

Mutant anaplastic lymphoma kinase inhibitor identification by integrated in silico approaches

Nadeem Ahmad, Salman Ali Khan, Madiha Sardar, Mamona Mushtaq, Ali Raza Siddiqui, Sajida Munsif, Mohammad Nur-e-Alam, Dmitry Nerukh & Zaheer Ul-Haq

To cite this article: Nadeem Ahmad, Salman Ali Khan, Madiha Sardar, Mamona Mushtaq, Ali Raza Siddiqui, Sajida Munsif, Mohammad Nur-e-Alam, Dmitry Nerukh & Zaheer Ul-Haq (2024) Mutant anaplastic lymphoma kinase inhibitor identification by integrated in silico approaches, *Molecular Simulation*, 50:5, 404-419, DOI: [10.1080/08927022.2024.2316828](https://doi.org/10.1080/08927022.2024.2316828)

To link to this article: <https://doi.org/10.1080/08927022.2024.2316828>

 View supplementary material [↗](#)

 Published online: 17 Feb 2024.

 Submit your article to this journal [↗](#)

 Article views: 76

 View related articles [↗](#)

 View Crossmark data [↗](#)



Mutant anaplastic lymphoma kinase inhibitor identification by integrated in silico approaches

Nadeem Ahmad^a, Salman Ali Khan^a, Madiha Sardar^a, Mamona Mushtaq^b, Ali Raza Siddiqui^a, Sajida Munsif^a,
Mohammad Nur-e-Alam^c, Dmitry Nerukh^d and Zaheer Ul-Haq^b

^aH. E. J. Research Institute of Chemistry, International Center for Chemical and Biological Sciences, University of Karachi, Karachi, Pakistan; ^bDr. Panjwani Center for Molecular Medicine and Drug Research, International Center for Chemical and Biological Sciences, University of Karachi, Karachi, Pakistan; ^cDepartment of Pharmacognosy, College of Pharmacy, King Saud University, Riyadh, Kingdom of Saudi Arabia; ^dDepartment of Mathematics, Aston University, Birmingham, UK

ABSTRACT

Non-small-cell lung cancer (NSCLC) is the primary form of lung cancer globally and remains a leading cause of mortality. Anaplastic lymphoma kinase (ALK) mutations, such as I1171N + L1198H, have been discovered to confer resistance to current ALK inhibitors, reducing their therapeutic effectiveness. Addressing drug resistance necessitates exploring selective inhibitors for innovative therapeutic approaches. In this study, a structure-based pharmacophore model, using ALK-approved inhibitors, was developed to screen an in-house database for potential mutant ALK inhibitors. Compounds with requisite pharmacophoric features were evaluated for binding potential against the I1171N + L1198H ALK mutant phenotype. Selected hits underwent assessment for chemical reactivity, and dynamics stability. The study identified five chemical scaffolds (NS1-5) with favorable binding modes and pharmacokinetic properties. The conformational ensembles featured the average RMSD values, ranging from 0.4 to 0.6 nm. RMSF analysis revealed consistent side chain fluctuations with reduced flexibility, while Rog analysis indicated convergence of most complexes. NS1 and NS5, in particular emerged as promising candidates, exhibiting remarkable performance than others, with binding free energies of -210.12 ± 9.94 and -163.68 ± 11.14 kcal/mol, respectively. These findings thus suggest further exploration and optimisation of NS1 and NS5 for mutant ALK inhibitors for the treatment of NSCLC.

ARTICLE HISTORY

Received 20 October 2023
Accepted 3 February 2024

KEYWORDS

Structure-based pharmacophore; non-small-cell lung cancer; anaplastic lymphoma kinase; virtual screening; MD simulation

1. Introduction

Cancer, a complex and pervasive disease, in its broadest meaning, encompasses more than 277 distinct types. These cancer types are differentiated based on the specific organ involved and the unique histological features they possess [1]. Notably, lung cancer emerged as a leading cause of cancer-related deaths in developed countries. Particularly, non-small cell lung cancer (NSCLC) [2] reigns supreme, accounting for over 85% of all lung cancer cases. The NSCLC encompasses a diverse array of subtypes including adenocarcinoma, large-cell carcinoma and squamous-cell carcinoma [1]. Squamous cell carcinoma accounts for approximately 25% of all lung cancers, characterised by the presence of abnormal squamous cells in the lung lining. Adenocarcinoma, the most prevalent subtype, constitutes approximately 40% of lung cancers and originates in the cells that produce mucus in the lungs. Large cell carcinoma, representing around 10% of lung cancers, is a less common subtype with features that do not align with squamous cell carcinoma or adenocarcinoma. Since they differ in many features these subtypes respond differently to the treatment, particularly, where specific genetic mutations are involved. Therefore, personalised medicine approaches aim to tailor treatments based on the specific characteristics of

the tumour, and ongoing research is providing valuable insights into optimising therapeutic strategies for subtypes of NSCLC particularly Anaplastic Lymphoma Kinase (ALK). Although arising from several factors, the fusion event of ALK with a partner gene representing a diverse range of molecular aberrations is regarded as a prevalent cause. To date, no less than 19 different fusion partners for ALK have been reported in NSCLC including KLC1, EML4, KIF5B and TPR, consequently leading to an abnormal ALK fusion protein [3]. The ALK receptor belongs to the family of tyrosine kinase receptors, catalysing the phosphorylation of specific sites on the protein. In a pioneering work, Iwahara and colleagues provided pivotal insights into the molecular architecture and distinctive functional attributes of the Anaplastic Lymphoma Kinase Receptor Tyrosine Kinase (ALK RTK) [4,5]. The structural architect of ALK revealed three main parts: an extracellular binding domain (amino acids 19–1038), a region spanning the membrane (amino acids 1039–1059) and an intracellular kinase domain (amino acids 1060–1620). The kinase domain harbours catalytic activity and participates in cellular signalling pathways by phosphorylating its targets. Internally, the domain comprises a small N-terminal and a large C-terminal lobe connected by a flexible hinge region to form a cleavage

that functions as an ATP-binding pocket [6]. As mentioned previously, the fusion events, involving ALK, are involved in the pathogenesis of cancer. Of the documented fusions, the EML4-ALK fusion exhibits significant association with NSCLC. At least 3%–7% of all NSCLC cases worldwide are presented with EML4-ALK rearrangement underscoring its notable presence within individuals afflicted from the discussed disease [6,7]. NSCLC exhibits inversion of chromosome 2P, fusing the N-terminal of EML4 with ALK's kinase domain subsequently generating a ligand-independent oncogene with altered enzyme activity, resulting in increased cell division, proliferation and tumour formation [8,9]. Owing to the well-documented role of targeting EML4-ALK fusion in reducing NSCLC, the pursuit of identifying inhibitors against EML4-ALK has emerged as a promising solution to combat the associated malfunction [10]. In line with this, significant efforts have been dedicated to discovering ALK inhibitors during the past few decades. Starting with Crizotinib, the first ALK inhibitor, was sanctioned by the FDA for the first-line treatment of patients with ALK-positive NSCLC. The mechanism of action involves binding to the ATP-binding site of the kinase domain of ALK, immobilising the associated function and locking the downstream activity [11]. Unfortunately, the effectiveness of Crizotinib treatment is challenged by clinically acquired mutations involving the gatekeeper mutation, L1196M, as well as the other reported mutations i.e. F1174C, I1171T, S1206Y, G1202R and G1269A [12]. Such drug-resistant mutations have prompted the interest in the development of second-generation ALK inhibitors. Ceritinib, Alectinib and Brigatinib are effectively used for the treatment of ALK-positive NSCLC patients who did not respond to Crizotinib [13]. However, clinical experiences and further research have highlighted the mutations including G1202R and F1174L, G1202R, I1171G, V1180L and L1196M and G1202R responsible for conferring resistance to Crizotinib, Alectinib and Brigatinib, respectively [14]. The following scientific efforts resulted in the development of third-generation drugs such as Lorlatinib, a dual inhibitor of ALK and ROS1 [15,16]. Clinical trials demonstrated remarkable therapeutic efficacy in ALK-positive NSCLC patients, overcoming the known resistant mutations, including the G1202R mutation, the commonest mutation associated with second-generation drugs [17]. As discussed earlier, the therapeutic strategies for ALK-positive NSCLCs involve the use of ALK inhibitors, with documented success and favourable response rate. While extensive efforts have been made, the intricacy of cancer pathogenesis and the drive of cancer cells to survive, unmasked the novel mutations i.e. I1171N + L1198H, conferring resistance to even third-generation drugs. Therefore, there is a dire need to develop innovative therapeutic agents to overcome these emerging diverse resistance mutations, particularly high drug-resistant mutations. Furthermore by providing a tailored approach, therapeutics with improved efficacy, good tolerability profile and reduced side effects, can be designed [7]. Several experimental studies were dedicated to shedding light on the drug response mechanism in the presence of the acquired mutations in ALK, for identifying inhibitors of drug resistance [18–20]. Karabencheva and colleagues [21] revealed the significant conformational alterations

affecting the dynamics of interactions imperative for the activation of enzymes in mutant types of clinical significance [21]. The distortion in the activation loop and the altered phosphorylation in two cases of the compound mutations i.e. I1171N + L1198H and I1171N + F1174I have also been documented [7]. Another example is the work of He and the authors, where insight into the development of alectinib resistance following the mutations of I1171, V1180 and L1198, was provided via molecular simulations [22]. On the other hand, there is a growing interest in the development of new chemotypes such as gilteritinib as an effective candidate for lorlatinib-resistant ALK (I1171N/F1174I) (Liang et al., [23]). The Computer-aided drug design (CADD) has emerged as powerful and a reliable approach to accelerating the process of drug discovery. Numerous cutting-edge computational techniques have uncovered novel inhibitors which hold promise for improved treatment. Advanced methodologies enabled researchers to solve complex biological queries and diversified the repertoire of inhibitors against hot spot drug targets which consequently broadened therapeutic options [24–26]. Intrigued by this, the present study was set to identify potential lead compounds against ALK mutants utilising a structure-based virtual screening approach. The structural features of the ALK bound to reported drugs were used to map the essential features responsible for the observed biological activity, and in subsequent steps, the features were assembled in a pharmacophore model which was then used to identify new compounds against the target protein. In the following steps, the binding modes were envisaged via molecular docking and the compounds possessing promising affinities, desired intermolecular interaction patterns and good pharmacokinetic profiles were progressed to detailed mechanistic insights via dynamic studies. This study provides valuable information to significantly contribute to the development of effective inhibitors for the resistant phenotypes of ALK, paving the way towards therapeutic intervention against NSCLC.

2. Material and methods

Virtual screening (VS) is an important tool in identifying bioactive compounds and filtering out hundreds of non-binders from the available chemical space. Herein a multistage VS protocol was implemented. The approaches used include structure-based pharmacophore mapping, molecular docking, molecular dynamic simulation and MMPBSA analysis to discover potential inhibitors against mutant ALK.

2.1. Dataset preparation

Utilising a dataset of biologically active chemical compounds, prospective virtual hits against ALK were performed. In particular, the study employed 4 datasets, comprising a dataset of 53 active compounds, four experimentally tested inactive compounds, fetched from the literature, a set of 50 decoys generated for each active compound by the DUD-E web server [27] and a screening library of ~14,000 small molecules. These compounds were subjected to initial preparatory steps utilising the Structure Preparation module implemented in a Molecular Operating Environment [27]. These compounds

were then subjected to the correction of lone pairs followed by the addition of hydrogen atoms as needed. In subsequent steps to optimise energy and to introduce the partial charges, the compounds were treated with MMFF94x force field [28]. The resulting compounds underwent drug-like filtration using Lipinski's rule of drug-likeness before proceeding with further execution.

