Journal of the Serbian Chemical Society
JSCS@tmf.bg.ac.rs • www.shd.org.rs/JSCS
J. Serb. Chem. Soc. *89* (4) S162–S180 (2024) *Supplementary material*

SUPPLEMENTARY MATERIAL TO

A recent tactic for searching CDK-7 kinase inhibitor by NCI database screening

MOHAMMAD RASHID[1]*, MD TANWIR ATHAR[1], AFZAL HUSSAIN[2], NORAH M. ALMADANI[3] and ASHFAQ HUSSAIN[4]

[1]Department of Pharmacognosy and Pharmaceutical Chemistry, College of Dentistry and Pharmacy, Buraydah Private Colleges, Buraydah-51418, Saudi Arabia, [2]Department of Bioinformatics, Maulana Azad National Institute of Technology Bhopal, Madhya Pradesh-462003, India, [3]Biochemistry Department, Faculty of Science, University of Tabuk, Tabuk-47914, Saudi Arabia and [4]Department of Electronics Engineering, Rajasthan Technical University, Kota-324010, India

J. Serb. Chem. Soc. *89* (4) (2024) 505–519

INTRODUCTORY DETAILS

A cyclin-dependent protein kinase called cyclin-dependent kinase-7 (CDK7) regulates cell cycle progression by phosphorylating to other CDKs and altering their activity. Together with cyclin H and a third protein (a RING finger protein termed MAT1), the phosphorylation of CDK7 is necessary for both its stability and function. Cell cycle progression and transcription in the form of transcription factor TFIIH are regulated by the CDK7 catalytic subunit of CDK-activating kinase (CAK). Cell cycle progression is influenced by the threonine phosphorylation of cyclin-connected kinases CDK1, CDK2, CDK4, and CDK6 by CDK-activating kinase (CAK). RNA polymerase II (Pol II) is stimulated by CAK through serine phosphorylation (Ser5) of the repetitive C-terminal domain (CTD) of its large subunit (POLR2A), allowing escape from the promoter and transcript elongation. Transcription elongation is promoted by CDK7 phosphorylating CDK9, and CDK9 phosphorylating Pol II CTD Ser2.

Furthermore, in model systems, combinations of CDK7i with other targeted cancer medicines, including as BET inhibitors, BCL2 inhibitors, and hormone therapy, have demonstrated success. A promising clinical candidate, ICEC0942 potently reduced the growth of a variety of cancer cell lines and ER+ breast cancer xenografts. This was due to ICEC0942's good ADME and pharmacokinetic (absorption, distribution, metabolism, and excretion) characteristics. The medication is currently being tested in a Phase I/II clinical trial for patients with

* Corresponding author. E-mail: Rashid.Mohd@bpc.edu.sa; rashidpharm2008@gmail.com

advanced solid tumours, including a concentrated cohort of those with breast and prostate cancer. The medicine was licensed to Carrick Therapeutics.

EXPERIMENTAL DETAILS

National Cancer Institute (NCI) Database Compounds and Reference Ligand Preparation

A research division of the National Cancer Institute's chemical biology lab is known as CADD (computer-aided drug design) (https://cactus.nci.nih.gov, 2022). The NCI's (National Institutes of Health) Chemical Biology Laboratory (CBL), located at the Frederick National Laboratory for Cancer Research (FNLCR), formerly NCI-Frederick, houses the computer-aided drug design (CADD) group.

Preparation of the Grid

The receptor grid generation panel is used to specify a receptor structure as well as configure the grid generation task. This operation generates the grid files that depict the receptor's active site for glide ligand docking jobs. A prepared structure is required for receptor grid generation: an all-atom structure with acceptable bond ordering and formal charges.

Molecular Docking & Screening

All the generated NCI compounds were subjected to Lipinski filtration and reactive functionality analyses. To get the best result, HTVS (High-throughput virtual screening), SP (standard precision), XP (extra precision) screening, and docking were employed, and the XP descriptor information was also picked for pose viewing.

Simulation of Molecular Dynamics (MD)

In order to restrict the covalent bonds, the LINCS (linear constraint solver) methodology was used, and in order to deal with the long-range coulombic interactions, the particle mesh Ewald (PME) method was applied. Production MD runs lasted 200 ns at a constant temperature of 300 K, and MD trajectories were captured every 2.0 ps. Gromacs' built-in utilities, including gmx rms, hbond, rmsf, gyrate, and others, as well as the H-bonds plugin found in the VMD (Visual molecular dynamics) tool, were used to post-molecularly dynamically evaluate the trajectories. With a donor-acceptor distance cutoff of 3.5Å and a bond angle cutoff of 20Å the H-bonds plugin of VMD was employed to evaluate the involvement of intermolecular hydrogen bonds established by the ligand throughout the MD run during each of the two complexes, in addition to specific contacts with the key binding site residues of the CDK7 Kinase.

A Molinspiration Algorithm for Predicting Drug Similarity

Based on the five Lipinski criteria, the Molinspiration web software examined the physical characteristics of molecules. The smiles' representation was created using the ChemBioDraw Ultra software (11.0 version). A range of molecular characteristics, including topological polar surface area (TPSA), partition coefficient (LogP), molecular weight, hydrogen bond donors and acceptors, rotatable bonds, atom count, and departures from Lipinski's rule of five, were used to assess the compound's drug-likeness. To determine the percentage of absorption, the formula %Ab = 109 - [0.345 x TPSA] was employed. A substance meets Lipinski's criteria for a drug if it has a partition coefficient below 5, a polar surface area below140 $Å^2$, a H bond acceptor and donor below 10, and a molecular weight below 500 daltons. The Rule of 5 is another name for the rule because the border values are 5,

2*5, 500, and 5. A problem with a molecule's bioavailability may arise if it violates more than one of these rules.

