Journal of Taibah University Medical Sciences (2023) 18(2), 295–309

Taibah University

Journal of Taibah University Medical Sciences

www.sciencedirect.com

Original Article

2D-QSAR, molecular docking, drug-likeness, and ADMET/pharmacokinetic predictions of some non-small cell lung cancer therapeutic agents

M.T. Ibrahim, PhD * and A. Uzairu, PhD

Computational and Theoretical Chemistry, Department of Chemistry, Faculty of Physical Science, Ahmadu Bello University, Zaria, Kaduna State, Nigeria

Received 26 May 2022; revised 23 July 2022; accepted 2 September 2022; Available online 16 September 2022

الملخص

أهداف البحث: كان سرطان الرئة ذو الخلايا غير الصغيرة هو النوع الشائع والقاتل من سرطانات الرئة مع ما يقرب من 1.8 مليون حالة ومعدل البقاء على قيد الحياة أقل من 20 ٪ في كل 5 سنوات بعد التشخيص. الهدف من هذا البحث هو تحديد العوامل العلاجية المحتملة لخلايا سرطان الرئة ذو الخلايا غير الصغيرة الحاملة لمستقبلات عامل النمو الأدمي من خلال استخدام بعض التقنيات بمساعدة الكمبيوتر.

طرق البحث: تم استخدام تقنية النمذجة الجزيئية التي تقيس علاقة الهيكل بالنشاط الكمي ثنائية الأبعاد على بعض العوامل العلاجية المحتملة على سرطان الرئة ذو الخلايا غير الصغيرة لتطوير نموذج تنبؤي عالي للغاية. تم إجراء فحص افتراضي للالتحام الجزيئي على نفس مجموعة المركبات لتحديد مركبات الإصابة المحتملة. علاوة على ذلك، تم تقييم الميزات الشبيهة بالدواء والحركية الدوائية لأفضل النتائج باستخدام خوادم الويب "سويس أدمي" و بي كي سي اس ام"؛ على التوالي.

النتائج: تم العثور على النموذج الذي تم إنشاؤه عبر تقنية النمذجة الجزيئية التي تقيس علاقة الهيكل بالنشاط الكمي ثنائية الأبعاد على مجموعة البيانات بدرجة عالية من التنبؤية. حدد الفحص الافتراضي للالتحام الجزيئي الذي تم إجراؤه ، المركبات 25 و 32 و 15 و 21 و 23 التي حصلت على أعلى درجات الإرساء كأفضل نتيجة. يحتوي المركب 25 من بينها على أعلى درجة في الإرساء تبلغ -138.329 كيلو كالوري / مول وإعادة تصنيف درجة -95.449. شوهد أن جميع المركبات التي تم تحديدها لديها نتائج أفضل وأعلى في الإرساء ودرجات إعادة ترتيب من العقار القياسي أزد9291. تم التأكد من أن أفضل مركبات الإصابة هي ذات طبيعة شبيهة بالعقاقير من خلال عدم وجود أكثر من انتهاك واحد لشروط الترشيح المستخدمة في تقييم تشابه الدواء لجزيء صغير. وعرضت ميزات الحركية الدوائية الخاصة بهم متوسط ملامح الحرائك الدوائية.

الاستنتاجات: يمكن أن تعمل مركبات الإصابة المذكورة كعوامل علاجية محتملة لسرطان الرئة ذو الخلايا غير الصغيرة بسبب سلامتها وفعاليتها بعد اجتياز التجربة قبل السريرية باستثناء المركب 23 الذي وجد أنه سام. وأيضا، يمكن أن تكون بمثابة نموذج لتصميم عوامل علاجية جديدة لسرطان الرئة ذو الخلايا غير الصغيرة.

الكلمات المفتاحية: تقنية النمذجة الجزيئية التي تقيس علاقة الهيكل بالنشاط الكمي ثنائية الأبعاد؛ الالتحام الجزيئي؛ الأدوية الشبيهة؛ الحركية الدوائية؛ سرطان الرئة ذو الخلايا غير الصغيرة

Abstract

Objectives: Non-small cell lung cancer (NSCLC) is the most common type of lung cancer, with nearly 2 million diagnoses and a 17% 5-year survival rate. The aim of this study was to use computer-aided techniques to identify potential therapeutic agents for NSCLC.

Methods: The two dimensional-quantitative structure–activity relationship (2D-QSAR) modeling was employed on some potential NSCLC therapeutic agents to develop a highly predictive model. Molecular docking-based virtual screening were conducted on the same set of compounds to identify potential hit compounds. The pharmacokinetic features of the best hits were evaluated using SWISSADME and pkCSM online web servers, respectively.

Results: The model generated via 2D-QSAR modeling was highly predictive with R^2= 0.798, R^2adj = 0.754, Q^2CV = 0.673, R^2 test = 0.531, and cRp^2 = 0.627

* Corresponding address: Department of Chemistry, Ahmadu Bello University, P.M.B 1045, Zaria, Kaduna State, Nigeria.
E-mail: muhdtk1988@gmail.com (M.T. Ibrahim)

Peer review under responsibility of Taibah University.

Production and hosting by Elsevier

 https://doi.org/10.1016/j.jtumed.2022.09.002

assessment parameters. Molecular docking-based virtual screening identified compounds 25, 32, 15, 21, and 23 with the highest MolDock scores as the best hits, of which compound 25 had the highest MolDock score of −138.329 kcal/mol. All of the identified hits had higher MolDock scores than the standard drug (osimertinib). The best hit compounds were ascertained to be drug-like in nature following the Lipinski's rule of five. Also, their ADMET features displayed average pharmacokinetic profiles.

Conclusion: After successful preclinical testing, the hit compounds identified in this study may serve as potential NSCLC therapeutic agents due to their safety and efficacy with the exception of compound 23, which was found to be toxic. They can also serve as a template for designing novel NSCLC therapeutic agents.

Keywords: 2D-QSAR; ADMET; Drug-like; Molecular docking; Non-small cell lung cancer

Introduction

Stanley Cohen discovered epidermal growth factor receptor (EGFR) in 1956 while working at Vanderbilt University (Nashville, TN, USA).[1] EGFR is a member of the ErbB family of receptors, which controls crucial cellular functions such as cell evolution, separation, union, variation, breakdown, motility, and death.[2] EGFR is overexpressed in several cancer types including gastric, breast, ovarian, colon, and lung cancers. This overexpression is also seen in malignances including metastatic colon cancer, head and neck cancer, and non-small cell lung cancer (NSCLC). EGFR overexpression has been identified in about 40−85% of NSCLC patients,[3] making it the main and principal target for treating cancers such as NSCLC.[4]

Lung cancer remains the leading cause of cancer-related death globally, and is divided into oat cell lung cancer (known as small-cell lung cancer), which accounts for 15−20% of lung cancer cases; and NSCLC, which accounts for 80−85% of cases.[5,6] NSCLC is the most common type of lung cancer with 2 million diagnoses and 5-year survival rates of 17% for all stages and 2% for stage 4 NSCLC.[7,8] The main subtypes of NSCLC are large cell carcinoma, adenocarcinoma, and squamous cell carcinoma. The main procedures utilized for the treatment of NSCLC are surgery, radiotherapy, targeted therapy, and chemotherapy. Despite the development of treatment modalities, the diagnosis of NSCLC patients has not meaningfully improved.[9]

EGFR tyrosine kinase inhibitors (TKIs) targeting NSCLC can inhibit tumor growth by binding to the EGFR ATP-binding site. EGFR TKIs have been first-line therapy for the treatment of NSCLC patients with activating, classical, and gatekeeper mutations.[10,11]

Gefitinib and erlotinib are United States Food and Drug Administration [FDA]-approved first-generation EGFR TKIs for the treatment of patients with NSCLC harboring EGFR activating (L858R) and classical (del. E746-A750) mutations in exon 19 and exon 21, respectively.[12] Furthermore, patients harboring L858R EGFR mutations usually respond well to gefitinib and erlotinib. However, the time frame for their efficacy is very limited due to the resistance established by the $EGFR^{T790M}$ (gatekeeper mutation), which is present in about 50% of NSCLC patients with such EGFR mutation ($EGFR^{T790M}$).[13]

Afatinib and dacomitinib (FDA-approved irreversible second-generation EGFR TKIs) were developed to resolve the induced $EGFR^{T790M}$ mutation-related resistance to first-generation EGFR-TKIs.[12,14] Unfortunately, their clinical effectiveness is limited due to serious side effects such as skin rashes and gastrointestinal toxicity as well as nonexistence of choices between EGFR mutant and wild-type ($EGFR^{WT}$).[14,15]

AZD9291 (osimertinib) and CO-1686 (rociletinib) (FDA-approved irreversible third-generation EGFR TKIs) were designed and established to inhibit the $EGFR^{T790M}$ resistance mutation (second-generation EFGR TKIs) while being more selective for $EGFR^{WT}$, and also to overcome the observed toxicity with irreversible second-generation EGFR-TKIs. These irreversible third-generation EGFR-TKIs cannot achieve the stated goal due to emergence of the C797S mutation.[14,16,17]

The process of bringing an effective drug to market is a huge investment in time and money. Specifically, it is projected to require about $2.6 billion dollars (the cost has increased in the last decade by almost 150%) to develop and bring a drug to market and takes an average time of 10−15 years. Similarly, the increase in failure rate is nearly 90%, which indicates that virtually 70% of funds are invested in drugs that fail clinical trials due to inadequate efficacy or adverse effects.[18,19]

At the present time, advances in computer-aided drug discovery (CADD) have made it possible to rapidly evaluate, improve, and screen thousands or millions of molecules and choose a candidate drug against a target protein, thereby saving a lot of money and time.[19,20] Quantitative structure−activity relationship (QSAR) is a CADD method that entails the generation of a mathematical model, which relates the biological activities of a set of molecules to their molecular descriptors (physicochemical properties).[21] These properties have a key impact on the drug's activity. Molecular docking is also a CAAD technique utilized to reveal the interaction between a drug (ligand) and a receptor and how this drug binds to the target. It also contributes to the virtual screening of a library of compounds at the preclinical stage of drug development.[22] Predicting drug-likeness and absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties of a drug in hit-to-

lead and lead-optimization programs have played an essential role in drug research and development.[23]

This study identified potential NSCLC therapeutic agents by employing computer-aided techniques such as two-dimensional (2D)-QSAR modeling, molecular docking-based virtual screening, and prediction of pharmacokinetic (drug-likeness and ADMET) features.

