Medicinal Chemistry Research (2019) 28:1726–1739
https://doi.org/10.1007/s00044-019-02410-9

MEDICINAL CHEMISTRY RESEARCH

ORIGINAL RESEARCH

Check for updates

Implementation of pharmacophore-based 3D QSAR model and scaffold analysis in order to excavate pristine ALK inhibitors

Ramanathan K.[1,2] · Sayoni Maiti[1] · Shanthi V.[1] · Woong-Hee Shin[2,3] · Daisuke Kihara[2,4,5,6]

Received: 17 May 2019 / Accepted: 19 July 2019 / Published online: 31 July 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

The utilisation of anaplastic lymphoma kinase (ALK) inhibitors has unveiled a magnificent clinical activity in ALK-positive non-small cell lung cancer (NSCLC), as well as against the sanctuary site of CNS in selected patients. However, the unsatisfactory survival rates along with unaccomplished overall cure for NSCLC (specifically in metastatic diseases), create an importunity for superior and perpetuating research for the establishment of novel ALK inhibitors in order to ameliorate the consequences of NSCLC. Intriguingly, a few plant-based drugs have paved their way to phase II clinical trial, inspired by which, the present study essayed to unearth novel ALK inhibitors from the NPACT database which comprises 1574 plant-derived compounds that exhibit anti-cancerous activity, using 3D QSAR model (AAADD.1882). Furthermore, multiple docking algorithms (PL-PatchSurfer2 and Glide) were employed to eliminate the false positive prediction. In essence, the strength of the association between the IC_{50} values and docking score was measured by Pearson's correlation coefficient (r). Altogether, our anatomisation yielded three hits, namely; obovaten (NPACT00821), pinoresinol (NPCT00008) and (3S)-3′,7-dihydroxy-2′,4′,5′,8-tetramethoxyisoflavan (NPACT00018) with higher docking scores, predicted anti-cancer and pharmaceutically appurtenant properties with greater CNS involvement. Ultimately, molecular dynamic (MD) simulation highlights the real time evidence for stability of these hit compounds. It is noteworthy to mention that all the hits constitute of particular scaffolds which play a major role in the downregulation of some ALK-positive lung cancer pathways. We speculate that the outcomes of this research are of substantial prominence in the rational designing of novel and efficacious ALK inhibitors.

Keypoints

1. A total of 1574 plant-derived compounds was explored for their ALK inhibitory activity.
2. Possible mechanistic action of the hits was proposed.
3. Pearson's correlation coefficient was used to examine the statistical significance of the computational analysis.

Keywords NSCLC · ALK-EML4 · Polyphenols · Tetrahydrofuran · Benzofuran · 3D QSAR · MD simulations

Introduction

Lung cancer is scrutinised as a pandemic with exacerbating financial and social consequences (Hardavella et al. 2016). It elucidates for nearly 1.59 million deaths worldwide,

Supplementary information The online version of this article (https://doi.org/10.1007/s00044-019-02410-9) contains supplementary material, which is available to authorised users.

✉ Ramanathan K.
kramanathan@vit.ac.in

[1] Department of Biotechnology, School of Bio Sciences and Technology, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India

[2] Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA

[3] Department of Chemistry Education, Sunchon National University, Suncheon 57922, Republic of Korea

[4] Department of Computer Science, Purdue University, West Lafayette, IN 47907, USA

[5] Purdue University Center for Cancer Research, West Lafayette, IN 47907, USA

[6] Department of Pediatrics, University of Cincinnati, Cincinnati, OH 45229, USA

which comprises 19.4% of all cancers (Peters and Kerr 2018). Amidst this, NSCLC being ubiquitous, accounts for about 80–85% of the cases (Navada et al. 2006). Poor prognosis and development of drug resistance poses a serious complication rendering the survival rate to <18% after 5 years of treatment (Zappa and Mousa 2016). These statistical data emphasise the need of the hour to search for ALK inhibitors. ALK is a member of the insulin receptor superfamily, where the ALK gene encodes for the receptor tyrosine kinase. This protein comprises an extracellular domain which includes two Meprin, A5 proteins and protein tyrosine phosphatase Mu domains, one on either side of low-density lipoprotein receptor domain class A, an hydrophobic stretch which corresponds to a single pass trans-membrane region, and an intracellular kinase domain (Bayliss et al. 2016). It plays an important role in the development of brain and exerts its effect on specific neurons in the nervous system (Vernersson et al. 2006). In ALK fusion products, the partner regulates gene expression by multimerization of ALK kinase domain and promotion of autophosphorylation. It also determines the sub-cellular location of fusion protein (Soda et al. 2007). Within the confines of NSCLC, the partner belongs to the echinoderm microtubule-associated protein-like family. A member of this family, echinoderm microtubule-associated protein like 4 (EML4), comprises the HELP (hydrophobic EMAP-like proteins) motif and the WD (Trp-Asp) repeats (Dubey et al. 2017).

EML4-ALK fusion protein was identified to be the oncogenic driver of NSCLC (Soda et al. 2007), which arises due to a paracentric inversion within the short arm of chromosome 2 that joins the 5′-end (encoding the NH2-terminal portion, including the coiled-coil domain) of the *EML4* gene to the 3′-end (encoding the COOH-terminal portion, including the tyrosine kinase domain) of the *ALK* gene, fabricating a chimeric tyrosine kinase with an intracellular kinase domain (Sasaki et al. 2010). An elementary illustration of the above explanation is depicted in Fig. 1. It also expresses that the oncogene undergoes constitutive dimerisation, auto-phosphorylation and the activation of the ALK tyrosine kinase. When there is a gain of function mutation in ALK protein, it gets activated in a ligand independent manner and in turn triggers the downstream cancer pathways. These include the SRC, Janus kinase (JAK)/signal transducer and activator of transcription pathway (STAT), GRB2, RAS/mitogen-activated protein kinase, phosphatidylinositol-4,5-bisphosphate 3-kinase (PI3K)/protein kinase B (AKT), ERK pathways. These cause the activation of transcription factors like MYCN, JUNB, BCL-2A1, INK4A, HIF1-2alpha, VEGF (Hallberg and Palmer 2016), which promote cell proliferation, differentiation, and provide anti-apoptotic signals. Literature review also suggests that the ALK-EML4 fusion product stimulates the PI3K/AKT, JAK/STAT and RAS/ERK pathway very specifically (Chan and Hughes 2015). An overview has been illustrated in Fig. 2. Hence, the ALK fusion product possesses immense oncogenic potential, thus suggesting, its inhibition could represent an effective therapeutic strategy (Soda et al. 2008). One of the pioneer strategical outcomes is the small molecule inhibitor, Crizotinib, a first-generation drug, approved by US Food and Drug Administration (FDA) for the treatment of NSCLC with ALK rearrangements. It is also known to inhibit ALK, c-Met, ROS1 (Christensen et al. 2007; Zou et al. 2007; Bergethon et al. 2012). But within a short span of treatment, the patients develop resistance to crizotinib due its toxicity profiles and inadequate central nervous system (CNS) activity (Płużański et al. 2012).

Lately, computational methods have arrogated most of the other drug discovery strategies on account of lesser time-consuming tasks and subordinate expenses (Preethi et al. 2015). For instance, PHASE pharmacophore modelling has been implemented for the discovery of novel D2 antagonists for the treatment of Huntington's chorea and Schizophrenia (Dash et al. 2012). Moreover, 3D QSAR models have been used to investigate the statistical correlation between the structure and function of molecules which helps to expound the relation between ligand and