The sum of the contributions based on fragments and the correction factors is known as the LogP (Octanol/water partition coefficient). Topological Polar Surface Area (TPSA), which describes drug absorption, is the sum of the contributions produced by O- and N-centred polar fragments and includes intestinal absorption, bioavailability, Caco-2 permeability, and blood-brain barrier penetration. The entire quantity of a training set's fragment contributions to the actual 3D volume is the molecule's volume. The n-rotb (number of rotatable bonds), which is distinct from any individual non-ring bond and attached to a non-terminal heavy (non-hydrogen) atom, is a gauge of molecular flexibility.

Toxicity Assessment

The results of virtual screening are valued and colour coded as green/red/ yellow for properties such as effect on mutagenic, reproductive system, irritant effect, and tumorigenicity. Properties shown in red indicate high risks of undesired effects, while properties shown in green indicate drug conformity, compatibility, and safety. For the results prediction, the compound charges must be balanced and atom valences must not be exceeded for a proper chemical structure. A single bond connecting a positive charge on the nitrogen with a negative charge on one of the oxygens is required for drawing nitro-groups. Mutagenic or poor intestinal absorption are depicted in red, indicating a significant chance of unwanted consequences, whereas a green colour indicates drug-like action.

Higher-molecular-weight compounds are less likely to be absorbed and so never reach the site of action. Because low hydrophilicities and consequently large logP values induce poor absorption or penetration, the logP value must be less than 5.0. With more than 80% of drugs on the market having an estimated logS value greater than -4, a compound's water solubility has a significant impact on its absorption and distribution qualities. The capacity of the molecule to traverse membranes is improved if the inputs of all polar atoms exposed to the surface of the molecule amount to more than 80 or 100. To assess a compound's pharmacological similarity, many methods use topological descriptors, fingerprints of MDL structure keys, or other features like cLogP and molecular weights. The drug score, which can be used to determine if a molecule would be suitable for use as medication, is a condensed combination of rug similarity, cLogP, logS, molecular weight, and toxicity concerns.

Examining Osiris's Potential as a Drug and Its Relevant Properties

To prevent destructive compounds from further processing in the drug discovery and development unit, toxicity risk assessment is essential. The toxicity evaluation risk alarms indicate that the depicted structure may be hazardous for the stated risk category, and toxicity hazards are calculated using a colour code. The red hue signifies a greater danger of side effects or negative qualities like mutagenic or poor intestine absorption, whereas the green colour suggests drug-like pharmacological behaviour. Although the green hue reflects the desirable benefits of the chemical, the hazardous hazards (unwanted) consequences of the molecule are displayed in red. The mutagenic, tumorigenic, irritating, and reproductive toxicity hazards were assessed using a pre-computed set of structural fragments based on compound categorization from the Registry of Toxic Effects of Chemical Substances (RTECS) database.

Acute toxicity prediction by Pre ADMET program and T.E.S.T. tool

To avoid the 50% failure rate of drug candidates due to insufficient ADME/Toxicity inadequacies during the development stage, it is necessary to conduct ADME/Toxicity

screening at an early stage in the discovery phase. The Pre ADMET program is divided into four sections, as follows: This section comprises the determination of physicochemical characteristics such as lipophilicity (logP), molecular weight, polar surface area, and water solubility, all of which are directly connected to ADME/Toxicity qualities. Drug similarity prediction; The drug-similarity prediction module is based on Lipinski's and Lead-like. ADME prediction; Caco2 cell model and the MDCK (Madin Darby canine kidney) cell model are indicated as accurate in-vitro models for oral medication absorption prediction. Pre-ADMET also includes an in-silico human intestinal absorption (HIA) and skin permeability model for identifying possible medication candidates for oral and transdermal drug administration. Infiltration of the blood-brain barrier (BBB) provides information on therapeutic drugs in the central nervous system (CNS), plasma protein binding models in drug disposition, and drug effectiveness in the distribution stage.

The quantitative structure-activity relationship (QSAR) and Quantitative Structure-Property Relation (QSPR) of dangerous compounds were predicted using molecular properties of compounds as a descriptor using T.E.S.T software (version 4.2.1). These molecular features are crucial for predicting QSAR of harmful compounds. By using a type of approach such as consensus, hierarchical clustering, The United States food and drug administration (US-FDA), and closest neighbour method, the T.E.S.T. software program evaluated the acute toxicity of compounds, such as oral lethal dose for 50% of test rats (LD50), and mutagenic of compounds. The descriptors of synthetic and natural compounds in the T.E.S.T tool are based on the molecular structures of the substances.

Dr. Ames developed the straightforward Ames' test to determine whether a material is mutagenic or not. It uses several strains of the bacterium Salmonella typhimurium, which needs histidine to develop due to gene changes in histidine production. Ames TA100 (+S9) findings from the in-vitro test on the TA100 strain (metabolic activation by rat liver homogeneity). Results of the in-vitro test on the TA100 strain by Ames TA100 (-S9) (No metabolic activation). Results of the in-vitro test on the TA1535 strain by Ames, strain TA1535 (+S9) (Metabolic activation by rat liver homogeneity). Ames TA1535 (-S9); the TA1535 strain's in-vitro test results (No metabolic activation). By leaving the strictness of fingerprint matching at default and choosing all models, the pre-ADMET web-based application was able to forecast the first phase of metabolism for the chemical. Cytochrome P450 (CYP450), also known as the isozymes group and engaged in the metabolism of medicines, fatty acids, steroids, bile acids, and other carcinogens, is the most significant parameter. A family of heme proteins called cytochrome P450 enzymes are involved in the metabolism of numerous pharmacologically active substances. They can interact with other medications when taken together and result in undesirable side effects.