2.2. Pharmacophore-based virtual screening

2.2.1. Generation of structure-based pharmacophore model

The pharmacophore model, an ensemble of essential features required for the given biological activity, allows rapid screening of millions of compounds, with accuracy. The structure-based pharmacophore model, in particular, accounts for the crucial residues involved in binding compared to the ligand-based model which is restricted to the chemical features of the ligands only. In this connection, we generated a pharmacophore model utilising structural coordinates of the kinase domain of ALK bound with Crizotinib, Brigatinib and Lorlatinib with corresponding PDB identifiers of 2XP2 [11], 6MX8 [29] and 4CLI [30], respectively. The crystal structures were superimposed based on the alignment of their amino acids and the EHT annotation scheme, a most comprehensive approach was selected for pharmacophore modelling. The protein–ligand interaction profiling module allows comprehensive analysis and visualisation of intermolecular interaction patterns and protein–ligand binding modes. Considering the essential interactions, four distinct features, aromatic hydrophobic (Aro/Hyd), hydrogen bond acceptor (Acc), hydrogen bond donor (Don2) and hydrophobic (Hyd), were selected and a model was generated. The pharmacophore model was then optimised by the addition of exclusion volumes to exclude the false-negative rate without raising the false-positive one and to enhance the specificity and selectivity of the model. All steps concerning the pharmacophore model generation were carried out on MOE.

2.2.2. Model validation

The efficiency of a pharmacophore underlies its ability to distinguish the active dataset from the inactive one, determining the coherence of active compounds while excluding the inactive subset [31]. Concerning this, the efficiency of the designed pharmacophore was validated against a training set of actives, in-actives and decoys. The statistical metrics, such as enrichment factor (EF) and area under the curve (AUC), were also calculated.

2.3. Target protein preparation

2.3.1. Protein modelling and preparation

At present, ~ twelve entries are available for the ALK protein bound with first-, second- and third-generation drugs in the Protein Data Bank (PDB). However, for the present study, the crystal structure of the kinase domain of ALK in complex with the third-generation drug, Lorlatinib assigned with PDB ID: 4CLI was used. Initial preparatory steps involved the removal of irrelevant non-standard residues such as water

molecules and ligands that are not pertinent to our investigation. The missing residues in the structure were modelled followed by structure correction and protonation using MOE. The optimised structure was then attained by energy minimisation and the addition of partial charges using the AMBER10EHT force field [32]. The rest of the parameters were kept at default values.

2.3.2. Computational mutagenesis

Since the present work was aimed at the identification of a novel chemotype against mutant ALK the mutations were introduced into the wild-type ALK's kinase domain employing the 'Residue Scan wizard' in MOE. The process of mutation entails the substitution of Isoleucine and Leucine residues at positions 1171 and 1198 into Asparagine and Histidine, respectively, resulting in the I1171N + L1198H variant (SI-Figure 1), which confers resistance against all FDA-approved drugs targeting ALK [7]. Since the goal was to identify new chemotypes targeting the resistant phenotype, the mutated protein was used as a target protein for further investigations.

2.4. Molecular docking

Molecular docking is an efficient and extensively used method to explore the behaviour of ligands within a binding site of the protein. The hit compounds obtained from the initial pharmacophore-based search were subjected to molecular docking to envisage the exact binding mode. However, before the actual docking run, bench marking was done for the software and the docking protocol (SI-Figure 2). In specific, the co-crystallised drug, Lorlatinib, was extracted and re-docked into the ATP-binding site of the ALK (PDB: 4CLI) using MOE software [27]. The Triangle Matcher algorithm and the London dG were utilised as a placement method and a scoring function, respectively, with an induced fit approach. Afterwards, the generated poses were refined using the GBVI/WSA scoring function. The accuracy of the docking protocol was then evaluated by calculating the root mean square deviation (RMSD), a standard measure for method evaluation, to determine the difference between the coordinates of the crystallised pose and the re-docked pose. In subsequent steps, the test dataset was submitted for molecular docking simulation with a validated protocol. The docking grid was centred on the active site of the ALK variant protein with grid dimensions of 30.79, 46.20 and 8.34 Å for x, y and z coordinates. The compounds were assessed based on their binding affinities and the inter-molecular interaction pattern was analysed using the protein-ligand interaction fingerprint (PLIF) module in MOE and protein-ligand interaction profiler (PLIP) [33].

2.5. SwissADME analysis

SwissADME is an online web server that can be utilised to calculate several parameters including pharmacokinetics, bioavailability, physicochemical, drug-likeness, synthetic accessibility and toxicity of shortlisted compounds [34]. To predict the pharmacokinetics, profiles of the canonical SMILES of the test compounds obtained from docking were

submitted to the web server. The results were then manually compiled.

2.6. DFT calculation

Density Functional Theory (DFT) is a computational method based on quantum mechanics used extensively in the fields of chemistry, physics and material science to determine the electronic or nuclear structure of many-body systems, atoms and molecules, particularly in the ground state. Owing to the extreme usefulness of DFT in studying the electronic structures of molecules and providing chemical insights using the relationship between the molecular orbitals, the DFT calculations were carried out. The geometries of all the compounds were optimised using the density functional theory method, B3LYP with a 6-31G** basis set based on its reliability and accuracy in optimising the geometries as documented in the literature [35–37]. Frequency analysis confirmed the structures as true minima. Single-point energies were also calculated at the same level. Global reactivity parameters were determined not only to check the chemical reactivity of the compounds but also to investigate the inhibitory action of the compounds. All the calculations were performed using the Gaussian16 program package [38].

Koopmans's theory was utilised to calculate various electronic properties of the molecule, including electronegativity (χ), ionisation potential (IP), electron affinity (EA), global chemical softness (σ) and hardness (η).

$$\begin{aligned} \text{EA} &= -E_{\text{LUMO}} \\ \eta &= (\text{IP} - \text{EA}) \\ \sigma &= 1/\eta \\ \chi &= -\eta \end{aligned}$$

2.7. MD simulation analysis

Molecular dynamics (MD) simulation is a computational tool that functions similarly to a virtual experiment, allowing for the observation and study of the temporal development of molecular systems and offering insights into the dynamic behaviour of molecules. In this study, freely available software called GROMACS v2021.1 [39] and a GROMOS96 force field were used to simulate the interaction analysis of 5 selected compounds in complex with the mutated ALK system. Force field defines potential energy and intermolecular forces, defining molecular dynamic trajectories by controlling bounded and non-bonded interactions. The topologies of small molecules were generated via ACPYPE [40] and a simulation box with dimensions of $10.71 \times 10.71 \times 10.71 \text{ \AA}$, was built up around the protein–ligand complex, ensuring it was 10 \AA from the center of the complex. The box was filled with SCP water model for solvation of the system and counter ions were added to neutralise the system [41]. The steepest descent algorithm with a maximum force of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ was employed to minimise the system. The NVT and NPT ensembles were used to equilibrate the system for 100 and 10,000 ps at the physiological temperature of 300 K and

the pressure of 1 atm, respectively. The Parrinello-Rahman pressure coupling and the Berendsen thermostat were used to maintain temperature and pressure during the simulation [42,43]. Once the equilibrated state was attained, a 100 ns of MD production run was executed for all the test systems. Visual inspection of the simulated trajectories after simulation was done using VMD [44]. To acquire insight into the stability and flexibility of the protein–ligand complex, post-simulation data were evaluated using metrics such as root mean square deviation (RMSD), root mean square fluctuation (RMSF) and radius of gyration (RoG). To enhance reliability and transparency, simulated trajectories were utilised to plot graphs through Xmgrace [45] to illustrate our results while the interaction diagrams were rendered with Chimera [46].

2.8. Molecular Mechanics-Poisson Boltzmann Surface Area (MMPBSA)

The MM-PBSA approach is a widely used, efficient and reliable free energy simulation method to model molecular recognition, including protein–ligand binding interactions. Molecular dynamics, with free energy calculation methods, can yield energetic determinants to binding. The binding strength is determined by the binding free energy, ΔG_{bind} . Henceforth, the fundamental objective of computer-aided drug designing is to calculate the accurate and efficient free energy (ΔG_{bind}). In the current investigation, the `g_mmpbsa` module (http://rashmikumari.github.io/g_mmpbsa) a plugin for GROMACS [47] was used to calculate Gibb's free energy. For each protein–ligand complex, 1000 frames were considered from the overall simulated trajectories, representing the last snapshots of 10 ns MD simulation. The study investigated the contribution of different components of interaction such as polar and non-polar solvation energy, van der Waals interactions and electrostatic energy contributing to the total binding free energy.

3. Result and discussion

The multifaceted role of ALK, ranging from cell signalling and maintaining physiological homeostasis to the involvement in diverse pathological processes particularly carcinogenesis, has sparked substantial interest in the pursuit of developing effective anticancer therapeutics. Since the traditional drug discovery methods are time-intensive and labourious, virtual screening by harnessing the remarkable potential of computational algorithms enables the rapid identification of promising drug candidates. Virtual screening not only expedites the identification of potential ALK inhibitors but also aids in the design and optimisation of these compounds. The present piece of work demonstrates the virtual screening strategy directed to identify promising small molecules targeting the ALK, particularly for the variant exhibiting resistance against third-generation drugs.

3.1. Pharmacophore modelling and virtual screening

In pursuit of identifying potential lead candidates against the target protein, the information regarding the necessary

physicochemical features responsible for the observed biological activity can greatly help in future identification and optimisation. The pharmacophore modelling in this regard provides an 'abstract of chemical features' which differentiates between the active and inactive compounds with modest accuracy and correctly identifies many new active compounds from a screening dataset. The quality of the pharmacophore model, when structural coordinates are considered, greatly depends on the fidelity of the crystal structure and the accuracy of the annotation scheme employed to discern molecular features.

3.1.1. Construction of model

Herein, the pharmacophore model was constructed using three well-resolute structures of ALK co-crystallised with known small molecule inhibitors, Crizotinib [11], Brigatinib [29] and Lorlatinib [15]. The visual inspection revealed that the binding site of ALK featuring the native pocket for ATP reveals essential molecular interactions including π - π , cation- π and electrostatic interactions between ligands and amino acid residues, namely Met1199, Glu1197, Ala1200, Asp1203, Arg1253, Asn1254 and Asp1270. Employing a systematic approach, crucial and distinctive features were incorporated, and a refined pharmacophore model was developed. The model encompasses essential features such as hydrogen bond acceptors (Acc), hydrogen bond donors (Don2), hydrophobic (Hyd) and aromatic hydrophobic (Aro/Hyd). It is predicted that the developed model harbours the necessary molecular attributes required for ligand recognition and binding, thus providing a valuable framework for the pattern of inter-molecular interaction between the ligands and the target receptor. In particular, the Acc feature was mapped onto Met1199,

which contributes to the structural integrity by linking the large C and small N terminals of the ATP-binding pocket. Similarly, the characteristics of Don2 are presented by Glu1197, playing a crucial role in ATP binding to the ALK receptor. Conversely, the Asp1203 and Asn1254, involved in the catalytic activity of the receptor, displayed Hyd and Aro/Hyd characteristics, respectively.

3.1.2. Validation of the model

A validation process to test the generated model is considered a crucial step for a pharmacophore-based virtual screening. In this regard, the specificity and sensitivity of a pharmacophore model serve as a fundamental parameter in assessing the performance's reliability. The datasets utilised for validation comprised 53 actives, 4 in-actives [48] and a dataset of 3927 decoys, to determine the recognition accuracy of both active and inactive compounds. The optimised 3D pharmacophore model reserved 43 out of 53 actives (81%), correctly excluding all inactive compounds, and 372 out of 3927 (10%) decoys. These fruitful results led us to the hypothesis that the pharmacophoric characteristics found in these drugs could be useful in the development of new inhibitors targeting mutant ALK. The pharmacophoric features are presented in Figure 1.

