



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

Rahul Pawara¹ · Iqrar Ahmad¹ · Sanjay Surana¹ · Harun Patel¹

Received: 29 April 2021 / Accepted: 24 September 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Pharmacophore modelling studies have been performed for a series of 2,4-disubstituted-pyrimidines derivatives as EGFR L858R/T790M tyrosine kinase inhibitors. The high scoring AARR.15 hypothesis was selected as the best pharmacophore model with the highest survival score of 3.436 having two hydrogen bond acceptors and two aromatic ring features. Pharmacophore-based virtual screening followed by structure-based yielded the six molecules (ZINC17013227, ZINC17013215, ZINC9573324, ZINC9573445, ZINC24023331 and ZINC17013503) from the ZINC database with significant in silico predicted activity and strong binding affinity towards the EGFR L858R/T790M tyrosine kinase. In silico toxicity and cytochrome profiling indicates that all the 06 virtually screened compounds were substrate/inhibitors of the CYP-3A4 metabolizing enzyme and were non-carcinogenic and devoid of Ames mutagenesis. Density functional theory (DFT) and molecular dynamic (MD) simulation further validated the obtained hits.

Keywords 2,4-disubstituted-pyrimidines · Pharmacophore · Molecular docking · Virtual screening · MMGBSA · Molecular dynamic simulation

Introduction

Non-small cell lung cancer (NSCLC) is the most common lung cancer globally, with less than 20% of 5-year survival after diagnosis (Molina et al. 2008). The epidermal growth factor receptor (EGFR) belongs to the protein kinase family, a clinically established NSCLC treatment target, implicated in various cell signalling cascades that are critical for cell growth, proliferation, survival, and migration (Vansteenkiste and Schildermans 2005; Hirsch et al. 2003; Ohsaki et al. 2000; Sharma et al. 2007). In NSCLC, EGFR mutations primarily induce overexpression of EGFR or ‘Classic Mutation’ of exon 19 deletions (delE746-A750) and L858R substitutions in exon21 frame, contributing for 90% of activating

mutations (Kawahara et al. 2010). FDA approved first-generation tyrosine kinase inhibitors (TKI) were gefitinib (2009) and erlotinib (2013) for the EGFR activating mutation in positive NSCLC patients (Cohen et al. 2003; Bonomi 2003). However, a single point acquired drug resistance of threonine790-to-methionine790 (T790M) is often found in around 50% of patients with EGFR alteration, significantly restricting the effectiveness of these medications in clinical usage (Kobayashi et al. 2005; Shaikh et al. 2021; Engelman and Jänne 2008; Pao et al. 2005). To resolve the resistance induced by the T790M, second-generation EGFR-TK inhibitors (afatinib and dacomitinib) were developed with an electrophilic Michael acceptor head, capable of covalently alkylating the cysteine 797 residue (Cys797) near the EGFR ATP binding site (Yun et al. 2008; Patel et al. 2017). However, their clinical effectiveness has so far been limited due to skin rash and gastrointestinal toxicity because of the lack of selectivity between WT EGFR and mutant EGFR (Patel et al. 2017; Kim et al. 2012).

Third-generation EGFR TKIs were therefore developed (e.g. WZ4002, rociletinib/CO-1686, Osimertinib/AZD9291,

✉ Harun Patel
hpatel_38@yahoo.com

¹ Division of Computer-Aided Drug Design, Department of Pharmaceutical Chemistry, R. C. Patel Institute of Pharmaceutical Education and Research, District Dhule, Shirpur, Maharashtra 425 405, India

olmutinib/HM61713, etc.), which covalently bind to Cys797 with greater selectivity towards the T790M mutation as compared to the WT EGFR (Patel et al. 2017; Song et al. 2016; Chen et al. 2017). The first recorded third-generation EGFR-TKI, with 30–100 times more potency against EGFR L858R/T790M than to WT EGFR was WZ4002 (Zhou et al. 2009). Osimertinib has similar amino-pyrimidine scaffold to WZ4002. USFDA issued accelerated approval to Osimertinib in 2015 and full approval in March 2017 for the treatment of patients with EGFR-T790M mutations and whose disease has advanced in the first and second generation of EGFR-TKI therapy (Jänne et al. 2015; Goss et al. 2016; <https://www.drugs.com/history/tagrisso.html>). It has almost 200 times higher occupancy towards L858R/T790M than the WT-type EGFR, confirming its specificity for the mutant EGFR (Song et al. 2016; Chen et al. 2017; Lu et al. 2018). However, C797S mutation to Osimertinib has recently been identified, suggesting the need to find a new drug molecule to resolve T790M/C797S mutation problem (Thress et al. 2015; Chabon et al. 2016). In context to the above-mentioned facts and in continuation of our research on EGFR-TKIs (Patel et al. 2018a, 2020a, b), the current article deals with identifying the new L858R/T790M EGFR TK inhibitor based on the pharmacophore and structural based virtual screening approaches.

Methodology

Computational details

All computational work was performed on Schrodinger, LLC, New York, USA, 2008. LigPrep, PHASE, Glide, Protein Preparation Wizard QikProp, Desmond 3.1 used for pharmacophore hypothesis, docking study, ADME study and molecular dynamics simulation study (Chaudhari and Bari 2016; Kausar and Falcao 2018). The crystal structures of T790M mutant EGFR PDB ID: 2JIU (<https://www.rcsb.org/structure/2JIU>) and WT EGFR PDB ID: 4ZAU were retrieved from RCSB Protein Data Bank (PDB) (<https://www.rcsb.org/structure/4ZAU>).

Dataset

Twenty-four derivatives of 2,4-substituted-pyrimidines previously reported as L858R/T790M EGFR-TK inhibitors were taken for generation of pharmacophore (Chan et al. 2015). These selected compounds shared the same assay procedure with significant variations in their structures and potency profiles. The inhibitory activity of the compounds reported as IC₅₀ values were considered for the whole process (Table 1) (Choubey and Jeyaraman 2016; Noolvi and Patel 2013).

Ligand preparation

For the development of an accurate pharmacophore model, a 3D structure is necessary. The 3D ligand structures were built using Maestro's builder panel and optimized with the LigPrep module (version 2.2, Schrodinger, LLC, New York, USA, 2008). LigPrep was carried out to desalt and to generate all possible tautomers and pH 7.0 states with Epik. Ligands were minimized with OPLS 2005 by retaining chiralities (Tawari et al. 2008; Wang et al. 2015; Dong and Zheng 2008; Golbraikh et al. 2003; Patel et al. 2018c).

Pharmacophore modelling

Virtual screening is a very useful application when it comes to identifying hit molecules as a beginning for medicinal chemistry (Patel et al. 2018b). Pharmacophore mapping and 3D QSAR screening are useful tool for the ligand based virtual screening. Phase module of Schrodinger is very useful module for the development and validation of Pharmacophore and 3D QSAR model (Dixon et al. 2006a, b). 3D pharmacophore modelling study was carried out using the optimized structures. In the current analysis, we have used PHASE (v3.0) for the shape screening and alignment of the L858R/T790M EGFR inhibitors. PHASE recognizes the three-dimensional arrangement of functional groups that are prevalent and essential for the biological activity of the ligands. The PHASE system also offers an integrated set of six pharmacophore properties that include hydrogen bond donor (D), hydrogen bond acceptor (A), hydrophobic group (H), positively ionizable group (P), negatively charged group (N) and aromatic ring (R) (Dixon et al. 2006a, b). Using a set of pharmacophore features, the pharmacophore model was developed to generate sites for all the compounds (Singh et al. 2011). To recognize common pharmacophore hypotheses, an active analogue approach was used (Sharma et al. 2016). Common pharmacophores were derived from active ligands' conformations utilizing a tree-based partitioning strategy, which grouped similar pharmacophores based on inter-site distances (Pan et al. 2013). A series of the chemical structure has described the prominent pharmacophoric features. The allocated structural patterns are defined as SMARTS queries that enable the physical features of the site to reflect one of three potential geometric points, vector and group points (Chekkara et al. 2017). Common pharmacophoric hypothesis (CPHs) were verified using a score function to achieve the best alignment of active ligands employing a RMSD value of 1.2 Å with the default distance tolerance. The main components of an algorithm are the alignment of site points and vectors, relative conformational energy, volume overlap, selectivity, number of ligands matched and activity.

Table 1 2,4-disubstituted-pyrimidines along with their EGFR L858R/T790M inhibition used for pharmacophore development

Sr. No.	Compounds	IC ₅₀ (μM)	Sr. No.	Compounds	IC ₅₀ (μM)
1		0.150	13		0.159
2		0.191	14		0.056
3		0.066	15		0.032
4		0.041	16		0.023
5		0.038	17		0.073
6		0.018	18		0.102
7		0.038	19		0.023
8		0.044	20		0.031
9		0.031	21		0.024
10		0.032	22		0.056
11		0.047	23		0.029
12		0.293	24		0.026

The alignment quality is calculated using three parameters: the alignment score, the vector score and the volume score. Following Eq. (1) has been used for survival score calculation.

$$S = W_{\text{site}} S_{\text{site}} + W_{\text{vec}} S_{\text{vec}} + W_{\text{vol}} S_{\text{vol}} + W_{\text{sel}} S_{\text{sel}} + W_{\text{m}} \text{rew} \quad (1)$$

where W = weights and S = scores, S_{site} = alignment score; S_{vec} = vector score S_{vol} = volume score and S_{sel} = selectivity score. W_{site} , W_{vec} , W_{vol} , and W_{rew} have default values of 1.0, while W_{sel} has a default value of 0.0. In hypothesis generation, default values have been used. $W_{\text{m}} \text{rew}$ represents reward weights defined by $m-1$, where m is the number of actives that match the hypothesis (Prabhu et al. 2014). The PHASE-identified hypothesis was graded according to how the active ligands superimpose on different features of the hypothesis (Kandakatla et al. 2014). There are two options for 3D molecule alignments in PHASE: atom-based alignment and the pharmacophore-based alignment. In the current analysis, the atom-based alignment is used for the development of pharmacophoric model (Fig. 1).

Molecular docking

The protein crystal structure of the mutant L858R/T790M EGFR-TK (PDB ID: 2JIU) and WT EGFR-TK (PDB ID: 4ZAU) were retrieved from the protein data bank. The preparation of the protein was carried out with the Protein Preparation Wizard in Maestro (Bhadoriya et al. 2015; Teli and Rajanikant 2012; Ugale et al. 2017; Tropsha et al. 2003; Golbraikh and Tropsha 2002). After verifying chemical correctness, the protein was prepared, assigning bond orders, eliminating water molecules, and adding hydrogens for pH 7.0 using Epik (Mysinger et al. 2012; <http://dude.docking.org/targets/egfr>; Jafari et al. 2018; Khadikar et al. 2001; Verma et al. 2019; Shen et al. 2004). To complete missing side chains and loops, Prime was used, and termini were capped. The protein structure was minimized by OPLS 2005

force field and default constraint of 0.30 Å RMSD (Khan et al. 2019; Roy et al. 2015; Ugale and Bari 2016; Sastry et al. 2013; Kumar and Elizabeth Sobhia 2012; Elokely and Doerksen 2013). The binding site was defined around the co-crystallized ligand, and the receptor grid was prepared. Molecular docking was accomplished in SP (Standard Precision) mode using the Glide ligand docking module (Patel et al. 2020b, c).

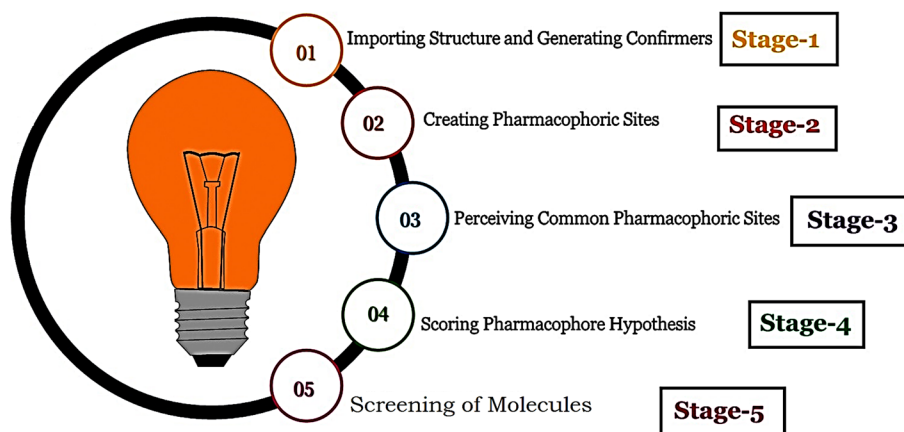
Lipinski's rule for drug likeliness and in silico ADME prediction

The primary concern about the failure of drug candidates in clinical trials is poor pharmacokinetics (Kennedy 1997). The integration of ideal ADME properties in early stages will produce the best candidates to prevent subsequent attrition and efficiently pass-through clinical trials. With this objective, the drug likeliness of the compounds was further predicted by analyzing the pharmacokinetic profile of compounds using the Qikprop module (QikProp 2010). The QikProp v4.3 module uses the Jorgensen method for the estimation of pharmacokinetic properties. Physically important descriptors and pharmaceutically significant properties of all test compounds, such as molecular weight, H-bond donors, $\log p$ and H-bond acceptors, were analyzed per Lipinski's rule of five. Lipinski's Five rule is a thumb rule for evaluating drug-likeness, or deciding whether a chemical compound with specific pharmacological or biological activities possesses properties that would make it orally active drug (Lipinski et al. 1997; Benet et al. 2016; Doak and Kihlberg 2017).