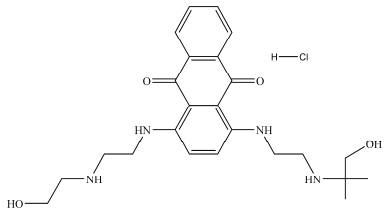
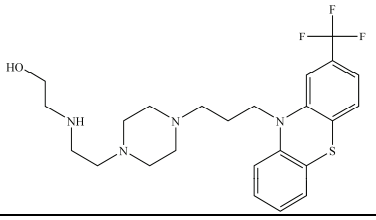
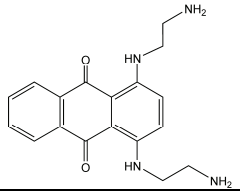
ADDITIONA RESULT AND DISCUSSION

Target Protein Docking Score and Analysis

The NCI database (https://cactus.nci.nih.gov/ncicadd/about.html, 2022) was used to ascertain the binding affinity and interactions of several molecules with the selected target protein, CDK7 kinase. Ten compounds were found after reviewing the NCI database; their names are NCI613391 (-13.06 kJ/mol), NCI169676 (-11.91 kJ/mol), NCI281246 (-11.74 kJ/mol) and NCI339580 (-11.58 kJ/mol). Among these top NCI compounds, the NCI613391 molecule has the highest docking score (-13.06 kJ/mol) in comparison to the other molecules. The

same residues that the ATP molecule binds to are found in the active site of the target protein (CDK7 kinase), where all drugs bind. This indicates that the molecules have potent inhibitory properties and is firmly docked in the target protein's active region. The docking score and binding interaction of NCI-screened compounds with the target protein's active site residues are displayed in Tables S-I and S-II, respectively. This interaction with the target protein is shown in Fig S-1. by the NCI-screened top molecular docking score and their Lig plot interaction diagram and for other best finding molecules shown in Fig. S-1. Fig. S-2 displays the plot chart of compounds that the NCI screened using a docking score comparison.

Table S-I: Docking scores, smiles, formula and 2D structure of NCI-screened compounds.

NCI Compd.	Docking score (kJ/mol)	2D Chemical structure
613391	-13.06	
169676	-11.91	
281246	-11.74	

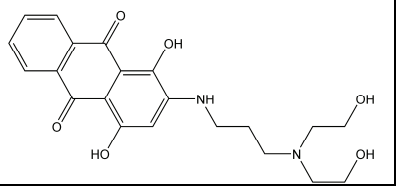
339580	-11.58	

Fig. S-1. The docking score comparison of NCI screened ten best finding compounds in the active binding site residues of CDK7 kinase protein.

Drug-Likeliness properties and ADME/T analysis of best NCI screened compounds

The physiochemical characteristics of the compounds that passed the NCI screening, as determined by the QikProp module of Schrödinger, are listed in Tables S-III and S-IV. All of the substances that passed the NCI screening were found to have oral absorption that was between 25% less and >80% higher than the human oral absorption limit, or 36.306 to 73.742. Compound NCI281246 had the lowest oral absorption (36.306 and 25% less) while compound NCI613391 had the highest oral absorption (73.74%). The hydrogen bond donor (HBD) count of every component was determined to be between 2 and 4, which is less than the allowed limit of 5. The total number of hydrogen bond acceptors (HBA) was also reported to be below the permitted limit (10), which is in the range of 6 to 9 (Table S-III). All NCI-screened compounds have solvent accessible surfaces in total below the permissible range (300-1000), which may be found in 636.839 and 832.142. All compounds under investigation had aqueous solubility (QPlogS) values between -1.162 and 4.213, which was within the permitted range (-6.5 to 0.5).

Table S-II: Drug-like properties analysis by Lipinski's Rule of Five of NCI screened compounds.

Comp. Id[a]	MW[b]	QPlogS[c]	SASA[d] (Å)	HB Acceptor[e]	HB Donor[f]	HOA[g] (%)
NCI613391	477.0021	-2.856	832.142	9.9	4	39.882
NCI169676	480.5903	-3.246	798.355	7.7	2	73.745
NCI281246	324.3816	-1.162	604.145	6	4	36.306
NCI339580	400.4304	-2.212	700.349	9.9	3	49.074

Comp. Id[a]; Compound Id, Mol. Wt.[b]; Molecular weight in daltons (permissible scale is: ≤500), QPlog S[c]; Prediction of aqueous solubility (permissible scale is: -6.5 to 0.5), SASA[d]; Total solvent accessible surface area in Å (permissible scale is: 300-1000), HB Acceptor[e]; Hydrogen bond acceptor (permissible scale is:≤10), HB donor[f]; Hydrogen bond donor (permissible scale is: ≤5), Human oral absorption[g] (permissible scale is:<25% less and >80% high)

The NCI-screened compounds were discovered to have molecular weights between 308 and 480 Dalton, which is less than 500 Dalton (within the Lipinski limit) and indicates that they are more easily transported, dispersed, and absorbed than heavier molecules. Comparatively, NCI169676 had a larger molecular weight (480), and NCI281246 had the lowest (324). All molecules scored well in this investigation, however NCI169676 shown the exceptional human oral absorption (73.74%), with the human oral absorption percentage for the NCI tested compounds ranging from 36 to 73%. On the other hand, the number of hydrogen bond donors (n-OHNH, NH, and OH) of the molecules was found to be below 5 (1 to 4), which is compatible, and the number of hydrogen bond acceptors (n-ON, O, and N atoms) was less than 10 (under the Lipinski limit), which indicated good binding strength of molecules with CKD7 receptor binding pocket (Table S-III).

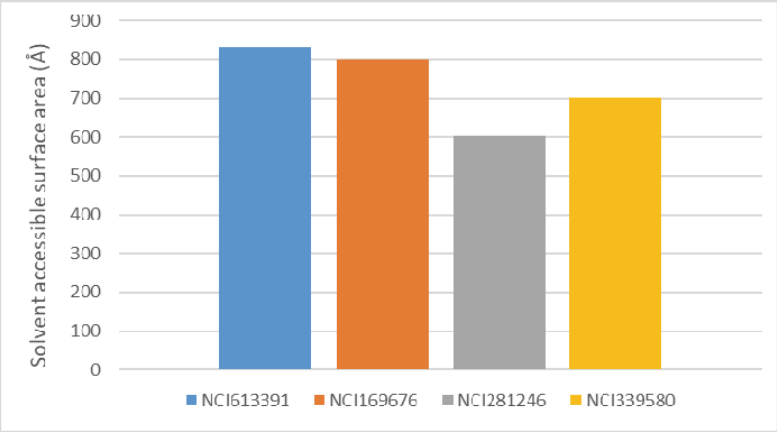


Fig. S-3. The solvent accessible surface area of NCI screened ten best finding compounds.