The other statistical matrices used to test the internal validation and efficacy of the constructed pharmacophore are the enrichment factor (EF) and the area under the ROC curve (AUC). The ROC curve analyses the rate of false positives and true positives to determine the model's specificity and sensitivity, which, in general, exhibits an inverse relation. The differentiating potential of the model for positive and negative datasets improves as the AUC value increases. The

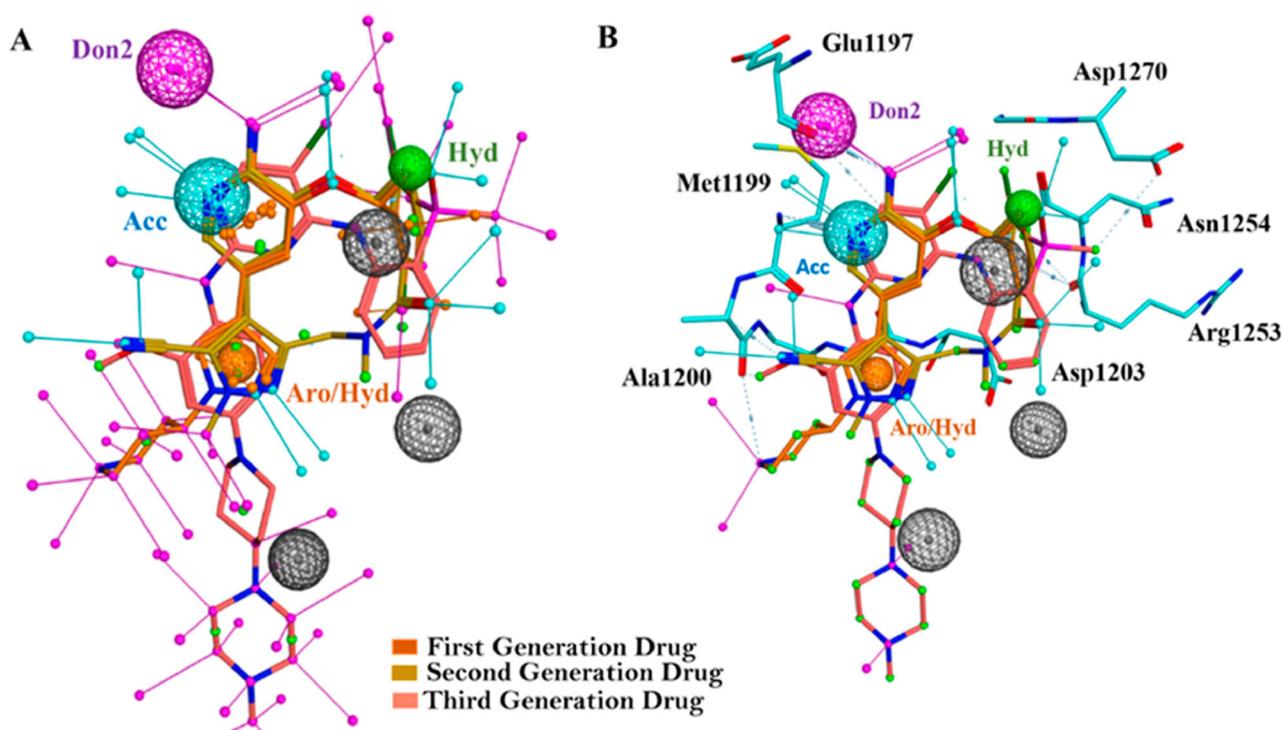


Figure 1. (Colour online) Construction of a structure-based pharmacophore model (A) The alignment of structural coordinates of ALK corresponding to PDB IDs 2XP2, 6MX8 and 4CLI complexed with first-, second- and third-generation drugs, respectively. The common features extracted following superimposition, including Acc: cyan, Don2: pink, Hyd: green and Aro/Hyd: orange colour. The exclusion volumes are shown in grey spheres.

false positives were plotted against the true positives at varying threshold values. The area under the curve (AUC) of the structure-based pharmacophore model in the current investigation as depicted in Figure 2 was 0.79. The AUC between 0.5 and 0.6 signifies a biased model whereas the one ranging between 0.7 and 0.8 is considered acceptable for further utilities. The acquired AUC value in the current study for the constructed model suggests lies in an acceptable region hence, highlighting the model's efficacy in accurately distinguishing the actives from in-actives.

The quality of the generated pharmacophore model was further evaluated by calculating the EF, indicating the model's potential to enrich the activities within a given database. The mathematical expression used for the EF calculation is given by

$$EF = D \times Ha/A \times Ht$$

where D and A represent the total number of decoys and active compounds, respectively; Ht is the hit molecules from the decoy dataset and Ha shows the hit rate from the active dataset. The EF for the tested pharmacophore model was 8.6 implying its remarkable potential in retrieving the actives by random selection.

3.1.3. Pharmacophore-based virtual screening

Once validated, the pharmacophore model was used to screen the In-house database comprising ~14,000 different types of small molecules from synthetic as well as natural sources. The screening aimed to find new ALK mutant inhibitors by identifying compounds possessing similar pharmacophoric properties as predicted by the model. The search resulted in 198 small molecules featuring identical properties to the developed model. These hits were then retained for further molecular docking simulation studies to envisage the binding mode and to determine the binding affinities.

3.2. Protein modelling and mutational analysis

The compound mutation, I1171N + L1198H, in ALK protein is associated with resistance to all approved Anaplastic Lymphoma Kinase and Tyrosine Kinase Inhibitors (ALK-TKI)

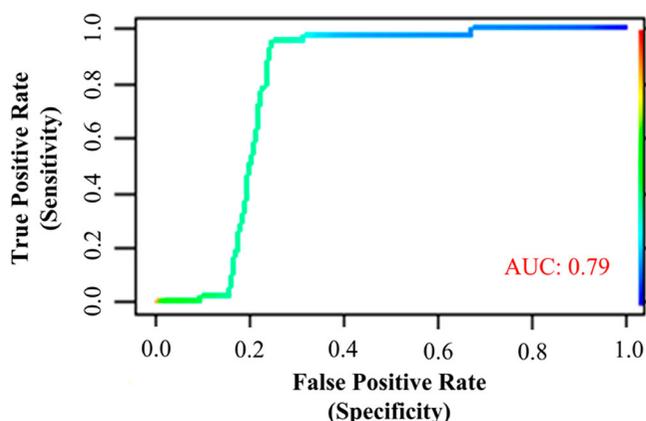


Figure 2. (Colour online) The AUC-ROC Plot, illustrating the classification performance achieved by the constructed pharmacophore model. The AUC values were generated using the ROC function in the R package.

inhibitors. The L1198H mutation occurs within the ATP binding pocket, which is a specific region of the kinase domain of the ALK. In contrast, the I1171N mutation is located adjacent to the ATP binding pocket. The L1198H and I1171N mutations are predicted to cause structural alteration in the ALK, rendering it in a state which is not favourable for the binding of the reported drugs. It is endorsed by the work conducted by Salifu and colleagues [7], which indicated that the wild type of protein exhibits a stable state and more consistent backbone conformation with smaller fluctuations compared to the I1171N + L1198H mutant protein. Therefore, the development of a potent inhibitor capable of effectively targeting the ALK mutant for the treatment of NSCLC is needed. To accomplish this, further docking studies were conducted on the mutant ALK.

3.3. Molecular docking analysis

Lorlatinib was redocked into the ATP binding pocket of the crystal structure of the kinase domain of the ALK receptor for the validation of molecular docking software and protocol. The redocking studies demonstrate the accuracy and efficiency of the docking approach employing MOE in reproducing the known binding conformation with an RMSD value of <1. The intermolecular interaction pattern revealed that a hydrogen atom is donated by the backbone atoms of Met1199 and accepted by the carbonyl's oxygen from Glu1197 distant at 3.10 and 2.77 Å respectively. The additional interactions observed were hydrophobic and involved Val1130, Lys1150, Asp1203 and Leu1256 from the binding cavity. The validation was followed by the docking of the test library and reference drug, Lorlatinib in the ATP-binding pocket of I1171N + L1198H mutant of ALK's Kinase domain. The binding mode of Lorlatinib, towards wild-type and mutant ALK protein, is depicted in Figure 3(A) and the interaction pattern for Lorlatinib and wild-type ALK complex is shown in Figure 3(B).

The findings indicate the displacement of Lorlatinib from the position corresponding to the wild type of target protein and projected a docking score of -5.0 kcal/mol. It further demonstrated the loss of interactions with hinge residue including Met1199 and Glu1197 in the I1171N + L1198H mutant system (Figure 4(A)), which might be a contributing factor in conferring resistance to the drug [7].

Thus, in the present work, the prerequisite for the selection of the compounds was based on the potential of establishing the interaction with the mutated residues, Asn1171 and/or His1198. The interaction analysis resulted in a dataset of 28 compounds with good affinities, ranging from -6.0 to -8.0 kcal/mol and desired interaction profile. Finally, five compounds were shortlisted for detailed investigation, the structural details in 2D format are mentioned in supplementary information (SI-Figure 1).

3.3.1. Binding modes and interaction pattern of shortlisted hits

Compound NS1, a Benzofuran derivative, presented with strong intermolecular interactions, as evidence from the docking score of -7.04 kcal/mol, compared to Lorlatinib. Specifically, two hydrogen bonds, one with mutated residue

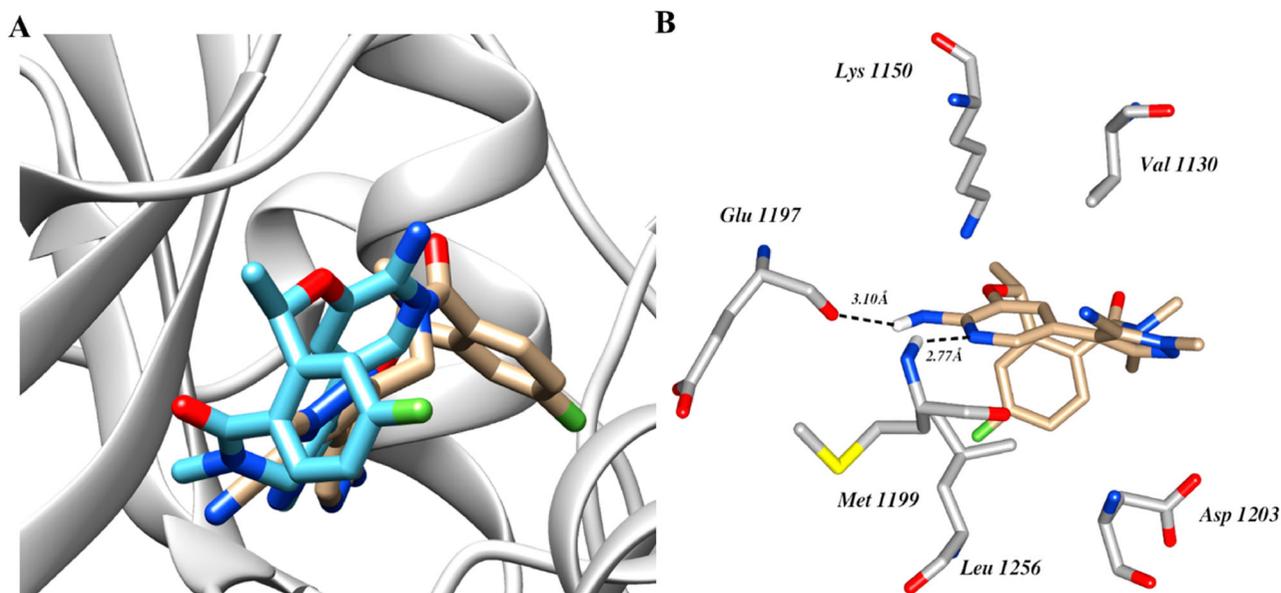


Figure 3. (Colour online) Docking studies of Lorlatinib, a third-generation ALK inhibitor (A) Docked pose of Lorlatinib, towards two distinct conformations of the ALK protein, the wild-type (Tan sticks) and mutated (Cyan sticks) ALK protein harbouring the I1171N+L1198H mutations (B) Detailed overview of inter-molecular interaction pattern for Lorlatinib and wild-type ALK complex.