Sites of metabolism (SOM) prediction using SMARTCyp

SOM of the virtual hits was predicted by SMARTCyp software developed by Patrik Rydberg and Lars Olsen (Rydberg et al. 2010, 2013). SMARTCyp predicts the sites in a

Fig. 1 Stages in pharmacophore generation and screening of the molecules



molecule that are labile for metabolism by Cytochromes. The SMARTCyp tool accepts molecules in MOL, SDF or smiles format and assesses the SOM by detecting the two (dimensional (2D) molecule motifs) and P450 isoform 3A4 mediated metabolic sites. This tool evaluates the oxidation states of aliphatic carbons and aromatic sites by applying density functional theory (Rydberg et al. 2013). SMARTCyp is freely available from the internet at https://smarcyp.sund.ku.dk/mol_to_som.

Binding free energy calculation using prime MMGBSA approach

Using the default settings of the Prime MM-GBSA modules implemented in the Schrödinger program, binding free energies (ΔG_{Bind}) of selected protein–ligand complexes were calculated by MM-GBSA approach. Glide pose viewer file was used for the Prime MM-GBSA analysis. To determine the relative binding affinity of ligands to the receptor, the MM-GBSA calculations are used. Since the MM-GBSA binding energies are estimated to be free binding energies, the more negative value indicates stronger binding affinity (Lyne et al. 2006; Chen et al. 2016; Ahmad et al. 2020).

The binding free energies for the protein–ligand complex is given by

$$\Delta G_{\text{bind}} = G_{\text{Complex}} - G_{\text{Ligand}} - G_{\text{Protein}} \quad (2)$$

Where $G = EMM + GSB + GNP$

EMM is the molecular mechanics energy; *GSB* is the surface generalized born solvation model for polar solvation, and *GNP* is the solvation term for the non-polar part.

Molecular dynamic simulation

Molecular dynamics (MD) simulations were performed using the Desmond software, an explicit solvent MD program and OPLS 2005 force field, for the top dock protein–ligand complex (Chow et al. 2008; Bowers et al. 2006; Raghu et al. 2014; Shivakumar et al. 2010). The protein–ligand complex was prepared in the protein preparation wizard with the predefined SPC (simple point charge) water model and an orthorhombic box shape. Sodium chloride with a physiological concentration of approximately 0.15 M was put between the protein atoms and the simulation box to set the ionic strength in 10 Å buffer regions (Deniz et al. 2016; Evans and Holian 1985). Minimization tasks have been carried out to relax the system to a local energy minimization; afterwards, this model system was submitted to 200 ns MD simulation steps using the OPLS_2005 force field. Noose-Hover chain thermostat algorithm at 300 K, Coulombic cut-off at 0.9 nm, Martyna-Tobias-Klein barostat algorithm at 1.01325 bar were employed for the MD simulation. The rest

of the parameters were default (Rydberg et al. 2013; Lyne et al. 2006; Chen et al. 2016; Ahmad et al. 2020; Chow et al. 2008; Bowers et al. 2006; Raghu et al. 2014; Shivakumar et al. 2010; Deniz et al. 2016; Evans and Holian 1985; Cho et al. 1993; Patel et al. 2018d; Shinoda and Mikami 2003). The trajectories of MD simulations evaluated for ligand-receptor interactions were identified using the simulation interaction diagram (SID).

Density functional theory (DFT) calculations

It is essential to determine the structural behaviour of the active compound after the completion of the ligand-based and structure-based virtual screening and to explore how structural orientation biologically influences, and what parameters can affect the molecule's biological activity. For this reason, single-point energy calculations using DFT were performed to explore the detailed facets in terms of structure, electronics, and energy states of every atom of compound. Virtually screened compounds were imported into the Jaguar platform in Schrodinger to compute the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) by using Lee–Yang–Parr correlation functional theory (B3LYP), incorporation of basis set 6-31G* level and hybrid DFT with Beckes 3-parameter exchange potential (Bochevarov et al. 2013; Panwar and Singh 2020; Jordaan et al. 2020). These factors play a significant role in explaining the magnitude of compounds interaction in the binding pocket of EGFR-TK.

Results and discussion

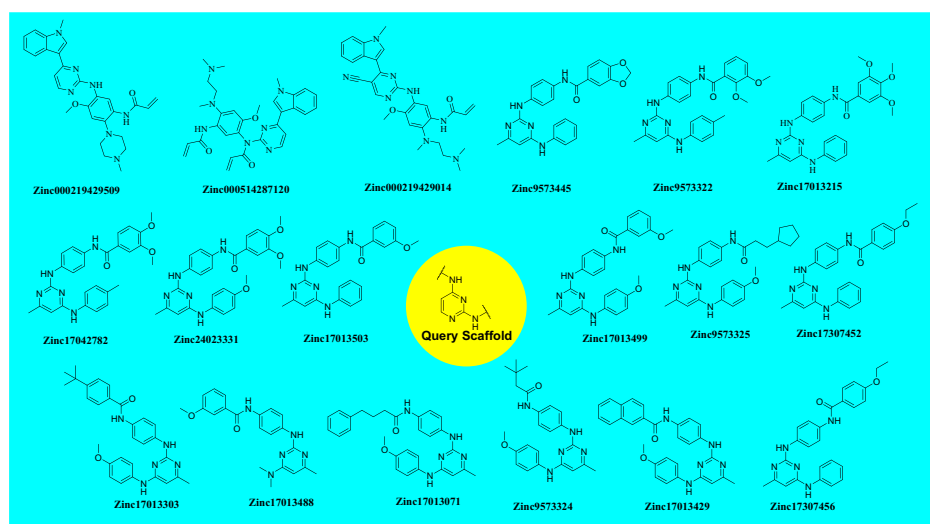
Zinc database mining

In an attempt to identify new hits that could potentially inhibit EGFR L858R/T790M, the Zinc database was mined using the query scaffold of 2,4-diminopyrimidine (basic ring of Osimertinib). The Osimertinib is the only drug approved by the FDA for the treatment of EGFR L858R/T790M mutation that contains a 2,4-diminopyrimidine scaffold, and to identify molecules with similar structures, we have set the Tanimoto coefficient of > 80% to mine the Zinc database that results in 98 compounds (Fig. 2a–e).

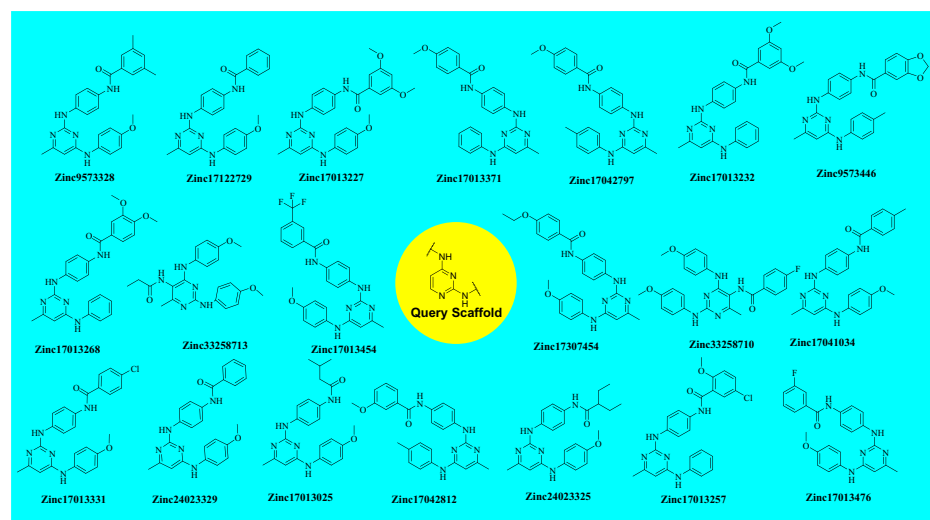
Pharmacophoric mapping

Virtual screening campaign started with identifying 3D pharmacophoric features in a molecule that are selective for L858R/T790M EGFR-TK inhibition. A systematic diagram of virtual screening protocol is shown in Fig. 3. To create the pharmacophoric hypotheses based on the sites and numbers of pharmacophore points (common in all 24 ligands),

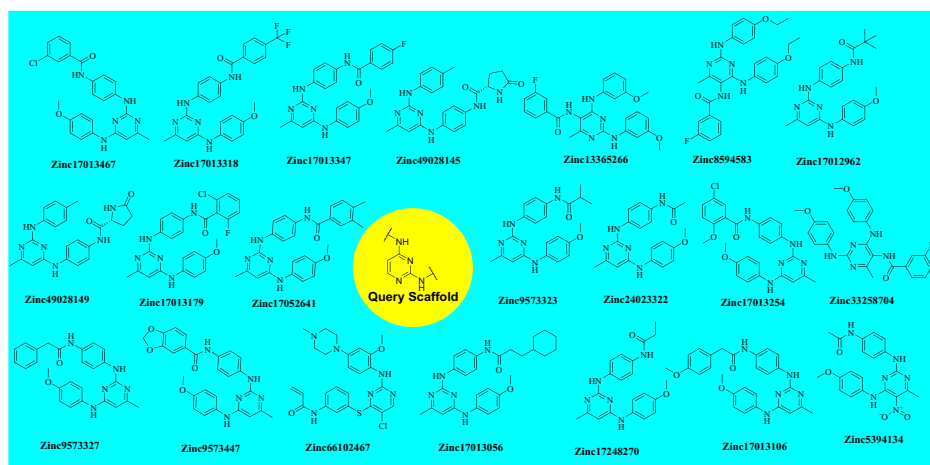
Fig. 2 **a** Zinc database mining using 2,4-diamino pyrimidine as query scaffold. **b** Zinc database mining using 2,4-diamino pyrimidine as query scaffold. **c** Zinc database mining using 2,4-diamino pyrimidine as query scaffold. **d** Zinc database mining using 2,4-diamino pyrimidine as query scaffold. **e** Zinc database mining using 2,4-diamino pyrimidine as query scaffold



A Zinc database mining using 2,4-diamino pyrimidine as query scaffold

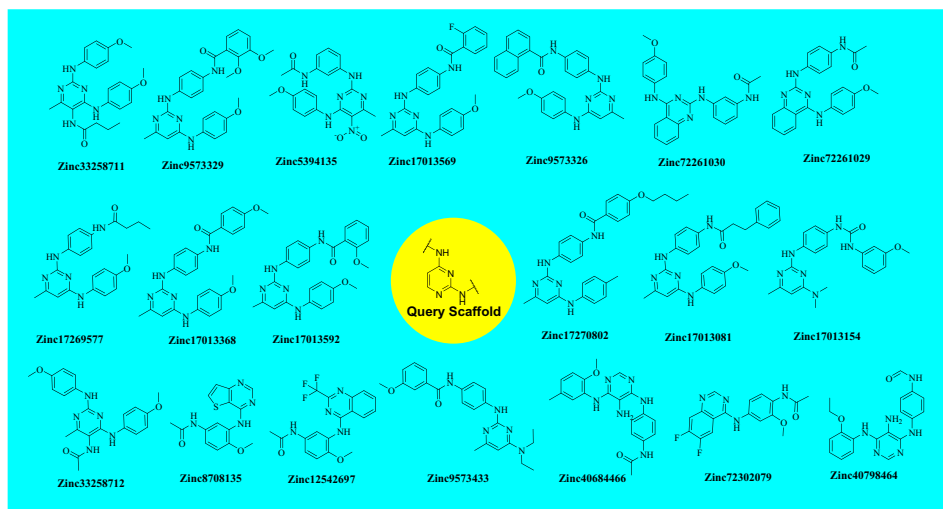


B Zinc database mining using 2,4-diamino pyrimidine as query scaffold

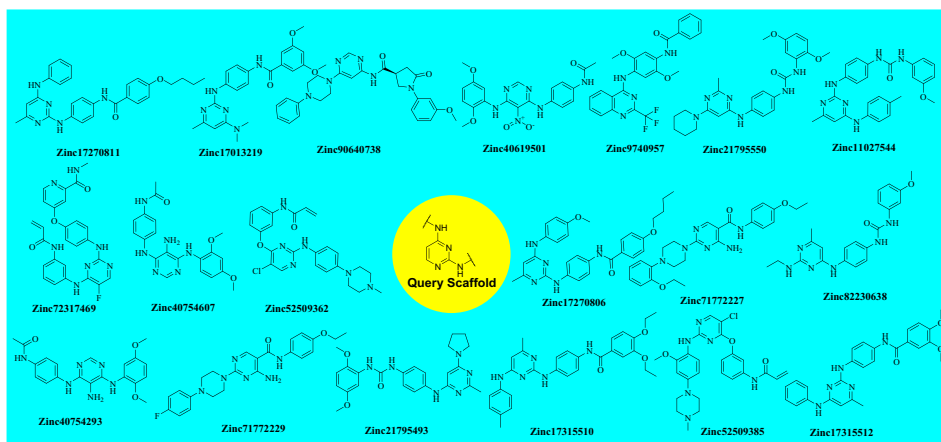


C Zinc database mining using 2,4-diamino pyrimidine as query scaffold

Fig. 2 (continued)



D Zinc database mining using 2,4-diamino pyrimidine as query scaffold



E Zinc database mining using 2,4-diamino pyrimidine as query scaffold

maximum four features were permitted for developing the hypotheses. 94 hypotheses have been developed, scored and classified according to their vector, scale, site scores, survival scores and survival activities (Table 1S).