The partition coefficient (QPlogPo/w) and water solubility (QPlogS) of NCI-screened compounds were also found to be within the range limits of 0.8 to 3.729 and -2.005 to -3.647, respectively. This is a sign of a drug's interaction with cell membranes, as well as its absorption and distribution in the body. It was also discovered that all molecules' projected water-octanol coefficient fell within the permitted range (16 to 26), indicating excellent permeability through cell membranes. The SASA of NCI-screened compounds was found to be between 604 and 832 under the limit (300-1000) in the current investigation, indicating good oral bioavailability (Table S3 and Fig. S3). Additionally, between 9 and 17, within the permissible range (4.0 - 45.0), the skin permeability (QPlogPw) of compounds subjected to NCI screening was discovered. Due to the fact that all NCI tested compounds satisfied the pharmacokinetic characteristics within the permissible range of Lipinski's guidelines set for human usage, this suggests that they have the potential to be drug-like molecules.

Table S-IV. Drug-like properties analysis of the NCI-screened compounds.

Compd. Id	QP polrz	QPlogP oct	QPlogP o/w	QPlog PC16	QPlogP w	QPlog Kp	CIQ PlogS	QPlog Khsa
NCI613391	45.882	26.284	1.74	16.161	17.178	-7.207	-2.858	-0.064
NCI169676	47.868	23.372	3.729	14.099	12.183	-5.387	-3.647	0.576
NCI281246	32.904	19.887	0.8	11.796	14.027	-7.577	-2.005	-0.157
NCI339580	37.417	22.427	0.762	13.763	15.798	-6.289	-3.208	-0.45

The polarizability value is predicted in cubic angstroms (Å), QPplorz; Predicted in octanol/gas partition coefficient (the permissible scale in between 13.0 – 70.0), QPlogPoct; Predicted in octanol/water partition coefficient (permissible scale in between 8.0 – 35.0), QplogPo/w; Predicted in hexadecane/gas partition coefficient (permissible scale is: –2.0 – 6.5), QPlogPC16; Predicted in water/gas partition coefficient (permissible scale is: 4.0 – 18.0), QPlogPw; Predicted in skin permeability (permissible scale is: 4.0 – 45.0), QPlogKp;

Prediction of binding to human serum albumin (log *K*p;. permissible scale is: –8.0 to 1.0), CIQPlogS; permissible scale is: –6.5 to 0.5), QPlogKhsa; Conformation-independent predicted aqueous solubility (permissible scale is: –1.5 to 1.5).

Prime MM/GBSA module for Drug-Target Binding Energy Assessment

Prime MM/GBSA module for Drug-Target Binding Energy Assessment of NCI Screened Molecules

The NCI-screened medication with the target CDK7 kinase protein was scored using the MM/GBSA-based rescoring method. The NCI compounds that were screened, notably NCI613391 (complex with PDB ID: 1UA2), exhibit a significant amount of binding free energy (G bind = -73.25 kJ/mol). The interaction between the ligand and receptor molecule was shown by the complex's binding free energy using a simulation-based methodology. Table S5 lists the specifics and remaining substances that bind free energy as complexes. Finding the protein-ligand binding affinity can be made much easier with the use of the main MM/GBSA module. The protein-ligand complexes that were chosen had this module added to them. A surface generalized born solvation model for polar solvation (GSGB), OPLS molecular mechanics energies (EMM), and a nonpolar solvation term (GNP) are all included in it. Total binding free energy is calculated using the equation: ΔG bind = G complex - (G protein + G ligand), where ΔG bind is the total binding free energy of the complex, G complex is the total energy of the complex, G protein is the energy of the receptor without ligand, and G ligand is the energy of the unbound ligand, and G = EMM+ GSGB+ GNP. In Fig. S3, the projected binding energy of the best-finding compounds from the NCI-screened data base is shown.

Table S-V. Prime MM/GBSA calculated binding energy in kcal/mol of NCI screened compounds.

Compd Id[a]	Gevdw[b] (kJ/mol)	GsolLipo[c] (kJ/mol)	Gcoul[d] (kJ/mol)	Gcovalent[e] (kJ/mol)	GsolGB[f] (kJ/mol)	ΔG bind[g] (kJ/mol)
NCI613391	-53.19	-23.18	-101.45	4.674	103.23	-73.25
NCI169676	-42.96	-28.13	-95.97	5.51	101.15	-64.36
NCI281246	-38.22	-21.39	-73.34	1.89	79.28	-53.57
NCI339580	-47.92	-18.42	-72.26	8.92	73.29	-59.84

Compd Id[a]; Compound Id, Gevdw[b]; Van der Waal energy, GsolLipo[c]; lipophilic energy, Gcoul[d]; Coulomb energy, Gcovalent[e]; Covalent energy, GsolGB[f]; Generalized born electro-static solvation energy, ΔG bind[g]; Free binding energy.

Molecular mechanics generalized born surface area (MM-GBSA) algorithm was used to calculate the free binding energy (ΔG) of the receptor-ligand system. The best finding compounds from the NCI screening were found to have free binding energies ranging from -73.25 to -53.57 kJ/mol. Among all the molecules, the ligand (NCI613391) complex had the largest negative binding free energy (-73.25 kJ/mol), while NCI281246 had the lowest (-53.57 kJ/mol). Despite having

a higher hydrogen occupancy than other compounds, compound NCI613391 complex showed a more stable engagement in the active free binding site of receptor. The top-scoring NCI screened compounds, such as NCI613391, had average ΔG values of -73.25 kJ/mol, respectively. Surprisingly, compound NCI613391 was found to have superior binding affinity to the CDK7 receptor binding site when compared to other substances. Compounds' ΔG values were further broken down into their component parts, which are shown in Table S5 and Fig. S4. Van der Waals interactions have a significant role in the binding of substances to the CDK7 receptor in the disintegration component of ΔG, followed by electrostatic interactions and SASA energy. Furthermore, the polar solvation energy of the compounds found in the NCI screening process favourably influenced the binding site and resisted the formation of the complex.