Materials and Methods

2D-QSAR methodology

Dataset sourcing and antiproliferative activity

Forty-five sets of tetrahydropyrazolo-quinazoline and tetrahydropyrazolo-pyrimidocarbazole derivatives, as anti-prostate cancer agents and Pim-1 proto-oncogene, serine/threonine kinase inhibitors, were synthesized and assessed for their antiproliferative activities against six different cancer cell lines: A549, HT-29, MKN-45, U87MG, SMMC-7721 and H460.[24] Antiproliferative activity (the defined biological activity for this study) against the A549 NSCLC cell line was selected and used in accordance with the Organization for Economic Co-operation and Development (OECD) principle 1, which specifies that a QSAR model must have defined biological activities. Antiproliferative activities against the A549 NSCLC cell line (half-maximal inhibitory concentration [IC_{50}]) reported in micromolar (μM) were subsequently changed to their logarithmic form (pIC_{50}) using equation (1) [25]. Table 1 shows the 1D molecular formula, IC_{50}, and transformed pIC_{50} of the dataset.

$$pIC_{50} = -\log (IC_{50} * 10^{-6}) \quad (1)$$

Drawing of structure of the sourced dataset and optimization

The ChemDraw software developed by the University of Cambridge was used to draw the 2D structures of the sourced dataset.[26] In this work, the Merck molecular force field with the B3LYP/6-311G* density functional theory were used for the optimal conformational search for all 45 sets of tetrahydropyrazolo-quinazoline and tetrahydropyrazolo-pyrimidocarbazole derivatives.[14,27]

Computation, pretreatment, and splitting of molecular descriptors

The Pharmaceutical Data Exploration Laboratory descriptor tool kit was used for the computation of molecular descriptors in this study.[28,29] About 1875 descriptors, which included 1D, 2D, and 3D molecular descriptors, were computed in this study. Data pretreatment software version 1.2 for eliminating uninformative, redundant, and constant molecular descriptors, which was developed by the Drug Theoretics and Cheminformatics (DTC) laboratory, was used to pretreat the computed molecular descriptors.[30] Data division software version 1.2 also developed by the DTC laboratory was used for splitting the pretreated molecular descriptors into model building training and validation test sets.[31]

Model development and assessment

Model development is of utmost significance and critical when performing QSAR molecular modeling research. In this study, the experimentally determined activities (antiproliferative activities against the A549 cell line) were retained as the response parameter and the computed and treated molecular descriptors were retained as independent parameters. The genetic algorithm combined with multiple linear regression was used as the variable selection method due its high relevance in QSAR molecular modeling, in accordance with OECD principle 2 (which means an ambiguous algorithm was avoided).[32,33]

Table 1: The 1D molecular formula, IC_{50}, and pIC_{50} of the dataset.

Entry	Formula	IC_{50}	pIC_{50}
1	$C_{22}H_{19}N_5O_2$	1.47	5.83
2	$C_{23}H_{21}N_5O_3$	3.28	5.48
3	$C_{22}H_{18}C_lN_5O_2$	1.77	5.75
4	$C_{23}H_{21}N_5O_2$	3.82	5.42
5	$C_{24}H_{23}N_5O_3$	2.9	5.54
6	$C_{23}H_{20}C_lN_5O_2$	0.83	6.08
7	$C_{22}H_{18}C_lN_5O_2$	1.47	5.83
8	$C_{23}H_{20}C_lN_5O_3$	0.44	6.36
9	$C_{22}H_{17}C_{l2}N_5O_2$	0.18	6.74
10	$C_{25}H_{19}N_7OS$	6.02	5.22
11	$C_{27}H_{24}N_6O_3S$	4.92	5.31
12	$C_{26}H_{21}N_7O_2S$	1.49	5.83
13	$C_{28}H_{26}N_6O_4S$	2.48	5.61
14	$C_{25}H_{18}C_lN_7OS$	2.42	5.62
15	$C_{27}H_{23}C_lN_6O_3S$	2.09	5.68
16	$C_{26}H_{21}N_7OS$	4.33	5.36
17	$C_{28}H_{26}N_6O3S$	3.4	5.47
18	$C_{27}H_{23}N_7O_2S$	3.59	5.44
19	$C_{29}H_{28}N_6O_4S$	6.61	5.18
20	$C_{26}H_{20}C_lN_7OS$	2.48	5.61
21	$C_{28}H_{25}C_lN_6O_3S$	2.82	5.55
22	$C_{25}H_{18}C_lN_7OS$	1.82	5.74
23	$C_{27}H_{23}C_lN_6O_3S$	2.19	5.66
24	$C_{26}H_{20}C_lN_7O_2S$	0.93	6.03
25	$C_{28}H_{25}C_lN_6O_4S$	1.89	5.72
26	$C_{25}H_{17}C_{l2}N_7OS$	0.6	6.22
27	$C_{27}H_{22}C_{l2}N_6O_3S$	0.14	6.85
28	$C_{28}H_{22}N_6O$	5.28	5.28
29	$C_{29}H_{24}N_6O_2$	3.97	5.40
30	$C_{28}H_{21}C_lN_6O$	1.49	5.83
31	$C_{29}H_{24}N_6O$	2.66	5.58
32	$C_{30}H_{26}N_6O_2$	2.27	5.64
33	$C_{29}H_{23}C_lN_6O$	2.37	5.63
34	$C_{28}H_{21}C_lN_6O$	3.15	5.50
35	$C_{29}H_{23}C_lN_6O_2$	1.6	5.80
36	$C_{28}H_{20}C_{l2}N_6O$	3.53	5.45
37	$C_{24}H_{23}N_5O_2$	1.07	5.97
38	$C_{25}H_{25}N_5O_3$	2.58	5.59
39	$C_{24}H_{22}C_lN_5O_2$	2.07	5.68
40	$C_{25}H_{25}N_5O_2$	1.02	5.99
41	$C_{26}H_{27}N_5O_3$	2.39	5.62
42	$C_{25}H_{24}C_lN_5O_2$	0.8	6.10
43	$C_{24}H_{22}C_lN_5O_2$	1.88	5.73
44	$C_{25}H_{24}C_lN_5O_3$	0.8	6.10
45	$C_{24}H_{21}C_{l2}N_5O_2$	0.22	6.66

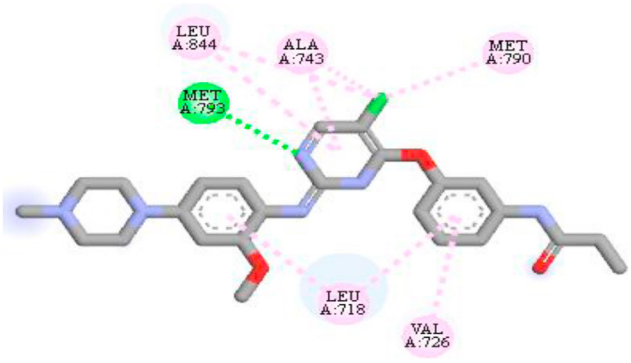

Figure 1: The **WZ4002** in the binding poses of the **3IKA** receptor in 2D view.

A good QSAR model to predict the activities of new chemicals was not used in building the model. According to OECD principle 3, the applicability domain (AD) for a QSAR model must be shown to ensure the quality of prediction of the QSAR model. The leverage method was implemented in this study to show the chemical space or AD of the reported QSAR model.[34]

Assessment/validation of QSAR models remain paramount and critical before any QSAR model can be considered effective and useful (OECD principle 4, which is a model assessment/validation for productivity). Therefore, it is necessary to subject any model to various assessments prior to its consideration as useful, and is expected to pass those assessments. The several validation parameters employed in this study include the $R^2_{internal}$, Q^2_{cv}, $R^2_{external}$, and cR^2p.[35]

Table 2: The statistical parameters of model 1 and accepted values for assessment of QSAR models.

