Briefings in Bioinformatics, 00(00), 2021, 1–15

doi: 10.1093/bib/bbab098
Case Study

Pharmacoinformatics and molecular dynamics simulation-based phytochemical screening of neem plant (*Azadiractha indica*) against human cancer by targeting MCM7 protein

Foysal Ahammad, Rahat Alam†, Rasel Mahmud†, Shahina Akhter, Enamul Kabir Talukder, Al Mahmud Tonmoy, Salman Fahim, Khalid Al-Ghamdi, Abdus Samad and Ishtiaq Qadri

Corresponding author: Abdus Samad, Laboratory of Computational Biology, Biological Solution Centre (BioSol Centre), Jashore-7408, Bangladesh. E-mail: kazisamad50@gmail.com; Ishtiaq Qadri, Department of Biological Science, Faculty of science, King Abdul-Aziz University, Jeddah-21589, Saudi Arabia. E-mail: ishtiaq80262@yahoo.com

†These authors contributed equally to this work.

Abstract

Minichromosome maintenance complex component 7 (MCM7) belongs to the minichromosome maintenance family that is important for the initiation of eukaryotic DNA replication. Overexpression of the MCM7 protein is relative to cellular proliferation and responsible for aggressive malignancy in various cancers. Mechanistically, inhibition of MCM7 significantly reduces the cellular proliferation associated with cancer. To date, no effective small molecular candidate has been identified that can block the progression of cancer induced by the MCM7 protein. Therefore, the study has been designed to identify small molecular-like natural drug candidates against aggressive malignancy associated with various cancers by targeting

Foysal Ahammad is a research assistant of the Department of Biological Science, Faculty of Science at King Abdul-Aziz University, Saudi Arabia. He is the chief executive officer (CEO) of Biological Solution Centre (BioSol Centre). His research mainly focusses on bioinformatics, microbiology, cancer-informatics, molecular biology, system biology, molecular modeling, natural products, vaccine design and computer-aided drug design.
Rahat Alam is a researcher and undergraduate student at the Department of Genetic Engineering and Biotechnology, Jashore University of Science and Technology, Bangladesh. His research interest focuses on molecular modeling as well as computer-aided drug and vaccine design, cancer-informatics and dynamics simulation.
Rasel Mahmud is an undergraduate student at the pharmacy, Mawlana Bhashani Science and Technology University, Bangladesh. His research interest focuses on molecular modeling as well as computer-aided drug and molecular dynamics simulation.
Dr. Shahina Akhter is an Assistant Professor at the Department of Biochemistry and Biotechnology, University of Science and Technology Chittagong, Bangladesh. Her research interest focuses on neurodegenerative, antimicrobial, chemoinformatics, molecular modeling as well as computer-aided drug and molecular dynamics simulation.
Md. Enamul Kabir Talukder is a research and undergraduate student at the Department of Genetic Engineering and Biotechnology, Jashore University of Science and Technology, Bangladesh. His research interest focuses on molecular modeling as well as computer-aided drug and vaccine design and molecular docking.
Al Mahmud Tonmoy is a research assistant at BioSol Centre, Bangladesh. His research interest focuses on chemoinformatics, and computer-aided drug. Also worked on animal biotechnology
Md. Salman Fahim is a research assistant at BioSol Centre, Bangladesh. His research interest focuses on virology, chemoinformatics, computer-aided drug and molecular dynamics simulation.
Dr Khalid Al-Ghamdi is the Professor and head of Biological Sciences, Faculty of Science at King Abdul Aziz University, Saudi Arabia. His work mainly focuses on cancer biology, sustainable science, thermoplastic, drug design and many more.
Abdus Samad is researcher at the Department of Genetic Engineering and Biotechnology, Jashore University of Science and Technology, Bangladesh. His research interest includes computer-aided drug, and vaccine design, bioinformatics, cancer-informatics, biomarker studies and molecular modeling.
Dr Ishtiaq Qadri is the Professor and head of virology, Department of Biology, Faculty of Science at King Abdul Aziz University. He is a molecular virologist with interests in the pathogenesis of viral diseases, antiviral developments, bioinformatics and computer-aided drug design. Dr Qadri is investigating the underlying molecular mechanism(s) that govern the onset of viral-induced diseases.
Submitted: 2 February 2021; **Received (in revised form):** 2 March 2021