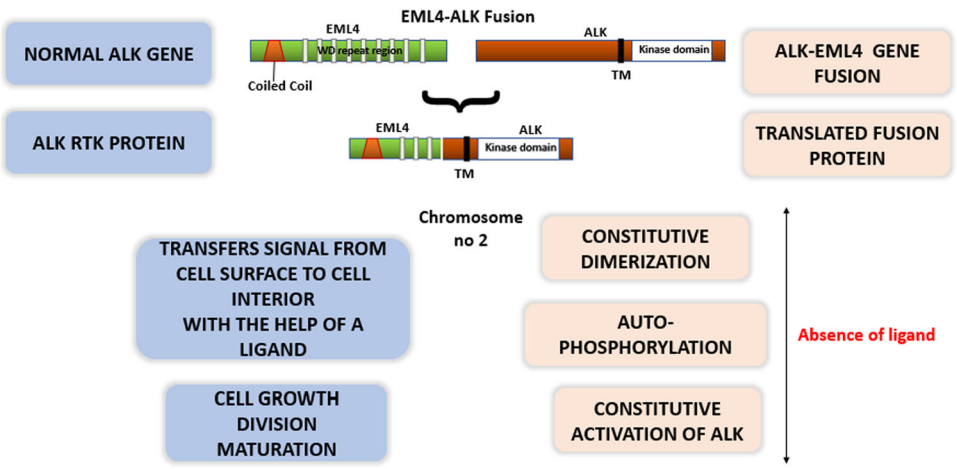

Fig. 1 Schematic representation of normal and fused ALK gene

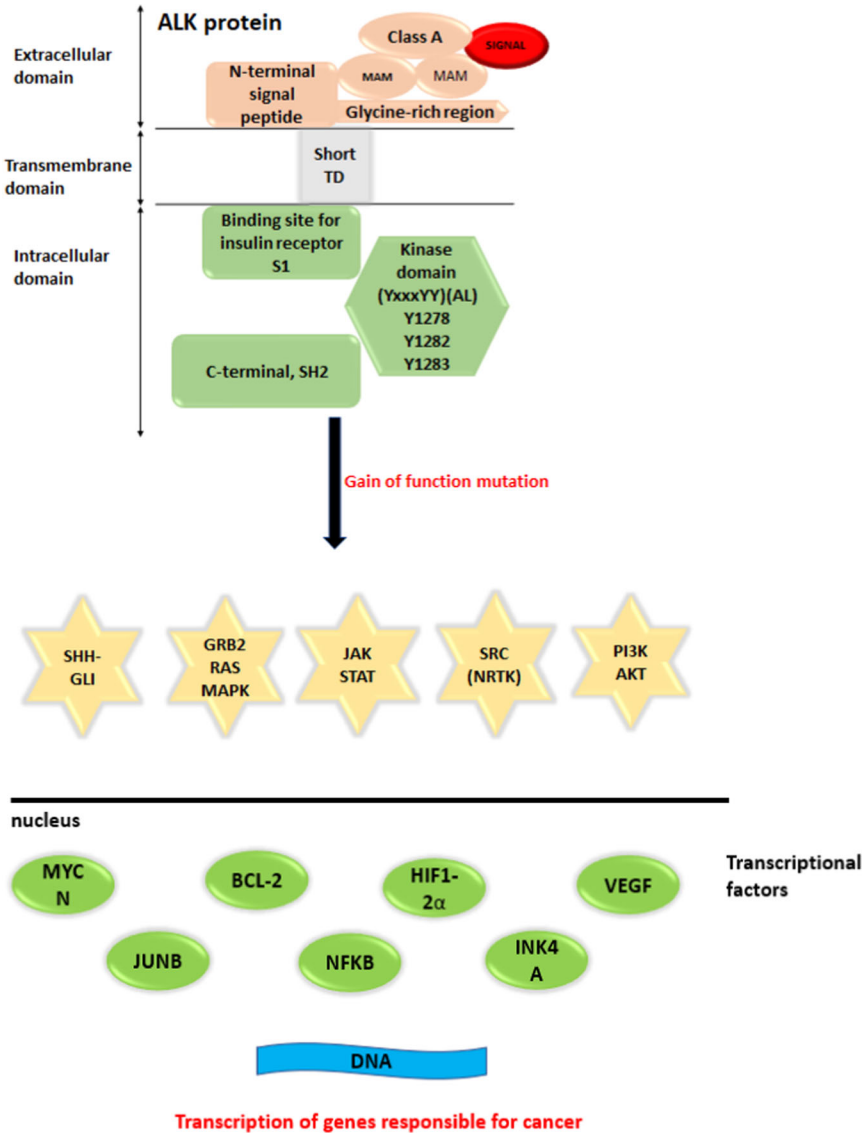

Fig. 2 Overview of downstream pathways of ALK gene

receptor. Exceptionally, the concomitance of Pharmacophore modelling and 3D QSAR has been executed for the discovery of c-kit tyrosine kinase inhibitor (Almerico et al. 2012). Besides virtual screening (VS) strategies have also shown optimistic results for the identification of lead compounds. Hence, in our present research work, we have essayed to figure out novel ALK inhibitors using the integration of 3D QSAR-based screening with molecular docking and simulation studies.

Materials and methods

Dataset

The X-ray crystal structure of native ALK was retrieved from Protein Data Bank (PDB) (Berman et al. 2000), with the identification number 2XP2 having a resolution of 1.9 Å. The Naturally Occurring Plant-based Anti-cancer Compound-activity-target (NPACT) database consists of 1574 molecules, which were downloaded in MOL format and utilised for VS process (Mangal et al. 2012). Further, a 3D QSAR-based pharmacophore (AAADD.1882) model, generated using a set of 50 aminopyrimidine derivatives (inclusive of crizotinib) was extricated from the previous literature (James et al. 2018). Indeed, the robustness of the model was validated against the training and test set of molecules and found to be significant with R^2 value of 0.9696 along with a tremendous predictive accuracy, Q2 value of 0.7652. Therefore, the retrieved model was passed down for database screening to discriminate the actives from the non-binders. The high-throughput screening workflow implemented in our study is shown in Fig. 3.