Symbol	Name	Threshold value	Selected model
R^2	Squared correlation coefficient	$\geq$ 0.6	0.798
R^2_{adj}	Adjusted squared correlation coefficient	$\geq$ 0.6	0.754
Q^2_{cv}	Cross-validation coefficient	$\geq$ 0.5	0.673
R2 − Q2	Difference between R^2 and Q^2	$\leq$ 0.3	0.125
N (ext., & test set)	Minimum number of external and test set	$\geq$ 5	45
R^2_{test}	Squared correlation coefficient of test set	$\geq$ 0.5	0.531
cRp^2	Y-scrambling parameter	$\geq$ 0.5	0.627

Selected model: $\mathbf{pIC_{50}}$ = 8.574102927 * **SpMax4_Bhs** − 1.680588701 * **FPSA-2** + 0.221336263 * **RDF155u** + 0.061781318 * **RDF150m** − 0.085108039 * **RDF155s** − 5.869064737 * **P1i** − 21.2418.

Table 3: The representation, explanation, category, VIF, and Me of the molecular descriptors contained in the chosen model.

S/No	Representation	Explanation	Category	VIF	ME
1	**SpMax4_Bhs**	Largest absolute eigenvalue of Burden modified matrix − n 4/weighted by relative I-state	2D	1.246915	1.237826
2	**FPSA-2**	PPSA-2/total molecular surface area	3D	2.520802	−0.09515
3	**RDF155u**	Radial distribution function −155/unweighted	3D	4.339996	0.008929
4	**RDF150m**	Radial distribution function −150/weighted by relative mass	3D	1.115442	0.004393
5	**RDF155s**	Radial distribution function −155/weighted by relative I-state	3D	2.886863	−0.00229
6	**P1i**	1st component shape directional WHIM index/weighted by relative first ionization potential	3D	2.058089	−0.1537

Table 4: Experimental, calculated antiproliferative activities and residual values.

S/No	Experimental antiproliferative activities	Theoretical antiproliferative activities	Residual
1	5.83	5.82	0.009139
2	5.48	5.54	−0.05524
3[VTS]	5.75	5.82	0.074671
4	5.42	5.48	−0.06292
5	5.54	5.53	0.012415
6	6.08	5.90	0.182519
7	5.83	5.78	0.047877
8 [VTS]	6.36	5.69	−0.6735
9	6.74	6.42	0.324962
10	5.22	5.31	−0.08845
11 [VTS]	5.31	5.43	0.123691
12	5.83	5.51	0.324364
13	5.61	5.55	0.058694
14	5.62	5.76	−0.13513

Table 4 (*continued*)

S/No	Experimental antiproliferative activities	Theoretical antiproliferative activities	Residual
15	5.68	5.66	0.019428
16	5.36	5.25	0.112203
17	5.47	5.65	−0.17748
18	5.44	5.41	0.029369
19	5.18	5.21	−0.03385
20	5.61	5.58	0.029188
21 [VTS]	5.55	5.34	−0.21321
22	5.74	5.94	−0.20044
23	5.66	5.79	−0.1329
24 [VTS]	6.03	6.10	0.071283
25	5.72	5.54	0.176777
26	6.22	6.21	0.011181
27 [VTS]	6.85	6.14	−0.71062
28 [VTS]	5.28	5.32	0.037654
29	5.4	5.62	−0.2184
30	5.83	5.62	0.211837
31	5.58	5.64	−0.05512
32 [VTS]	5.64	5.89	0.245716
33	5.63	5.59	0.035312
34	5.5	5.48	0.018097
35	5.8	5.78	0.022486
36	5.45	5.68	−0.22764
37 [VTS]	5.97	5.76	−0.2101
38	5.59	5.52	0.067905
39	5.68	5.97	−0.28752
40	5.99	5.75	0.242506
41	5.62	5.61	0.008074
42 [VTS]	6.1	6.09	−0.01181
43	5.73	5.97	−0.24292
44	6.1	6.19	−0.08651
45	6.66	6.60	0.060169

[VTS], Validation test set.

Molecular docking-based virtual screening methodology

The investigation of the protein-ligand interactions in study was established by utilizing Molegro Virtual Docker (MVD) as a result of the higher precision of the software compared to other molecular docking software.

Ligand preparation

The optimum conformations of the 45 sets of tetrahydropyrazolo-quinazoline and tetrahydropyrazolo-pyrimidocarbazole derivatives obtained were saved in Mol2 file format. The 45 sets of tetrahydropyrazolo-quinazoline and tetrahydropyrazolo-pyrimidocarbazole derivatives

Table 5: Randomization chance correlation test.

Model	R	R^2	Q^2
Original	0.825977	0.682238	0.448129
Random 1	0.245117	0.060082	−0.18538
Random 2	0.355447	0.126342	−0.29712
Random 3	0.192702	0.037134	−0.34844
Random 4	0.366005	0.13396	−0.19734
Random 5	0.440836	0.194337	−0.48966
Random 6	0.449798	0.202318	−0.73032
Random 7	0.445395	0.198377	−0.23294
Random 8	0.239219	0.057226	−0.7207
Random 9	0.285254	0.08137	−0.39411
Random 10	0.243496	0.05929	−0.46677
Average r:	0.326327		
Average r^2:	0.115044		
Average Q^2:	−0.40628		
cRp^2:	0.626736		

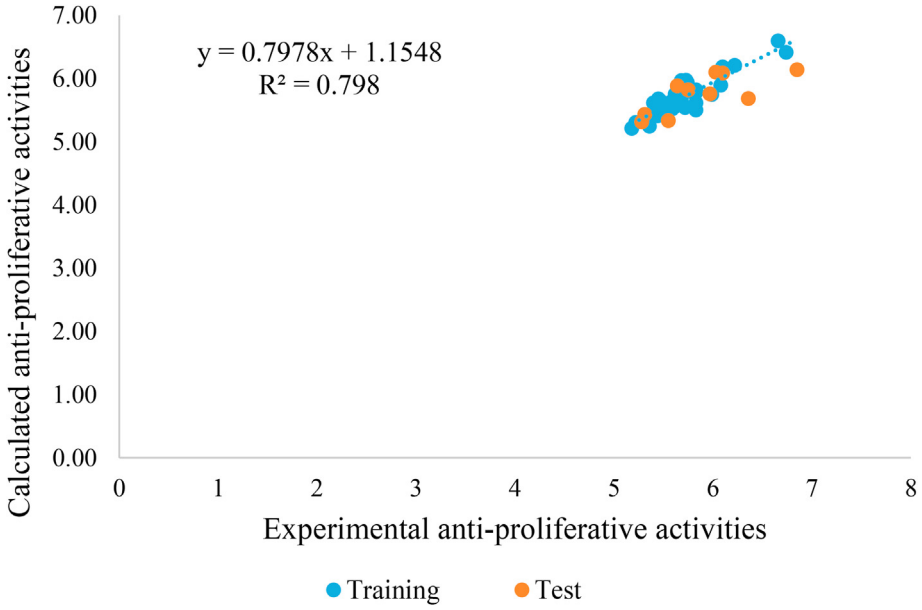

Figure 2: Experimental antiproliferative activities against calculated antiproliferative activities.

were further prepared using the default setting of the MVD by selecting "if missing" to all set of the parameters.[36,37]

Identification of the amino acids in the active site of the 3IKA receptor

The **3IKA** receptor and **WZ4002** (its co-crystalized ligand) were retrieved from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (PDB) database (https://www.rcsb.org/), respectively.[38] In this work, for identification of the group of amino acids in the binding poses of the **3IKA** receptor, the native (co-crystalized) ligand (WZ4002) fused to **3IKA** receptor were visualized with Discovery Studio Visualizer, version 16.1.0.15350. The group of amino acids recognized in the binding poses of the **3IKA** receptor was ALA743, LEU844, LEU718, MET793, MET790, and VAL726. The 2D views of the native ligand in the binding poses of **3IKA** receptor are presented in Figure 1.

3IKA receptor preparation

Before investigating the protein-ligand interactions, the **3IKA** receptor was prepared by importing it into the 3D view space (interface) of the MVD, and then the groups of amino acids with errors in their structures were reconstructed and restored. Furthermore, after rebuilding and repairing the group of amino acids with errors in their structures, the

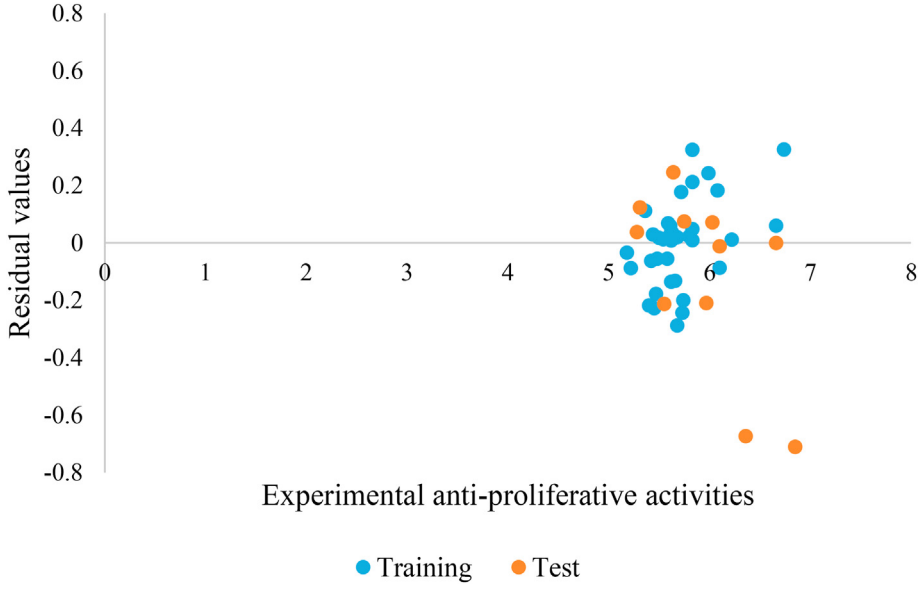

Figure 3: Experimental antiproliferative activities against residual values.