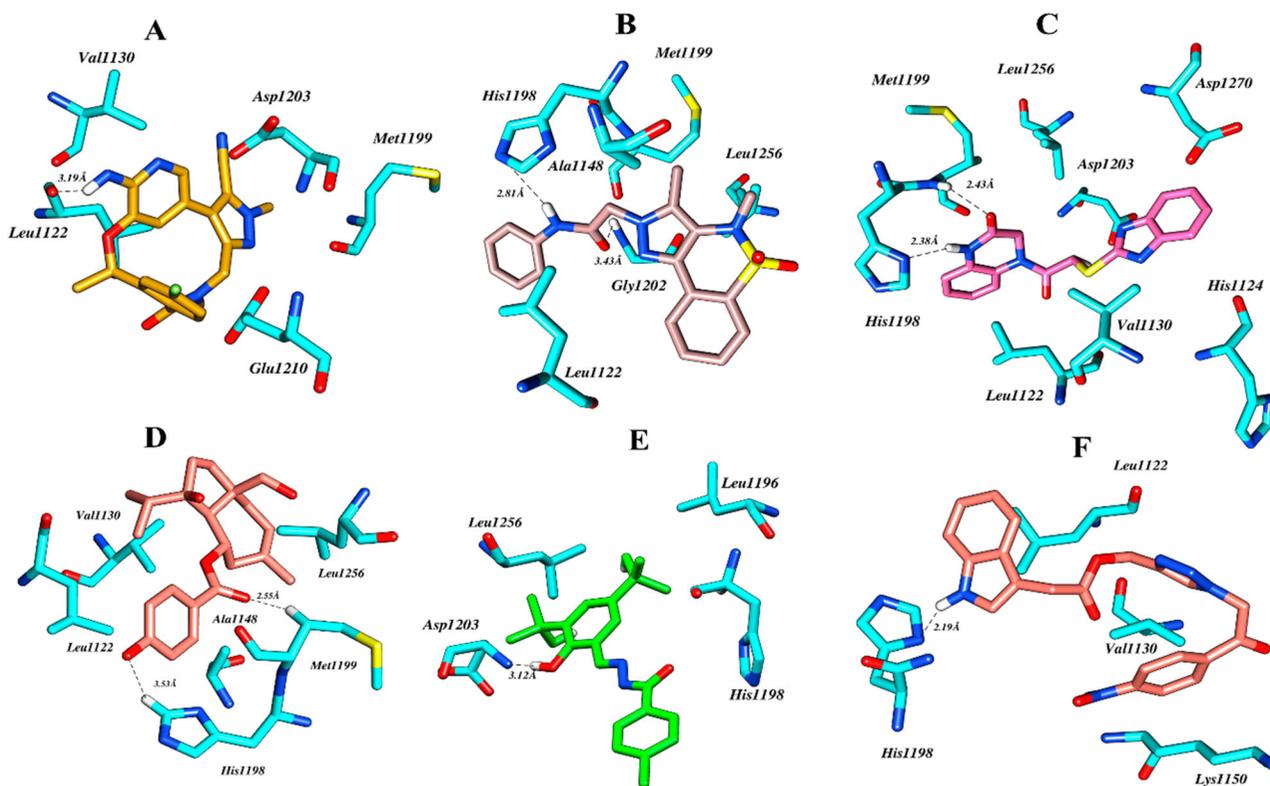


Figure 4. (Colour online) The panels (A–F) show the static modes of the Lorlatinib and shortlisted hits, labelled Compound-NS1–NS5. The crucial residues are depicted as cyan sticks, dotted lines in black representing hydrogen bonds, while the test compounds are present in differently coloured sticks. The figures were generated using UCSF Chimera 1.14.

His1198 and the other with Gly1202, at distances of 2.81 and 3.43 Å, respectively, were observed. The hydrophobic interactions with Leu1122, Ala1148 and Leu1256 provided further stability to the complex (Figure 4B). Leu1122 residue is in the glycine-rich loop of the ALK. The G-rich loop of the protein anchors the α and β phosphates of ATP, allowing the γ

phosphate to be properly aligned for catalysis [49]. The interaction with Leu1122 thus interferes with the ATP access to the G-rich loop consequently, inhibiting the catalysis of ALK and arresting cell proliferation. Crizotinib, an FDA-approved pyrazolopyridine-based inhibitor is the first-generation inhibitor of the target protein. The mechanism of action of Crizotinib

involves binding to the Leu1122 residue by hydrophobic interactions [50] highlighting the importance of this specific interaction and the inhibitory potential of NS1.

The compound NS2 is a Dihydroquinoxaline derivative and revealed a binding affinity of -6.87 kcal/mol. A combination of interactions was observed including hydrogen bonds and hydrophobic contacts (Figure 4(C)). Met1199 and His1198 established the hydrogen bonds with the ligand atom at a distance of 2.43 Å and 2.38 Å, respectively. Met1199 plays an important role in linking the C- and N-lobe of the kinase domain forming an ATP binding pocket known as the hinge region [51]. This hinge region coordinates the binding of 6-amino nitrogen of ATP to the binding site, initiating the subsequent catalytic process, thus triggering downstream signalling pathways responsible for cell growth, proliferation and survival. Additionally, the crucial role of the Met1199 in providing stability to the ligand has been reported in the literature [52]. Thus, the oxygen of the benzimidazole moiety from Compound NS2 may hinder the catalysis and effectively block the downstream processing. The hydrophobic contacts were observed with Leu1122, His1124, Val1130, Asp1203 and Leu1256 residues contributing to the stable binding. Compound NS3, a Benzodioxol derivative, was held in place through several interactions with the target protein, resulting in a binding score of -7.04 kcal/mol. The ligand under consideration entails the formation of two hydrogen bonds involving His1198 and the Met1199 positioned at the distance of 3.53 Å and 2.55 Å, respectively, as illustrated in Figure 4(D). Furthermore, the crucial residues i.e. Leu1122, Val1130, Ala1148 and Leu1256 [53], have hydrophobic contacts with the ligand. In Compound NS4, a benzo hydrazide derivative hydroxyl group established a hydrogen bond with Asp1203 having a bond length of 3.12 Å. The hydrophobic interactions have Leu1196, His1198 and Leu1256 (Figure 4(E)). The binding affinity of the corresponding compound with the targeted system was -6.94 kcal/mol.

The carboxamide derivative, Compound NS5, with a quinoline moiety, with a binding affinity of -7.55 kcal/mol, established a hydrogen bond with His1198 at a distance of 2.19 Å, as presented in Figure 4(F). Moreover, the compound engages in hydrophobic interactions with catalytic residues i.e. Leu1122, Val1130 and Lys1150. Lys1150 is also a critical residue involved in the catalytic process of ALK by forming a salt bridge with the glutamate residue of the alpha C helix located in the C-terminal of the protein. The region is responsible for

controlling the different conformations of the protein, specifically the active and inactive states. The mentioned salt bridge is imperative for the active state of kinase [50] hence, binding of an inhibitor to the Lys1150 is predicted to interfere with the kinase activity and associated function.

3.4. Swiss ADME

The pharmacokinetic profile of a drug plays an essential role in assessing its effectiveness and safety during the drug development process. *In silico* methods, specifically, ADME (absorption, distribution, metabolism, excretion and toxicity) were employed for predicting key pharmacokinetic properties, for identifying compounds possessing favourable properties. Bioavailability, an indicator of a drug which gets to the bloodstream, is an essential consideration in determining the efficacy of an orally administered drug. It has been established that predicting ADME properties early in the discovery phase significantly lowers the proportion of pharmacokinetics-related failure in the clinical stages [54]. The assessment of physicochemical and pharmacokinetic properties, involving molecular weight (MW), number of hydrogen bond acceptors (nHA), number of hydrogen bond donors (nHD), gastrointestinal (GI) absorption, aqueous solubility (LogS), lipophilicity (Log P) and blood-brain barrier (BBB) permeability permitted to sort out compounds that might face challenges in terms of absorption and distribution. In this pursuit, the hits 28 in number, obtained from the preceding docking studies were followed for the analysis of physicochemical and pharmacokinetic properties via Swiss ADME. The factors evaluated include Lipinski's rule of five, compound's solubility and gut absorption. The compounds permeating the blood-brain barrier were eliminated, resulting in a set of nine compounds. Of these, the comprehensive insight considering the docking score, inter-molecular interaction pattern and Swiss ADME profile leads to a refined dataset of five compounds. The physicochemical and pharmacokinetic properties of the compounds are summarised in Table 1. The five virtual hits fall within an acceptable range for all the aforementioned properties. The accepted range for MW reported in the literature varies from 150 and 500 g/mol. All the five hits identified in the present work fall within this range. The prediction suggested that diminished absorption or permeation is probable in instances where there exist more than 5 hydrogen bond donors and 10 hydrogen bond acceptors [55]. Particularly, the investigated

Table 1. Comprehensive evaluation of shortlisted compounds through Swiss ADME.

ADME Analysis	NS1	NS2	NS3	NS4	NS5
Mol. wt. (g/mol)	285.24	338.38	374.47	386.91	419.39
H.B. acceptors	5	3	5	3	7
H.B. donors	1	2	3	2	1
Log P (lipophilicity)	1.80	2.08	3.00	3.66	1.52
Water solubility	Soluble	Soluble	Moderate	Moderate	Soluble
Log S (ESOL)					
GI absorption	High	High	High	High	High
BBB permanent	No	No	No	No	No
Lipinski Violation	0	0	0	0	0
PAINS alert	0	0	0	0	0
Synthetic accessibility	3.00	2.70	4.00	3.18	3.12

compounds displayed favourable outcomes, as each of them holds fewer than 5 hydrogen bond donors and less than 10 hydrogen bond acceptors. Log P, on the other hand, is a measure of lipophilicity although a higher value might indicate potential challenges in solubility, it's essential to carefully evaluate this parameter in the context of overall drug design. The optimal values have been stated depending upon the hypothesis and the desired outcome, though the generally recommended range falls between -0.5 and 7.0 . A soluble molecule is essential to optimised drug development, ensuring ease in handling and formulation [56]. Additionally, a drug intended for intramuscular administration must exhibit high water solubility to deliver an adequate amount of the active ingredient in the limited volume of the pharmaceutical dosage [58]. The obtained result indicated that the studied compounds are soluble in water as shown in Table 1. As discussed earlier, one of the key aspects of CADD is to aid in the selection of the most promising virtual compounds for synthesis and testing, with synthetic accessibility (SA) being a significant factor. The optimal synthetic accessibility (SA) range is from 1 (showing ease of synthesis) to 10 (representing difficulty in synthesis [34]). The studied compounds fall within a range of 2–4, indicating a favourable correlation. Moreover, high gastrointestinal absorption was predicted for all compounds, further highlighting their potential utility via oral administration. Further insight into hits concerning the chances of undesired off-site effects and potential cross-reactivity was checked, employing pan assay interface compounds (PAINS) prediction. The PAINS predicts molecules that exhibit nonspecific interactions with multiple biological targets rather than the selective mode of action. Remarkably, our analysis revealed that none of the compounds displayed PAINS-like behaviour and was specific and non-cross reactive, emphasising again, their potential to be used as a drug candidate.

3.5. DFT study

At present, DFT is accepted as a popular post-Hartree–Fock (HF) approach for the ab initio computation of molecular structures and the energies of molecules (Kurt et al., 2008 [59]). It is extremely useful in the study of the electronic structures of molecules. Moreover, an attempt is made to attain chemical insights using the relationship between the highest occupied molecular orbital (HOMO)–lowest unoccupied molecular orbital (LUMO) gaps in the considered compounds. The obtained results were justified with global reactivity descriptor studies to give a deeper insight into the chemical reactivity or inhibitory capacity of a given compound.

3.5.1. Optimised molecular structures

To get the molecular and electronic insights of the hit molecules, geometric optimisation was done using hybrid DFT functional, B3LYP with a 6-31G** basis set available in Gaussian 16. The optimised molecular configurations of the hit compounds are presented in Figure 5, whereas the corresponding energies are depicted in Table 2. All the optimised structures gave no negative vibrational modes indicating that all the structures were stationary points in the geometry

optimisation procedures. Frequency vibrational analysis was applied to confirm that the obtained structures are stable with no imaginary frequencies.