The high scoring AARR.15 hypothesis was selected as the best pharmacophore model with the highest survival rate of 3.436 having two hydrogen bond acceptors and two aromatic ring features. In Figs. 4 and 5, pharmacophoric distances, angles and alignment of active compounds were depicted. The detail of pharmacophoric distance and angles of model AARR.15 is given in Table 2.

98 zinc mined compounds were screened through the developed pharmacophore and their alignment score was calculated. The alignment score filter (Alignment score ≥ 1.75) returned 17 molecules from 98 molecules, and the top hits were ZINC9573445, ZINC9573322, ZINC17013215, ZINC17042782, ZINC24023331, ZINC17013503, ZINC17013499, ZINC9573325, ZINC17307452, ZINC17013303, ZINC17013071,

ZINC9573324, ZINC17013429, ZINC17307456, ZINC9573328, ZINC17122729 and ZINC17013227 (Table 3).

Molecular docking against the L858R/T790M EGFR and WT EGFR

The top-scoring 17 compounds were further subjected to the structure-based virtual screening, where they have been docked on mutant (L858R/T790M) and WT EGFR. The molecular docking protocol was validated by re-docking the co-crystallized ligand at the receptor binding site. The co-crystallized ligands were docked with mutant (2JIU) and wild (4ZAU) EGFR binding pockets, and the docked poses were compared to the crystal structure pose by computing the RMSD value, which was found to be 1.2883 Å and 1.0203 Å, respectively (Fig S1, supplementary material). Docking method is generally considered legitimate if the RMSD value is less than 2 Å (Ahmad et al. 2021a,

Fig. 3 In silico modus operandi of virtual screening

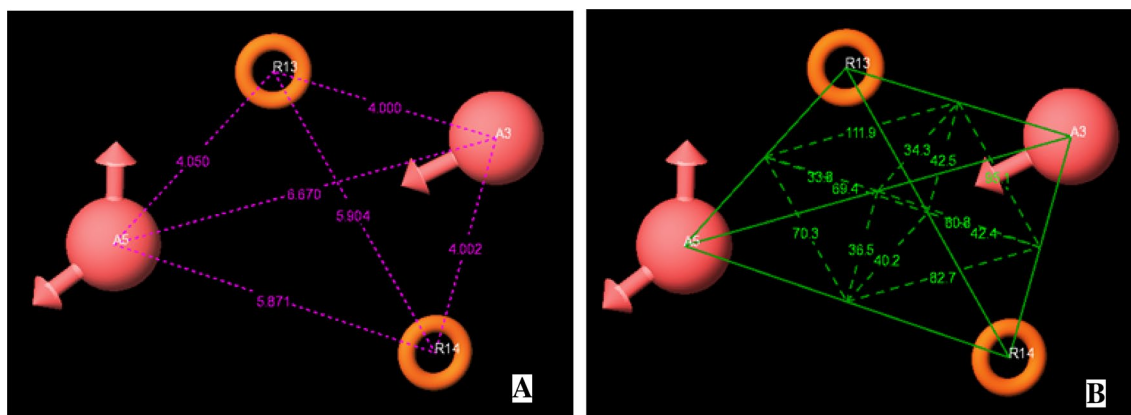
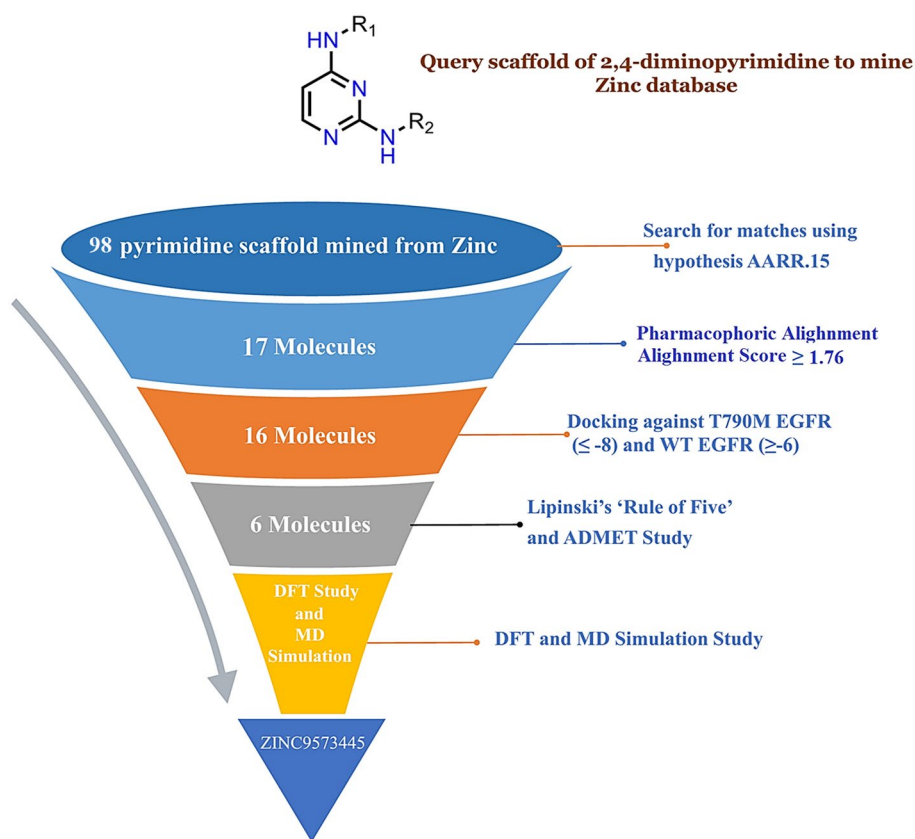


Fig. 4 Distances and angles between the pharmacophoric points of hypothesis AARR.15 (Site 1, 2, 3 and 4). Red ball shows hydrogen bond acceptor site, while the brown ring R demonstrates the ring feature

b). In virtual screening, the Cutoff Glide score on mutant EGFR of ≥ -8.0 and on WT EGFR of ≤ -6.0 yielded the 16 ligands with ZINCID; ZINC9573324, ZINC09573325, ZINC09573328, ZINC17013071, ZINC17013215, ZINC17013227, ZINC17013303, ZINC17013429, ZINC17013499, ZINC17013503, ZINC17042782, ZINC17122729, ZINC17307452, ZINC24023331, ZINC9573322, ZINC9573445 (Table S2). Virtually potent hit ZINC9573445, showed the H-bond interaction with the

ASP 800 of the DFG motif of the L858R/T790M EGFR. The hit ZINC17013227 binds to the hinge region via Met793 and it also forms another two H-bond with Asp855 (ligand-NH-) and Leu745 (ligand terminal methoxy group) of L858R/T790M EGFR. The hit compound ZINC17013215 demonstrated a bidentate hydrogen binding profile, where terminal 3,4,5-trimethoxy amino phenyl moiety showed two hydrogen bonds with a positively charged amino group of Lys716 and Lys728 of L858R/T790M EGFR (Fig. 6). All of

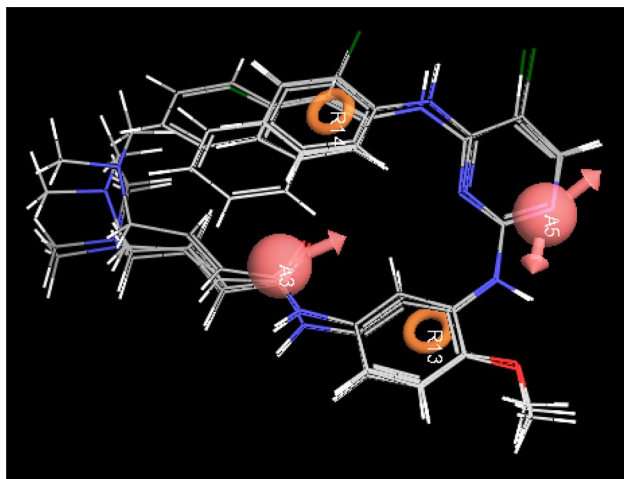


Fig. 5 Alignment of the molecules on the AARR.15 hypothesis

these obtained hits were further screened employ the next filter of “Lipinski’s rule of five” and “Jargan’s Rule of Three”.

ADMET calculations

QikProp 3.2 was used to examine drug-likeness (Lipinski’s Rule of Five), and results are given in Table S2 (supplementary material) and Table 4. It was found that only 06 of the 16 hits comply with these rules. Different pharmacokinetic parameters of these 6 compounds were further subjected to ADME predictions. The partition coefficient (QPlog Po/w) and water solubility coefficient (QPlogS), which are important for predicting the drug absorption and distribution were within ranged from 4.232 to 4.984 and -6.816 to -5.531 , respectively. Intestinal absorption or permeation is one of the significant factors to be studied concerning the absorption of the drug molecule. Further, Caco-2 cells (QPPCaco) predicted permeability indicates excellent results in determining the intestinal absorption.

MDCK cells (QPPMDCK) prediction gives an idea about the permeability of the blood brain barrier and QPlogkhsa denotes the human serum albumin binding. All the 06 compounds were falling within the prescribed range. Overall, the percentage of human oral absorption

is 100%. These data indicate the drug-like potential of screened 06 compounds (Table 4).

In silico toxicity study and cytochrome profiling

The screened molecules can only be used safely, if they are not harmful to humans. Therefore, in-silico toxicity study and cytochrome profiling for virtual hits were done using the admetSAR tool (Table 5). All the 06 virtually screened compounds were substrate/inhibitors of the CYP-3A4 metabolizing enzyme. Apart from CYP-3A4, ZINC17013227 is inhibitor of CYP-1A2; ZINC9573324 is substrate/inhibitors of CYP-2C9 and CYP-2C19; ZINC9573445 is inhibitors of CYP-1A2 and CYP-2C19; ZINC17013503 is inhibitors of CYP-1A2. Toxicity study indicates that all the six hits were non-carcinogenic and devoid of Ames mutagenesis.

The identification of a drug’s potential metabolic sites can provide crucial information about its pharmacokinetic and pharmacodynamic properties. Mostly drugs are metabolized by a special class of enzymes which are known as cytochrome P450 (CYP) enzymes. Among CYP enzymes, CYP3A4 is the most abundantly expressed CYP and accounts for approximately 30–40% of the total CYP content in human adult liver and small intestine. SMARTCyp was used to predict the CYP3A4-derived site of metabolism. The prediction of the best three sites of metabolism of the six virtual hits is presented in Fig. 7. The terminal methyl group ($-OCH_3$) of compounds ZINC17013227, ZINC17013215, ZINC9573324, ZINC17013503 and ZINC24023331 are ranked as the best SOM. This is followed by the 6-methyl group attached to the pyrimidine ring. For the compound ZINC17013227 and ZINC9573324, the amino group attached to pyrimidine ring at second position are regarded as the SOM. For the compound ZINC9573445, the methylene group ($-CH-$) between the oxygen of the benzodioxole ring was identified as the SOM. In compound ZINC9573445 and ZINC17013503 the $-CH-$ of amino phenyl are predicted as the SOM.