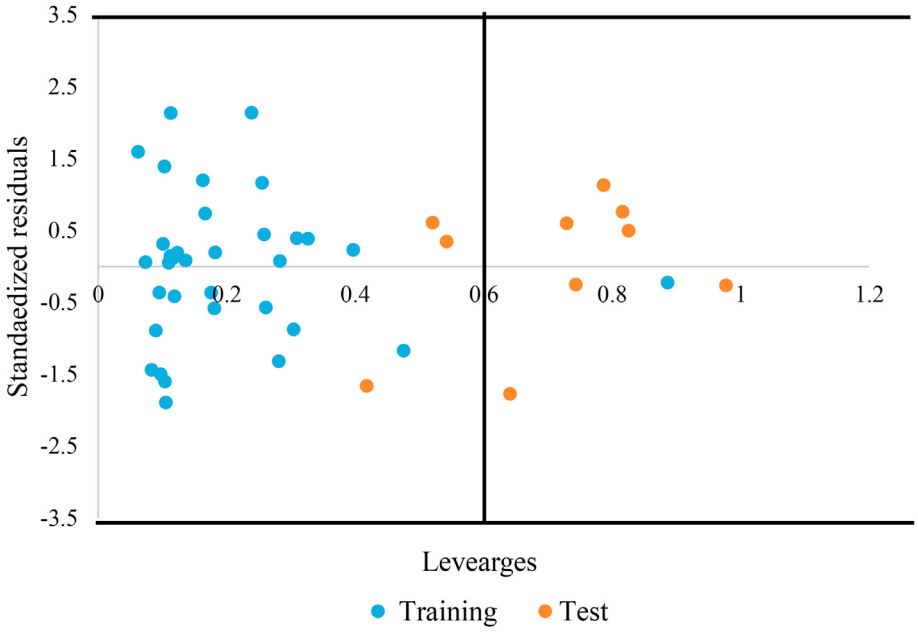

Figure 4: Williams' sketch of the best model.

Table 6: Docking results of the investigated compounds.

Name	MolDock Score	Re-rank Score
25	−138.329	−95.449
32	−133.326	−98.910
15	−131.387	−92.345
21	−129.081	−98.652
23	−128.867	−96.305

surface was formed and cavities were noticed (by right clicking on the **3IKA** receptor and clicking the surface as well as detecting cavities) prior to removal of the **WZ4002** from the **3IKA** receptor.

Molecular docking calculations

Implementation of the molecular docking protocol was completed by choosing the molecular docking algorithm to be the plant score and the scoring function to be the

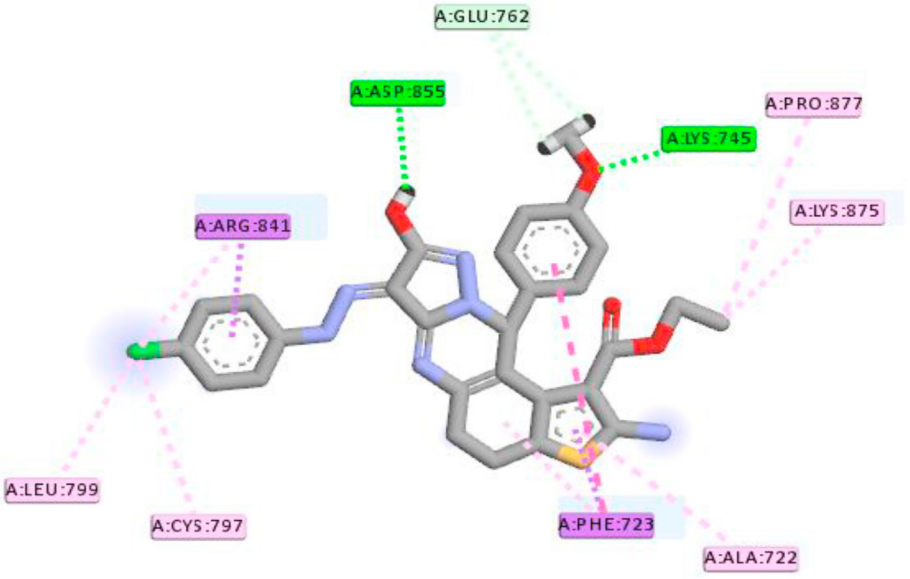

Figure 5: 2D interaction between compound 25 and the 3IKA EGFR enzyme.

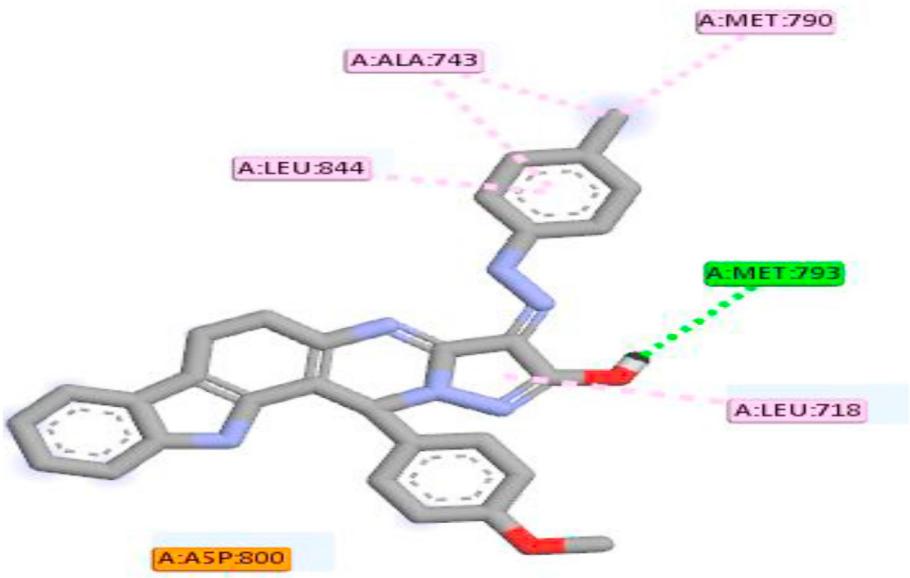

Figure 6: 2D interaction between compound 32 and the 3IKA EGFR enzyme.

MolDock score. The grid box for defining the binding poses that consist of the binding cavities in the protein was set to 24 Å. For all other calculations, the default settings were sustained and preserved.[39]

Validation of the molecular docking protocol

Justification of the protocol of molecular docking was performed by re-docking the native ligand (WZ4002) into the binding poses of the receptor. The groups of amino acids of the re-docked native ligand were compared with those of the investigated molecules in this study.

Drug-like and ADMET/pharmacokinetic studies

The modeling of the drug-like and ADMET features of these investigated compounds was evaluated using the SWISSADME (http://www.swissadme.ch/index.php) and

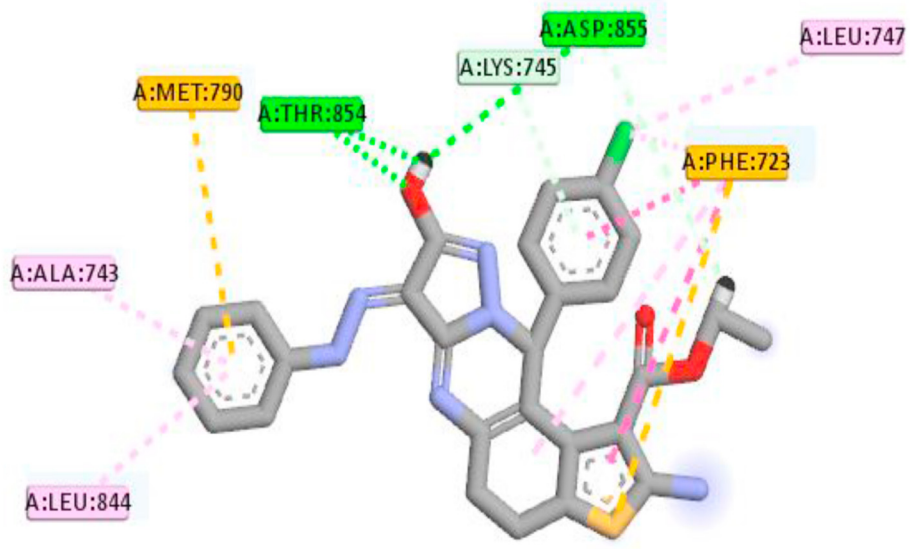

Figure 7: 2D interaction between compound 15 and the 3IKA EGFR enzyme.