3.5.2. Frontier molecular orbital analysis

DFT calculations were utilised to quantify the chemical reactivity of the analysed compounds. Hence, frontier molecular orbital (FMO) analysis was performed. In molecular systems, FMO plays a crucial role in drug design. The energy of the highest occupied molecular orbital (E_{HOMO}) and lowest unoccupied molecular orbital (E_{LUMO}) were determined. The E_{HOMO} measures a molecule's ionisation potential, while the E_{LUMO} measures its electron affinity [60]. The difference in HOMO and LUMO energies represents the energy gap ($E_{\text{H-L}}$). The $E_{\text{H-L}}$ serves as an indicator of the structural stability of a molecule. The HOMO and LUMO, often known for molecular descriptive energy gap computation, provide information on a compound's electrical, optical, and biological reactivity characteristics and stability. A narrow energy gap signifies high chemical reactivity and low kinetic stability, leading to easy excitation from the HOMO to the LUMO and vice versa. The results of various parameters, including total energy (E_{Total}), electronegativity, ionisation energy, electron affinity, E_{HOMO} , E_{LUMO} , $E_{\text{H-L}}$, chemical hardness and chemical softness are presented in Table 2.

Based on the results, out of the five tested compounds, NS1 and NS5 showed the highest chemical reactivity (i.e. these are soft compounds), as evidenced by the lowest energy gap of 3.51 and 2.12 eV, respectively, compared to Lorlatinib (4.08 eV) and other compounds. However, NS2 exhibited the lowest chemical reactivity due to its highest $E_{\text{H-L}}$ value (4.98 eV). This information is useful in predicting the properties and behaviour of molecules in various chemical reactions and environments. In the generated HOMO–LUMO counter plot, the distribution of red and green colours represents the positive and negative phases, respectively, of the molecular orbitals [61], as shown in Figure 6. The spatial distributions of HOMO and LUMO of a ligand play a critical role in determining the nature of its interaction with a potential receptor. The HOMO of the ligand interacts with the LUMO of the receptor, and vice versa. Therefore, elevating the energy of the HOMO level of the ligand reduces the energy gap between the HOMO and LUMO levels of the ligand and receptor, respectively, thereby increasing the probability of binding. Conversely, decreasing the energy of the LUMO level of the ligand also increases the likelihood of binding [61]. It is clear from the molecular orbital diagrams of the NS1–NS5 the electron densities in the HOMOs of all these 5 compounds were largely located on donor moiety, and electron densities on the LUMOs were found localised on the acceptor. The excitation from HOMO to LUMO mostly consists of charge transfer from the donor to the acceptor end. The HOMO–LUMO energy gap explains the charge transfer interactions within these compounds.

3.5.3. Global reactive indices

Global reactivity indices offer insights into the chemical reactivity or inhibitory capacity of a given compound, providing valuable information from a chemical perspective. Theoretical static

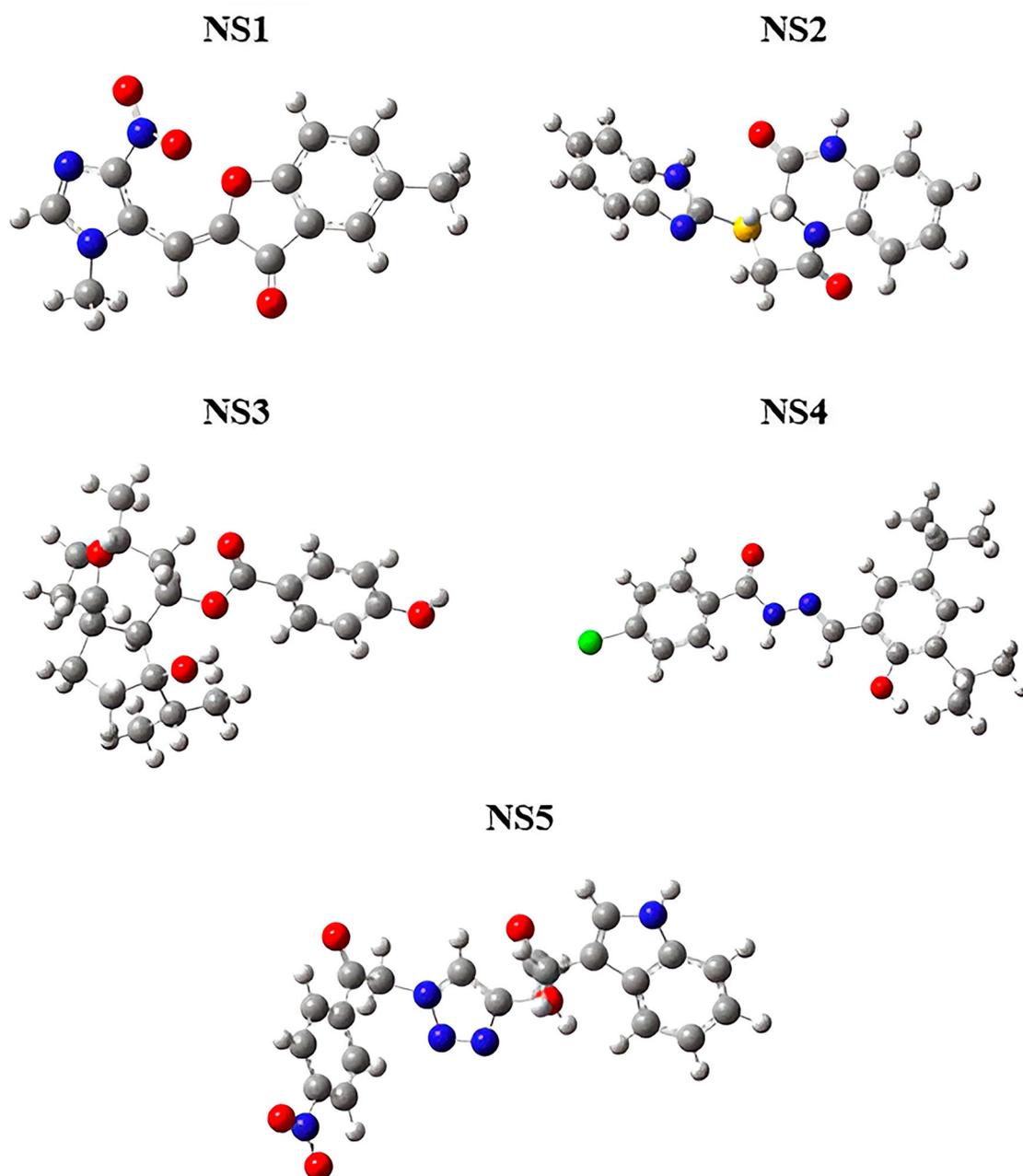


Figure 5. (Colour online) Optimised geometric structures of the shortlisted hits obtained by the DFT method.

Table 2. Thermodynamics parameters and molecular orbital energy values of selected hit compounds.

Compounds	E_{Tot} (ev)	E_{LUMO} (ev)	E_{HOMO} (ev)	$E_{\text{H-L}}$ (ev)	Electronegativity (x)	Ionisation Potential (ev)	Electron Affinity (ev)	Chemical hardness (η) (ev)	Chemical softness (σ) (eV^{-1})
NS1	-27,351.40	-2.69	-6.20	3.51	4.44	6.20	2.69	3.31	0.30
NS2	-38,746.83	-0.82	-5.81	4.98	3.32	5.81	0.82	4.99	0.20
NS3	-33,538.91	-1.05	-5.90	4.85	3.48	5.90	1.05	4.85	0.21
NS4	-42,851.33	-1.50	-5.66	4.15	3.58	5.66	1.50	4.16	0.24
NS5	-39,721.80	-3.51	-5.63	2.12	4.57	5.63	3.51	2.12	0.47
STD	-37,836.79	-1.38	-5.46	4.08	3.42	5.46	1.38	4.08	0.24

calculations are employed to obtain the reactivity indices that pertain to molecular orbitals [62]. These indices include chemical hardness (η) and softness (σ) that describe a molecule's potential to be stable or reactive. The high softness and low hardness value of a molecule indicate that it is a potent inhibitor.

In other words, molecules with higher softness values are more likely to react readily and form bonds, while those with lower hardness values require more energy to undergo a chemical reaction. The global reactivity indices values calculated for the hit compounds indicated that NS1 and NS5 exhibit the lowest

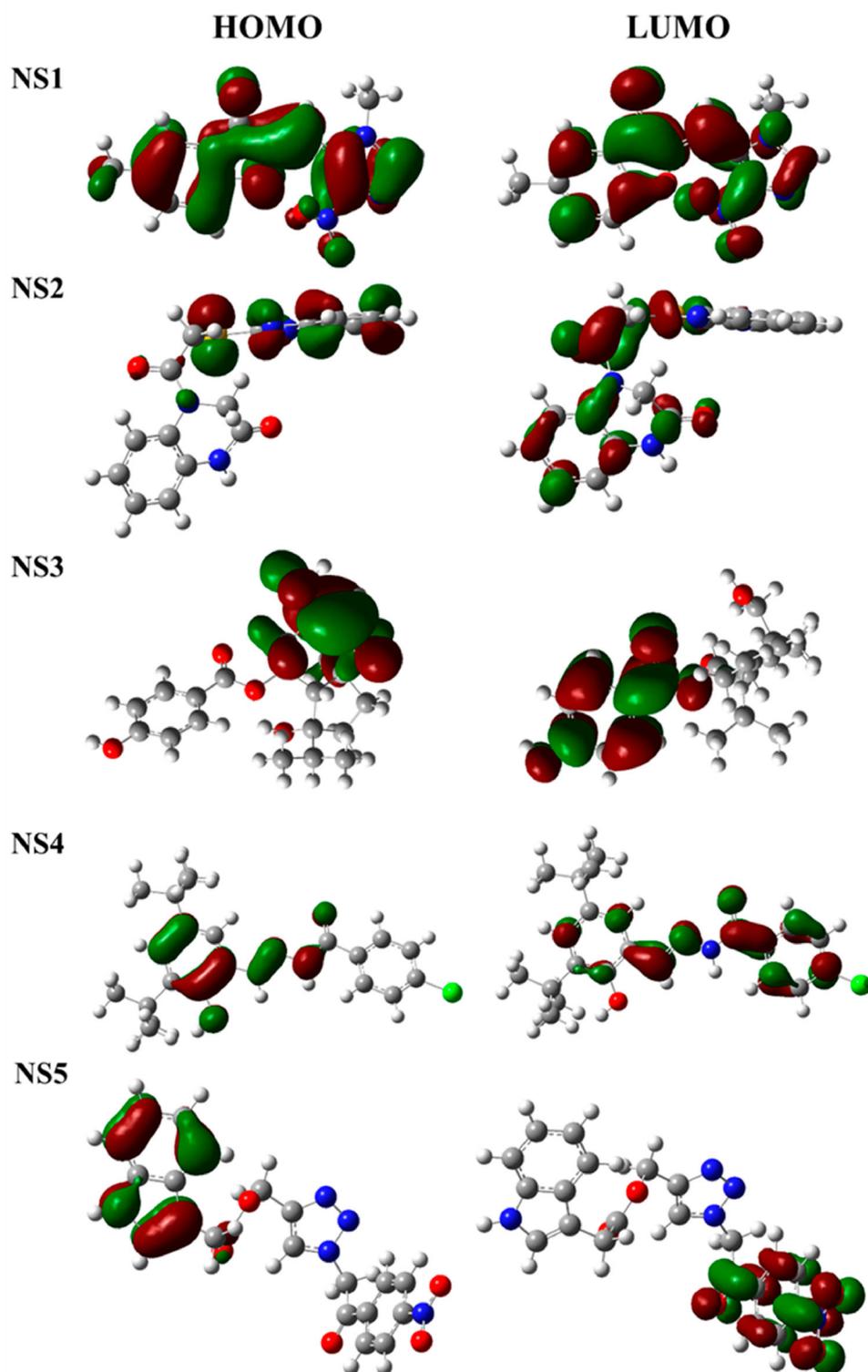


Figure 6. (Colour online) Spatial distribution of HOMO and LUMO molecular orbitals for NS1, NS2, NS3, NS4, NS5 and NS6 calculated via Gaussian 16 software package. Pictures were rendered using Gauss view 6.

hardness value, and the highest softness value, which indicates their high reactivity. Additionally, their lowest hardness value and the highest softness value indicate that they are potent inhibitors. On the other hand, compounds NS2, NS3 and NS4 have shown lower softness values than the standard compound, indicating its lower reactivity.