Table 2 Distances and angles of pharmacophoric AARR.15 hypothesis (Model-1)

Hypothesis	Site 1	Site 2	Distance (Å)	Hypothesis	Site 1	Site 2	Site 3	Angles (°)	Hypothesis	Site 1	Site 2	Site 3	Angles (°)
AARR.15	A5	A3	6.670	AARR.15	A3	A5	R13	33.8	AARR.15	A5	R13	A3	111.9
AARR.15	A5	R13	4.050	AARR.15	A3	A5	R14	36.5	AARR.15	A5	R13	R14	69.4
AARR.15	A5	R14	5.871	AARR.15	A13	A5	R14	70.3	AARR.15	A3	R13	R14	42.5
AARR.15	A3	R13	4.000	AARR.15	A5	A3	R13	34.3	AARR.15	A5	R14	A3	82.7
AARR.15	A3	R14	4.002	AARR.15	A5	A3	R14	60.8	AARR.15	A5	R14	R13	40.2
AARR.15	A13	R14	5.904	AARR.15	R13	A3	R14	95.1	AARR.15	A3	R14	R13	42.4

Table 3 Pharmacophoric fitness and alignment of the screened molecules

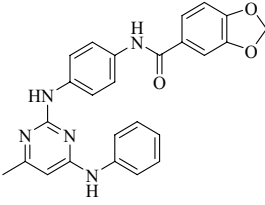
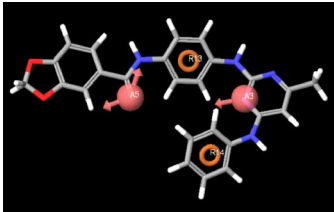
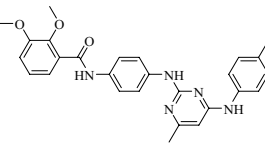
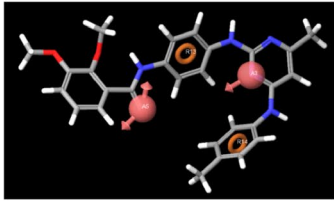
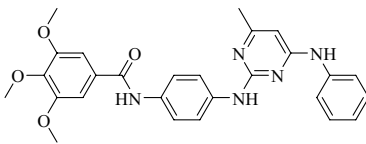
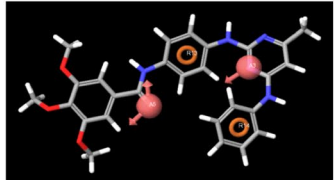
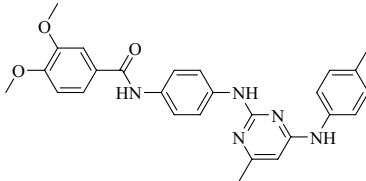
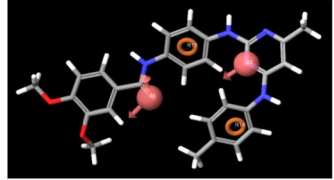
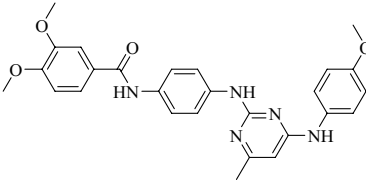
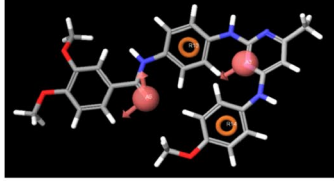
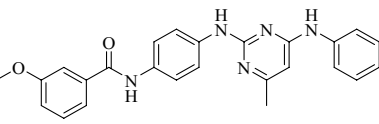
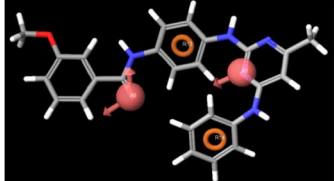
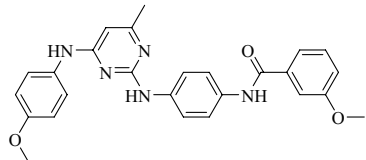
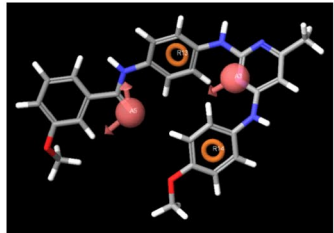
Sr. No.	ZINC ID	Structures	Pharmacophoric Fitness	Alignment Score
1	ZINC9573445			2.1045
2	ZINC9573322			2.0457
3	ZINC17013215			1.9449
4	ZINC17042782			1.9045
5	ZINC24023331			1.8896
6	ZINC17013503			1.8678
7	ZINC17013499			1.8567

Table 3 (continued)

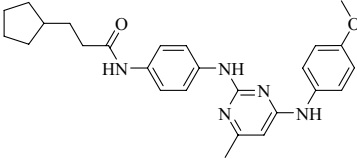
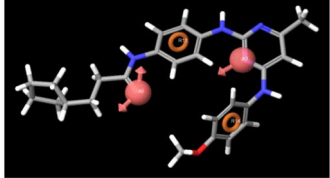
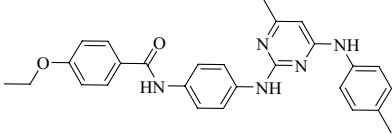
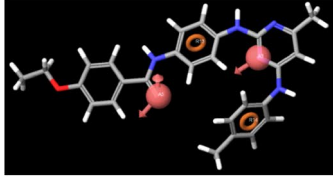
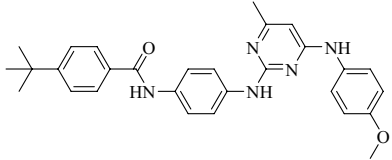
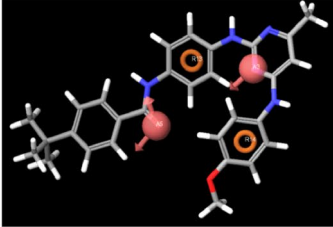
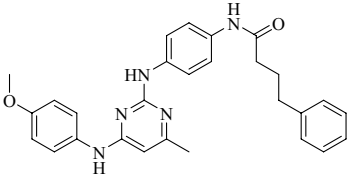
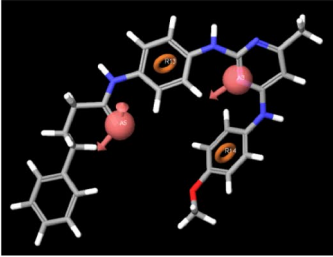
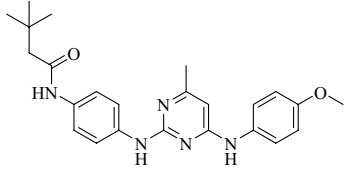
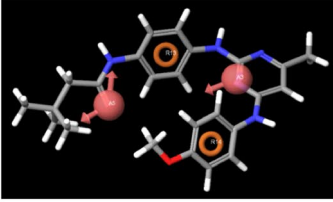
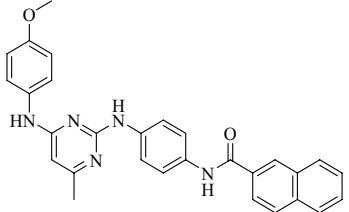
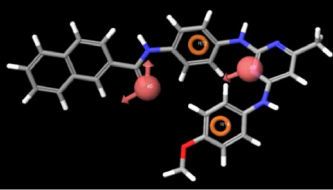
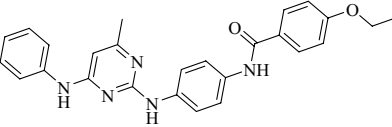
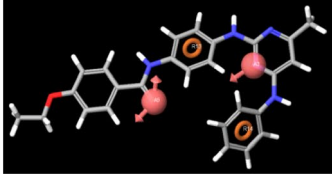
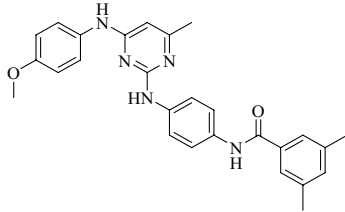
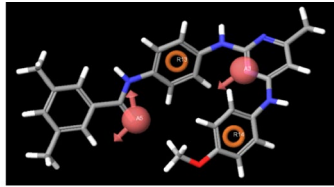
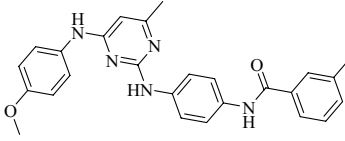
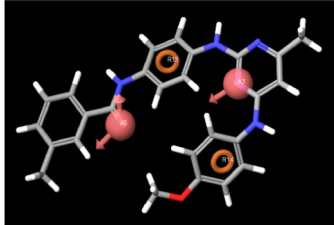
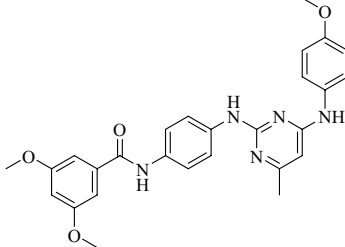
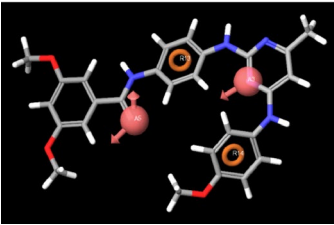
8	ZINC9573325			1.8423
9	ZINC17307452			1.8323
10	ZINC17013303			1.8311
11	ZINC17013071			1.8023
12	ZINC9573324			1.8011
13	ZINC17013429			1.7945
14	ZINC17307456			1.7856

Table 3 (continued)

15	ZINC9573328			1.7756
16	ZINC17122729			1.7645
17	ZINC17013227			1.7568

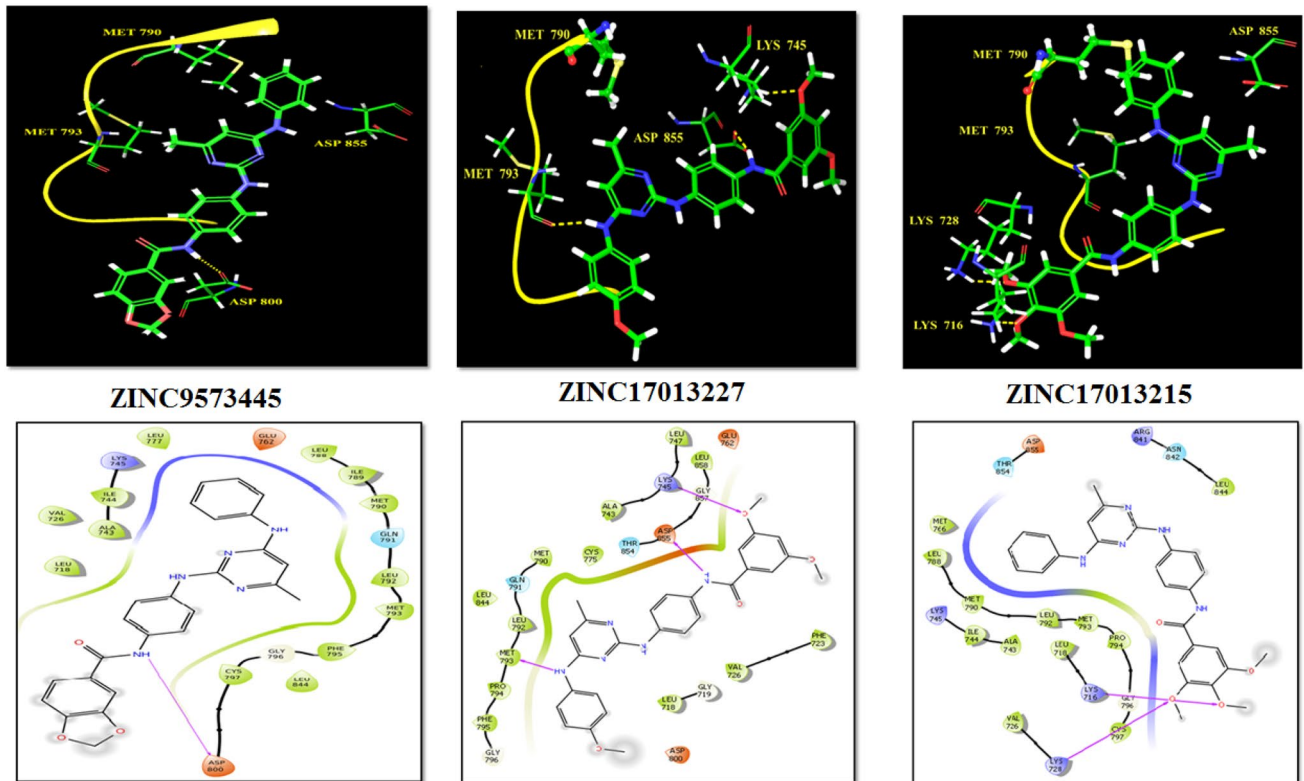


Fig. 6 Docking interaction of the virtually screened compounds with L858R/T790M EGFR TK

Table 4 Glide docking score (against the Mutant and WT EGFR), Lipinski's rule and ADME parameters of the final six ZINC hits