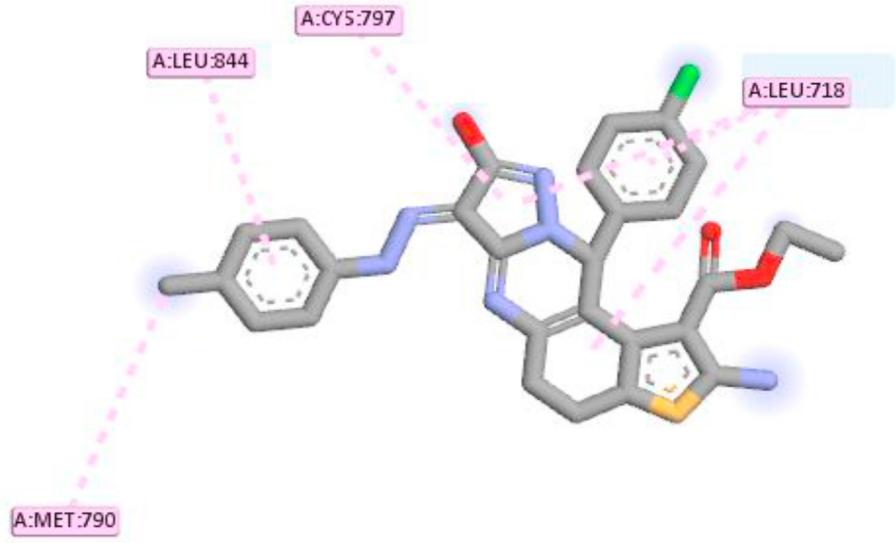

Figure 8: 2D interaction between compound 21 and the 3IKA EGFR enzyme.

pkCSM (http://structure.bioc.cam.ac.uk/pkcsm) online web servers, respectively.[23,26,40]

Results

2D-QSAR

The results of the 2D-QSAR modeling on the 45 sets of tetrahydropyrazolo-quinazoline and tetrahydropyrazolo-pyrimidocarbazole derivatives under investigation are shown in Tables 2–5 and Figs. 2–4.

Molecular docking-based virtual screening

The molecular docking-based virtual screening results of the top five hit tetrahydropyrazolo-quinazoline and tetrahydropyrazolo-pyrimidocarbazole derivatives under investigation are presented in Table 6 and Figs. 5–11.

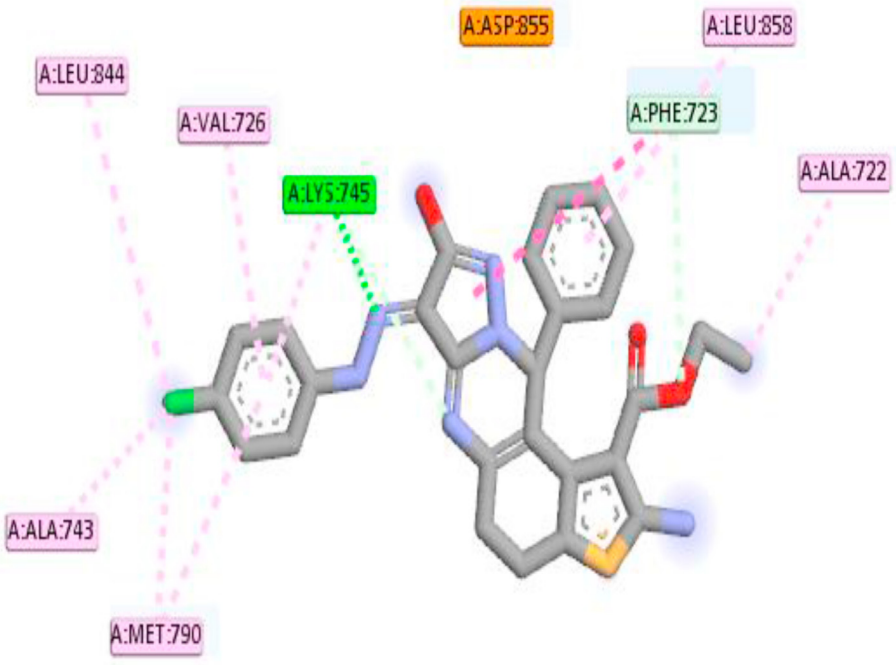

Figure 9: 2D interaction between compound 23 and the 3IKA EGFR enzyme.

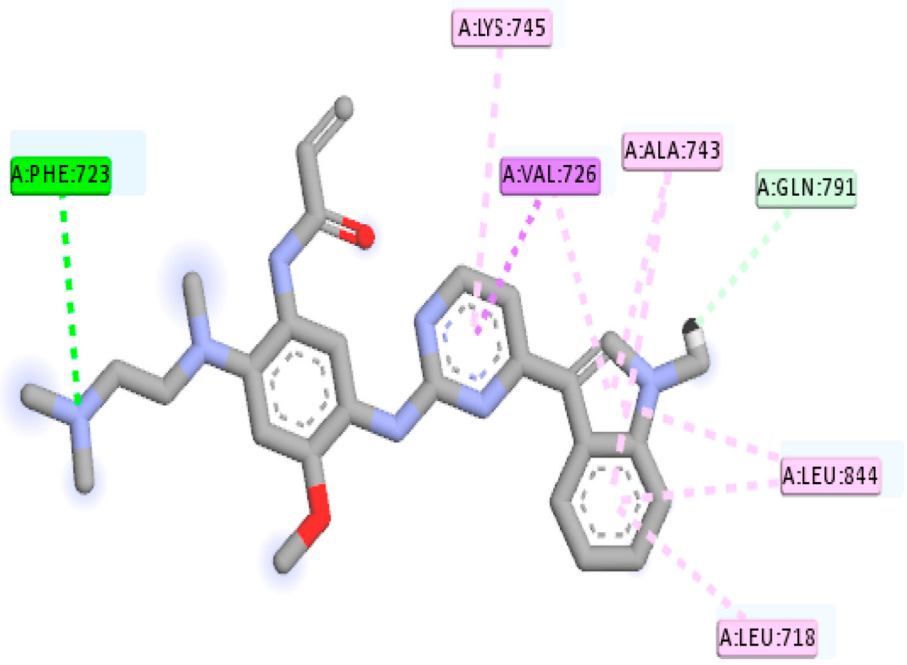

Figure 10: 2D interaction between AZD9291 and the 3IKA EGFR enzyme.

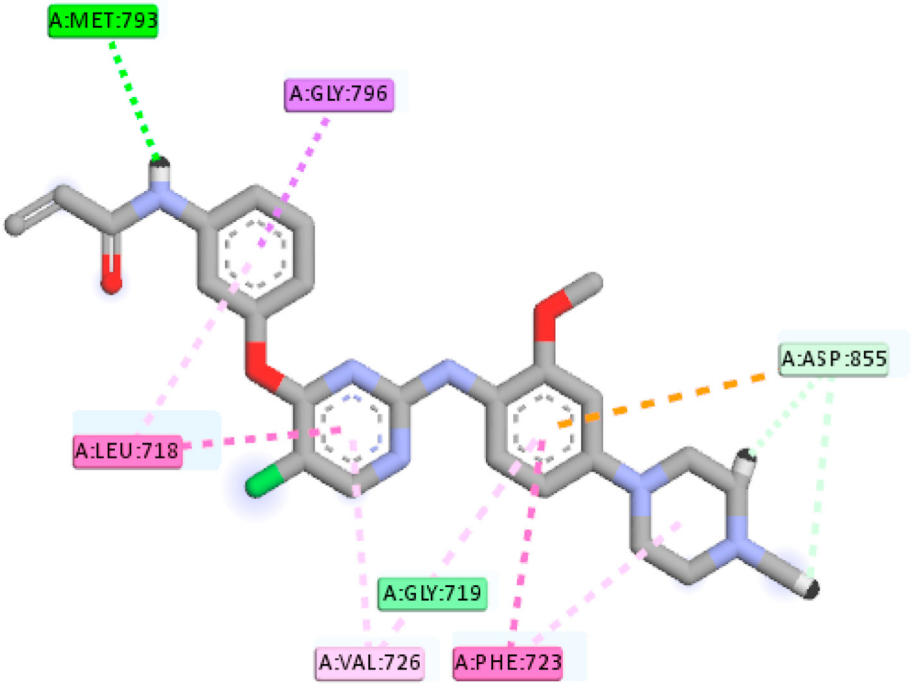

Figure 11: 2D interaction between WZ4002 and the 3IKA EGFR enzyme.

Table 7: Drug-like features of the best hit compounds using Lipinski's RO5 filtering criteria.

S/No.	MW	No. H-bond donors	No. H-bond acceptors	MLOGP	Lipinski RO5 violations	Bioavailability Score
25	577.05	3	7	3.68	1	0.55
32	502.57	3	5	3.97	1	0.55
15	547.03	3	6	3.98	1	0.55
21	561.05	3	6	4.18	1	0.55
23	547.03	3	6	3.98	1	0.55

Table 8: ADMET/pharmacokinetic features of the best hit compounds.