3.6. MD simulation

A detailed exploration of binding strength in a dynamic state was studied at the atomic level via MD simulation. Six systems were considered, including the five ligand–protein complexes and an unbound target protein. The rigorous examination of

both the free and bound states provided a comprehensive overview of the inter-molecular interactions and the structural perturbation upon ligand binding. In specific the stability matrices considered herein quantified the overall deviation, the movement at the residue level, and the compactness over time about initial conformation.

3.6.1. Root mean square deviation

The Root Mean Square Deviation (RMSD) offers an insight into the conformational changes which, in turn, determine the structural stability of the system under study. As mentioned earlier, RMSD is one of the commonly used matrices when studying structural deviations and is regarded as a potential predictor in assessing the system's quality. In the case of protein–ligand complexes, the RMSD sheds light on the stability of the system under study as a function of time. Herein, the protein's backbone atoms were included while calculating the RMSD, providing a quantitative measure of structural variations. The findings indicated that the RMSD value for the mutated protein in the ligand-free (apo) form falls between 0.5 and 0.6 nm. In general, the RMSD values for the test compounds lie below 0.5 nm except for the NS1-bound protein suggesting the stable binding mode in the targeted pocket. The most stable pattern of deviation was observed for the protein bound to the NS5 with an average RMSD value of 0.4 nm. An overview of the RMSD pattern is presented in Figure 7. The average deviation was 0.44 ± 0.02 , 0.55 ± 0.03 ,

0.46 ± 0.03 , 0.48 ± 0.03 , 0.52 ± 0.01 , 0.46 ± 0.03 nm for the ligand unbound mutant ALK and NS1-5 + ALK systems, respectively.

The observed variations in the RMSD values during the initial intervals stem from the necessary structural adjustments made by the ligand molecules to accommodate and bind the targeted cavity. However, the convergence was achieved afterwards as evidenced by a consistent pattern resulting from the strong and stable interactions established by the ligand within the active site. These findings indicate the potential of selected hits in maintaining the structural integrity with time consequently, providing evidence of favourable binding and reinforcing the potential therapeutic relevance.

3.6.2. Root mean square fluctuation

The dynamic variability and insight into per-residue movements as a function of time were assessed via root-mean-square fluctuation (RMSF). The extent of fluctuation of the side chains of protein residues was nearly identical across the tested complexes, as illustrated in Figure 8. In general, the studied systems demonstrated reduced fluctuations throughout the simulations compared to their initial placement. Meanwhile, the loop region displayed a minor degree of fluctuation across all systems compared to free protein, while the areas encompassing the active site residues exhibited marked stability throughout the simulation period.

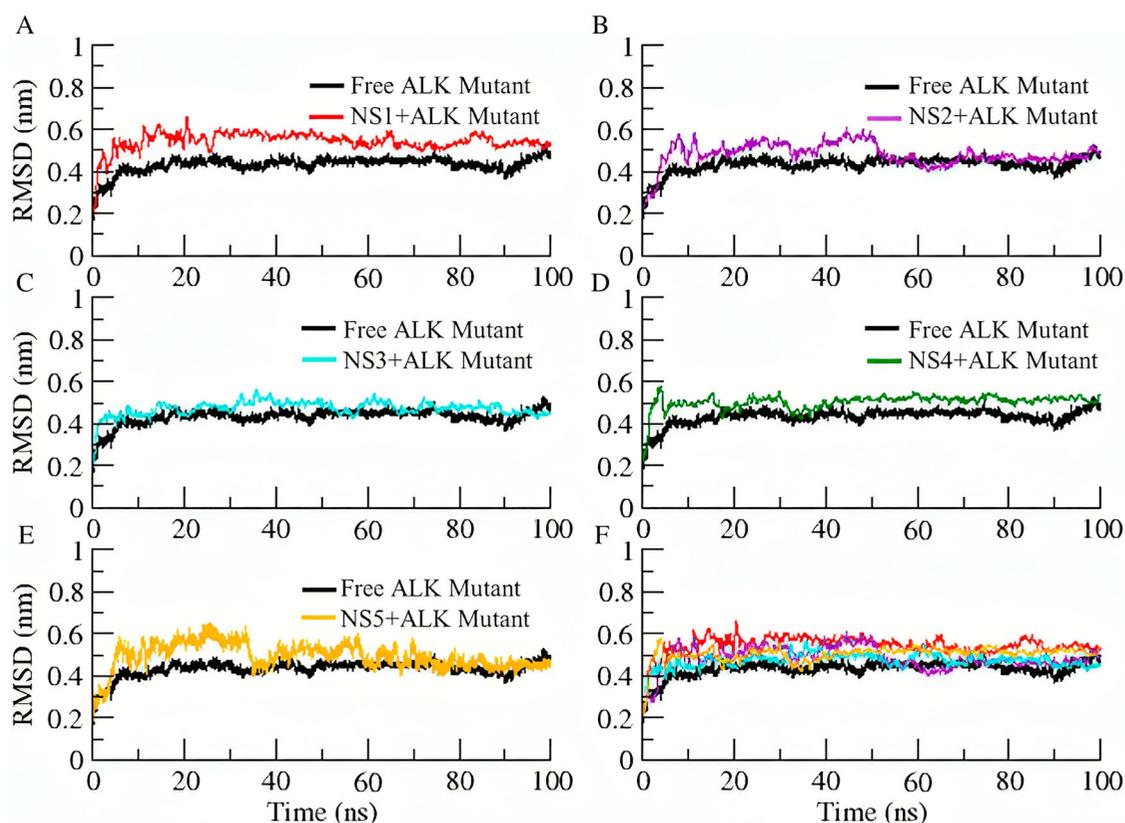


Figure 7. (Colour online) A depiction of the RMSD profile for backbone atoms of the ligand unbound or Apo protein (Black) and the ligand bound ALK for five identified hits (A–E) calculated as a function of time, (F) A comparative overview for the stability of backbone atoms observed for simulated systems. The colour scheme aligns with the hits (A–E), facilitating visual understanding.

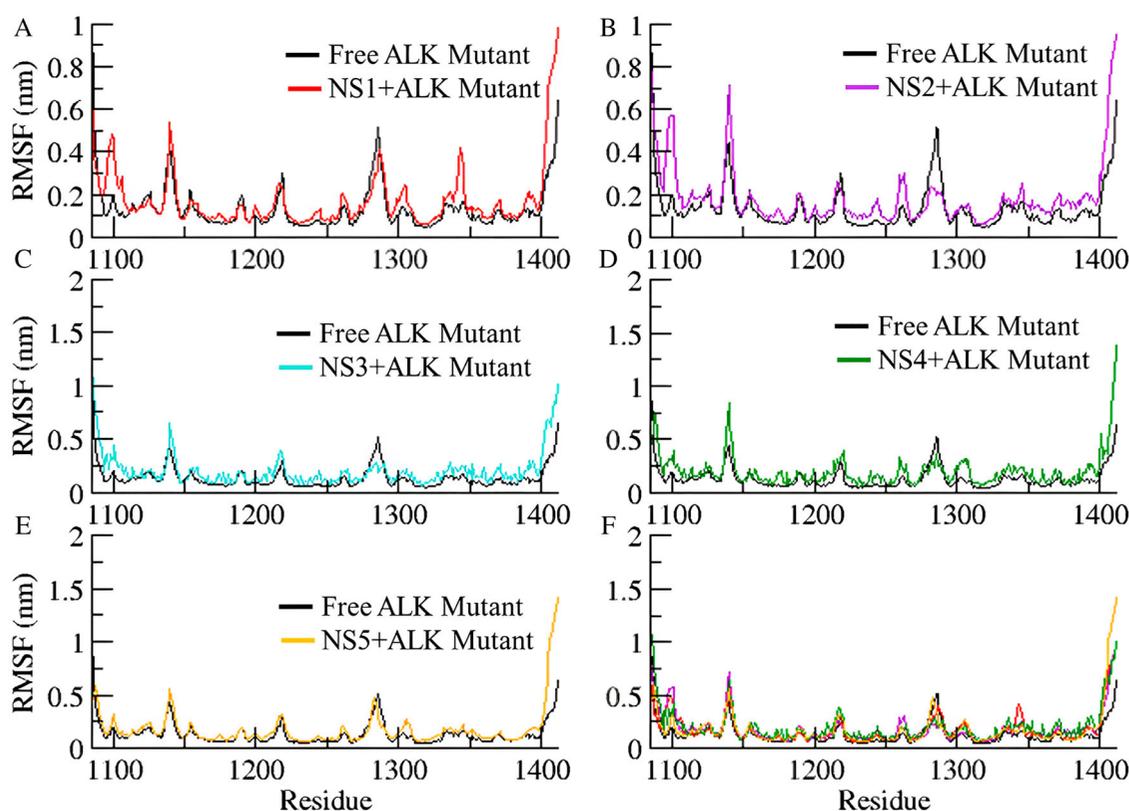


Figure 8. (Colour online) Time-dependent RMSF analysis presenting fluctuations of every amino acid residue within the simulated systems (A–E) during the targeted time frame alongside (F) a comparative overview employing a similar colour scheme as for the individual hits.

A mixed pattern of fluctuation was noticed for the binding pocket residues as displayed in SI-Figure 4. NS1 and NS3 imparted rigidity to some of the residues while inducing significant flexibility to other residues compared to the free state of the protein. The pocket residue for instance Leu1122 yielded an RMSF value of 0.170 nm in the apo state which decreased to 0.144, 0.143 and 0.156 nm in the case of protein bound to NS1, 3 and 4, respectively. However, the increased fluctuation was noticed for the backbone atom of the target protein bound to NS5 (0.216 nm).

3.6.3. Radius of gyration

The influence of ligand binding on the folding behaviour of a protein was monitored by the Radius of Gyration (Rog). The Rog evaluates the compactness of the system over time, with higher values indicating an unfolded conformation and vice versa. The system studied herein demonstrated a compact state during the terminal stages of the simulation, except for NS1 [Figure 9](#). In most cases, the complex system required 20 ns for initial adjustment except NS1, which reached equilibrium after 25 ns and showed insignificant fluctuations during the 75–85 ns time frame. The average deviation observed for the ligand unbound mutant ALK and NS1-5 + ALK was 2.12 ± 0.01 , 2.13 ± 0.02 , 2.12 ± 0.01 , 2.11 ± 0.02 , 2.08 ± 0.01 and 2.13 ± 0.02 nm, respectively. Overall, the data indicate that all systems maintained a compact conformation throughout

the simulation, implying that the systems have achieved convergence.

3.7. MM-PBSA binding free energy study of shortlisted hits

Molecular mechanics Poisson–Boltzmann surface area (MM/PBSA) is arguably the most extensively employed approach to assess the energetic factors involved in binding. Via integrating the energetic calculations employing molecular mechanics and implicit solvent models, the method describes the collective contribution of non-bonded interactions i.e. van der Waals, electrostatic, polar solvation and SASA between the heterocomplex over the simulation run. The low value (more negative) of free energy indicates a stable and thermodynamically favoured state. However, a higher value signifies a weaker binding and consequently a less stable state.