Zinc ID	Glide dock score (L858R/T790M EGFR)		Glide dock score (WT EGFR)		Lipinski's rule of five (drug likeness)			In silico ADME by QikProp, Schordinger 9.0					
	Docking Score	Glide Energy	Docking Score	Glide Energy	MW ^a	Donor HB ^b	Accept HB ^c	QPlogPo/w (log p) ^d	QPlogS ^e	QPPCaco ^f	QPPMDCK ^g	QlogKhsa ^h	Percent human oral absorption ⁱ
ZINC17013227	- 8.136	- 57.267	- 5.642	- 54.836	485.541	3	7.75	4.941	- 6.816	1563.187	801.776	0.646	100
ZINC17013215	- 8.78	- 56.309	- 6.062	- 56.786	485.541	3	7.75	4.984	- 6.526	1831.572	951.55	0.649	100
ZINC9573324	- 8.264	- 50.666	- 6.196	- 52.064	419.525	3	6.25	4.871	- 5.83	2449.099	1302.626	0.674	100
ZINC9573445	- 9.012	- 60.808	- 6.297	- 54.436	439.473	3	7	4.232	- 5.531	1853.529	963.886	0.446	100
ZINC24023331	- 8.916	- 58.150	- 6.395	- 59.549	485.541	3	7.75	4.944	- 6.179	2153.064	1133.301	0.622	100
ZINC17013503	- 8.858	- 56.066	- 6.439	- 55.645	425.489	3	6.25	4.838	- 6.341	1792.627	929.7	0.639	100
Rocilatinib	- 7.41	- 52.568	- 5.342	- 53.123	499.614	2	8.75	5.302	- 6.381	674.068	357.337	0.919	95.6
Osimertinib	- 8.921	- 58.868	- 4.870	- 48.432	555.558	3	10.25	4.289	- 6.314	560.362	2127.42	0.267	88.9

^aMolecular weight

^bHydrogen bond donor

^cHydrogen bond acceptor

^dPredicted octanol/water partition co-efficient log p (acceptable range: - 2.0 to 6.5)

^ePredicted aqueous solubility in mol/L (acceptable range: - 6.5 to 0.5)

^fPredicted Caco-2 cell permeability in nm/s (acceptable range: < 25 is poor and > 500 is great)

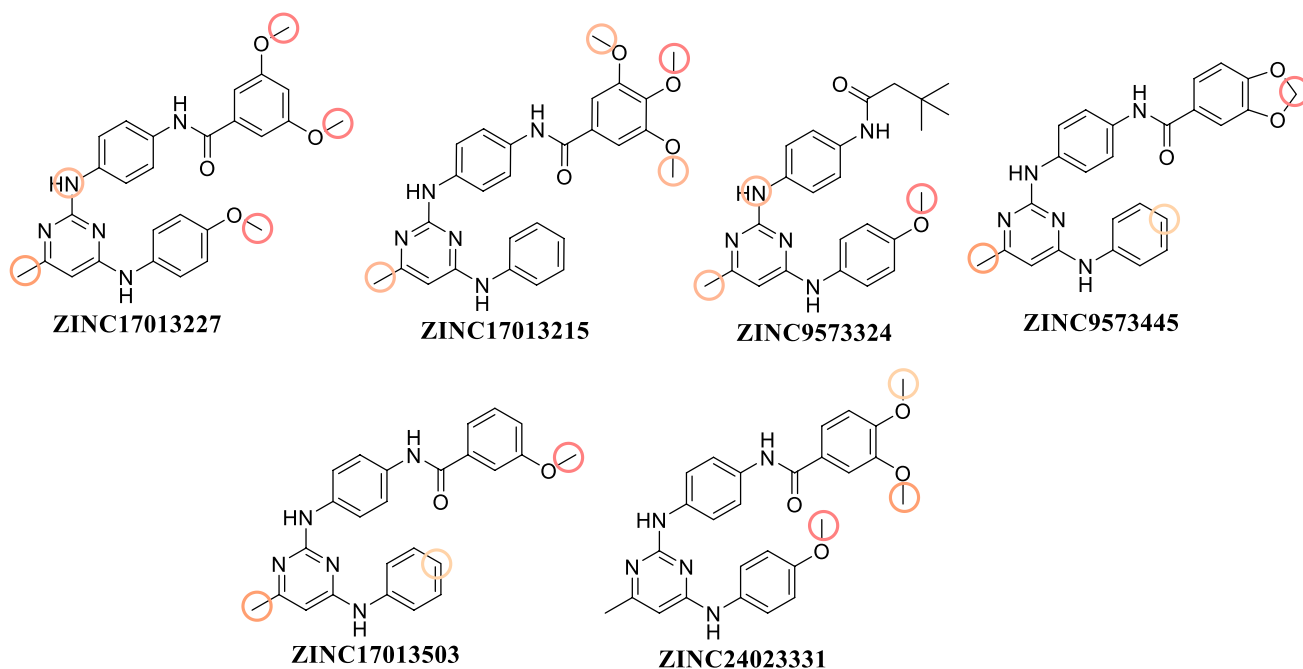
^gPredicted apparent MDCK cell permeability in nm/s

^hPrediction of binding to human serum albumin

ⁱPercentage of human oral absorption (< 25% is poor and > 80% is high)

Table 5 Effect on various metabolic cytochromes enzymes and toxicity assessment obtained from the admetSAR server of the final six ZINC hits

Zinc ID	CYP-3A4 substrate/inhibitor	CYP-2C9 substrate/inhibitor	CYP-2D6 substrate/inhibitor	CYP-1A2 inhibitor	CYP-2C19 inhibitor	Ames mutagenesis	Carcinogenicity
ZINC17013227	Substrate/inhibitor	Non-substrate	Non-substrate	Inhibitor	Non-inhibitor	Non-toxic	Non-carcinogen
ZINC17013215	Substrate/inhibitor	Non-substrate	Non-substrate	Non-inhibitor	Non-inhibitor	Non-toxic	Non-carcinogen
ZINC9573324	Substrate/inhibitor	Substrate/inhibitor	Non-substrate	Non-inhibitor	Inhibitor	Non-toxic	Non-carcinogen
ZINC9573445	Substrate/inhibitor	Non-substrate	Non-substrate	Inhibitor	Inhibitor	Non-toxic	Non-carcinogen
ZINC24023331	Substrate/inhibitor	Non-substrate	Non-substrate	Non-inhibitor	Non-inhibitor	Non-toxic	Non-carcinogen
ZINC17013503	Substrate/inhibitor	Non-substrate	Non-substrate	Inhibitor	Non-inhibitor	Non-toxic	Non-carcinogen

**Fig. 7** Representation of the site of metabolism (SOM) of six virtual hits predicted by SMARTCy

MM-GBSA (ΔG_{bind}) calculation

The six virtual hit protein–ligand complexes were analyzed by MM-GBSA (ΔG_{bind}), to determine ligands' affinities with the protein receptors. The binding energies calculated by this approach are more effective than the Glide Score values for protein–ligand complex selection (Table 6). Among the six complexes studied, ZINC9573445–L858R/T790M EGFR showed high binding free energy ($\Delta G_{\text{bind}} = -37.104$ kcal/mol) while ZINC17013227–L858R/

T790M EGFR had the second-highest binding free energy ($\Delta G_{\text{bind}} = -36.855$ kcal/mol).

DFT calculation

To validate the pharmacophore based and structure based virtual screening, DFT calculation was performed by B3LYP/6-31G** basic set (single point energy calculation) of the six-hit compounds. The HOMO and LUMO energy describe how the molecule interacts with other species and helps to understand the chemical reactivity

Table 6 Binding free energy calculations of the protein–ligand complexes obtained by MM-GBSA analysis

Zinc ID	Prime MMGBSA Ligand Energy	Prime MMGBSA ΔG bind	Prime MMGBSA ΔG bind vdW	Prime MMGBSA ΔG bind Solv GB
ZINC17013227	– 135.954	– 36.855	– 24.769	26.850
ZINC17013215	– 128.904	– 34.095	– 26.265	40.096
ZINC9573324	– 132.624	– 30.681	– 19.054	24.829
ZINC9573445	– 137.977	– 29.522	– 28.334	36.283
ZINC24023331	– 128.195	– 37.104	– 23.043	29.874
ZINC17013503	– 134.347	– 34.352	– 25.369	26.518
Co-crystalized Ligand	– 145.156	– 51.20	– 38.88	25.52

Ligand Energy: Prime energy; ΔG bind: Gibbs energy of binding; ΔG vdW: Van der Waals energy; ΔG Solv GB: Generalized Born electrostatic solvation energy

and kinetic stability of the molecule, while electronic and optical properties of the molecule can be derived from the HOMO–LUMO energy gap (ΔE) (Panwar and Singh 2020; Amala et al. 2019). The HOMO is the orbital of highest energy (electron-rich), has the potential to give electrons, and LUMO is the lowest-lying orbital that is empty (lack of the electron) has the potential to accept electrons (Chinnasamy et al. 2019). The energy of HOMO, LUMO, their energy gap (ΔE) and molecular electrostatic potential surface (MESP) of best six-hit is tabularized in Table 7. Presence of negative values of HOMO–LUMO all compounds implies good stability, which is essential to form the stable ligand–protein complex. Additionally, the energy gap represents a useful tool to determine the most active compounds.

Less energy gap between the HOMO and LUMO energies has a considerable impact on the intermolecular charge transfer and bioactivity of compounds (Amala et al. 2019). Consequently, a more energy gap observed in the compounds negatively affect the electron to move from the HOMO to the LUMO, which subsequently led to a weak affinity of the inhibitor for L858R/T790M EGFR TK.

Hence, the reactivity order increases according to: ZINC9573445 (– 0.147633 eV) > ZINC17013503 (– 0.15094 eV) > ZINC 2402331 (– 0.152296 eV) > ZINC 17013215 (– 0.153695 eV) > ZINC 09573324

(– 0.175761 eV) > ZINC17013227 (– 0.214379 eV), where the most reactive is clearly ZINC9573445 (– 0.147633 eV). Among the identified six hits, less energy gap was observed with ZINC9573445 ($\Delta E = -0.147633$ eV), where the HOMO orbitals are located in the phenyl aminopyrimidine group and LUMO are located on the carbonyl benzodioxole group (Fig. 8).

The molecular electrostatic potential surface (MESP) provides details about charge distribution (positive and negative) and also identify the reactive sites for electrophilic and nucleophilic attack in a compound for binding to protein in protein substrate interactions. The different charges are denoted in different colours. The red colour indicates negative potential (negative regions) which has the affinity to attract protons; the blue colour represents the positive potential (positive regions) which has the affinity to repulse proton and green colour represents zero potential. The negative region plays a vital role in forming a hydrogen bonding with the protein (Panwar and Singh 2020; Amala et al. 2019; Chinnasamy et al. 2019). In ZINC9573445, the most negative potential is located at the carboxamide, benzodioxole ring (Fig. 8). This region may be liable for hydrogen bonding interaction with the protein. This is well correlated with pharmacophore-based and structure-based virtual screening.