MCM7 protein. To identify potential compounds against the targeted protein a comprehensive *in silico* drug design including molecular docking, ADME (Absorption, Distribution, Metabolism and Excretion), toxicity, and molecular dynamics (MD) simulation approaches has been applied. Seventy phytochemicals isolated from the neem tree (*Azadiractha indica*) were retrieved and screened against MCM7 protein by using the molecular docking simulation method, where the top four compounds have been chosen for further evaluation based on their binding affinities. Analysis of ADME and toxicity properties reveals the efficacy and safety of the selected four compounds. To validate the stability of the protein–ligand complex structure MD simulations approach has also been performed to the protein–ligand complex structure, which confirmed the stability of the selected three compounds including CAS ID:105377-74-0, CID:12308716 and CID:10505484 to the binding site of the protein. In the study, a comprehensive data screening process has performed based on the docking, ADMET properties, and MD simulation approaches, which found a good value of the selected four compounds against the targeted MCM7 protein and indicates as a promising and effective human anticancer agent.

Key words: Active site; *in silico* screening; *Azadirachta indica*; MCM7; virtual screening; ADMET; MD simulation

Introduction

DNA replication licensing factor MCM7 is one of the members of the minichromosome maintenance (MCM) protein family that assists to form a double trimeric complex with MCM4 and MCM6 [1]. The MCM7 protein is a part of the hexameric (MCM2-7) protein complex and acts as a key component of the pre-replication complex (pre-RC) forms at the origin of replication of a DNA. The pre-RC led to the formation of replication forks and regulated the activity of DNA helicase by utilizing different DNA unwinding enzymes [2]. Helicase activity regulated by the MCM7 protein help to initiate DNA replication resulting in elongation of eukaryotic cells [3]. The assembly of the hexamer (MCM2-7) complex and regulation of helicase activity that initiates DNA replication is controlled in a tightly orchestrated manner during the four (G1, S, G2 and M) different stages of the cell cycle [4, 5].

G1 is an intermediate and the first stage of four phases represents a critical step during cell-cycle progression, where the inactive double hexamer (DH) helicase is loaded sequentially onto double-strand DNA (dsDNA) origin of replication [6]. Before activation of the S phage, the hexamer formation is blocked, which is maintained by the initial holo-helicase Cdc45-Mcm2-7-GINS (CMG) complex (Figure 1) [7]. Later, the DNA with two helicase rings is dimerized through the N termini of MCM2-7 subunits and undergoes separation resulting initiation of bidirectional replication [8]. During the bidirectional replication process, the dimerized MCM2-7 subunits acquired critical target sites for different cyclin-dependent kinase (CDK). CDK4 is an MCM7 associate cyclin D1-dependent kinase enzyme that regulates the binding activity of this protein with different tumor suppressors and overexpression of the protein promotes cancer progression [9, 10]. Moreover, phosphorylation of MCM7 mediated by two splice isoforms of the tyrosine kinase Lyn (p53 and p56 Lyn) promotes the hexamer MCM complex formation and chromatin loading resulting in enhance DNA synthesis and cancer cell proliferation (Figure 1) [11]. High expression of the MCM7 protein is a remarkable biomarker for cervical cancer and prognostic factor for colorectal, lung and ovarian cancer [12, 13]. Therefore, the protein can be utilized as a diagnostic and prognostic tool in the clinical setting for human malignancy.