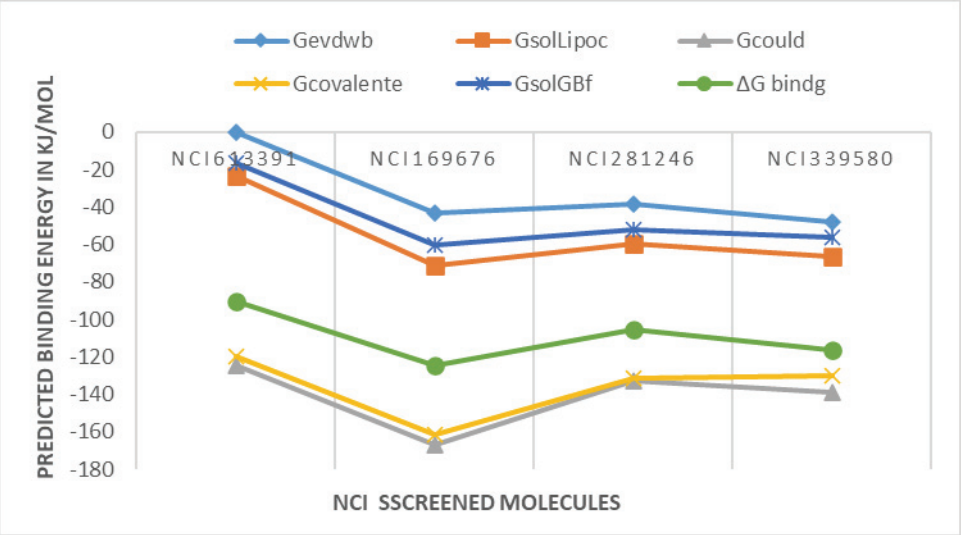


Fig. S-4. The predicted binding energy MM/GBSA calculation of best finding NCI screened molecules.

Structural analysis of NCI Screened Ligands by Molecular Dynamic Simulation (MD)

The importance of molecular dynamics simulation in computer-assisted drug design is for the assessment of dynamic and thermodynamic parameters of biological systems under specific circumstances involving physiological factors. The NCI compound NCI613391 (-13.06 kJ/mol) that was evaluated for molecular docking analysis was found to have the greatest docking scores and affinities with the target CDK7 kinase protein, therefore it was selected for the additional analysis using MD simulation. In order to evaluate the stability of the ligand-protein complex as well as the primary intermolecular interactions over the simulation period, an MD simulation study was conducted on the best docked

pose of a chosen NCI chemical in complex with the target CDK7 kinase protein. The radius of gyration (rGyr) during MD simulation was used to determine the target protein's structurally compact residues. As can be seen, there was no discernible variation in the rGyr readings, and the compactness persisted for the entire 200 ns. Observing the variations in the examined complex's molecular surface area (MolSA), polar surface area (PSA), and solvent accessible surface area (SASA) further suggested the presence of stable ligand-protein complexes.

The importance of Molecular Dynamic simulation in computer-aided drug design is in the estimation of dynamic and thermodynamic parameters of living systems under specific situations involving physiological parameters. In MM/GBSA computation, the screened NCI compounds NCI613391 (-13.06 kJ/mol) was found to have the highest docking score and affinity with the target CDK7 kinase protein, so these were chosen for further analysis by MD simulation. An MD simulation study was performed on the best docked pose of a selected NCI compound in complex with the target CDK7 kinase protein in order to investigate the stability of the ligand-protein complex as well as the main intermolecular interactions during the simulation period. The simulated complex was inspected for different standard simulation parameters such as backbone root-mean-square deviations (RMSDs) for all Ca carbon atoms of protein, the root means square fluctuations (RMSFs) of individual amino acid residues, intermolecular interactions involved, radius of gyration (rGyr) and hydrogen bond analysis. However, minor fluctuation was visible in both proteins during the initial period, which then stabilized for the remainder of the simulation period. A system showing steady fluctuations is usually considered stable and deemed to be properly equilibrated. Stable RMSF plots were obtained during the entire simulation period.

The residues of structural compactness of the target protein were ascertained by the radius of gyration (rGyr) during MD simulation. As can be seen, no significant deviation was noticed in the values of rGyr and the compactness was maintained throughout the 200 ns. In addition, the existence of stable ligand-protein complexes was also implied by observing the changes in various surface areas, such as molecular surface area (MolSA), polar surface area (PSA), and solvent accessible surface area (SASA) of the studied complex. The structural disturbances and real movement of a receptor and ligand complex under possible biological settings are simulated via MD simulation. During the MD simulation analysis, we confirmed the docking orientations using the RMSD, a well-established indicator for protein stability as well as equilibrating, and the averaged behavioural pattern for the complete protein structure could be seen.

The RMSD (Root-Mean-Square Deviation)

Fig. 5-A displays the RMSDs of the free CDK7 kinase receptor and the CDK7+NCI613391 complexes. For the complexes CDK7+NCI613391, the

RMSD values of the analyzed parameters gradually increased from 0 to 50 ns and remained stable after reaching equilibration during simulation duration. The complex CDK7+NCI613391 was shown lower RMSD value and found equal to the receptor CDK7 Kinase protein. These drug candidates complex naming CDK7+NCI613391 was found stable in active residue of CDK7 Kinase receptor and provide a solid foundation for research work and illustrated in Fig. 5-A.

The RMSF (Root-Mean-Square fluctuations)

The RMSF value of backbone atoms was determined, and the RMSF plot indicates that the variations are at the atom scale. The atoms' fluctuations data was found to be substantially equal to the free receptor (CDK7 kinase protein), and the complexes (CDK7+NCI613391, was found to have low RMSF values. Thus, the RMSF plot revealed that the binding interactions of all the tested drugs were observed to be stable within the receptors, and that there was no significant effect on the receptor's flexibility during the simulation analysis study. The RMSF values of CDK7 kinase alone and in combination with NCI613391, was displayed in Fig. 5-B.