	Absorption	Distribution		Metabolism CYP							Excretion	Toxicity
				Substrate		Inhibitors						
	Intestinal absorption	BBB permeability	CNS permeability	2D6	3A4	1A2	2C19	2C9	2D6	3A4	Total clearance	AMES toxicity
25	99.343	−1.316	−2.061	No	Yes	Yes	Yes	Yes	No	Yes	−0.058	No
32	100	−0.959	−1.502	Yes	Yes	Yes	Yes	Yes	No	Yes	0.625	No
15	97.823	−1.279	−1.887	No	Yes	Yes	Yes	Yes	No	Yes	−0.09	No
21	98.472	−1.143	−1.813	No	Yes	Yes	Yes	Yes	No	Yes	−0.147	No
23	98.004	−1.132	−1.887	No	Yes	Yes	Yes	Yes	No	Yes	−0.091	Yes

Drug-like/ADMET prediction and evaluation

The results of the drug-like and ADMET prediction and evaluation of the 45 sets of tetrahydropyrazolo-quinazoline and tetrahydropyrazolo-pyrimidocarbazole derivatives under investigation are presented in Tables 7 and 8, respectively.

Discussion

2D-QSAR

The 2D-QSAR molecular modeling technique was employed for the 45 sets of tetrahydropyrazolo-quinazoline and tetrahydropyrazolo-pyrimidocarbazole derivatives under investigation to generate a model with high predictive capability. The 45 sets of tetrahydropyrazolo-quinazoline and tetrahydropyrazolo-pyrimidocarbazole derivatives under investigation were split into a training model building set and test validation set of the data. The training model building set was utilized to generate four diverse models. In the four diverse QSAR models generated, model 1 was the first model selected and used due to its high $R^2 = 0.798$, $R^2_{adj} = 0.754$, $Q^2_{CV} = 0.673$, $R^2_{test} = 0.531$, and $cRp^2 = 0.627$ assessment parameters. Then these assessment parameters (reported values) for the selected model (model 1) were compared with the accepted threshold values (Table 2), used for the assessment of QSAR models, and found to be greater than the accepted threshold values, which clearly showed the internal predictive capability of model 1 (the selected model).[41,42]

The R^2 value of model 1 illustrates that it is able to describe about 79.8% of the gap detected in the antiproliferative pIC_{50} against the A549 NSCLC cell line of the tetrahydropyrazolo-quinazoline and tetrahydropyrazolo-pyrimidocarbazole derivatives under investigation. Moreover, the theoretical R^2 value of the model was a little bit better and higher than its equivalent R^2_{adj}, which depicted the importance of the model and showed that the model was free from overparameterization.

The variation inflation factor (VIF) values of the molecular descriptors contained in the training model building set were calculated to determine whether there was an occurrence of problems associated with multicollinearity in the model. These computed values were expected to be less/smaller than 10, signifying the competence of the model. If the computed values were less than 10, the molecular descriptors contained in the model were considered statistically significant and impertinent/orthogonal (Table 3). The VIF values of the molecular descriptors contained in the model computed were all less than 10, which displayed the goodness of the model. Also, the model was free from problems associated with multicollinearity, as such a model can be accepted and used.[43]

The mean effect on the molecular descriptors contained in the model building training set was calculated to determine the impact and significance of a molecular descriptor in comparison to others in the model on the antiproliferative activities against the A549 NSCLC cell line. The positive or negative coefficients of each value of a molecular descriptor indicated the magnitude of the impact and the significance of such properties in the main structures of the investigated compounds on the antiproliferative activities against A549 NSCLC cell line. Based on the calculated mean effect of the molecular descriptors contained in the model (Table 3), molecular descriptor with a symbol SpMax4_Bhs had the highest value of +1.237826, followed by molecular descriptor RDF155u with a value of +0.008929, RDF150m with a value of +0.004393, RDF155s with a value of −0.00229, FPSA-2 with a value of −0.09515, and lastly P1i with a value of −0.1537. The coefficient and sign of the SpMax4_Bhs molecular descriptor showed that if such a property was increased to the main structures of the investigated compounds, there might be improvement in their antiproliferative activities against the A549 NSCLC cell line and vice versa. This applies to all molecular descriptors with positive signs. For molecular descriptors with negative signs such as RDF155s, when such a property is reduced from the main structures of the investigated compounds, there might be improvement in their antiproliferative activities against the A549 NSCLC cell line and vice versa. Also, this applies to all molecular descriptors with negative signs. The order in terms of the impact and significance of these molecular descriptors is as follows[44]:

SpMax4_Bhs > RDF155u > RDF150m > RDF155s > FPSA-2 > P1i

Figure 2 depicts the plot of experimental antiproliferative activities against calculated antiproliferative activities of both the training model building and test validation sets for the chosen model, which was used to ascertain the reliability of the selected model. The internal R^2 value of 0.798 of the selected model was observed to agree and be the same with the one deduced from Figure 2 (R^2 value of

0.798), which further established the reliability of the model. Figure 3 depicts the plot of experimental antiproliferative activities against calculated residuals of the training model building and test validation sets for the chosen model, which is used to determine the presence of methodological and systematic error of the chosen model. The chosen and selected model was further established to be methodological and systematic error free by unexpected dispersal/spreading of the calculated residuals on Figure 3. Figure 4 depicts the Williams' sketch of the best model, which clearly portrayed the chemical space (also known as the applicability domain) of the chosen model. From the Williams' sketch, no outlier compound (a compound with a standardized residuals $> \pm 3\ \delta$) was identified from the whole data. Eight (8) influential compounds (leverage of a compound $> h^* = 0.6.0$ [h^*: the threshold leverage]) were identified in the data used (seven from the test set and one from the training set). This showed the model's quality of estimation and foil the waste of the outcomes acquired by it (the model).

Table 4 shows the experimental antiproliferative pIC_{50}, calculated antiproliferative pIC_{50}, and calculated residual values of the training model building and test validation sets for the model. The residual values are calculated to determine the predictive performance and capability of the model. The model is said to have high predictive performance if the calculated residuals are very low. Based on this, the model was seen and established to have high predictive performance by the low calculated residuals on Table 4. Table 5 presents the randomization chance correlation test on the model building training test for the chosen model. In this randomization chance correlation test, 10 distinct models were developed by changing the actual experiment activities while leaving the molecular descriptors unaffected. The newly randomized models developed are assumed to have low R^2 and Q^2 values for numerous counts. Based on this, the chosen model was found to be robust with low values of the R^2 and Q^2 for numerous counts. Furthermore, the chosen model was not established by chance since the ${}^{C}R^2{}_{P}$ value of 0.626736 obtained was greater than 0.5.

Molecular docking-based virtual screening

MVD was used to carry out virtual screening to investigate the ligand−protein interactions in this research. This virtual screening was performed between the 45 sets of tetrahydropyrazolo-quinazoline and tetrahydropyrazolo-pyrimidocarbazole derivatives under investigation and EGFR enzyme crystal structure with PDB entry identifier: **3IKA**. The reason for carrying out this virtual screening was to confirm whether the investigated compounds can have enzymatic inhibitory activities against the EGFR enzyme and as well as to identify the best hit that can serve as potential EGFR TKIs. The investigated compounds were scored based on their MolDock score values. The MolDock score values of the investigated compounds ranged between −85.4875 and −138.329 kcal/mol (Supplementary Table 1).

Based on the virtual screening performed on the investigated compounds, compound 25 with a MolDock score of −138.329 kcal/mol and re-rank score of −95.449 was found to be the best hit, followed by compound 32 with a MolDock score of −133.326 kcal/mol and re-rank score of −98.910, compound 15 with a MolDock score of −131.387 kcal/mol and re-rank score of −92.345, compound 21 with a MolDock score of −129.081 kcal/mol and re-rank score of −98.652, and finally compound 23 with a MolDock score of −128.867 kcal/mol and re-rank score of −96.305 (Table 6).

The interactions between compound 25 and the binding poses of the 3IKA receptor observed were hydrogen bonding with these set of amino acids ALYS745, ASP855, LYS745 and GLU762; hydrophobic (Pi-alkyl, Pi−Pi stacked, amide-Pi stacked and alkyl) interactions with these set of amino acids PHE723, ARG841, PHE723, LYS875, PRO877, CYS797, LEU799 and ALA722; and electrostatic (Pi-anion) interactions with amino acids CYS797 and PHE723 as shown in Figure 5. The interactions of compound 32 with the binding poses of the 3IKA receptor were via hydrogen bonds with amino acid residue MET793 and hydrophobic (alkyl and Pi-alkyl) interactions with amino acids MET790, ALA743, LEU844, and LEU718 as shown in Figure 6. The interactions between compound 15 and the active site of the 3IKA EGFR enzyme were observed to be with these group of amino acids THR854 and ASP855 through hydrogen bonds. Hydrophobic (Pi-Sigma, Pi−Pi T-shaped, alkyl, and Pi-alkyl) interactions with these set of amino acid residues PHE723, LEU747, ALA743, and LEU844 were also seen. Electrostatic (Pi-cation) interactions with these set of amino acids MET790 and PHE723 were respectively seen as depicted in Figure 7. Molecule 21 interacted with the binding poses of the 3IKA receptor with amino acids LEU844, LEU718, MET790, and CYS797 via only Pi-alkyl, Pi-sigma, and alkyl (hydrophobic interactions), as shown in Figure 8. The kinds of interactions detected for molecule 23 and the binding poses of the 3IKA receptor were hydrogen bonds with amino acid residues LYS745, and PHE723, and hydrophobic interactions with amino acids ALA743, PHE723, ALA722, MET790, VAL726, LEU844, LEU858, and LYS745 amino acids, as depicted in Figure 9.