The free energy calculation using the Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA) approach holds considerable prominence in computational chemistry and drug design. Since it provides definite information regarding the binding affinities of selected protein–ligand complexes, it thus aids in understanding the determinants of molecular recognition. Herein, the collective impact of energetic factors such as non-bonded interactions including van der Waals, electrostatic, polar solvation and SASA are calculated as molecular mechanics and implicit solvent models during

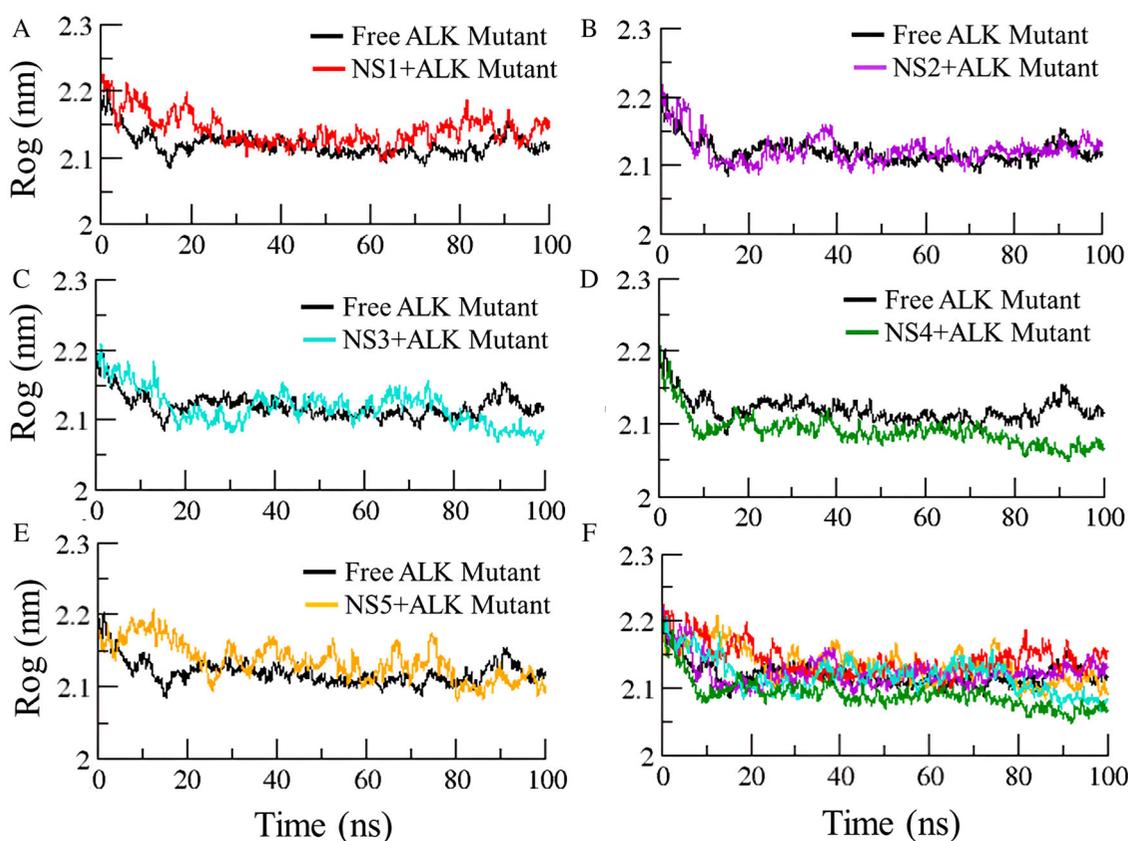


Figure 9. (Colour online) Radius of Gyration unveiling structural changes in the studied systems as a function of time for the identified hits (A–E), along with (F) an overlaid graph for comparison.

simulation (Table 3). The calculations revealed that the most stable and thermodynamically favourable state has been achieved by NS1 as evidenced by the highest ΔG_{bind} value of -154.96 ± 14.34 kcal/mol. Conversely, the lowest binding affinity (-33.06 ± 14.83 kcal/mol) was observed for NS2. These results agree with the preceding findings, highlighting the remarkable potential of NS1 to establish consistent electrostatic and hydrophobic interactions with the crucial residues of the target protein compared to the fewer stable inter-molecular contacts for the NS2 bound system. To gain further insight into the key elements determining the binding process, the total binding free energy was divided into several energy components. It was revealed that the electrostatic (Eelec) and van der Waals (EvdW) interactions favoured the ligand binding while antagonised by the polar solvation-free energy. In particular, the binding energy was significantly influenced by the van der Waals interactions, with NS1 having the highest van der Waals energy, followed by NS5 while NS2 bound ALK presented with the lowest van der Waals energy. As evident from the results, the electrostatic contributions were also

there but the impact was noticeably smaller than the hydrophobic component. Furthermore, Solvent-accessible surface area energy (SASA) was computed to evaluate the thermodynamic stability of the complex, with a lower SASA value suggesting more contracted centres and higher thermodynamic stability. As shown in Table 3, the system demonstrated stable conformation. Although all the hits exhibited stable binding mode and good binding energy profile, the NS1 and NS5 exhibited exceptional performance in all the tested parameters.

Considering the results obtained via an array of matrices studied in the present work, the identified hits emerge as potential candidates for future consideration against resistant phenotypes of ALK, a step towards the development of potent chemotypes to combat NSCLC. Although the pharmacophore-based virtual screening, particularly the structure-based models effectively prioritises potential candidates for drug discovery, the lack of experimental validation via in vitro and testing in animal models poses a limitation. Furthermore, computational mutagenesis was performed; however, a

Table 3. The binding free energy and different energy components calculated via the MMPBSA approach for the studied systems.

Energy component (kcal/mol)	NS1	NS2	NS3	NS4	NS5
Van der Waal (kcal/mol)	-210.12 ± 9.94	-130.46 ± 10.36	-161.18 ± 12.86	-133.23 ± 9.93	-163.68 ± 11.14
Electrostatic energy (kcal/mol)	-15.69 ± 4.38	-73.29 ± 10.29	-14.63 ± 6.68	-72.49 ± 10.42	-64.46 ± 14.74
Polar solvation energy (kcal/mol)	90.81 ± 14.85	188.57 ± 17.95	120.38 ± 18.24	177.57 ± 17.52	167.80 ± 12.68
SASA energy (kcal/mol)	-19.97 ± 1.18	-17.88 ± 0.99	-18.60 ± 1.19	-17.06 ± 0.99	-19.09 ± 1.42
Binding Energy (kcal/mol)	-154.96 ± 14.34	-33.06 ± 14.83	-74.08 ± 14.53	-52.79 ± 14.83	-79.42 ± 20.37

detailed investigation regarding mutation-induced alterations i.e. disturbance in interaction network, within the protein before the ligation with identified hits, as not pursued, representing another avenue for research.

4. Conclusion

A structure-based pharmacophore approach was used in this study to identify possible ALK mutant inhibitors. The pharmacophore model was validated using specificity and sensitivity measurements, and it could successfully identify compounds from our in-house database that shared similar pharmacophoric features. Molecular docking and interaction analysis helped narrow down the list of potential hits based on the binding energy and the interactions between proteins and ligands. The evaluation of the stability of protein–ligand complexes was conducted through RMSD and RMSF analyses after MD simulation, which provided insight into the capacity of the identified compounds to bind effectively within the active site and establish strong interactions. Overall, these findings provide valuable insights for the development of novel inhibitors targeting mutant ALK. These compounds stand out among the initial library of 14,000 small molecules, which, after further optimisation and experimental validation, are expected to offer promising research avenues, facilitating drug discovery pipeline in the future.

Acknowledgements

The authors extend their appreciation to the Researchers Supporting Project number (RSPD2024R994), King Saud University, Riyadh, Saudi Arabia.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Zaheer Ul-Haq  <http://orcid.org/0000-0002-8530-8711>

References

- [1] Zappa C, Mousa SA. Non-small cell lung cancer: current treatment and future advances. *Transl Lung Cancer Res.* 2016;5(3):288–300. doi: [10.21037/tlcr.2016.06.07](https://doi.org/10.21037/tlcr.2016.06.07)
- [2] Long F, Su J-H, Liang B, et al. Identification of gene biomarkers for distinguishing small-cell lung cancer from non-small-cell lung cancer using a network-based approach. *Biomed Res Int.* 2015;2015.
- [3] Du X, Shao Y, Qin HF, et al. ALK-rearrangement in non-small-cell lung cancer (NSCLC). *Thorac cancer.* 2018;9(4):423–430. doi: [10.1111/1759-7714.12613](https://doi.org/10.1111/1759-7714.12613)
- [4] Iwahara T, Fujimoto J, Wen D, et al. Molecular characterization of ALK, a receptor tyrosine kinase expressed specifically in the nervous system. *Oncogene.* 1997;14(4):439–449. doi: [10.1038/sj.onc.1200849](https://doi.org/10.1038/sj.onc.1200849)
- [5] Morris SW, Naevae C, Mathew P, et al. ALK, the chromosome 2 gene locus altered by the t(2;5) in non-Hodgkin's lymphoma, encodes a novel neural receptor tyrosine kinase that is highly related to leukocyte tyrosine kinase (LTK). *Oncogene.* 1997;14(18):2175–2188. doi: [10.1038/sj.onc.1201062](https://doi.org/10.1038/sj.onc.1201062)
- [6] Kong X, Pan P, Sun H, et al. Drug discovery targeting anaplastic lymphoma kinase (ALK). *J Med Chem.* 2019;62(24):10927–10954. doi: [10.1021/acs.jmedchem.9b00446](https://doi.org/10.1021/acs.jmedchem.9b00446)
- [7] Salifu EY, Rashid IA, Soliman ME. Impact of compound mutations I1171N+ F1174I and I1171N+ L1198H on the structure of ALK in NSCLC pathogenesis: atomistic insights. *J Biomol Struct Dyn.* 2022;41(10):4735–4743.
- [8] Rikova K, Guo A, Zeng Q, et al. Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell.* 2007;131(6):1190–1203. doi: [10.1016/j.cell.2007.11.025](https://doi.org/10.1016/j.cell.2007.11.025)
- [9] Soda M, Choi YL, Enomoto M, et al. Identification of the transforming EML4–ALK fusion gene in non-small-cell lung cancer. *Nature.* 2007;448(7153):561–566. doi: [10.1038/nature05945](https://doi.org/10.1038/nature05945)
- [10] Morán T, Quiroga V, de los Llanos Gil M, et al. Targeting EML4–ALK driven non-small cell lung cancer (NSCLC). *Transl Lung Cancer Res.* 2013;2(2):128.
- [11] Cui JJ, Tran-Dubé M, Shen H, et al. Structure based drug design of crizotinib (PF-02341066), a potent and selective dual inhibitor of mesenchymal–epithelial transition factor (c-MET) kinase and anaplastic lymphoma kinase (ALK). *J Med Chem.* 2011;54(18):6342–6363. doi: [10.1021/jm2007613](https://doi.org/10.1021/jm2007613)
- [12] Carpenter EL, Mossé YP. Targeting ALK in neuroblastoma—preclinical and clinical advancements. *Nat Rev Clin Oncol.* 2012;9(7):391–399. doi: [10.1038/nrclinonc.2012.72](https://doi.org/10.1038/nrclinonc.2012.72)
- [13] Roskoski R Jr. Anaplastic lymphoma kinase (ALK) inhibitors in the treatment of ALK-driven lung cancers. *Pharmacol Res.* 2017;117:343–356. doi: [10.1016/j.phrs.2017.01.007](https://doi.org/10.1016/j.phrs.2017.01.007)
- [14] Katayama R, Shaw AT, Khan TM, et al. Mechanisms of acquired crizotinib resistance in ALK-rearranged lung cancers. *Sci Transl Med.* 2012;4(120):120ra117–120ra117. doi: [10.1126/scitranslmed.3003316](https://doi.org/10.1126/scitranslmed.3003316)
- [15] Johnson TW, Richardson PF, Bailey S, et al. Discovery of (10R)-7-Amino-12-fluoro-2,10,16-trimethyl-15-oxo-10,15,16,17-tetrahydro-2H-8,4-(metheno)pyrazolo[4,3-h][2,5,11]-benzoxadiazacyclotetradecine-3-carbonitrile (PF-06463922), a macrocyclic inhibitor of anaplastic lymphoma kinase (ALK) and c-ros oncogene 1 (ROS1) with preclinical brain exposure and broad-spectrum potency against ALK-resistant mutations. *J Med Chem.* 2014;57(11):4720–4744. doi: [10.1021/jm500261q](https://doi.org/10.1021/jm500261q)
- [16] Zafar A, Wang W, Liu G, et al. Molecular targeting therapies for neuroblastoma: progress and challenges. *Med Res Rev.* 2021;41(2):961–1021. doi: [10.1002/med.21750](https://doi.org/10.1002/med.21750)
- [17] Yang J, Gong W. Lorlatinib for the treatment of anaplastic lymphoma kinase-positive non-small cell lung cancer. *Expert Rev Clin Pharmacol.* 2019;12(3):173–178. doi: [10.1080/17512433.2019.1570846](https://doi.org/10.1080/17512433.2019.1570846)
- [18] Pan Y, Deng C, Qiu Z, et al. The resistance mechanisms and treatment strategies for ALK-rearranged non-small cell lung cancer. *Front Oncol.* 2021;11:713530. doi: [10.3389/fonc.2021.713530](https://doi.org/10.3389/fonc.2021.713530)
- [19] Shiba-Ishii A, Johnson TW, Dagogo-Jack I, et al. Analysis of lorlatinib analogs reveals a roadmap for targeting diverse compound resistance mutations in ALK-positive lung cancer. *Nat Cancer.* 2022;3(6):710–722. doi: [10.1038/s43018-022-00399-6](https://doi.org/10.1038/s43018-022-00399-6)
- [20] Yoda S, Lin JJ, Lawrence MS, et al. Sequential ALK inhibitors can select for lorlatinib-resistant compound ALK mutations in ALK-positive lung cancer. *Cancer Discov.* 2018;8(6):714–729. doi: [10.1158/2159-8290.CD-17-1256](https://doi.org/10.1158/2159-8290.CD-17-1256)
- [21] Karabencheva TG, Lee CC, Black GW, et al. How does conformational flexibility influence key structural features involved in activation of anaplastic lymphoma kinase? *Mol. Biosyst.* 2014;10(6):1490–1495. doi: [10.1039/C4MB00141A](https://doi.org/10.1039/C4MB00141A)
- [22] He M, Li W, Zheng Q, et al. A molecular dynamics investigation into the mechanisms of alectinib resistance of three ALK mutants. *J Cell Biochem.* 2018;119(7):5332–5342. doi: [10.1002/jcb.26666](https://doi.org/10.1002/jcb.26666)
- [23] Liang Shuai, et al. Deciphering the mechanism of gilteritinib overcoming lorlatinib resistance to the double mutant I1171N/F1174I in anaplastic lymphoma kinase. *Front Cell Dev Biol.* 2021;23(9):808864. doi: [10.3389/fcell.2021.808864](https://doi.org/10.3389/fcell.2021.808864)
- [24] Bhardwaj V, Purohit R. Computational investigation on effect of mutations in PCNA resulting in structural perturbations and