The Radius of Gyration (Rg)

During the molecular simulation, the radius of gyration (Rg) is used to estimate the compactness of a protein; if the protein is very compact, it will not fold easily and the complex will be stable in folded or unfolded forms. The Rg plot revealed that the CDK7 kinase receptor, CDK7+NCI613391 complex have similar characteristics. In comparison to the protein complex, the Rg values of protein and protein-drug candidate complexes demonstrated a significant resemblance and a stable folded structure, as illustrated in Fig. 5-C.

The Hydrogen Bond Analysis

MD simulations of all complicated trajectories were examined for the number of hydrogen bonds generated. During MD simulations, it has been shown that the CDK7+NCI613391, complexes formed more hydrogen bonds, as illustrated in Fig. 5-D. Thus, throughout the simulation time, NCI613391 was able to sustain a robust contact with the binding pocket of CDK7 kinase. The ligand (NCI613391) was shown maximum hydrogen bond occupancy with CDK7 residues protein in overall simulation.

The Physicochemical Properties Prediction

All the compounds did not violate Lipinski's rule of five or Ghose filter, except NCI613391 and NCI281246 (only one violation detected), which showed 0-1 violations and are expected to be orally active. Molecular hydrophobicity or lipophilicity is indicated by the logP or partition coefficient. The logP values of all compounds were found to be less than 5 and were not in violation of Lipinski’s rule of five or the Ghose filter, suggesting good permeability across the cell membrane. The molecular weight of compounds was found to be less

than 500, and thus these molecules are anticipated to be easily transported, diffused, and absorbed in comparison to large molecules. The numbers of hydrogen bond donors (NH and OH) in the compounds were found in accordance with Lipinski's rule of five, which is less than 5. TPSA of the compound was observed in the range of 43.66–130.32Å, which is below the limit of 160Å. The percentages of absorption for the compounds calculated from TPSA were in the range of 64.04–93.94 and indicated good oral bioavailability (Table S6). The molar refractivity of compounds was found in the range 100.27–136.82 except for compounds NCI613391 and NCI169676 (131.46 and 136.82, respectively), and polar surface area was found in the range 34.14–37.38 that is obeyed by the Ghose filter.

Table S-VI: Physicochemical characteristics of NCI-screened compounds by Lipinski's rule of five.

Compd. ID	Abs %	miLogP (o/w)	TPSA (Å^2)	n atoms	n-ON	n-OHNH	n-violation	n-rotb	MV (cm3/m)
RO5		<5			<10	<5	≤1		
NCI613391	66.66	2.05	122.71	32	8	6	1	12	415.46
NCI169676	93.94	3.94	43.66	33	5	2	0	10	426.79
NCI281246	70.97	0.90	110.24	24	6	6	1	6	297.17
NCI339580	64.04	1.96	130.32	29	8	5	0	9	357.45

%Abs, TPSA, n atoms, n-rotb, and MV are the abbreviations for percentage of absorption, Topological polar surface area, number of atoms, number of rotatable bonds, and molecular volume (cm3/mol) respectively. n-OHNH is the number of hydrogen bond donors, n-ON is the number of hydrogen bond acceptors, and miLogP is the logarithm of the partition coefficient between n-octanol and water. n violations are number of violations of Rule of Five and RO5 is Rule of five.

Table S-VII: Ghose filter foresaw drug-like characteristics of NCI-screened compounds.

Compd.	Log-P (log10)	MR (m3/mol)	The number of atoms	PSA (Å^2 molecule-1)	C-Indication
The Ghose filter	-0.4-5.6	40-130	20-70	<140	Green
NCI613391	-0.581	131.46	32	34.14	Pink
NCI169676	1.877	136.82	33	35.02	Pink
NCI281246	-0.82	100.27	24	34.14	Pink
NCI339580	-0.179	112.77	29	37.38	Green

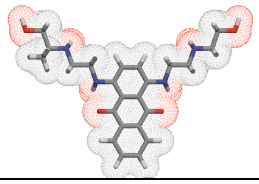
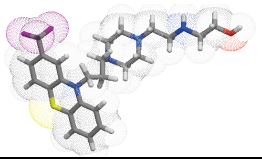
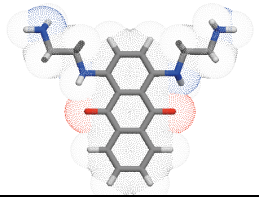
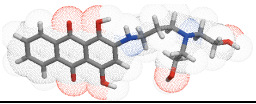
Log-P: Partition coefficient; MR: Molar Refractivity; PSA: Polar Surface Area; C- Indication: Color Indication.

The score of Bioactivity

Compound NCI613391 also demonstrated excellent binding with all targets other than nuclear receptor ligand (moderate activity). All compounds were found to have outstanding activity scores against GPCR ligand affinity, ion channel modulator, kinase inhibitor, and enzyme inhibitor (moderate activity).

Entire compounds demonstrated outstanding activity scores against protease inhibitor activity. Compound NCI613391 was found to have better ligands for protease inhibitors, nuclear receptor ligands, GPCR ligands, kinase inhibitors, enzyme inhibitors, ion channel modulators, and nuclear receptor ligands (Table S8) than the other examined compounds. The most promising compounds, including NCI169676, NCI281246 and NCI339580 which are projected to function across more than three pathways, were identified using the bioactivity ratings (Table S8). If a chemical's bioactivity score is greater than 0.00, it is thought to have significant biological activity; values between -0.50 and 0.00 are seen to be moderately active; and values less than -0.50 are thought to be inert.

Table S-VIII: Bioactivity score rating of NCI screened compounds against the various targets.

