrecht; 2011.
41. Zhou Z, Madura JD. CoMFA 3D-QSAR analysis of HIV-1 RT nonnucleoside inhibitors, TIBO derivatives based on docking conformation and alignment. J Chem Infor Comp Sci. 2004;44(6):2167-2178.
42. Buolamwini JK, Assefa H. CoMFA and CoMSIA 3D QSAR and docking studies on conformationally restrained cinnamoyl HIV-1 integrase inhibitors: exploration of a binding mode at the active site. J Med Chem. 2002;45:841-852.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history/27250