- inhibition of mismatch repair pathway. *J Biomol Struct Dyn*. 2019;38(7):1963–1974. doi: [10.1080/07391102.2019.1621210](https://doi.org/10.1080/07391102.2019.1621210)
- [25] Sharma J, Bhardwaj VK, Das P, et al. Identification of naturally originated molecules as γ -aminobutyric acid receptor antagonist. *J Biomol Struct Dyn*. 2021;39(3):911–922. doi: [10.1080/07391102.2020.1720818](https://doi.org/10.1080/07391102.2020.1720818)
- [26] Singh R, Bhardwaj VK, Purohit R. Computational targeting of allosteric site of MEK1 by quinoline-based molecules. *Cell Biochem Funct*. 2022;40(5):481–490. doi: [10.1002/cbf.3709](https://doi.org/10.1002/cbf.3709)
- [27] Mysinger MM, Carchia M, Irwin JJ, et al. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J Med Chem*. 2012;55(14):6582–6594. doi: [10.1021/jm300687e](https://doi.org/10.1021/jm300687e)
- [28] MOE, CCG. (2019). Molecular Operating Environment (MOE). In.
- [29] Halgren TA. MMFF VI. MMFF94s option for energy minimization studies. *J Comput Chem*. 1999;20(7):720–729. doi: [10.1002/\(SICI\)1096-987X\(199905\)20:7<720::AID-JCC7>3.0.CO;2-X](https://doi.org/10.1002/(SICI)1096-987X(199905)20:7<720::AID-JCC7>3.0.CO;2-X)
- [30] Huang W-S, Liu S, Zou D, et al. Discovery of brigatinib (AP26113), a phosphine oxide-containing, potent, orally active inhibitor of anaplastic lymphoma kinase. *J Med Chem*. 2016;59(10):4948–4964. doi: [10.1021/acs.jmedchem.6b00306](https://doi.org/10.1021/acs.jmedchem.6b00306)
- [31] Akamine T, Toyokawa G, Tagawa T, et al. Spotlight on lorlatinib and its potential in the treatment of NSCLC: the evidence to date. *Onco Targets Ther*. 2018;11:5093–5101. doi: [10.2147/OTT.S165511](https://doi.org/10.2147/OTT.S165511)
- [32] Motoc I, Dammkoehler RA, Mayer D, et al. Three-dimensional quantitative structure-activity relationships I. General approach to the pharmacophore model validation. *Quant Struct-Act Relat*. 1986;5(3):99–105. doi: [10.1002/qsar.19860050305](https://doi.org/10.1002/qsar.19860050305)
- [33] Case DA, Darden TA, Cheatham TE, et al. (2021). Amber 2021. University of California, San Francisco.
- [34] Adasme MF, Linnemann KL, Bolz SN, et al. PLIP 2021: expanding the scope of the protein–ligand interaction profiler to DNA and RNA. *Nucleic Acids Res*. 2021;49(W1):W530–W534. doi: [10.1093/nar/gkab294](https://doi.org/10.1093/nar/gkab294)
- [35] Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep*. 2017;7(1):42717. doi: [10.1038/srep42717](https://doi.org/10.1038/srep42717)
- [36] Ali M, Mansha A, Asim S, et al. DFT Study for the spectroscopic and structural analysis of p-Dimethylaminoazobenzene. *J Spectrosc*. 2018;2018.
- [37] Deshmukh MS, Sekar N. A combined experimental and TD-DFT investigation of three disperse azo dyes having the nitroterephthalate skeleton. *Dyes Pigment*. 2014;103:25–33. doi: [10.1016/j.dyepig.2013.10.035](https://doi.org/10.1016/j.dyepig.2013.10.035)
- [38] Kovach A, He J, Saris PJ, et al. Optically tunable microresonator using an azobenzene monolayer. *AIP Adv*. 2020;10(4). doi: [10.1063/1.5143253](https://doi.org/10.1063/1.5143253)
- [39] Frisch MJ, Trucks GW, Schlegel HB, et al. (2016). Gaussian 16 Rev. C.01. In.
- [40] Abraham MJ, Murtola T, Schulz R, et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015;1-2:19–25. doi: [10.1016/j.softx.2015.06.001](https://doi.org/10.1016/j.softx.2015.06.001)
- [41] Sousa da Silva AW, Vranken WF. ACPYPE-antechamber python parser interface. *BMC Res Notes*. 2012;5:1–8. doi: [10.1186/1756-0500-5-367](https://doi.org/10.1186/1756-0500-5-367)
- [42] Jorgensen WL, Chandrasekhar J, Madura JD, et al. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983;79(2):926–935. doi: [10.1063/1.445869](https://doi.org/10.1063/1.445869)
- [43] Martoňák R, Laio A, Parrinello M. Predicting crystal structures: the Parrinello-Rahman method revisited. *Phys Rev Lett*. 2003;90(7):075503. doi: [10.1103/PhysRevLett.90.075503](https://doi.org/10.1103/PhysRevLett.90.075503)
- [44] Martyna GJ, Klein ML, Tuckerman M. Nosé–Hoover chains: the canonical ensemble via continuous dynamics. *J Chem Phys*. 1992;97(4):2635–2643. doi: [10.1063/1.463940](https://doi.org/10.1063/1.463940)
- [45] Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph*. 1996;14(1):33–38. doi: [10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5)
- [46] Turner P. (2005). XMGRACE, Version 5.1. 19. Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, Beaverton, OR, 2.
- [47] Pettersen EF, Goddard TD, Huang CC, et al. UCSF chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25(13):1605–1612. doi: [10.1002/jcc.20084](https://doi.org/10.1002/jcc.20084)
- [48] Kumari R, Kumar R, Consortium OSDD, et al. g_mmpbsa—A GROMACS tool for high-throughput MM-PBSA calculations. *J Chem Inf Model*. 2014;54(7):1951–1962. doi: [10.1021/ci500020m](https://doi.org/10.1021/ci500020m)
- [49] Jiang X, Zhou J, Ai J, et al. Novel tetracyclic benzo [b] carbazolones as highly potent and orally bioavailable ALK inhibitors: design, synthesis, and structure–activity relationship study. *Eur J Med Chem*. 2015;105:39–56. doi: [10.1016/j.ejmech.2015.10.005](https://doi.org/10.1016/j.ejmech.2015.10.005)
- [50] James N, Shanthi V, Ramanathan K. Density functional theory and molecular simulation studies for prioritizing anaplastic lymphoma kinase inhibitors. *Appl Biochem Biotechnol*. 2020;190:1127–1146. doi: [10.1007/s12010-019-03156-1](https://doi.org/10.1007/s12010-019-03156-1)
- [51] Roskoski R Jr. Classification of small molecule protein kinase inhibitors based upon the structures of their drug-enzyme complexes. *Pharmacol Res*. 2016;103:26–48. doi: [10.1016/j.phrs.2015.10.021](https://doi.org/10.1016/j.phrs.2015.10.021)
- [52] Roskoski R Jr. Anaplastic lymphoma kinase (ALK): structure, oncogenic activation, and pharmacological inhibition. *Pharmacol Res*. 2013;68(1):68–94. doi: [10.1016/j.phrs.2012.11.007](https://doi.org/10.1016/j.phrs.2012.11.007)
- [53] Peddi SR, Sivan SK, Manga V. An integrated molecular modeling approach for in silico design of new tetracyclic derivatives as ALK inhibitors. *J Recept Signal Transduct*. 2016;36(5):488–504. doi: [10.3109/10799893.2015.1130057](https://doi.org/10.3109/10799893.2015.1130057)
- [54] Bossi RT, Saccardo MB, Ardini E, et al. Crystal structures of anaplastic lymphoma kinase in complex with ATP competitive inhibitors. *Biochemistry*. 2010;49(32):6813–6825. doi: [10.1021/bi1005514](https://doi.org/10.1021/bi1005514)
- [55] Hay M, Thomas DW, Craighead JL, et al. Clinical development success rates for investigational drugs. *Nat Biotechnol*. 2014;32(1):40–51. doi: [10.1038/nbt.2786](https://doi.org/10.1038/nbt.2786)
- [56] Lipinski CA, Lombardo F, Dominy BW, et al. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*. 2012;64:4–17. doi: [10.1016/j.addr.2012.09.019](https://doi.org/10.1016/j.addr.2012.09.019)
- [57] Ritchie TJ, Macdonald SJ, Peace S, et al. Increasing small molecule drug developability in sub-optimal chemical space. *MedChemComm*. 2013;4(4):673–680. doi: [10.1039/c3md00003f](https://doi.org/10.1039/c3md00003f)
- [58] Savjani KT, Gajjar AK, Savjani JK. Drug solubility: importance and enhancement techniques. *Int Sch Res Notices*. 2012;2012.
- [59] Kurt M, Sertbakan TR, Özduran M. An experimental and theoretical study of molecular structure and vibrational spectra of 3- and 4-pyridineboronic acid molecules by density functional theory calculations. *Spectrochim Acta A Mol Biomol Spectrosc*. 2008;70(3):664–673. doi: [10.1016/j.saa.2007.08.019](https://doi.org/10.1016/j.saa.2007.08.019)
- [60] Abraham CS, Prasana JC, Muthu S. Quantum mechanical, spectroscopic and docking studies of 2-amino-3-bromo-5-nitropyridine by density functional method. *Spectrochim Acta A Mol Biomol Spectrosc*. 2017;181:153–163. doi: [10.1016/j.saa.2017.03.045](https://doi.org/10.1016/j.saa.2017.03.045)
- [61] Shanmugam G, Lee SK, Jeon J. Identification of potential nematocidal compounds against the pine wood nematode, *Bursaphelenchus xylophilus* through an in silico approach. *Molecules*. 2018;23(7):1828. doi: [10.3390/molecules23071828](https://doi.org/10.3390/molecules23071828)
- [62] La Porta FA, Ramalho TC, Santiago RT, et al. Orbital signatures as a descriptor of regioselectivity and chemical reactivity: the role of the frontier orbitals on 1, 3-dipolar cycloadditions. *J Phys Chem A*. 2011;115(5):824–833. doi: [10.1021/jp108790w](https://doi.org/10.1021/jp108790w)