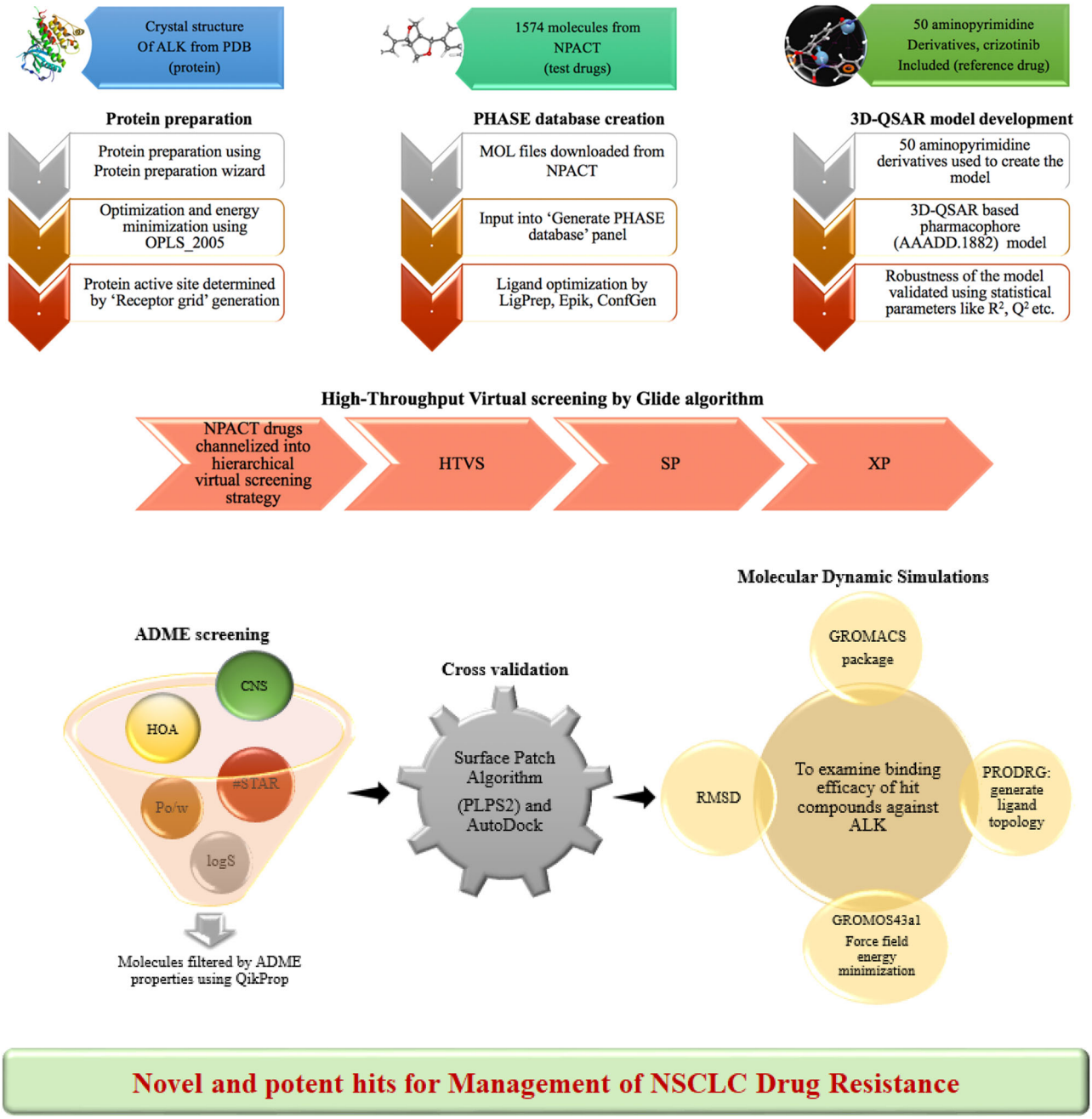

Fig. 3 Schematic representation of virtual screening workflow

Protein preparation

The X-ray crystal structure of the native ALK protein was prepared using the Protein preparation wizard of Schrödinger graphical user interface Maestro (De Falco et al. 2016). The crude PDB structural file consists of heavy atoms viz water molecules, metal ions, activators, cofactors and lacks ionisation states, tautomeric states as well as some of the side chains. Moreover, no information is accessible on the atomic charges and the bond orders. These structural complications were addressed by the Protein preparation wizard (Lionta et al. 2014). Therefore, the imported protein was pre-processed, modified and the undesirable water molecules beyond 5.0 Å were deleted. In addition, bond orders were designated, di-sulphide bonds were established and hydrogen bonds were added maintaining a neutral pH of 7.0. Finally, this structure was optimised and subjected to energy minimisation using OPLS_2005 force field (Gudipati et al. 2018). This prepared protein was carried forward for further docking analysis.

Phase database creation

A phase database was constructed using the MOL files downloaded from NPACT database. It comprises experimentally validated plant-derived natural compounds exhibiting anti-cancerous activity (in vitro and in vivo). These compounds were used as an input in the "Generate phase database" panel of PHASE module which preceded the database creation. During database creation, the NPACT molecules were prepared using LigPrep and were processed for redundancy. Further the ligands were converted to their 3D geometry, hydrogen atoms were added, charged structures were neutralised and proper chiralities were allotted using Epik (Shelley et al. 2007). Consequently, ConfGen performed conformation sampling on all the database molecules generating up to 100 conformers for each ligand, using a heuristic search algorithm (Dixon et al. 2006). Ultimately, an ADME filter was applied to eliminate all the compounds that violated the Lipinski rule.

Molecular docking

Molecular docking algorithms prove to be very helpful to facilitate the understanding of poly-pharmacology effects of the bioactive compounds. In essence, multiple docking algorithms were employed to eliminate the false prediction in the computational analysis. Consecutively, Grid-based (Glide) algorithm (Halgren et al. 2004); (Friesner et al. 2004) was employed, then a surface patch algorithm, PL-PatchSurfer2, [thereafter abbreviated as 'PLPS2'] was utilised to validate the resultant hits from Glide algorithm (Shin et al. 2016); (Shin and Kihara 2018). Of note, Pearson's correlation coefficient (r) method was used to assess the possible association between IC_{50} values and docking scores. Utilising the benchmark data set of known ALK inhibitors (Table S1) the results from both these algorithms were retrieved and further analysed.

In Glide, a receptor grid file was generated using 'Receptor Grid Generation' panel to determine the position and size of the protein active site. Further treating all the ligands as flexible, a hierarchical VS strategy was executed, commencing with high-throughput virtual screening (HTVS) succeeded by Standard Precision (SP) and Extra Precision (XP). Finally, the docked complexes were ranked using the Emodel scoring function (primarily defined by the protein-ligand coulomb-vdW energy) to separate the actives from non-binders (Halgren et al. 2004); (Friesner et al. 2004). It also comprises other terms to account hydrogen bond, rotatable bond penalty, hydrophobic enclosure, etc., that account for the physic of binding process.

On the other hand, PLPS2, represented the protein and ligand surfaces by local patches which is then characterised by its geometrical shape and the electrostatic potential, atom-based hydrophobicity, hydrogen bond donor and acceptor features. These functionalities were represented using three-dimensional Zernike descriptors. Subsequently, the score was computed not only based on the similarity features of the matched patches but also on the relative position of the patches between protein and ligand molecules. Finally, the ranking was carried out by identifying compatible patch pairs between protein pocket and the ligand molecules. Two scoring functions such as lowest conformer score (LCS) and Boltzmann—weighted ligand score (BS) was utilised to rank ligands for the ALK pocket.

ADME property analysis

Pharmacokinetics and drug metabolism assessment plays a crucial role during the early stages of a drug discovery process. Moreover, computer-based methods are gaining momentum in this area and are frequently used to comprehend the Adsorption, Distribution, Metabolism and Excretion (ADME) of a molecule, thus disposing those compounds that are likely to display unacceptable pharmacokinetic profiles (Honorio et al. 2013). Therefore, in the present investigation QikProp module was employed to predict the ADME properties of the hit molecules. Despite the higher frequency of brain metastasis in NSCLC patients, the ability of ALK-TKIs to control and prevent CNS metastasis remains obscure. In addition, these small molecule TKIs are administered through oral routes. Therefore, CNS and human oral absorption (HOA) were the two important properties that were taken into consideration during the current analysis. Of note, the ambit of CNS value lies between -2 (inactives) to $+2$ (actives) and HOA is said to be 100% if its value is 3 (Zhou et al. 2016). Yet another parameter, entitled as the #star descriptor, notifies about the number of properties of each compound that fail to abide by the recommended range, was also considered. Further, lipophilicity (Po/w) and aqueous solubility (logS) of the compounds were also determined.

In essence, the Lipinski's "rule of five" has been considered a standard rule of thumb to rapidly assess whether a molecule has a good balance of solubility and permeability, and has driven the rational design of novel chemical entities endowed with biological activity. However, a close analysis of the currently marketed drugs used for the treatment of several diseases that must be by definition harmless and effective shows that a small percentage of them present a panoply of "forbidden" functionalities (structural alerts) and "unacceptable" properties, including endoperoxides, compounds containing nitro- or isothiourea-moieties, large macrocyclic ring systems and high lipophilicity that fall beyond the drug-likeness

dogma (Doak et al. 2014; Hoagland et al. 2016). Moreover, adherence to the "rule of five" is only one of the characteristics embraced by the notion of drug-likeness. Indeed, other structural red flags, such as the inhibitory activity of the compounds, presence of functional groups known to be metabolically reactive etc. must also be considered when the screening of a drug-like molecule is pursued (Machado et al. 2018). Therefore, in the present investigation ADMET filter was applied as the final criterion of screening rather than stringent data reduction parameters in the initial stage of the process.

Molecular dynamic simulations

Molecular dynamics (MD) simulation was carried out with the help of the GROMACS package 4.6.3 (Lindahl et al. 2001). MD simulations have been very significant in procuring structural information for certain aspects, where experimental data were found to be exiguous. Therefore, in the present research, MD simulations have been employed to examine the binding efficacy of the hit compound against the ALK protein. The acquired data was analogised with the ALK-crizotinib structure data file to unveil energetics viz. stability. Initially, the PRODRG server (Schüttelkopf and Van Aalten 2004) was employed to generate the ligand topology and energy minimisation was conducted utilising GROMOS43a1 force field parameters. A 0.9-nm periodic cubic box was implemented to solvate the structures within the simple point charge water model (Meagher and Carlson 2005). Counter ions were added to the solvated system to achieve neutralisation. Ultimately, the system was subjected to simulation for 50 ns at a 300 K temperature and 1 bar pressure. The snapshots of every 1 ps were collected for the complex stability analysis, root mean square deviation (RMSD).

Results

Database screening using pharmacophore hypothesis

In order to discern potential ALK inhibitors, the CPH (AAADD.1882) was generated with the help of 3D QSAR studies which was utilised to screen against the prepared phase database. All the molecules in the database were required to explicitly match at least four out of five sites, without any partial matches. A total of 1574 molecules were screened using these criteria and it fetched back only 644 hit molecules. Since these molecules were screened using a CPH generated from active ligands, they authenticated the possession of potential ALK inhibitory activities (Kirubakaran et al. 2012). Thus, 644 hit molecules were subjected to the next level of filtration, multiple docking analysis.