The FDA-approved drug osimertinib (**AZD9291**) was also docked into the active site of the 3IKA EGFR receptor for comparisons between the standard drug (**AZD9291**) and the identified best hit compounds. **AZD9291** with a MolDock score of −118.872 kcal/mol and re-rank score of −91.2323 interacted via hydrogen bonds, and hydrophobic and electrostatic interactions with residues VAL726, LEU844, ALA743, LEU718, GLN791, LYS745 and PHE723, as presented in Figure 10. All of the identified hit compounds had better MolDock scores and re-rank scores than the standard drug (**AZD9291**). The current NSCLC drug (**AZD9291**) was designed and established to inhibit the $EGFRT^{790M}$ resistance mutation and also to overcome the observed toxicity with irreversible second-generation EGFR TKIs. However, it could not achieve the stated goal due to emergence of resistance by the C797S mutation. Furthermore, these best hit compounds (EGFR NSCLC therapeutic agents) could act as better therapeutic agents compared to the current NSCLC drug (**AZD9291**) by having higher affinity towards $EGFRT^{790M}$ mutation and thus, inhibiting the $EGFRT^{790M/C797S}$ resistance mutation.

WZ4002 was well re-docked into the binding poses of the 3IKA receptor to successfully validate the molecular docking procedure. **WZ4002** with a MolDock score of −103.728 kcal/mol and re-rank score of −94.249 interacted with amino acids MET793, ASP855, GLY796, PHE723, LEU718, GLY719, and VAL726 after successfully re-docking it into the binding poses of the 3IKA receptor as depicted in Figure 11. The amino acid LEU718, MET793, and VAL726 were common to the native ligand and the re-docked one.

Drug-likeness, ADMET prediction, and evaluation

The SwissADME web server was utilized to theoretically predict the drug-like features of the investigated compounds (Supplementary Table 2) including the best hit compounds identified by adopting the filtering conditions of the Lipinski's rule of five.[45] Any small molecule that violates more than one of the stated conditions may have issues related to its bioavailability. The drug-like features of the best hit compounds were first filtered using the Lipinski's rule of five filtering criteria. The evaluated drug-like features showed only one violation for all (that is their molecular weight was more than 500) (Table 7). All of the best hit compounds had three hydrogen bond donors, which was within the accepted range. For the number of hydrogen bond acceptors, between 5 and 7 is also within the accepted range. Furthermore, the calculated MlogP value was never beyond the accepted range. Based on these filtering conditions, the best hit compounds were ascertained to be drug-like in nature by not violating the threshold values set (or by not having more than one violation of the filtering conditions used). Furthermore, based on the value of their bioavailability score (all having 0.55), they were all further confirmed to be orally bioavailable. Generally, these compounds were predicted not to have any issue associated with their bioavailability as they followed all of the filtering criteria used.

The online web server pkCSM was used to theoretically evaluate the ADMET/pharmacokinetic features of the investigated compounds including the best hit compounds identified (Supplementary Table 3). The best identified hit compounds were all observed to have human intestinal absorption between the range of 98.004% and 100%. The values of their human intestinal absorption were found to be greater than the minimum recommended rate of 30% set for the evaluation of this property, which is an indication that these small molecules can be absorbed within the human intestine. The recognized threshold value set for the blood brain barrier (BBB) permeability is >0.3 to <−1 and that for central nervous system (CNS) permeability is >−2 to <−3. The BBB permeability for these small molecules were observed to be >−1 with the exception of compound 32, which was <−1, indicating that the compounds poorly infiltrate/penetrate the BBB with the exception of compound 32. Regarding CNS permeability value, it was <−2 for all except compound 25 (−2.061), which indicated that they poorly infiltrate/penetrate the CNS with the exception of compound 25. The cytochrome (CYP) is regarded as very important in the enzymatic breakdown/metabolism of small molecules in the body. As such, it is necessary to take into account the breakdown/metabolism of these small molecules in the human body. CYP1A2, 2C9, 2C19, 2D6 and 3A4 are responsible for the breakdown/metabolism of small molecules in the body, of which CYP3A4 is the most significant (a good small molecule is expected to be both a substrate and inhibitor of CYP3A4). The best recognized hits were all substrates and inhibitors of CYP3A4. Based on this, it was further confirmed that these small molecules can be broken down in the body. Another important factor is how these small molecules can be removed from the body, as excretion/total clearance defines the connection/relationship between the rate of their removal and concentration within the body. The best identified hit compounds displayed a greater value of excretion and were found to be in the accepted threshold for a drug. Based on the toxicity test, they were all found to be nontoxic except compound 23. Finally, the general ADMET features of these best identified hit compounds displayed average pharmacokinetic profiles (Tale 11).

Conclusion

The 2D-QSAR molecular modeling technique was employed on the 45 sets of tetrahydropyrazolo-quinazoline and tetrahydropyrazolo-pyrimidocarbazole derivatives under investigation to generate a model with high predictive capability. The training model building set was used to generate four diverse models. In the four diverse QSAR models generated, model 1 was the first model selected and was used due its high $R^2 = 0.798$, $R^2_{adj} = 0.754$, $Q^2_{CV} = 0.673$, $R^2_{test} = 0.531$, and $cRp^2 = 0.627$ assessment parameters. These assessment parameters for the selected model (model 1) were compared with the accepted threshold values used for the assessment of QSAR models and found to be greater than the accepted threshold values, which clearly showed the internal predictive capability of model 1.

Molecular docking-based virtual screening identified compounds 25, 32, 15, 21 and 23 as the best hit compounds, as they had the highest MolDock scores among the investigated compounds. The MolDock score values of the investigated compounds ranged between −138.329 and −128.867. Compound 25 had the highest MolDock score of −138.329 kcal/mol and re-rank score of −95.449. All of the identified hit compounds were better than the reference drug (AZD9291 with a MolDock score of −118.872 kcal/mol and re-rank score of −91.2323) in terms of their MolDock score and re-rank score. Furthermore, these best hit compounds may be better therapeutic agents than the current NSCLC drug (**AZD9291**) by having higher affinity towards $EGFRT^{790M}$ mutation, and thus inhibiting the $EGFRT^{790M/C797S}$ resistance mutation.

The best hit compounds were found to be drug-like in nature due to their compliance with the filtering criteria used for evaluation of the drug-likeness of a small molecule (or by not having more than one violation of the filtering conditions used). Furthermore, based on the value of their bioavailability score of 0.55, they were all confirmed to be orally bioavailable. Their ADMET features displayed their average pharmacokinetic profiles. These hit compounds can serve as

potential NSCLC therapeutic agents because of their safety and efficacy after successful preclinical trial testing, with the exception of compound 23 which was found to be toxic. Also, they can serve as a template for designing new NSCLC therapeutic agents.

Source of funding

This research received a specific grant from the Tertiary Education Trust Fund (Tetfund) (No. TETF/DR&D/UNI/ZARIA/IBR/2020/VOL.1/52).

Conflict of interest

The authors have no conflict of interest to declare.

Ethical approval

There are no ethical or financial issues or animal experiments related to this research.

Authors' contributions

MTI and AU formulated, conceptualized, and designed the research. MTI conducted the program using software, organized the data, and analyzed the results. MTI wrote the final draft of the paper. AU reviewed and edited the draft. All authors have critically reviewed and approved the final draft and are responsible for the content and similarity index of the manuscript.

Acknowledgments

The authors sincerely acknowledge the Tertiary Education Trust Fund (Tetfund) for sponsoring this research under the Tetfund IBR research project grant (No. TETF/DR&D/UNI/ZARIA/IBR/2020/VOL.1/52) as well as Ahmadu Bello University, Zaria for their technical support during the course of this research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jtumed.2022.09.002.