Compd. ID	Structure 3D	GPCR ligand	Ion channel modulator	Kinase inhibitor	Nuclear receptor ligand	Protease inhibitor	Enzyme inhibitor
NCI613391		0.15	0.05	0.15	-0.27	0.21	0.09
NCI169676		0.24	0.15	0.05	0.06	0.03	0.13
NCI281246		0.10	0.02	0.29	-0.31	0.15	0.19
NCI339580		0.06	0.00	0.20	-0.02	0.04	0.21

GPCR stands for "G-protein coupled receptor," which is active >0, moderately active -5.0-0.0, and inactive -5.0.

The Osiris Property Explorer

This program predicts the properties of investigated compounds on the basis of functional group similarity with the extensively in-vitro and in-vivo studied compounds present in its database. The obtained results clearly showed that the compound NCI613391, which showed green colour with Tumorigenic effect and Reproductive effect would be safe. All the tested compounds showed green colour indication with tumorigenic effect except NCI339580, and reproductive effect except NCI169676. Most of the compounds were observed to have an irritant effect, except NCI169676 (Table S9).

Table S-IX: Osiris property explorer toxicity and drug-relevant properties of NCI screened compounds.

Compd. ID	The Prediction of Toxicities				The Drug-relevant properties			
	Tumorigenic effect	Reproductive effect	Irritant effect	Mutagenicity	cLogP	Solubility	Drug-likeness	Drug Score
NCI613391	L	L	H	H	0.89	-4.49	-1.77	0.17
NCI169676	L	H	L	H	3.84	-3.99	3.54	0.24
NCI281246	L	L	H	H	0.49	-4.59	-5.12	0.17
NCI339580	H	L	H	H	1.10	-3.48	-1.71	0.11

L - low or no risk, H - high toxicity risk.

The Prediction of ADME/T Properties by pre-ADMET server

The NCI-screened compounds showed negative AMES toxicity results, indicating that the majority of compounds had little mutagenic potential. All compounds displayed negative carcinogenicity in rats as well as mice. Pharmacokinetic parameters like absorption, bioavailability, and distribution of NCI screened compounds were predicted by pre-ADMET tool through a type of test (Table S10). The obtained result from the pre-ADMET server revealed that the compounds NCI169676 showed excellent human intestinal absorption (HIA) scores of more than 90% in comparison to other compounds. The compounds NCI613391, NCI281246, NCI339580, were observed to have a better human intestinal absorption (HIA) score (Table S11) than the rest. The highest HIA score suggests that when these compounds are given orally, the substance may be better absorbed from the intestinal system. Entire compounds were found to have a low value of blood-brain barrier (BBB) penetration (<1). The skin permeability value in cm/hour (logKp) was also found within the limit range of NCI screened compounds. The compound (NCI613391) was found to have excellent HIA (76.08%), cell permeability by Caco-2 (20.65) and MDCK (0.06), solubility in

water (31.38) and buffer (0.27) in mg/mL, and a blood-brain barrier (BBB) penetration value of 0.05 (Table S11).

Table S-X: Pre ADMET toxicity prediction by Ames's test of NCI-screened compounds.

Compd. ID	The prediction of toxicity	Test Name	Test Values
NCI613391	Ames's test	+S9 (Ames TA100)	-
		-S9 (Ames TA100)	-
		+S9 (Ames TA1535)	-
		-S9 (Ames TA1535)	+
	Carcinogenicity	Mouse	-
		Rat	-
NCI169676	Ames's test	+S9 (Ames TA100)	+
		-S9 (Ames TA100)	-
		+S9 (Ames TA1535)	-
		-S9 (Ames TA1535)	-
	Carcinogenicity	Mouse	-
		Rat	-
NCI281246	Ames's test	+S9 (Ames TA100)	-
		-S9 (Ames TA100)	-
		+S9 (Ames TA1535)	-
		-S9 (Ames TA1535)	+
	Carcinogenicity	Mouse	-
		Rat	-
NCI339580	Ames's test	+S9 (Ames TA100)	-
		-S9 (Ames TA100)	-
		+S9 (Ames TA1535)	-
		-S9 (Ames TA1535)	-
	Carcinogenicity	Mouse	-
		Rat	-

-: Negative; +: Positive

Table S-XI: Pharmacokinetic parameters value of NCI screened compounds by pre-ADMET tool.

Compd. ID	Parameters	Test name for the prediction of ADME	Test values
NCI613391	Absorption	The % of Human intestinal abs. (HIA)	76.083419
		In-vitro cell perm. by Caco-2 in nm/sec	20.6528
		In-vitro cell perm. by MDCK in nm/sec	0.0642523
		In-vitro skin perm. in cm/h (logKp)	-4.31999
	Bioavailability	The solubility in buffer in mg/mL	0.271736
		The solubility in pure water in mg/mL	31.3839
	Distribution	*In-vitro* the % of plasma protein binding	28.655152
		In-vivo blood brain barrier (BBB) penetration (C. brain/C. blood)	0.0568443
NCI169676	Absorption	The % of Human intestinal abs. (HIA)	95.854664
		In-vitro cell perm. by Caco-2 in nm/sec	24.2669
		In-vitro cell perm. by MDCK in nm/sec	0.130145
		In-vitro skin perm. in cm/h (logKp)	-2.42092
	Bioavailability	The solubility in buffer in mg/mL	166.508
		The solubility in pure water in mg/mL	18.1855
	Distribution	*In-vitro* the % of plasma protein binding	51.308648
		In-vivo blood brain barrier (BBB) penetration (C. brain/C. blood)	0.989341
NCI281246	Absorption	The % of Human intestinal abs. (HIA)	88.329084
		In-vitro cell perm. by Caco-2 in nm/sec	21.0049
		In-vitro cell perm. by MDCK in nm/sec	2.95133
		In-vitro skin perm. in cm/h (logKp)	-4.59474
	Bioavailability	The solubility in buffer in mg/mL	0.768911
		The solubility in pure water in mg/mL	1.64088
	Distribution	*In-vitro* the % of plasma protein binding	52.762220
		In-vivo blood brain barrier (BBB) penetration (C. brain/C. blood)	0.0338409
NCI339580	Absorption	The % of Human intestinal abs. (HIA)	75.723043
		In-vitro cell perm. by Caco-2 in nm/sec	19.3925
		In-vitro cell perm. by MDCK in nm/sec	0.318305
		In-vitro skin perm. in cm/h (logKp)	-4.53595
	Bioavailability	The solubility in buffer in mg/mL	134.576
		The solubility in pure water in mg/mL	494.614
	Distribution	*In-vitro* the % of plasma protein binding	47.437082
		In-vivo blood brain barrier (BBB) penetration (C. brain/C. blood)	0.0633325