High-throughput screening

Initially, the Pearson's correlation coefficient between IC_{50} values and docking scores was computed. Accordingly, the Glide docking algorithm was employed as primary search algorithm to accelerate the scoring of ligands. Although, Glide XP scoring function returned a moderate correlation coefficient of 0.0842, the search is swift enough. It is the prime importance of HTVS especially when large libraries need to be investigated. Further, the resultant hits were examined with respect to its pharmacokinetics and dynamic properties. Finally, PLPS2 algorithm was used to eliminate the false positive prediction in the screening as it showed better correlation for the standard data set. The results of the correlation are depicted in Fig. 4. In particular, the results of PLPS2 were examined with respect to the BS scoring scheme in our analysis as it is reported to have better performance than LCS scheme (Paulsen and Anderson 2009).

Molecular docking by glide

The VS analyses commenced with the docking of crizotinib (reference ligand) and ALK using Glide XP scoring function. Primarily, the Glide score (−5.112 kcal/mol) was registered for the reference ligand and this score was kept as the threshold value for the analysis. Since HTVS is a fast process, the assemblage of 644 hits were initially exposed to this protocol. Consecutively, 322 (50%) high-scoring molecules were selected for further refining with the second run of docking in SP mode. Following SP docking, the top-scoring 146 molecules (50%) were subjected to XP protocol in order to weed out false positives and to refine hits with good binding scores. Thus, only 83 compounds with Glide gscore >−5.112 kcal/mol were recuperated ultimately. The glide score and glide energy of all the 83 hits are provided in supplementary file (Table S2).

ADME properties

The ADME properties significant for the assortment and development of drug molecules were estimated for the hits using the in silico QikProp module. Initially, for the selected 83 hits CNS activities were evaluated. The results of the ADME properties for all the investigated molecule are provided in Table S3. A positive CNS predicts the possibility of the drugs to cross the blood-brain barrier. It can be observed from the results that only 12 hit molecules had positive CNS value. These 12 hits were further analysed for their #stars property. The results like CNS and #stars property with the corresponding docking scores are reported in the form of heat map generated through a hierarchical clustering method (Fig. 5). #stars analysis revealed that eight hits molecules namely; NPACT00675,

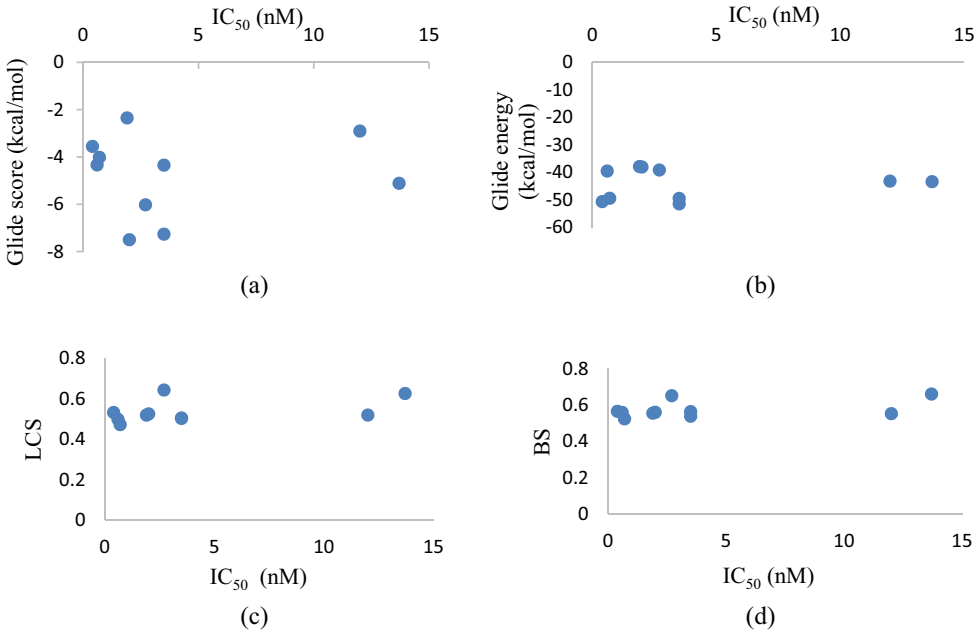

Fig. 4 Correlation between experimentally determined IC_{50} values and the docking scores. **a** Glide score **b** Glide energy **c** LCS and **d** BS

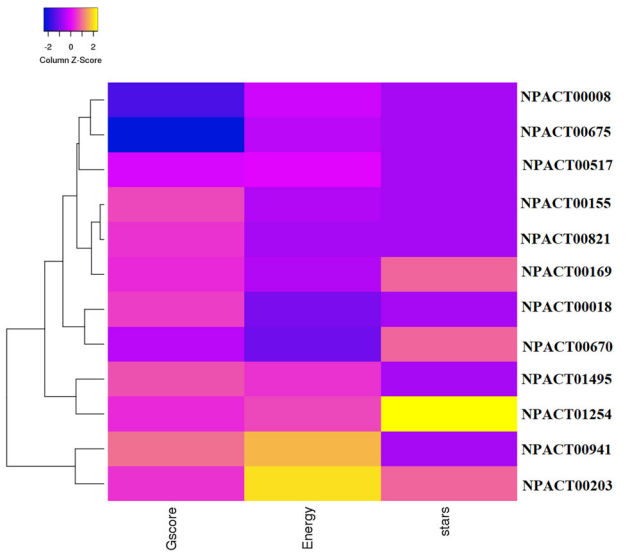

Fig. 5 A heat map depicting the glide gscores, glide energies and #stars properties of the 12 hit molecules. The colour codes in the heat map ranges from low (blue) to high (yellow) values for each of the properties

NPACT00008, NPACT00517, NPACT00821, NPACT00018, NPACT00155, NPACT01498 and NPACT00941, exhibited the value 0 indicating that all the eight hits possess satisfactory Lipinski's properties. Finally, the eight hit molecules were analysed for their HOA and other ADME parameters such as QikProp, QPlogKp, QPlogBB, QPlogS, QPlogPo/w, QPlogPw and QPlogPoct. Interestingly, all the eight hits were found to show zero deviation from the acceptable ranges of these parameters hence indicating their drug-likeness potential (Table S3).

Analysis of ligand interaction details

The knowledge on the protein–ligand interaction pattern plays a climacteric role in understanding the productivity of the drug molecule. Therefore, the molecular interaction patterns of the eight hit molecules with ALK were scrutinised using the ligand interaction diagram in the Schrödinger suite. It is evident from the literature that H-bond interactions with Leu1122, Val1149, Lys1150, Met1199 and Ala1200 were imperative for high ALK inhibitory activity.

For instance, the hinge region of ALK protein is embraced with Met1199 and Glu1197. Hinge region connects the C-lobe with N-lobe and serves as the binding pocket for ATP (Roskoski 2013). This key recognition motif of kinases is the target for majority of kinase inhibitors (Roskoski 2013). Most importantly, it is reported that ATP competitive inhibitors also makes hydrogen bonds with the backbone residues of hinge region (Roskoski 2013). On the other hand, Leu1122 residue is positioned in the Glycine rich (G-rich) loop of the protein. The G-rich loop helps in anchoring the α/β phosphates of ATP and thus appropriately positioning γ phosphate for its catalysis. Thus, binding of the inhibitors to the loop can encumber γ phosphate positioning, its catalysis and thereby the downstream signalling pathways (Roskoski 2013). In fact, crizotinib also makes interactions with the Leu1122, Met1199 and Glu1197 residues (Roskoski 2016).

Accordingly, the eight hit molecules were analysed for at least one of these vital interactions with the ALK protein. Thus among the eight hit molecules, three hits namely; NAPCT00008, NAPCT00675 and NAPCT01498 displayed

interactions with Met1199 and three other hits namely; NPACT00821, NAPCT00018 and NAPCT00155 interacted with Val1149, Met1199 and Lys1150 residues. Whereas the remaining two compounds NPACT00517 and NPACT00941 did not form any essential interactions. The interaction profiles are illustrated in Fig. 6.