Medicinal plants contain many highly valuable therapeutic bioactive compounds, which have been found effective against different cancer [14]. These compounds function through different pathways and demonstrate their anticancer ability by inhibiting several proteins involved in the growth and cell division [15]. Natural compounds are a type of chemical compound or substance produced by living organisms or medicinal plants as a rich resource for the discovery of drugs like small compounds [16, 17]. Previously many compounds originated from plant sources like *Taxus brevifolia* and *Cinchona spp.* are being used as anticancer drugs such as Taxol [18]. A quarter of plant-based medicines has been approved by the FDA and the European Medical Agency, which underlines the value of plant-based compounds in the biomedical arena [19]. Nowadays, it is believed that different medicinal plants are strong sources of anticancer compounds.

Neem plant (*Azadirachta indica*) is a member of the Meliaceae family and a well-known medicinal plant in the Indian subcontinent that has been found effective against various diseases [20]. The ethanol extracts of neem leaf have shown antimicrobial, free radical scavenging and anti-inflammatory activities, where the antineoplastic activity at alkaline pH conditions is more precise [14]. It has also been found that neem leaf extract is very effective against the human MDA-MB-231 breast cancer cell line, where 95.7% (percent) cancer cell mortality has been observed at the concentration of 1600 µg/ml extract with a pH 6.2. Additionally, a concentration of 200 µg/ml neem extract with a pH of 7.1 has found a strong cytotoxicity effect [21]. Depending on the antineoplastic, low toxicity and anticancer properties of neem extracts, molecules originating from the plant can be utilized as an excellent source for designing a potential drug against different cancer [20]. However, there are many significant barriers to producing candidates for synthetic drugs by mimicking naturally occurring compounds.

Difficulties in extraction and characterization of compounds, determination of potential targets, and developing successful assays for measuring drug effectiveness, and pharmacokinetics (PK) of the compounds are important barriers for the identification of natural sources as anticancer agents [22]. Recent advances in computational methods for designing a drug have dispelled the barrier and make it easier to characterize compounds, identifying potential drug targets, and screening or repurposing approved drugs against a specific target [23]. Nowadays, computational drug design methods include molecular docking, virtual screening and molecular dynamics (MD) simulation is the most favorite and popular approach to discover and analyze drugs and similar biologically active molecules against a specific target [24, 25]. For this reason, the study utilized computer-aided drug design (CADD) approaches like molecular docking, MD simulation and pharmacoinformatics approaches to identify and screening phytochemical compounds from neem plant to combat MCM7 related cancer.

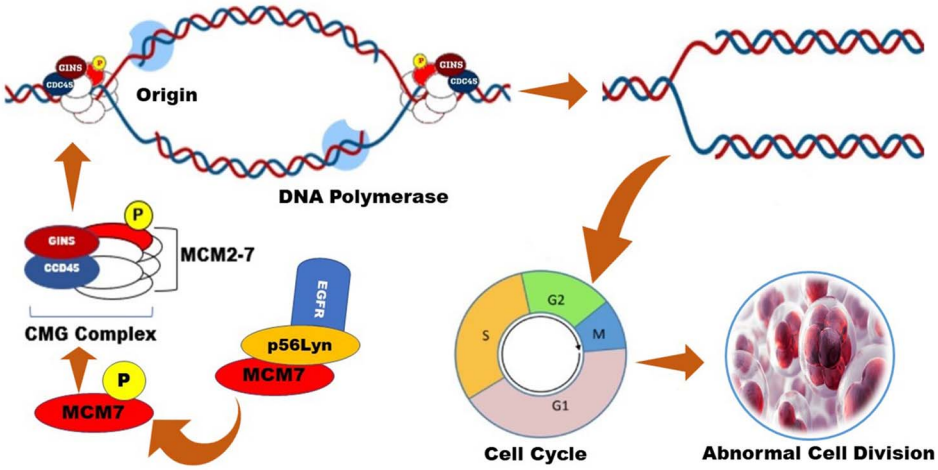


Figure 1. Schematic representation of p56Lyn mediate MCM phosphorylation and their role in cancer development. MCM7 undergo phosphorylation with the help of EGFR-p56Lyn that promotes hexamer MCM complex assembly and chromatin loading resulting in enhance DNA synthesis and development of cancer.