The Acute Toxicity Prediction by T.E.S.T. (US-EPA) web-based server

The acute toxicity of NCI screened compounds was predicted by the T.E.S.T. (US-EPA) web-based server by different approaches such as consensus, hierarchical clustering, United States food and drug administration (US-FDA),

and closest neighbour method. The obtained results were found within in the limit under the different approaches like the consensus method predicted an oral rat LD50 in the range of 2.10-4.33 (-Log10 (mol/kg), the hierarchical clustering method predicted an oral rat LD50 in the range of 1.94-2.36 (-Log10 (mol/kg), the FDA method predicted an oral rat LD50 in the range of 1.75-4.81 (-Log10 (mol/kg), and the nearest neighbour method predicted an oral rat LD50 in the range of 2.04-4.34. The consensus approach, Hierarchical clustering, FDA and Nearest neighbour determined that the drugs' mutagenic values fell within the ranges of 0.53-0.95, 0.66-.81, 0.53-0.97, and 0.33-0.67, whereas the FDA method and the nearest neighbour technique and other toxicity outcomes are shown in Table S12.

Table S-XII: Results of projected toxicities of NCI screened compounds from the T.E.S.T. (US-EPA).

Methods	Toxicity parameters	Toxicity results of NCI screened compounds			
		1	2	3	4
Consensus	LD50 value (Oral rat) in -log10 (mol/kg)	2.42	4.33	2.10	2.28
	The value of Mutagenicity	0.77	0.53	0.95	0.86
	Bioaccumulation factor Log10	0.37	N/A	0.75	0.22
	The value of developmental toxicity	0.80	1.07	0.64	0.79
Hierarchical clustering	LD50 value (Oral rat) in -log10 (mol/kg)	2.36	3.83	1.94	2.35
	The value of Mutagenicity	0.66	0.72	0.90	0.81
	Bioaccumulation factor Log10	0.71	N/A	0.64	0.65
	The value of developmental toxicity	0.67	1.07	0.83	0.77
FDA	LD50 value (Oral rat) in -log10 (mol/kg)	2.79	4.81	2.31	1.98
	The value of Mutagenicity	0.97	0.53	0.96	0.77
	Bioaccumulation factor Log10	1.18	N/A	N/A	0.52
	The value of developmental toxicity	1.01	0.96	0.23	0.68
Nearest neighbor	LD50 value (Oral rat) in -log10 (mol/kg)	2.12	4.35	2.04	2.50
	The value of Mutagenicity	0.67	0.33	1.00	1.00
	Bioaccumulation factor Log10	1.23	1.58	2.22	0.78
	The value of developmental toxicity	0.67	1.00	N/A	1.00

NCI screened compounds, 1 (NCI613391), 2 (NCI169676), 3 (NCI281246) and 4 (NCI339580)

The Prediction of Metabolism by pre-ADMET web server

Table S-XIII: Inhibition and substrate of NCI-screened compounds to cytochrome P450 enzymes and p-glycoprotein.

Compd. ID	Inh. of CYP450 2C19	Inh. of CYP450 2C9	Inh. of CYP450 2D6	Subst. of CYP450 2D6	Inh. Of CYP450 3A4	Subst. of CYP450 3A4	Inhibitor P-gp
NCI613391	NI	I	NI	S	NI	WS	NI
NCI169676	I	NI	I	S	NI	NS	I
NCI281246	NI	I	NI	S	NI	WS	NI
NCI339580	NI	I	NI	S	I	WS	NI

Cytochrome P450 enzyme (CYP450) & P-glycoprotein (P-gp), NI - Non-inhibitor, I - inhibitor, S – substrate, WS – weakly substrate, NS – non-substrate

Drug likeness and Rule Violation Prediction of NCI Screened Compounds

Drug-like characteristics of NCI-screened compounds were estimated by various drug-like rules, like Lipinski's rule, the Ghose filter, the lead-like rule, CMC, WDI, and MDDR, on a single platform that is a pre-ADMET web tool. The NCI screened compounds were found to have drug likeness as per the MDDR like rule, and zero violations were observed. The NCI screened compounds according to Lipinski's rule of five and found them suitable without crossing the violation. All compounds qualified for the CMC rule except compound NCI169676 (just one violation). The compound (NCI613391) was qualified for the CMC rule, suitable as Lipinski's rule, and to have drug-like properties by the MDDR rule with zero violation (Table S14). The well-known rule for the prediction of drug likeness of newly designed molecules is Lipinski's rule, usually called the rule of five, but another well-known rule, the lead-like rule, was developed from a quantitative survey based upon 18 lead and drug pairs of chemical structure.

Table S-XIV: Drug likeness and violation prediction for NCI-screened compounds by different rules.

Compd. ID	CMC Rule		Rule like Lead		Rule like MDDR		Lipinski's rule		Rule like WDI	
	R	nV	R	nV	R	nV	R	nV	R	nV
NCI613391	Q	0	V	1	DL	0	S	1	Out of 90% cutoff	3
NCI169676	NQ	1	V	2	DL	0	S	0	Out of 90% cut off	4
NCI281246	Q	0	V	1	MS	1	S	0	In 90% cutoff	0
NCI339580	Q	0	V	1	DL	0	S	1	Out of 90% cutoff	1

R – rule; nV – number of violations; Q – qualified; NQ – not qualified; S – suitable; V – Violated, MS – mid-structure; DL – drug like;