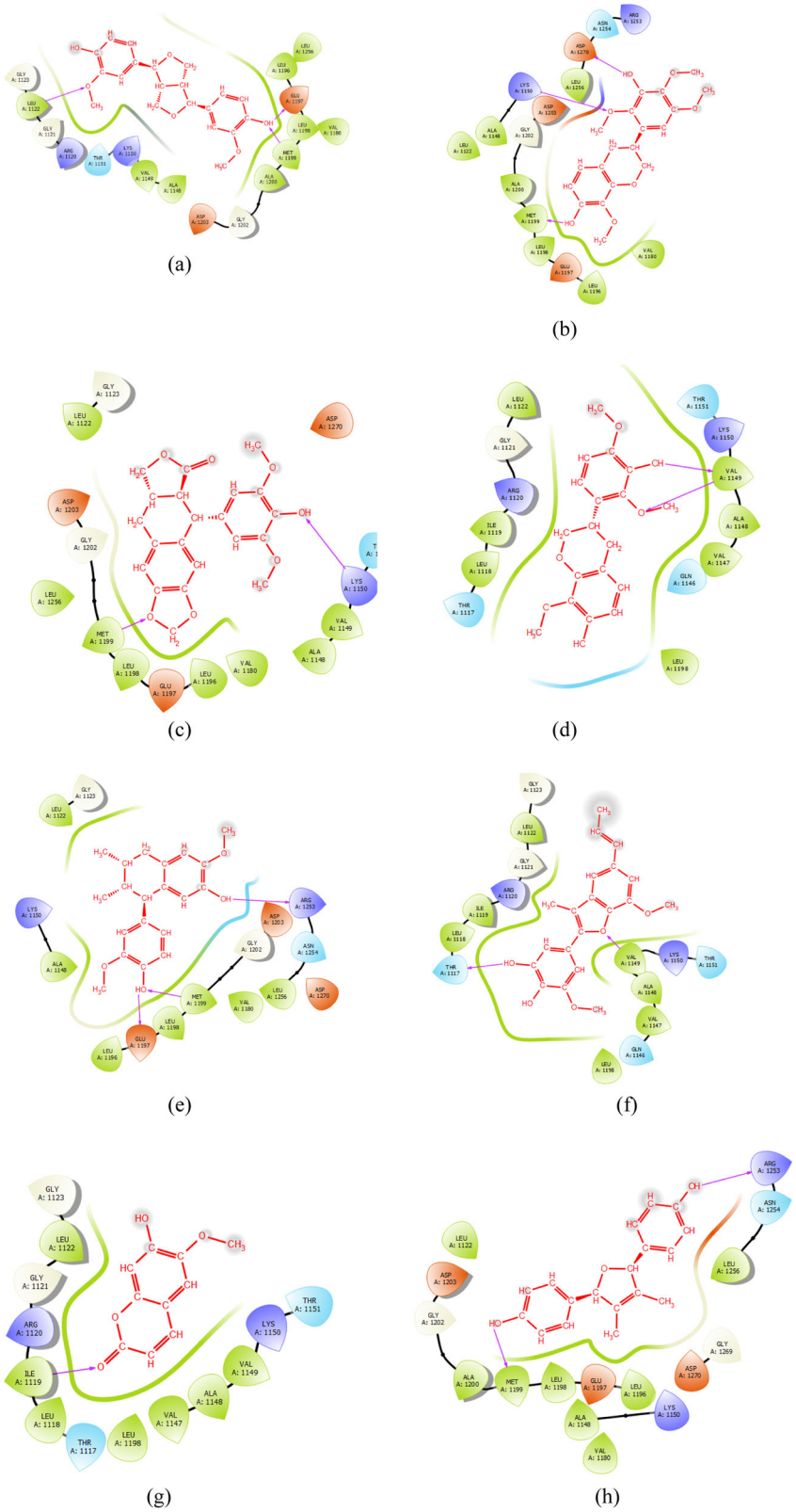

Fig. 6 The LID of the screened hit molecules. **a** NPACT00008, **b** NPACT00018, **c** NPACT00155, **d** NPACT00517, **e** NPACT00675, **f** NPACT00821, **g** NPACT00941, **h** NPACT01498

Table 1 Binding energies and interacting residues of screened hit molecules

S. No	Hits	Binding energy (kcal/mol)	Interacting residues
1.	NPACT00008	−6.60	Met1199, Glu1197
2.	NPACT00018	−6.71	Met1199, Arg1253
3.	NPACT00517	−4.99	Met1199, Arg1253
4.	NPACT00821	−6.49	Met1199, Lys1120, Asp1270
5.	NPACT00155	−5.03	Met1199, Lys1150, Arg1253
6.	NPACT01498	−6.62	Asp1270
7.	NPACT00675	−4.86	Asp1203
8.	NPACT00941	−5.14	Met1199, Glu1197
9.	Crizotinib	−5.53	Met1199, Glu1197, Asp1203, Ser1206

Of note, highly specific screening test is likely to produce false positive results. These circumstances could be eliminated by the implementation of multiple docking algorithm. Here, we have implemented AutoDock algorithm to explore and validate the hit compound binding pattern with ALK protein. The results are shown in Table 1. It is clear from the table that four compounds such as NPACT00008, NPACT00018, NPACT00821 and NPACT01498 showed good inhibitory activity against ALK protein than crizotinib. Most importantly, all the compounds (except NPACT01498 and NPACT00675) were able to display interaction with key residue (Met1199) in ALK protein. The interaction profiles of all the hits studied in our analysis were shown in Fig S1. These data are evident that methodology of VS used in this analysis was found to be adequate and worth for further studies.

Molecular docking by PLPS2

Finally, the prioritisation of known actives was accomplished using the BS-scoring function calculated through the PLPS2 algorithm. Initially, a maximum of 50 conformers were generated for all the eight lead compounds using OMEGA. The surface of each conformer of the ligand was converted into three-dimensional Zernike-descriptors, followed by, exploring the best complementarity with receptor pocket. The overall score of a pocket and a ligand is calculated based on the similarity features and relative position of the corresponding patches between the protein and ligand molecule. Finally, the resultant scores were weighted by Boltzmann distribution to maximise the accuracy of screening.

Of note, BS score of 0.67627, 0.69844, 0.69909, 0.70782, 0.71598, 0.72418, 0.75885 and 0.88277 was obtained for the compounds namely, NPACT00155, NPACT00018, NPACT00675, NPACT00821, NPACT00008, NPACT0 0517, NPACT01498 and NPACT00941, respectively.

This analysis also highlighted that five compounds, specifically, NPACT00155, NPACT00018, NPACT00675, NPACT00821 and NPACT00008, which performed similarly when evaluated on the basis of BS score of Crizotinib (0.69124).

It is worth mentioning that the decrement in performance of the other lead molecules such as NPACT00517, NPACT01498 and NPACT00941 in BS-scoring functionality could also be correlated on account of the absence of crucial-binding interactions of hit compounds.

The surface patch view generated by PLPS2 algorithm for both the ALK protein and lead compounds is shown in Fig. 7. Of note, the key patch positions of functionally important residues in the particular zone (<5 Å) is depicted in Fig. 7. These figures highlight the involvement of key-binding residues of ALK protein such as Leu1122, Val1149, Lys1150, Met1199 and Ala1200 particularly in the binding of hit compounds. Interestingly, the results correlate well with binding pattern derived by Schrödinger algorithm. Therefore, the five compounds which showed better score in both Glide and PL-PatchSurfer2 algorithm were further examined by MD simulation studies.

Molecular dynamics analysis

Molecular docking simulations contemplates only a static view of protein–ligand interactions. In order to study the dynamic stability, hit compounds were subjected to molecular simulations for 50 ns. The results of MD trajectories were analysed in terms of Cα atoms RMSD, the widely used measure to analyse the complex structure and dynamics. The RMSD trajectories corresponding to all the hit compounds namely, NPACT00018, NPACT00821, NPACT00675, NPACT00155 and NPACT00008 are depicted in Fig. 8. All the compounds (red) were examined on the basis of reference molecule, crizotinib (black).

It is evident from the figure that, during the initial simulation time of 20 ns ALK-crizotinib complex showed an average RMSD value of ~2.5 Å. Subsequently, the trajectory showed a rapid increase to ~3.5 Å between 20 and 38 ns. Finally, the system equilibrated after 38 ns with an RMSD value of ~3.5 Å. Consequently, the hit compounds stability was examined with respect to the crizotinib simulation data.

The RMSD of Cα atoms during the MD simulation for NPACT00018 is shown in Fig. 8a. From this, it can be observed that the system is equilibrated in 7 ns and the system remained stable throughout the simulation time with an RMSD value of ~2.8 Å. It is interesting to note that a similar kind of RMSD pattern was observed for NPACT00821 (Fig. 8b).