Materials and methods

Protein preparation

Inspection of the RCSB protein data bank (PDB) (https://www.rcsb.org/) revealed 3D experimental tertiary structures of the MCM7 protein in the PDB. The human MCM7 protein having a PDB identity 6XTX consisting of 719 amino acids (AA) length with a predicted molecular weight of 81.3 kDa [26]. The PDB structure of the proteins was retrieved and prepared by using the following criteria such as the water, metal ions and cofactors of the protein were removed, polar hydrogen atoms were added, and nonpolar hydrogen were merged, and gasteiger charges were calculated by using AutoDockTools.

Compounds retrieval and preparation

Phytochemicals from naturally originated medicinal plants encompass diverse chemical spaces for drug design and discovery processes. Indian Medicinal Plants, Phytochemistry, And Therapeutics (IMPPAT) is a manually curated database containing >1742 Indian medicinal plants along with >9500 phytochemicals compounds that utilize cheminformatic approaches to enhance natural product-based drug discovery [27]. From the database, the phytochemical of neem plant (*Azadirachta indica*) has been identified and retrieved for virtual screening. The compounds retrieved from the database were prepared by assigned correct AutoDock 4 atom types, nonpolar hydrogens have merged, aromatic carbons were detected, and 'torsion tree' was set. For most atoms, it has found that the AD4 atom type is the same as the compound's elements.

Active site identification and receptor grid generation

Active site (AS) of an enzyme can be defined as a region of an enzyme-containing a specific shape that allows it to bind with a specific molecular substrate resulting in a chemical reaction of the enzyme [28]. AS ensures the optimum and favorable catalytic microenvironments and helps chemical compounds to form enough contact points to generate strong binding with desired enzymes. Therefore, to obtain a strong binding affinity of our compound the AS of the protein has been determined through BIOVIA Discovery Studio Visualizer v19.1.0.18287(BIOVIA) [24]. The binding site which has been found from the complex AS of the protein has also been determined and utilized for the receptor grid generation process by using PyRx virtual screening tool AutoDock Vina [29].

Molecular docking

Nowadays, molecular docking technique is a key component in structural biology mainly utilized for CADD. The techniques help to predict the best binding mode of a small molecule (e.g. drug and enzyme or protein) to a targeted macromolecule [30]. To identify the binding mode of the desired protein with selected phytochemicals the molecular docking study was carried out by using the PyRx virtual screening tool AutoDock Vina [29]. PyRx is an open-source virtual screening tool mainly utilized for CADD approaches that can screen libraries of compounds against a specific drug target. PyRx includes both AutoDock 4 and AutoDock Vina as a docking wizard with an easy-to-use user interface that makes it a more reliable tool for CADD. This study utilized PyRx tools AutoDock Vina wizard for molecular docking to identify the best binding poses of the protein and ligand. Default configuration parameters of the PyRx virtual screening tools have been used for docking purposes and based on the highest binding energy (kcal/mol) with the negative sign has been chosen for further evaluation. Finally, the binding interaction of the protein–ligands complex has been observed by using the BIOVIA Discovery Studio Visualizer v19.1.0.18287(BIOVIA).

PK properties prediction

PK in CADD refers to the computational characterization of the time course of drug absorption, distribution, metabolism and excretion (ADME). Taken together, PK (ADME) properties are mainly the movement of drugs into, though and out of the body related to the intensity and time course [31]. In the early stages of CADD, pharmacokinetic properties support and define compounds' integrity and efficiency. To evaluate the early-stage PK properties of our selected compounds, SwissADME (www.swissadme.ch) server has been used in the study [32]. The SwissADME server is a web-based free tool that can predict the PK and drug-likeness properties of small molecules.