Table 7 Single point energy (Jaguar) calculations of frontier orbital energies

Compounds	HOMO (eV)	LUMO (eV)	Energy gap/ ΔE (eV)	MESP (kcal/mol)
ZINC 9573445	– 0.182627	– 0.034994	– 0.147633	– 46.1767 to 43.2420
ZINC 17013503	– 0.186684	– 0.035744	– 0.15094	– 43.8887 to 43.7169
ZINC 2402331	– 0.181064	– 0.028768	– 0.152296	– 45.4253 to 44.9598
ZINC 17013215	– 0.187215	– 0.03352	– 0.153695	– 42.6457 to 45.5245
ZINC 9573324	– 0.18544	– 0.009679	– 0.175761	– 37.2584 to 43.5196
ZINC 17013227	– 0.182098	0.032281	– 0.214379	– 45.0902 to 41.8367

The ΔE value decreases accordingly: ZINC9573445 (– 0.147633 eV) < ZINC17013503 (– 0.15094 eV) < ZINC 2402331 (– 0.152296 eV) < ZINC17013215 (– 0.153695 eV) < ZINC 09573324 (– 0.175761 eV) < ZINC17013227 (– 0.214379 eV)

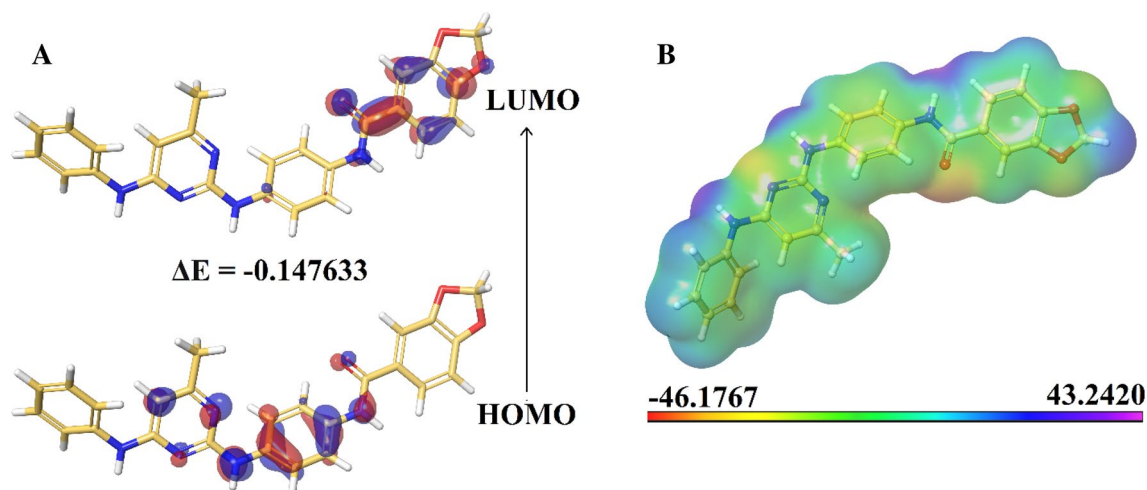


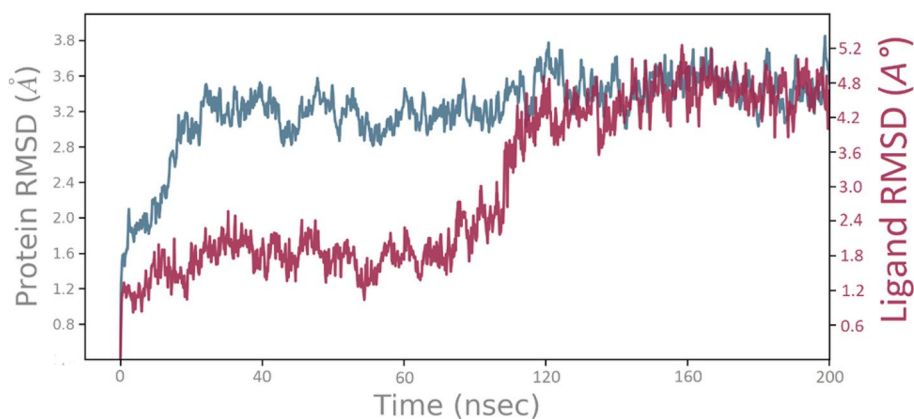
Fig. 8 HOMO, LUMO and MESP of ZINC9573445

Molecular dynamic simulation study

In the docking studies, the flexibility of the protein was not taken into consideration; therefore we have assessed drug-receptor interactions in the dynamic condition using molecular dynamic simulation to obtain the stable binding confirmation and further validate the docking result. ZINC9573445 in complex with EGFR T790M was considered for the molecular dynamic simulation for 200 ns in simple point charge (SPC) water mode. The stability of the protein–ligand complex was observed by comparing Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF) values with respect to unbound protein structure. RMSD is used to determine the average variation in the displacement of a selection of atoms for a particular frame with respect to the initial frame. As small as the RMSD

value in the simulation, it is correlated with the higher stability of the protein–ligand complex (Ahmad et al. 2020). The RMSD plot of ZINC9573445–EGFR^{L858R/T790M} complex is depicted in Fig. 9. The plot exhibited minor fluctuation of protein up to 3.7 Å, while ligand initially showed an increasing trend in RMSD from 60 to 120 ns having RMSD value range from 1.6 to 3.4 Å. Between the simulation timescale 50–120 ns, the ligand RMSD increased due to the rearrangement of ligand pose for better binding and stability within the EGFR binding Pocket. After 120 ns promising result was observed and the graph line became stable until 200 ns having a constant RMSD value of 4.8 Å. The RMSF plot represents the mobility and flexibility of each protein residue during the simulation. More RMSF values show more flexibility during the simulation while the lower value of RMSF reflects the good stability of the system (Ahmad et al. 2020).

Fig. 9 Time-dependent plot of protein–ligand RMSD (Angstrom) of the ZINC9573445–EGFR L858R/T790M complex



In this plot, Protein residues that contacts with the ligand are indicated with vertical green lines, Alpha-helical and beta-strand regions are displayed in red and blue backgrounds, respectively, while white background indicates loop region.

Secondary structural elements such as alpha helices and beta strands typically are rigid than the unstructured part of the protein and thus fluctuate less than the loop regions. High fluctuations were observed in N- and C terminal region compared to any other part of the protein. If the active site's fluctuation and the main chain atoms were mild, it indicated that the conformational change was slight (Hang et al. 2018). The RMSF plot confirmed that the ligand contacted residues showed less fluctuation from 0.8 to 1.7 Å (Fig. 10). Changes of the order of 1–3 Å are perfectly acceptable for small globular proteins.

Furthermore, protein interaction with the ligand can be monitored throughout the simulation. These interactions are categorized into hydrogen bonds, hydrophobic, and water bridges, as shown in Fig. 11. Hydrogen bond interaction

plays an imperative role in strong bond formation in ligand binding, which was seen with Ser720, Pro794, Asp800 and Asp 855.

Water mediated hydrogen bonds were also seen with Asp 855, Arg 841 and Lys 728. Hydrophobic interactions were also seen with Leu718, Ala743, Met790, and Leu844. Additionally, the (Fig. 12) shows total number of specific contacts protein makes with ligand over the course of trajectory. The contribution of amino acids in each trajectory frame of 200 ns MD simulation as shown in bottom panel of Fig. 12 which represent the number of contacts and their density, i.e., a darker shade of orange shows more than one contact in that frame. Key interactions seen during each frame were Asp800, which found consistent during the complete simulation process. Other interactions were also found Leu718, Lys728, Ala743, Met790, Leu841 and Asp85, which were not consistent during the simulation.

Fig. 10 Time-dependent Protein RMSF plots (Angstrom) of the ZINC9573445–EGFR L858R/T790M complex

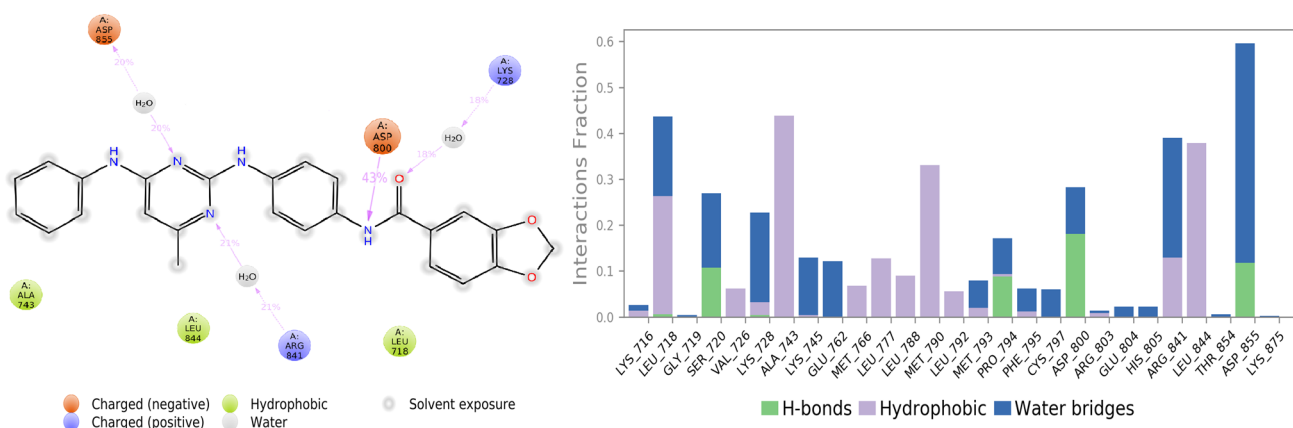
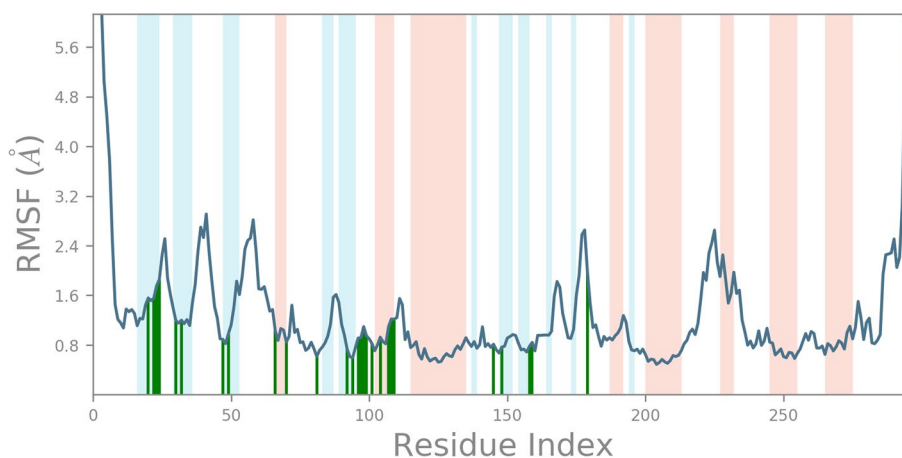
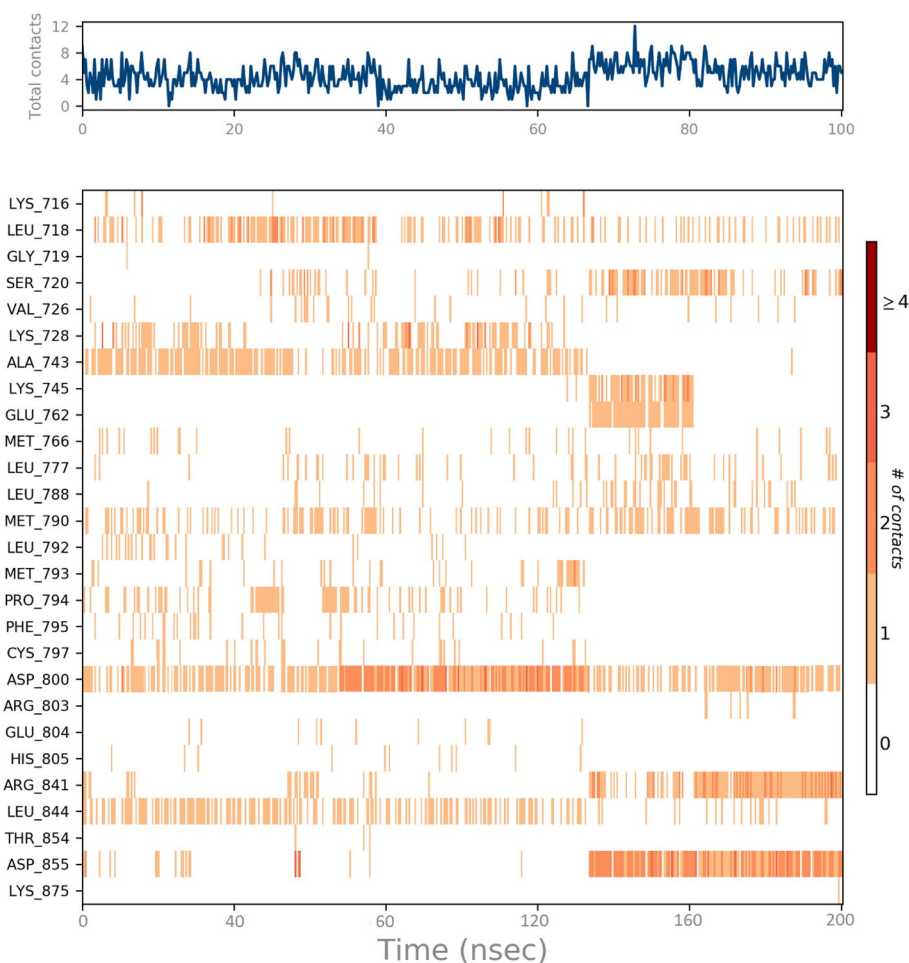


Fig. 11 Simulation Interactions Diagram, (2D binding interaction of ZINC9573445 along with bar diagram indicating the fold of interaction fraction and contacts, Hbonds, hydrophobic, ionic, water bridges)

Fig. 12 Interaction shown by the active site amino acids in each trajectory frame of ZINC9573445–EGFR T790M complex