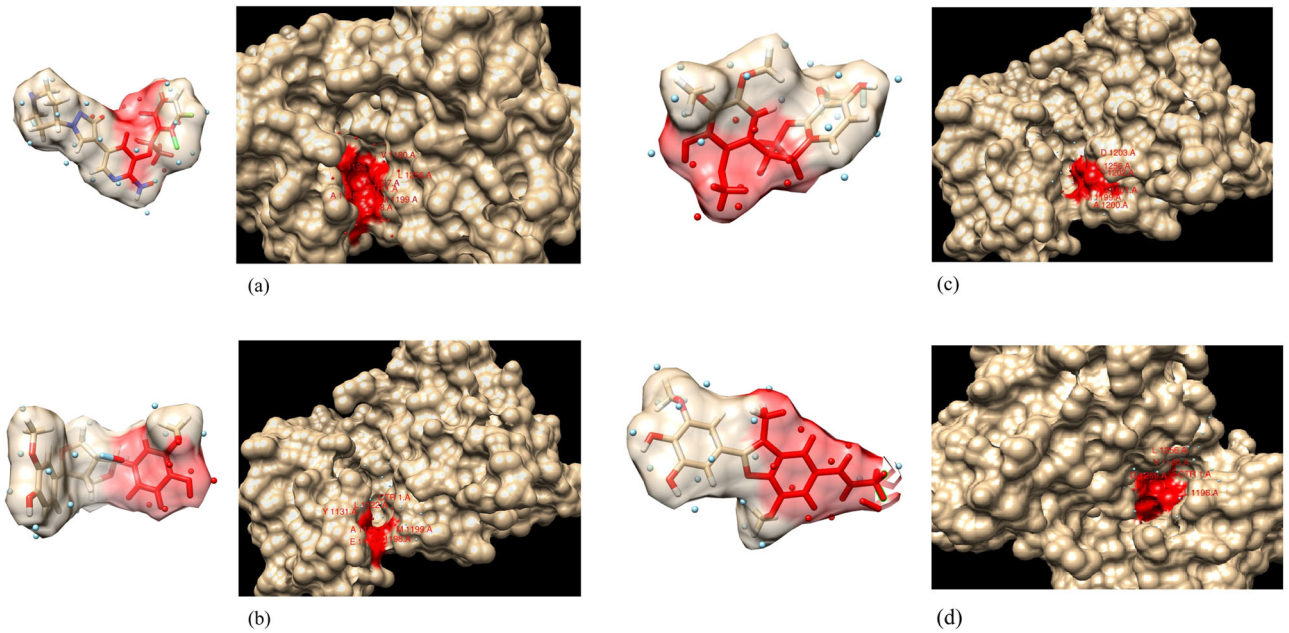

Fig. 7 ALK key residues involvement in the binding of hit compounds identified by PLPS2 algorithm. **a** Crizotinib **b** NPACT00008 **c** NPACT00018 **d** NPACT00821

On the other hand, the hits such as NPACT00675 (Fig. 8c) and NPACT00155 (Fig. 8d) showed higher RMSD fluctuations in our analysis. For instance, the hit compounds NPACT00155 did not form stable conformations with ALK protein even in the 50 ns simulation time. Though NPACT00675 was able to stabilise itself at 7 ns, fluctuations were observed over the end of simulation period and attained higher RMSD value of ~3.6 Å. The larger RMSD values of these hits indicate that the overall topologies of the structures have changed and not able to form stable conformation with ALK protein.

In the case of NPACT00008, equilibration was obtained at around 8 ns and the system remained stable after 10 ns with RMSD value of ~2.9 to ~3.0 Å. This highlights that NPACT00008 performs similar to crizotinib on the basis of RMSD. The results were also correlated well with BS score resulted from the PLPS2 algorithm.

Overall, the smaller backbone RMSD deviations resulted for the hits such as NPACT00018, NPACT00821 and NPACT00008 showed more stable binding with ALK structure than crizotinib and the other hits studied in our analysis. Thus, they have been hypothesised as the final hit molecules for ALK inhibition.

Mechanistic insights of the hit compounds

The panacea for drug discovery can be attained by scaffold analysis of the hit compounds as well as the role of the scaffolds in cancer pathways. Therefore, the hit compounds were analysed for its scaffolds to address the mechanistic aspects relevant to the potent inhibition of ALK. It is interesting to note that the hit compounds obtained in our study composed of crucial scaffolds, namely; benzofuran (in the case of Obovaten, NPACT00821) and polyphenols and tertrahydrofurans (in the case of Pinoresinol, NPACT00008) and flavans (in the case of (3S)-3′,7-dihydroxy-2′,4′,5′,8-tetramethoxyisoflavan, NPACT00018).

It was found that benzofuran is the backbone of obovaten and it has significant anti-tumour activity (Tsai et al. 1998). Literature survey suggests that benzofuran plays a crucial role in the inhibition of mTOR signalling pathway (Salomé et al. 2014) by dephosphorylation of S6 ribosomal protein and 4-EBP1 protein, induction of caspase-dependent apoptosis (Coskun et al. 2017) by the activation of Caspase-3/7 and inhibition of angiogenesis due to the down-regulation of the key angiogenic factor, VEGFR2 (Abdelhafez et al. 2015). This is depicted in Fig. 9. Collectively, all these mechanisms provide very dominating rationale for anti-tumour activity. Numerous preclinical investigations have also suggested that certain natural phytochemicals, such as curcumin (CUR) and resveratrol (RES), which have similar scaffolds, also inhibit a few of the above pathways (Tan et al. 2014), thus providing a strong support to our proposal. Altogether, obovaten which possesses benzofuran as the main scaffold is thus hypothesised to play a role in the abovementioned pathways.

Further, recent reports have unveiled that polyphenol and tertrahydrofuran scaffolds also play a key role in the downregulations of major cancer pathways like PI3K-Akt, NF-kB, HH/GLI pathways (Fantini et al. 2015). A few