Toxicity prediction

Toxicity can be evaluated by measuring the degree of the quality of a chemical substance being poisonous to humans or animals and that can damage an organ. It is very important to evaluate the harmful effect of chemicals compounds before undergoing a drug trial. Toxicity evaluation is one of the main and vital steps in the drug design process. Therefore, the toxicity of the selected compounds has been evaluated from admetSAR 2.0 (http://lmmd.ecust.edu.cn/admetsar2/) web-based server [33].

MD simulations

To determine the binding stability of the selected candidate compounds to the desired protein AS cavity the complex structure was evaluated through 50 ns MD simulations [24]. To determine the thermodynamic stability of the receptor-ligand complex the MD simulation of the complex structure has been performed by using 'Desmond v3.6 Program' in Schrödinger (Academic version) under Linux environment [34]. A predetermined TIP3P water model was used to solve the system, where orthorhombic periodic boundary box shape with distance 10 Å has assigned both sides to maintain a specific volume. Suitable ions like Na + and Cl− with a salt concentration of 0.15 M were chosen and placed randomly in the solvated system to neutralize the system electrically. The system was minimized and relaxed using the default protocol implemented within the Desmond module using OPLS-2005 force field parameters after constructing the solvated system containing protein in complex with the ligand [24]. NPT ensembles that utilized the Nose-Hoover temperature coupling and isotropic scaling method were maintained at 300 K and one atmospheric (1.01325 bar) pressure, followed by 50 PS recording intervals with an energy of 1.2.

Simulation trajectory analysis

All the MD simulation snapshots were rendered using Schrodinger's maestro interface v9.5. The quality of the MD simulation has been checked and the Simulation Interaction Diagram (SID) of the Desmond module in the Schrödinger package has been used to analyze the simulation event. Based on the trajectory output, the stability of the complex structure has been evaluated according to the root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), protein–ligand contacts (P–L contact), and hydrogen bond interaction.

RMSD analysis

RMSD in MD simulation is the measurement of average distance produced by displacement of a selected atom during a specific time compared to a reference time [34]. At the very beginning, RMSD of the protein structural atoms like Cα, backbone, sidechain and heavy atoms are counted, then RMSD of the protein fit ligand atoms from all the time frames is aligned and calculated compared to the reference time (in our case 50 ns). An MD simulation with a time frame *x* should have RMSD that can be calculated from the following equation (Eq: i).

$$RMSD_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(r'_i\ (t_x)) - r_i\left(t_{ref}\right)\right)^2} \quad \text{(i)}$$

Here, *N* expresses the number of selected atoms; t_{ref} is the reference or given time, and *r'* expresses the position of the selected atom in the frame *x* after superimposing on the reference frame, tx define the recording intervals.

RMSF analysis

The RMSF is mainly used to characterize and observe the local conformational change within a protein structure [35]. An MD simulation of a protein having the number of residues i, its RMSF value can be calculated by the following equation (Eq: ii).

$$RMSF_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T} < \left(r'_i\ (t)) - r_i\left(t_{ref}\right)\right)^2 >} \quad \text{(ii)}$$

Here, *T* mainly expresses the trajectory time; *tref* is the reference or given time, and *r'* expresses the position of the selected atoms in the frame i after superimposing on the reference frame, and (< >) expressed the average of the square distance taken over residue *b*.

Result of the study

Phytochemical retrieval and preparation

An Indian natural and medicinal phytochemical compound library namely the IMPPAT database has primarily searched to obtain the available compounds of the desired plant [27]. The database identified a list of 70 compounds from the Neem (*Azadirachta indica*) plant listed in Supplementary Table S1. The phytochemical compounds of neem plants have been retrieved and saved in a 2D (SDF) file format. The compounds have been prepared and optimized during the ligand preparation steps and converted into pdbqt file format for further evaluation.