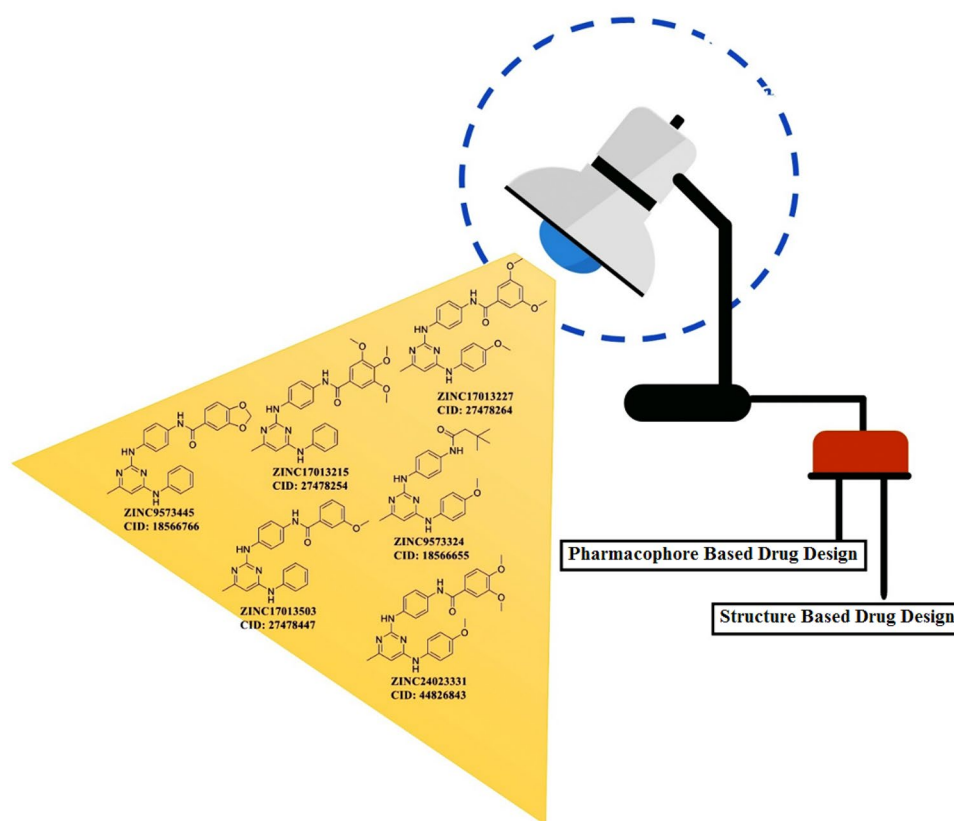
Conclusion

3D Pharmacophore identification was carried out using a reported 2,4-disubstituted-pyrimidines with well-defined EGFR L858R/T790M kinase inhibitory activity. Pharmacophore-based and structure-based virtual screenings identified six hits with potential in silico activity against EGFR L858R/T790M kinase. Molecular docking study indicates that hydrogen-bond interactions with Lys716, Lys728, and Leu745, Asp800, DFG motif of Asp855 and hinge residue M793 key for EGFR T790M domain affinity. The ADME properties of the hits were calculated and found to be within an acceptable range. In silico toxicity and cytochrome, profiling indicates that all the 06 virtually

screened compounds were substrate/inhibitors of the CYP-3A4 metabolizing enzyme and were non-carcinogenic and devoid of Ames mutagenesis. DFT calculation, using the B3LYP/6-31G** basic set further validates the ligand-based and structure-based virtual screening (Fig. 13).

The top potential hit, ZINC9573445 was subjected to molecular dynamics simulation in order to check the overall stability of the protein–ligand complex. Molecular dynamics simulation suggested that docked ZINC9573445–EGFR T790M complex was stable for 200 ns. Currently we are synthesizing these screened molecules at our lab and additional experimental validation will be planned and the results will be reported in due course.

Fig. 13 Structures and ZINC IDs of final 6 best hits obtained through ligand-based and structure-based virtual screening



Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40203-021-00113-x>.

Acknowledgements The authors would like to thank “Indian Council of Medical Research (ICMR), Govt. of India” (Grant No. ISRM/12(11)/2019) for funding the project.

Author contributions RP, IA and HP was involved in the idea generation and performing the computational chemistry work. SS have contributed for the manuscript writing and grammatical check.

Funding The authors would like to thank ‘Indian Council of Medical Research (ICMR) Ministry of Health and Family Welfare, Department of Health Research Govt. of India’ (Grant No. ISRM/12(11)/2019) for funding the project.

Availability of data and materials Additional data is provided in Supplementary Information.

Declarations

Conflict of interest All authors declare no actual or potential conflict of interest including any financial, personal, or other relationships with other people or organizations.

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

- Ahmad I, Shaikh M, Surana S, Ghosh A, Patel H (2020) p38 α MAP kinase inhibitors to overcome EGFR tertiary C797S point mutation associated with osimertinib in non-small cell lung cancer (NSCLC): emergence of fourth-generation EGFR inhibitor. *J Biomol Struct Dyn* 1–14
- Ahmad I, Kumar D, Patel H (2021a) Computational investigation of phytochemicals from *Withania somnifera* (Indian ginseng/ashwagandha) as plausible inhibitors of GluN2B-containing NMDA receptors. *J Biomol Struct Dyn* 10:1–13
- Ahmad I, Jadhav H, Shinde Y, Jagtap V, Girase R, Patel H (2021b) Optimizing Bedaquiline for cardiotoxicity by structure based virtual screening, DFT analysis and molecular dynamic simulation studies to identify selective MDR-TB inhibitors. *In Silico Pharmacol* 23(9):23
- Amala M, Rajamanikandan S, Prabhu D, Surekha K, Jeyakanthan J (2019) Identification of anti-filarial leads against aspartate semi-aldehyde dehydrogenase of *Wolbachia* endosymbiont of *Brugia malayi*: combined molecular docking and molecular dynamics approaches. *J Biomol Struct Dyn* 37(2):394–410
- Benet LZ, Hosey CM, Ursu O, Oprea TI (2016) BDDCS, the rule of 5 and drugability. *Adv Drug Deliv Rev* 101:89–98
- Bhadoriya KS, Sharma MC, Jain SV (2015) Pharmacophore modeling and atom-based 3D-QSAR studies on amino derivatives of indole as potent isoprenylcysteine carboxyl methyltransferase (Icmt) inhibitors. *J Mol* 1081:466–476
- Bochevarov AD, Harder E, Hughes TF, Greenwood JR, Braden DA, Philipp DM, Rinaldo D, Halls MD, Zhang J, Friesner RA (2013) Jaguar: a high-performance quantum chemistry software program with strengths in life and materials sciences. *Int J Quantum Chem* 113(18):2110–2142

- Bonomi P (2003) Erlotinib: a new therapeutic approach for non-small cell lung cancer. *Expert Opin Inv Drug* 12:1395–1401
- Bowers KJ, Chow DE, Xu H, Dror RO, Eastwood MP, Gregersen BA, Klepeis JL, Kolossvary I, Moraes MA, Sacerdoti FD, Salmon JK (2006) Scalable algorithms for molecular dynamics simulations on commodity clusters. In: SC'06: Proceedings of the (2006), ACM/IEEE Conference on Supercomputing, pp 43–43
- Chabon JJ, Simmons AD, Lovejoy AF, Esfahani MS, Newman AM, Haringsma HJ, Kurtz DM, Stehr H, Scherer F, Karlovich CA, Harding TC, Durkin KA, Otterson GA, Purcell WT, Camidge DR, Goldman JW, Sequist LV, Piotrowska Z, Wakelee HA, Neal JW, Alizadeh AA, Diehn M (2016) Circulating tumour DNA profiling reveals heterogeneity of EGFR inhibitor resistance mechanisms in lung cancer patients. *Nat Commun* 7:11815
- Chan S, Han K, Qu R, Tong L, Li Y, Zhang Z, Cheng H, Lu X, Patterson A, Smaill J, Ren X (2015) 2, 4-Diaryl-amino-pyrimidines as kinase inhibitors co-targeting IGF1R and EGFR L858R/T790M. *Bioorg Med Chem Lett* 25:4277–4281
- Chaudhari P, Bari S (2016) in silico exploration of c-KIT inhibitors by pharmaco-informatics methodology: pharmacophore modeling, 3D QSAR, docking studies, and virtual screening. *Mol Divers* 20:41–53
- Chekkara R, Kandakatla N, Gorla VR, Tenkayala SR, Susithra E (2017) Theoretical studies on benzimidazole and imidazo [1, 2-a] pyridine derivatives as Polo-like kinase 1 (Plk1) inhibitors: pharmacophore modeling, atom-based 3D-QSAR and molecular docking approach. *J Saudi Chem Soc* 21:S311–S321
- Chen F, Liu H, Sun H, Pan P, Li Y, Li D, Hou T (2016) Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. Capability to predict protein–protein binding free energies and re-rank binding poses generated by protein–protein docking. *Phys Chem Chem Phys* 18:22129–22139
- Chen L, Fu W, Zheng L, Liu Z, Liang G (2017) Recent progress of small-molecule epidermal growth factor receptor (EGFR) inhibitors against C797S resistance in non-small-cell lung cancer: miniperspective. *J Med Chem* 61:4290–4300
- Chinnasamy S, Selvaraj G, Kaushik AC, Kalliamurthi S, Chandrabose S, Singh SK, Thirugnanasambandam R, Gu K, Wei DQ (2019) Molecular docking and molecular dynamics simulation studies to identify potent AURKA inhibitors: Assessing the performance of density functional theory, MM-GBSA and mass action kinetics calculations. *J Biomol Struct Dyn* 38:1–11
- Cho K, Joannopoulos JD, Kleinman L (1993) Constant-temperature molecular dynamics with momentum conservation. *Phys Rev E* 47(5):3145
- Choubey SK, Jeyaraman J (2016) A mechanistic approach to explore novel HDAC1 inhibitor using pharmacophore modeling, 3D-QSAR analysis, molecular docking, density functional and molecular dynamics simulation study. *J Mol Graph Model* 70:54–69
- Chow E, Rendleman CA, Bowers KJ, Dror RO, Hughes DH, Gullingsrud J, Sacerdoti FD (2008) Desmond performance on a cluster of multicore processors. DE Shaw Research Technical Report DESRES. TR 01. <https://deshawresearch.com>
- Cohen MH, Williams GA, Sridhara R, Chen G, Pazdur R (2003) FDA drug approval summary: gefitinib (ZD1839)(Iressa) tablets. *Clin Oncol* 8:303–306
- Deniz U, Ozkirimli E, Ulgen KO (2016) A systematic methodology for large scale compound screening: a case study on the discovery of novel S1PL inhibitors. *J Mol Graph Mode* 63:110–124
- Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, Friesner RA (2006a) PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J Comput Aided Mol* 20:647–671
- Dixon SL, Smondyrev AM, Rao SN (2006b) PHASE: a novel approach to pharmacophore modeling and 3D database searching. *Chem Biol Drug Des* 67:370–372
- Doak BC, Kihlberg J (2017) Drug discovery beyond the rule of 5-Opportunities and challenges. *Expert Opin Drug Discov* 12:115–119
- Dong X, Zheng W (2008) A new structure-based QSAR method affords both descriptive and predictive models for phosphodiesterase-4 inhibitors. *Curr Chem Genom* 2:29
- Elokely KM, Doerksen RJ (2013) Docking challenge: protein sampling and molecular docking performance. *J Chem Inf Model* 53:1934–1945
- Engelman JA, Jänne PA (2008) Mechanisms of acquired resistance to epidermal growth factor receptor tyrosine kinase inhibitors in non-small cell lung cancer. *Clin Cancer Res* 14:2895–2899
- Evans DJ, Holian BL (1985) The nose–hoover thermostat. *J Chem Phys* 83:4069–4074
- Golbraikh A, Tropsha A (2002) Beware of q²? *J Mol Graph Model* 20:269–276
- Golbraikh A, Shen M, Xiao Z, Xiao YD, Lee KH, Tropsha A (2003) Rational selection of training and test sets for the development of validated QSAR models. *J Comput Aided Mol Des* 17:241–253
- Goss G, Tsai CM, Shepherd FA, Bazhenova L, Lee JS, Chang GC, Crino L, Satouchi M, Chu Q, Hida T, Han JY (2016) Osimertinib for pretreated EGFR Thr790Met-positive advanced non-small-cell lung cancer (AURA2): a multicentre, open-label, single-arm, phase 2 study. *Lancet Oncol* 17:1643–1652
- Hang H, Ma G, Zhu Y, Zeng L, Ahmad A, Wang C, Pang B, Fang H, Zhao L, Hao Q (2018) Active-site conformational fluctuations promote the enzymatic activity of NDM-1. *Antimicrob Agents Chemother* 62:e01579–18
- Hirsch FR, Varella-Garcia M, Bunn PA, Di Maria MV, Veve R, Bremnes RM, Barón AE, Zeng C, Franklin WA (2003) Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis. *Int J Clin Oncol* 21:3798–3807
- <http://dude.docking.org/targets/egfr>
- <https://www.drugs.com/history/tagrisso.html>
- <https://www.rcsb.org/structure/2IUU>
- <https://www.rcsb.org/structure/4ZAU>
- Jafari F, Nowroozi A, Shahlaei M (2018) Discovery of novel glucagon receptor antagonists using combined pharmacophore modeling and docking. *Iran J Pharm Res: IJPR* 17:1263
- Jänne PA, Yang JCH, Kim DW, Planchard D, Ohe Y, Ramalingam SS, Ahn MJ, Kim SW, Su WC, Horn L, Haggstrom D (2015) AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer. *N Engl J Med* 372:1689–1699
- Jordaan MA, Ebenezer O, Damoyi N, Shapi M (2020) Virtual screening, molecular docking studies and DFT calculations of FDA approved compounds similar to the non-nucleoside reverse transcriptase inhibitor (NNRTI) efavirenz. *Heliyon* 6(8):e04642
- Kandakatla N, Ramakrishnan G, Karthikeyan J, Chekkara R (2014) Pharmacophore modeling, atom based 3D-QSAR and docking studies of chalcone derivatives as tubulin inhibitors. *Orient J Chem* 30:1083–1098
- Kausar S, Falcao AO (2018) An automated framework for QSAR model building. *J Cheminformatics* 10:1
- Kawahara A, Yamamoto C, Nakashima K, Azuma K, Hattori S, Kashiwara M, Aizawa H, Basaki Y, Kuwano M, Kage M, Mitsudomi T (2010) Molecular diagnosis of activating EGFR mutations in non-small cell lung cancer using mutation-specific antibodies for immunohistochemical analysis. *Clin Cancer Res* 16:3163–3170
- Kennedy T (1997) Managing the drug discovery/development interface. *Drug Discov* 2:436–444