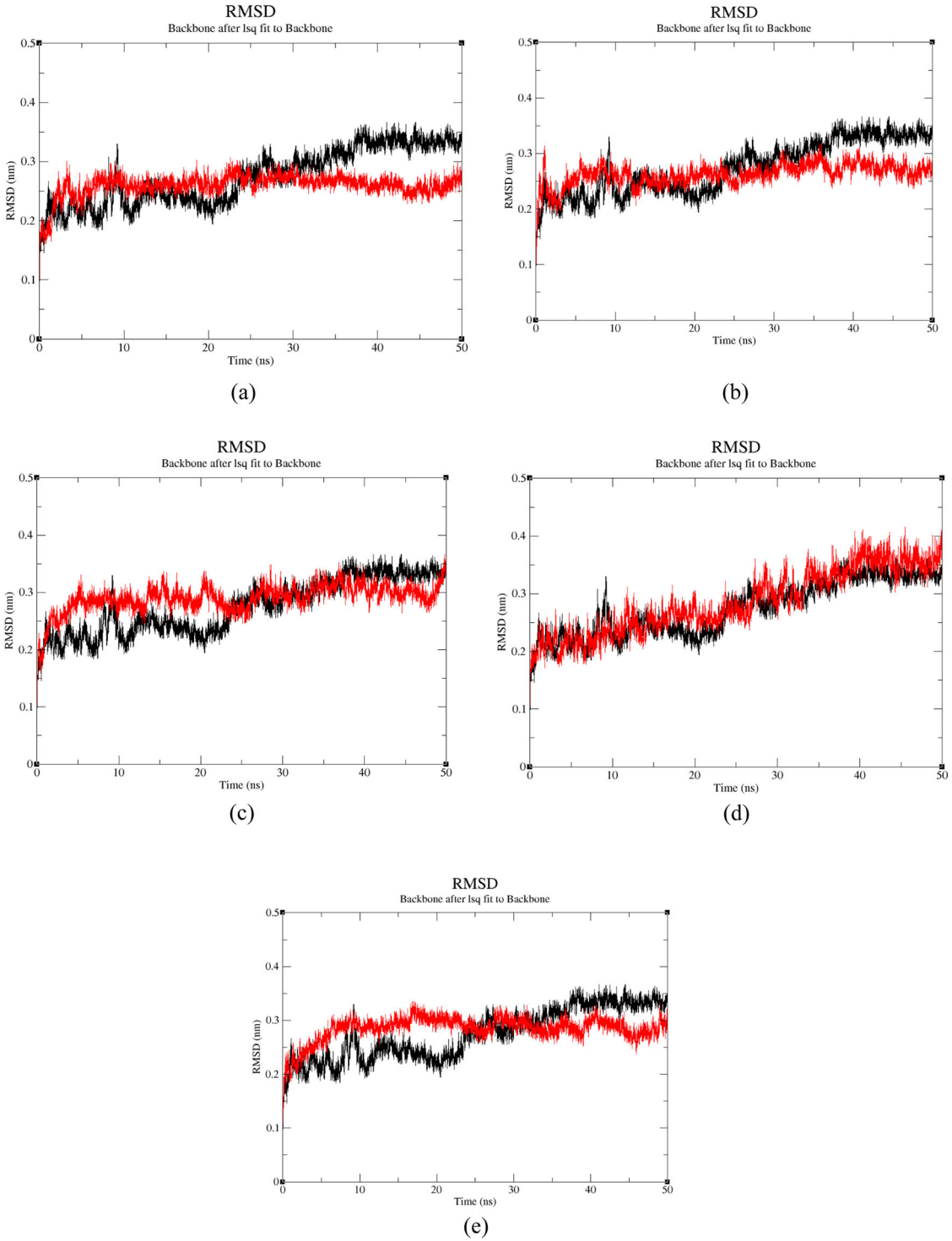

Fig. 8 RMSD analysis of the hit compounds from molecular dynamic simulation. **a** NPACT00018 **b** NPACT00821 **c** NPACT00675 **d** NPACT00155 and **e** NPACT00008. (Colour scheme: black: ALK-Crizotinib complex and red: ALK-Hit complex)

FDA-approved polyphenolic drugs namely CUR, Epigallocatechin-3-O-gallate and resveratrol, share a number of similarities with our hit compound; pinoresinol. The Fig. 9 also shows that polyphenol plays a role in most of the critical pathways. Upregulation of BAX/BCL2 ratio in the PI3K-Akt signalling pathway downregulates cell proliferation, downregulation of SHH reduces vascularisation, and downregulation of IKB-α phosphorylation reduces cell growth. Since pinoresinol has the same scaffold as CUR, it is expected to deliver similar mechanistic activities.

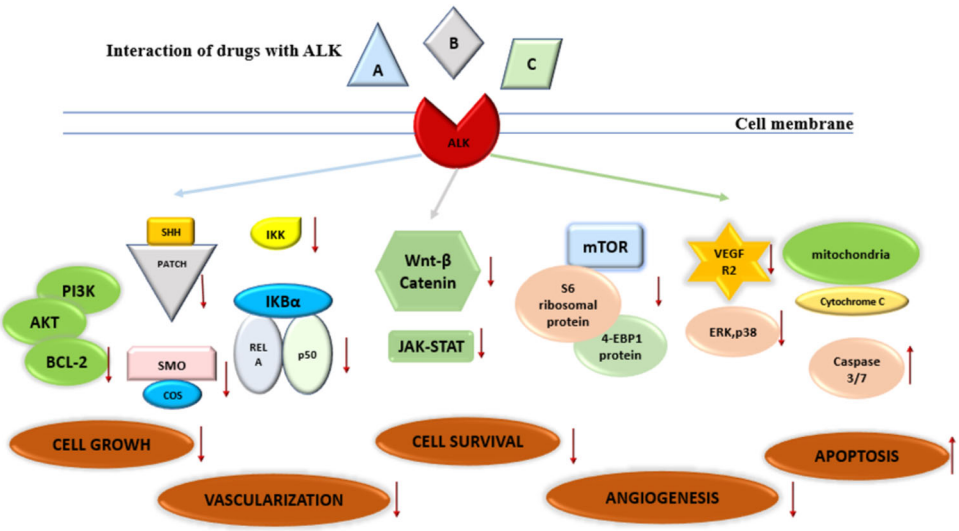

Fig. 9 Mechanistic action of scaffolds on a major cancer pathways. (A: NPACT00008; B: NPACT00018 and C: NPACT00821)

Subsequently, pinoresinol consists of two tetrahydrofuran rings fused to each other. Literature review suggests that acetogenins, structurally a tetrahydrofuran, trigger endoplasmic reticulum stress response which in turn blocks cell proliferation in human nasopharyngeal carcinoma (Juang et al. 2016). Literature survey also suggests that pinoresinol upregulates ATM/P_{53} and their downstream pathways and downregulates cdc2, consequently inducing G2 cell cycle arrest in P_{53} proficient cell lines (Fini et al. 2007). Ultimately, Literature review also suggests that flavans play an effective role in the downregulation of the wnt/βcatenin and JAK/STAT pathway, exhibiting similar anti-tumour properties (Chirumbolo et al. 2018). This has been depicted in Fig. 9.

Unequivocally, pinoresinol, obovaten and (3S)-3′,7-dihydroxy-2′,4′,5′,8-tetramethoxyisoflavan establish themselves as possible ALK inhibitor for the treatment of NSCLC. Indeed, this information could be utilised for further mechanistic research and for devising optimised therapeutic strategies against NSCLC.

Conclusions

In this study, an excursus over some relevant computational techniques in drug discovery has been performed to explore ALK inhibitors from natural sources. Incontrovertibly, the virtually screened compounds namely NAPCT00821, NPACT00018 and NPACT00008 were found best fit by clearing all filters and predicted to be cytotoxic against ALK-positive NSCLC. Overall, the results from our computational analysis predict that the identified hit compounds exhibit magnificent binding affinity, favourable CNS activity and exists within the standard limit of pharmacokinetic and dynamic properties. Prevalent experimental data supports that the scaffolds identified in the study exhibit anti-cancer activities, hence demonstrating the reliability of our results. Fascinatingly, it comprised mainly of benzofuran, polyphenols, tetrahydrofurans and flavans. Literature review suggests that these functional moieties play a pivotal role in the downregulation of dynamic cancer pathways and thus stands distinguished as potent ALK inhibitors. Overall, we believe that the outcomes of our study are of immense importance in NSCLC drug discovery from naturally occurring plant-derived compounds.

Acknowledgements The authors (KR, SM and VS) thank the management of VIT, Vellore for providing the facilities to carry out this work. KR thanks ICMR for their support by the International Fellowship for Young Biomedical Scientists Award.

Funding The authors (KR & VS) are grateful to Department of Science and Technology-Science and Engineering Research Board (DST-SERB) for funding the research project (File No. EMR/2016/001675). DK acknowledges supports from the National Institute of Health (R01GM123055), the National Science Foundation (DMS1614777, CMMI1825941) and the Purdue Institute of Drug Discovery.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

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

References

Abdelhafez OM, Ali HI, Amin KM, Abdalla MM, Ahmed EY (2015) Design, synthesis and anticancer activity of furochromone and benzofuran derivatives targeting VEGFR-2 tyrosine kinase. RSC Adv 5:25312–25324

Almerico AM, Tutone M, Lauria A (2012) Receptor-guided 3D-QSAR approach for the discovery of c-kit tyrosine kinase inhibitors. J Mol Model 18:2885–2895

Bayliss R, Choi J, Fennell DA, Fry AM, Richards MW (2016) Molecular mechanisms that underpin EML4-ALK driven cancers and their response to targeted drugs. Cell Mol Life Sci 73:1209–1224

Bergethon K, Shaw AT, Ou SHI, Katayama R, Lovly CM, McDonald NT, Mark EJ (2012) ROS1 rearrangements define a unique molecular class of lung cancers. J Clin Oncol 30:863–870

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28:235–242

Chan BA, Hughes BG (2015) Targeted therapy for non-small cell lung cancer: current standards and the promise of the future. Transl Lung Cancer Res 4:36–54

Chirumbolo S, Bjørklund G, Lysiuk R, Vella A, Lenchyk L, Upyr T (2018) Targeting cancer with phytochemicals via their fine tuning of the cell survival signaling pathways. Int J Mol Sci 19:3568

Christensen JG, Zou HY, Arango ME, Li Q, Lee JH, McDonnell SR, Los G (2007) Cytoreductive antitumor activity of PF-2341066, a novel inhibitor of anaplastic lymphoma kinase & c-met, in experimental models of anaplastic large-cell lymphoma. Mol Cancer Ther 6:3314–3322

Coskun D, Erkisa M, Ulukaya E, Coskun MF, Ari F (2017) Novel 1-(7-ethoxy-1-benzofuran-2-yl) substituted chalcone derivatives: synthesis, characterization and anticancer activity. Eur J Med Chem 136:212–222