References

1. Sato JD, Okamoto T, Barnes D, Hayashi J, Serrero G, McKeehan WL. A tribute to Dr. Gordon Hisashi Sato (December 24, 1927–March 31, 2017). **In Vitro Cell Dev Biol Animal 2018**; 54(3): 177–193.
2. Othman IM, Alamshany ZM, Tashkandi NY, Gad-Elkareem MA, Anwar MM, Nossier ES. New pyrimidine and pyrazole-based compounds as potential EGFR inhibitors: synthesis, anticancer, antimicrobial evaluation and computational studies. **Bioorg Chem 2021**; 114:105078.
3. Singh M, Jadhav HR. Targeting non-small cell lung cancer with small-molecule EGFR tyrosine kinase inhibitors. **Drug Discov Today 2018**; 23(3): 745–753.
4. Hanan EJ, Baumgardner M, Bryan MC, Chen Y, Eigenbrot C, Fan P, et al. 4-Aminoindazolyl-dihydrofuro [3, 4-d] pyrimidines as non-covalent inhibitors of mutant epidermal growth factor receptor tyrosine kinase. **Bioorg Med Chem Lett 2016**; 26(2): 534–539.
5. Bistrović A, Krstulović L, Harej A, Grbčić P, Sedić M, Koštrun S, et al. Design, synthesis and biological evaluation of novel benzimidazole amidines as potent multi-target inhibitors for the treatment of non-small cell lung cancer. **Eur J Med Chem 2018**; 143: 1616–1634.
6. Khuri F. Lung cancer and other pulmonary neoplasms. **Goldman Cecil Med 2016**; 2: 1303–1313.
7. Tièche CC, Peng R-W, Dorn P, Froment L, Schmid RA, Marti TM. Prolonged pemetrexed pretreatment augments persistence of cisplatin-induced DNA damage and eliminates resistant lung cancer stem-like cells associated with EMT. **BMC Cancer 2016**; 16(1): 1–15.
8. Ibrahim MT, Uzairu A, Shallangwa GA, Uba S. Molecular docking investigation and pharmacokinetic properties prediction of some anilinopyrimidines analogues as egfr t790m tyrosine kinase inhibitors. **Egypt J Basic Appl Sci 2021**; 8(1): 203–213.
9. Wang D, Zhou J, Fang W, Huang C, Chen Z, Fan M, et al. A multifunctional nanotheranostic agent potentiates erlotinib to EGFR wild-type non-small cell lung cancer. **Bioact Mater 2022**; 13: 312–323.
10. Passaro A, Mok T, Peters S, Popat S, Ahn M-J, De Marinis F. Recent advances on the role of EGFR tyrosine kinase inhibitors in the management of NSCLC with uncommon, non exon 20 insertions, EGFR mutations. **J Thorac Oncol 2021**; 16(5): 764–773.
11. Hsu W-H, Yang J-H, Mok T, Loong H. Overview of current systemic management of EGFR-mutant NSCLC. **Ann Oncol 2018**; 29: i3–i9.
12. Song Z, Huang S, Yu H, Jiang Y, Wang C, Meng Q, et al. Synthesis and biological evaluation of morpholine-substituted diphenylpyrimidine derivatives (Mor-DPPYs) as potent EGFR T790M inhibitors with improved activity toward the gefitinib-resistant non-small cell lung cancers (NSCLC). **Eur J Med Chem 2017**; 133: 329–339.
13. Ibrahim MT, Uzairu A, Shallangwa GA, Uba S. Lead identification of some anti-cancer agents with prominent activity against non-small cell lung cancer (NSCLC) and structure-based design. **Chem Africa 2020**; 3(4): 1023–1044.
14. Pawara R, Ahmad I, Surana S, Patel H. Computational identification of 2, 4-disubstituted amino-pyrimidines as L858R/T790M-EGFR double mutant inhibitors using pharmacophore mapping, molecular docking, binding free energy calculation, DFT study and molecular dynamic simulation. **Silico Pharmacol 2021**; 9(1): 1–22.
15. Patel H, Pawara R, Ansari A, Surana S. Recent updates on third generation EGFR inhibitors and emergence of fourth generation EGFR inhibitors to combat C797S resistance. **Eur J Med Chem 2017**; 142: 32–47.
16. Chen L, Fu W, Feng C, Qu R, Tong L, Zheng L, et al. Structure-based design and synthesis of 2, 4-diaminopyrimidines as EGFR L858R/T790M selective inhibitors for NSCLC. **Eur J Med Chem 2017**; 140: 510–527.
17. Karnik KS, Sarkate AP, Tiwari SV, Azad R, Burra PV, Wakte PS. Computational and synthetic approach with biological evaluation of substituted quinoline derivatives as small molecule L858R/t790m/C797S triple mutant EGFR inhibitors targeting resistance in non-small cell lung cancer (NSCLC). **Bioorg Chem 2021**; 107:104612.
18. Khanna I. Drug discovery in pharmaceutical industry: productivity challenges and trends. **Drug Discov Today 2012**; 17(19–20): 1088–1102.
19. Wang X, Song K, Li L, Chen L. Structure-based drug design strategies and challenges. **Curr Top Med Chem 2018**; 18(12): 998–1006.

20. Šinko G, Maraković N. The lock is the key: development of novel drugs through receptor based combinatorial chemistry. **Acta Chim Slov 2017**; 64(1): 15–39.
21. Ibrahim MT, Uzairu A, Umar BA, Sadiq BA, Isyaku Y. Molecular modelling, docking and pharmacokinetic studies of N-arylidenequinoline-3-carbohydrazides analogs as novel β-glucuronidase inhibitors. **J Mexican Chem Soc 2020**; 64(1): 30–40.
22. Bello A, Uzairu A, Ibrahim M, Gatugel Y. Quantum modelling analysis of some potent indole derivatives on ns5b polymerase inhibitors. **Sci World J 2019**; 14(1): 32–37.
23. Daina A, Michielin O, Zoete V. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. **Sci Rep 2017**; 7: 42717.
24. Mohareb RM, Abbas NS, Mohamed AA. Synthesis of tetrahydropyrazolo-quinazoline and tetrahydropyrazolo-pyrimidocarbazole derivatives as potential anti-prostate cancer agents and Pim-1 kinase inhibitors. **Med Chem Res 2017**; 26(6): 1073–1088.
25. Ibrahim MT, Uzairu A, Uba S, Shallangwa GA. Design of more potent quinazoline derivatives as EGFRWT inhibitors for the treatment of NSCLC: a computational approach. **Future J Pharmaceut Sci 2021**; 7(1): 1–11.
26. Mills N. *ChemDraw ultra 10.0 CambridgeSoft, 100 CambridgePark drive, Cambridge, MA 02140*. ACS Publications; 2006. Commercial Price: 1910fordownload, 2150 for CD-ROM; Academic Price: 710fordownload, 800 for CD-ROM, www.cambridgesoft.com.
27. Kohn W, Becke AD, Parr RG. Density functional theory of electronic structure. **J Phys Chem 1996**; 100(31): 12974–12980.
28. Yap CW. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. **J Comput Chem 2011**; 32(7): 1466–1474.
29. Grisoni F, Ballabio D, Todeschini R, Consonni V. Molecular descriptors for structure–activity applications: a hands-on approach. In: *Computational toxicology*. Springer; 2018. pp. 3–53.
30. Ambure P, Aher RB, Gajewicz A, Puzyn T, Roy K. "Nano-BRIDGES" software: Open access tools to perform QSAR and nano-QSAR modeling. **Chemometr Intell Lab Syst 2015**; 147: 1–13.
31. Kennard RW, Stone LA. Computer aided design of experiments. **Technometrics 1969**; 11(1): 137–148.
32. Ibrahim MT, Uzairu A, Shallangwa GA, Ibrahim A. Computational studies of some biscoumarin and biscoumarin thiourea derivatives as alfa-glucosidase inhibitors. **J Eng Exact Sci 2018**; 4(2): 276–285.
33. Hansch C, Fujita T. p-σ-π Analysis. A method for the correlation of biological activity and chemical structure. **J Am Chem Soc 1964**; 86(8): 1616–1626.
34. Tropsha A, Gramatica P, Gombar VK. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. **QSAR Comb Sci 2003**; 22(1): 69–77.
35. Bajorath ATJ. *Computational methods for drug discovery and design*. ACS Publications; 2015.
36. Ghamali M, Chtita S, Hmamouchi R, Adad A, Bouachrine M, Lakhlifi T. The inhibitory activity of aldose reductase of flavonoid compounds: combining DFT and QSAR calculations. **J Taibah Univ Sci 2016**; 10(4): 534–542.
37. Ibrahim MT, Uzairu A, Shallangwa GA, Uba S. Structure-based design and activity modeling of novel epidermal growth factor receptor kinase inhibitors; an in silico approach. **Sci African 2020**:e00503.
38. Chang J, Ren H, Zhao M, Chong Y, Zhao W, He Y, et al. Development of a series of novel 4-anlinoquinazoline derivatives possessing quinazoline skeleton: design, synthesis, EGFR kinase inhibitory efficacy, and evaluation of anticancer activities in vitro. **Eur J Med Chem 2017**; 138: 669–688.
39. Abdullahi M, Uzairu A, Shallangwa GA, Arthur DE, Umar BA, Ibrahim MT. Virtual molecular docking study of some novel carboxamide series as new anti-tubercular agents. **Eur J Chem 2020**; 11(1): 30–36.
40. Hadni H, Elhallaoui M. 3D-QSAR, docking and ADMET properties of aurone analogues as antimalarial agents. **Heliyon 2020**; 6(4):e03580.
41. Veerasamy R, Rajak H, Jain A, Sivadasan S, Varghese CP, Agrawal RK. Validation of QSAR models-strategies and importance. **Int J Drug Des Discov 2011**; 3: 511–519.
42. Tropsha A, Golbraikh A. Predictive quantitative structure-activity relationships modeling. In: *Handbook of chemoinformatics algorithms*33; 2010. p. 211.
43. Beheshti A, Pourbasheer E, Nekoei M, Vahdani S. QSAR modeling of antimalarial activity of urea derivatives using genetic algorithm–multiple linear regressions. **J Saudi Chem Soc 2016**; 20(3): 282–290.
44. Adedirin O, Uzairu A, Shallangwa GA, Abechi SE. QSAR and molecular docking based design of some n-benzylacetamide as?-aminobutyrate-aminotransferase inhibitors. **J Eng Exact Sci 2018**; 4(1): 65–84.
45. Lipinski CA. Lead-and drug-like compounds: the rule-of-five revolution. **Drug Discov Today Technol 2004**; 1(4): 337–341.

How to cite this article: Ibrahim MT, Uzairu A. 2D-QSAR, molecular docking, drug-likeness, and ADMET/pharmacokinetic predictions of some non-small cell lung cancer therapeutic agents. J Taibah Univ Med Sc 2023;18(2):295–309.