- Khadikar PV, Karmarkar S, Agrawa VK (2001) A novel PI index and its applications to QSPR/QSAR studies. *J Chem Inf Comput Sci* 41:934–949
- Khan MF, Verma G, Akhter W, Shaquiquzzaman M, Akhter M, Rizvi MA, Alam MM (2019) Pharmacophore modeling, 3D-QSAR, docking study and ADME prediction of acyl 1, 3, 4-thiadiazole amides and sulfonamides as antitubulin agents. *Arab J Chem* 12:5000–5018
- Kim Y, Ko J, Cui Z, Abolhoda A, Ahn JS, Ou SH, Ahn MJ, Park K (2012) The EGFR T790M mutation in acquired resistance to an irreversible second-generation EGFR inhibitor. *Mol Cancer Ther* 11:784–791
- Kobayashi S, Boggon TJ, Dayaram T, Jänne PA, Kocher O, Meyerson M, Johnson BE, Eck MJ, Tenen DG, Halmos B (2005) EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N Engl J Med* 352:786–792
- Kumar V, Elizabeth Sobhia M (2012) Implication of crystal water molecules in inhibitor binding at ALR2 active site. *Comput Math Method Med*. <https://doi.org/10.1155/2012/541594>
- Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 23:3–25
- Lu X, Yu L, Zhang Z, Ren X, Smaill JB, Ding K (2018) Targeting EGFR L858R/T790M and EGFR L858R/T790M/C797S resistance mutations in NSCLC: Current developments in medicinal chemistry. *Med Res Rev* 38:1550–1581
- Lyne PD, Lamb ML, Saeh JC (2006) Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *J Med Chem* 49:4805–4808
- Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA (2008) Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin Proc* 83:584–594
- Mysinger MM, Carchia M, Irwin JJ, Shoichet BK (2012) Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J Med Chem* 55:6582–6594
- Noolvi MN, Patel HM (2013) A comparative QSAR analysis and molecular docking studies of quinazoline derivatives as tyrosine kinase (EGFR) inhibitors: a rational approach to anticancer drug design. *J Saudi Chem Soc* 17:361–379
- Ohsaki YO, Tanno SA, Fujita Y, Toyoshima E, Fujiuchi S, Nishigaki Y, Ishida S, Nagase A, Miyokawa N, Hirata S, Kikuchi K (2000) Epidermal growth factor receptor expression correlates with poor prognosis in non-small cell lung cancer patients with p53 overexpression. *Curr Oncol Rep* 7:603–610
- Pan Y, Wang Y, Bryant SH (2013) Pharmacophore and 3D-QSAR characterization of 6-arylquinazolin-4-amines as Cdc2-like kinase 4 (Cdk4) and dual specificity tyrosine-phosphorylation-regulated kinase 1A (Dyrk1A) inhibitors. *J Chem Inf Model* 53:938–947
- Panwar U, Singh SK (2020) Atom-based 3D-QSAR, molecular docking, DFT, and simulation studies of acylhydrazones, hydrazine, and diazene derivatives as IN-LEDGF/p75 inhibitors. *J Struct Chem* 32:1–16
- Pao W, Miller VA, Politi KA, Riely GJ, Somwar R, Zakowski MF, Kris MG, Varmus H (2005) Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med* 2:e73
- Patel H, Pawara R, Ansari A, Surana S (2017) Recent updates on third generation EGFR inhibitors and emergence of fourth generation EGFR inhibitors to combat C797S resistance. *Eur J Med Chem* 142:32–47
- Patel H, Ansari A, Pawara R, Ansari I, Jadhav H, Surana S (2018a) Design and synthesis of novel 2, 4-disubstituted aminopyrimidines: reversible non-covalent T790M EGFR inhibitors. *J Recept Signal Transduct Res* 38(5–6):393–412
- Patel H, Pawara R, Surana S (2018b) In-silico evidences for binding of Glucokinase activators to EGFR C797S to overcome EGFR resistance obstacle with mutant-selective allosteric inhibition. *Comput Biol Chem* 74:167–189
- Patel H, Dhanger K, Sonawane Y, Surana S, Karpoomath R, Thapliyal N, Shaikh M, Noolvi M, Jagtap R (2018c) In search of selective 11 β -HSD type 1 inhibitors without nephrotoxicity: an approach to resolve the metabolic syndrome by virtual based screening. *Arab J Chem* 11:221–232
- Patel S, Modi P, Chhabria M (2018d) Rational approach to identify newer caspase-1 inhibitors using pharmacophore based virtual screening, docking and molecular dynamic simulation studies. *J Mol Graph Model* 8:106–115
- Patel H, Ahmad I, Jadhav H, Pawara R, Lokwani D, Surana S (2020a) Investigating the impact of different acrylamide (electrophilic warhead) on osimertinib's pharmacological spectrum by molecular mechanic and quantum mechanic approach. *Comb Chem High Throughput Screen*. <https://doi.org/10.2174/1386207323666201204125524>
- Patel HM, Shaikh M, Ahmad I, Lokwani D, Surana (2020b) BREED based de novo hybridization approach: generating novel T790M/C797S-EGFR tyrosine kinase inhibitors to overcome the problem of mutation and resistance in non-small cell lung cancer (NSCLC). *J Biomol Struct Dyn* 39:1–19
- Patel HM, Ahmad I, Pawara R, Shaikh M, Surana SJ (2020c) In silico search of triple mutant T790M/C797S allosteric inhibitors to conquer acquired resistance problem in non-small cell lung cancer (NSCLC): a combined approach of structure-based virtual screening and molecular dynamics simulation. *J Biomol Struct Dyn* 39:1–15
- Prabhu K, Manoj Kumar M, Gopalakrishnan VK (2014) Pharmacophore modeling and QSAR study of Thieno [3, 2-b] pyrimidine analogs as VEGFR-2 inhibitors. *Int J Pharm Pharm Sci* 6:200–207
- QikProp (2010) version 9.0, Schrodinger, LLC. New York, NY
- Raghu R, Devaraji V, Leena K, Riyaz SD, Baby Rani P, Kumar Naik P, Dubey PK, Velmurugan D, Vijayalakshmi M (2014) Virtual screening and discovery of novel aurora kinase inhibitors. *Curr Top Med Chem* 14:2006–2019
- Roy K, Kar S, Ambure P (2015) on a simple approach for determining applicability domain of QSAR models. *Chemometr Intell Lab* 145:22–29
- Rydberg P, Gloriam DE, Zaretski J, Breneman C, Olsen L (2010) SMARTCyp: a 2D method for prediction of cytochrome P450-mediated drug metabolism. *ACS Med Chem Lett* 1(3):96–100
- Rydberg P, Rostkowski M, Gloriam DE, Olsen L (2013) The contribution of atom accessibility to site of metabolism models for cytochromes P450. *Mol Pharm* 10:1216–1223
- Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W (2013) Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des* 27:221–234
- Shaikh M, Shinde Y, Pawara R, Noolvi M, Surana S, Ahmad I, Patel H (2021) Emerging approaches to overcome acquired drug resistance obstacles to osimertinib in non-small-cell lung cancer. *J Med Chem*. <https://doi.org/10.1021/acs.jmedchem.1c00876>
- Sharma SV, Bell DW, Settleman J, Haber DA (2007) Epidermal growth factor receptor mutations in lung cancer. *Nat Rev Cancer* 7:169–181
- Sharma V, Kumar H, Wakode S (2016) Pharmacophore generation and atom based 3D-QSAR of quinoline derivatives as selective phosphodiesterase 4B inhibitors. *RSC Adv* 6:75805–75819
- Shen M, Béguin C, Golbraikh A, Stables JP, Kohn H, Tropsha A (2004) Application of predictive QSAR models to database mining: identification and experimental validation of novel anticonvulsant compounds. *J Med Chem* 47:2356–2364

- Shinoda W, Mikami M (2003) Rigid-body dynamics in the isothermal-isobaric ensemble: a test on the accuracy and computational efficiency. *J Comput Chem* 24:920–930
- Shivakumar D, Williams J, Wu Y, Damm W, Shelley J, Sherman W (2010) Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *J Chem Theory Comput* 6:1509–1519
- Singh KD, Karthikeyan M, Kirubakaran P, Nagamani S (2011) Pharmacophore filtering and 3D-QSAR in the discovery of new JAK2 inhibitors. *J Mol Graph Model* 30:186–197
- Song Z, Ge Y, Wang C, Huang S, Shu X, Liu K, Zhou Y, Ma X (2016) Challenges and perspectives on the development of small-molecule EGFR inhibitors against T790M-mediated resistance in non-small-cell lung cancer: miniperspective. *J Med Chem* 59:6580–6594
- Tawari NR, Bag S, Degani MS (2008) Pharmacophore mapping of a series of pyrrolopyrimidines, indolopyrimidines and their congeners as multidrug-resistance-associated protein (MRP1) modulators. *J Mol* 14:911–921
- Teli MK, Rajanikant GK (2012) Pharmacophore generation and atom-based 3D-QSAR of N-iso-propyl pyrrole-based derivatives as HMG-CoA reductase inhibitors. *Org Med Chem Lett* 2:25
- Thress KS, Paweletz CP, Felip E, Cho BC, Stetson D, Dougherty B, Lai Z, Markovets A, Vivancos A, Kuang Y, Ercan D (2015) Acquired EGFR C797S mutation mediates resistance to AZD9291 in non-small cell lung cancer harboring EGFR T790M. *Nat Med* 21:560–562
- Tropsha A, Gramatica P, Gombar VK (2003) The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *Qsar Comb Sci* 22:69–77
- Ugale VG, Bari SB (2016) Identification of potential Gly/NMDA receptor antagonists by cheminformatics approach: a combination of pharmacophore modelling, virtual screening and molecular docking studies. *Sar Qsar Environ Res* 27(2):125–145
- Ugale VG, Patel HM, Surana SJ (2017) Molecular modeling studies of quinoline derivatives as VEGFR-2 tyrosine kinase inhibitors using pharmacophore based 3D QSAR and docking approach. *Arab J Chem* 10:S1980–S2003
- Vansteenkiste JF, Schildermans RH (2005) The future of adjuvant chemotherapy for resected non-small cell lung cancer. *Expert Rev Anticancer Ther* 5(1):165–175
- Verma G, Khan MF, Akhtar W, Alam MM, Akhter M, Alam O, Hasan SM, Shaquiquzzaman M (2019) Pharmacophore modeling, 3D-QSAR, docking and ADME prediction of quinazoline based EGFR inhibitors. *Arab J Chem* 12:4815–4839
- Wang J, Li Y, Yang Y, Zhang J, Du J, Zhang S, Yang L (2015) Profiling the interaction mechanism of indole-based derivatives targeting the HIV-1 gp120 receptor. *RSC Adv* 5:78278–78298
- Yun CH, Mengwasser KE, Toms AV, Woo MS, Greulich H, Wong KK, Meyerson M, Eck MJ (2008) The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc Natl Acad Sci* 105:2070–2075
- Zhou W, Ercan D, Chen L, Yun CH, Li D, Capelletti M, Cortot AB, Chirieac L, Iacob RE, Padera R, Engen JR (2009) Novel mutant-selective EGFR kinase inhibitors against EGFR T790M. *Nature* 462:1070–1074

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.