Dash RC, Bhosale SH, Shelke SM, Suryawanshi MR, Kanhed AM, Mahadik KR (2012) Scaffold hopping for identification of novel D2 antagonist based on 3D pharmacophore modelling of iloperidone analogs. Mol Divers 16:367–375

Doak BC, Over B, Giordanetto F, Kihlberg J (2014) Oral druggable space beyond the rule of 5: insights from drugs and clinical candidates. Chem Biol 21:1115–1142

De Falco F, Di Giovanni C, Cerchia C, De Stefano D, Capuozzo A, Irace C, Iuvone T, Santamaria R, Carnuccio R, Lavecchia A (2016) Novel non-peptide small molecules preventing IKKß/NEMO association inhibit NF- κB activation in LPS-stimulated J774 macrophages. Biochem Pharmacol 104:83–94

Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, Friesner RA (2006) PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. J Comput Aided Mol Des 20:647–671

Dubey AP, Pathi N, Viswanath S, Rathore A, Pathak A, Sud R (2017) New insights into anaplastic lymphoma kinase-positive nonsmall cell lung cancer. Indian J Cancer 54:203–208

Fantini M, Benvenuto M, Masuelli L, Frajese GV, Tresoldi I, Modesti A, Bei R (2015) In vitro and in vivo antitumoral effects of combinations of polyphenols, or polyphenols and anticancer drugs: perspectives on cancer treatment. Int J Mol Sci 16:9236–9282

Fini L, Hotchkiss E, Fogliano V, Graziani G, Romano M, De Vol EB, Ricciardiello L (2007) Chemopreventive properties of pinoresinol-rich olive oil involve a selective activation of the ATM–p53 cascade in colon cancer cell lines. Carcinogenesis 29:139–146

Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 47:1739–1749

Gudipati S, Muttineni R, Mankad AU, Pandya HA, Jasrai YT (2018) Molecular docking based screening of Noggin inhibitors. Bioinformation 14:15–20

Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL (2004) Glide: a new approach for rapid, accurate docking and scoring. J Med Chem 47:1750–1759

Hallberg B, Palmer RH (2016) The role of the ALK receptor in cancer biology. Ann Oncol 27:4–15

Hardavella G, George R, Sethi T (2016) Lung cancer stem cells—characteristics, phenotype. Transl Lung Cancer Res 5:272–279

Hoagland DT, Liu J, Lee RB, Lee RE (2016) New agents for the treatment of drug-resistant mycobacterium tuberculosis. Adv Drug Deliv Rev 102:55–72

Honorio KM, Moda TL, Andricopulo AD (2013) Pharmacokinetic properties and in silico ADME modeling in drug discovery. Med Chem 9:163–176

James N, Shanthi V, Ramanathan K (2018) Drug design for ALK-positive NSCLC: An integrated pharmacophore-based 3D QSAR and virtual screening strategy. Appl Biochem Biotechnol 185:289–315

Juang SH, Chiang CY, Liang FP, Chan HH, Yang JS, Wang SH, Lin YC, Kuo PC, Shen MR, Thang TD, Nguyet BT, Kuo SC, Wu TS (2016) Mechanistic study of tetrahydrofuran- acetogenins In triggering endoplasmic reticulum stress response-apotoposis in human nasopharyngeal carcinoma. Sci Rep 6:39251

Kirubakaran P, Muthusamy K, Singh KHD, Nagamani S (2012) Ligand-based pharmacophore modeling; atom-based 3D-QSAR analysis and molecular docking studies of phosphoinositide-dependent kinase-1 inhibitors. Indian J Pharm Sci 74:141–151

Lindahl E, Hess B, Van Der Spoel D (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. Mol Model Annu 7:306–317

Lionta E, Spyrou G, Vassilatis DK, Cournia Z (2014) Structure-based virtual screening for drug discovery: Principles, applications & recent advances. Curr Top Med Chem 14:1923–1938

Machado D, Girardini M, Viveiros M, Pieroni M (2018) Challenging the "drug-likeness" dogma for new drug discovery in tuberculosis. Front Microbiol 9:1367

Mangal M, Sagar P, Singh H, Raghava GP, Agarwal SM (2012) NPACT: naturally occurring plant-based anti-cancer compound-activity-target database. Nucleic Acids Res 41:1124–1129

Meagher KL, Carlson HA (2005) Solvation influences flap collapse in HIV-1 protease. Proteins 58:119–125

Navada S, Lai P, Schwartz AG, Kalemkerian GP (2006) Temporal trends in small cell lung cancer: analysis of the national surveillance, epidemiology, and end-results (SEER) database. J Clin Oncol 24:7082–7082

Paulsen JL, Anderson AC (2009) Scoring ensembles of docked protein–ligand interactions for virtual lead optimization. J Chem Inf Model 49:2813–2819

Peters S, Kerr KM (2018) Stahel RPD-1 blockade in advanced NSCLC: a focus on pembrolizumab. Cancer Treat Rev 62:39–49

Płużański A, Piórek A, Krzakowski M (2012) Crizotinib in the treatment of non-small-cell lung carcinoma. Contemp Oncol 16:1195–1201

Preethi B, Shanthi V, Ramanathan K (2015) Investigation of nalidixic acid resistance mechanism in salmonella enterica using molecular simulation techniques. Appl Biochem Biotechnol 177:528–540

Roskoski R (2013) Anaplastic lymphoma kinase (ALK): structure, oncogenic activation, and pharmacological inhibition. Pharmacol Res 68:68–94

Roskoski R (2016) Classification of small molecule protein kinase inhibitors based upon the structures of their drug-enzyme complexes. Pharmacol Res 103:26–48

Salomé C, Ribeiro N, Chavagnan T, Thuaud F, Serova M, de Gramont A, Faivre S, Raymond E, Désaubry L (2014) Benzofuran derivatives as anticancer inhibitors of mTOR signaling. Eur J Med Chem 81:181–191

Sasaki T, Rodig SJ, Chirieac LR, Jänne PA (2010) The biology and treatment of EML4-ALK non-small cell lung cancer. Eur J Cancer 46:1773–1780

Schüttelkopf AW, Van Aalten DM (2004) PRODRG: a tool for high-throughput crystallography of protein–ligand complexes. Acta Crystallogr D Biol Crystallogr 60:1355–1363

Shelley JC, Cholleti A, Frye LL, Greenwood JR, Timlin MR, Uchimaya M (2007) Epik: a software program for pKa prediction & protonation state generation for drug-like molecules. J Comput Aided Mol Des 21:681–691

Shin WH, Christoffer CW, Wang J, Kihara D (2016) PL-PatchSurfer2: improved local surface matching-based virtual screening method that is tolerant to target and ligand structure variation. J Chem Inf Model 56:1676–1691

Shin WH, Kihara D (2018) Virtual ligand screening using PL-PatchSurfer2, a molecular surface-based protein–ligand docking method. Methods Mol Biol 1762:105–121

Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Bando M (2007) Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature 448:561–566

Soda M, Takada S, Takeuchi K, Choi YL, Enomoto M, Ueno T, Sugiyama Y (2008) A mouse model for EML4-ALK-positive lung cancer. Proc Natl Acad Sci USA 105:19893–19897

Tan HK, Moad AI, Tan ML (2014) The mTOR signalling pathway in cancer and the potential mTOR inhibitory activities of natural phytochemicals. Asian Pac J Cancer Prev 15:6463–6475

Tsai IL, Hsieh CF, Duh CY (1998) Additional cytotoxic neolignans from Persea obovatifolia. Phytochemistry 48:1371–1375

Vernersson E, Khoo NK, Henriksson ML, Roos G, Palmer RH, Hallberg B (2006) Characterization of the expression of the ALK receptor tyrosine kinase in mice. Gene Expr Patterns 6:448–461

Zappa C, Mousa SA (2016) Non-small cell lung cancer: current treatment and future advances. Transl Lung Cancer Res 5:288–300

Zhou W, Wang Y, Lu A, Zhang G (2016) Systems pharmacology in small molecular drug discovery. Int J Mol Sci 17:246

Zou HY, Li Q, Lee JH, Arango ME, McDonnell SR, Yamazaki S, Nambu MD (2007) An orally available small-molecule inhibitor of c-met, PF-2341066, exhibits cytoreductive antitumor efficacy through antiproliferative and antiangiogenic mechanisms. Cancer Res 7:4408–4417