ASs identification and receptor grid generation

AS of an enzyme is a complex form of different AA residues in a specific region that help to form a temporary bond with the substrate known as the binding site [35]. AS allows to make binding with a chemical substrate and undergo a catalyzed reaction, it also helps to stabilize the intermediates of the reaction, where the binding site is a position of a protein or nucleic acid that can recognize ligand and can make a strong binding interaction with the protein. The study first identified AS of the MCM7 then retrieved binding site position from the complex AS pocket.

Analysis of the MCM7 found four AS (AS1, AS2, AS3 and AS4) of the enzymes then the complex AS pocket analysis retrieved binding site residue of the protein (Figure 2). AS1 pocket that was represented in ball shape with red color showed four AA residues (MET369, ARG514, ARG604 and LEU607) as binding sites of the MCM7 protein. AS2 has also found four binding site residues (CYS184, CYS187, CYS206 and CYS211) represented in ball shape with yellow color. For AS3 in yellow color displayed most binding site residues of MCM7 protein with 12 AA residues (GLU343, ILE344, TYR345, PRO383, GLY384, VAL385, ALA386, LYS387, SER388, GLN389, ASN489 and LEU533), where only one residue (SER388) as a binding site formed at the AS4 represented in ball shape with blue color (Figure 2).

Ligand-based molecular docking with the desired protein by using the PyRx is not possible without fixing a receptor grid [29]. Receptor grid generation requires a 'prepared' protein with adequate bond orders and formal charges that have been done in the study at the very first stage of the protein preparation. Grid generation during the molecular docking process provides progressively more reliable scoring of the ligand poses. Therefore, to

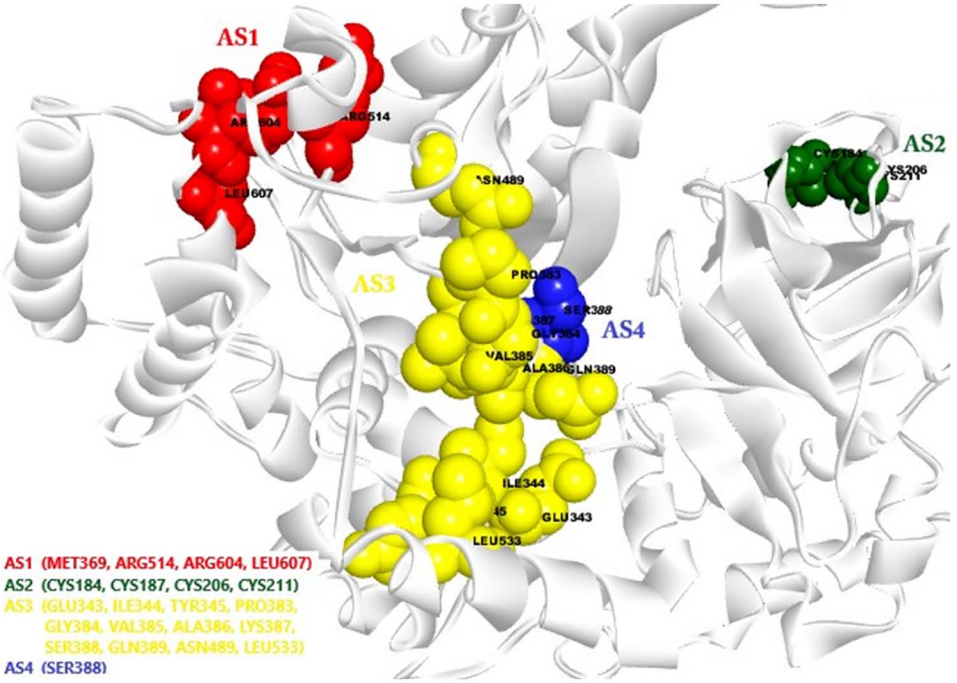


Figure 2. Showing the active site and correspondence binding site of MCM7 protein. Ball shape with red, green, yellow and blue color representing AS1, AS2, AS3 and AS4, respectively with their binding site position of the MCM7 protein.

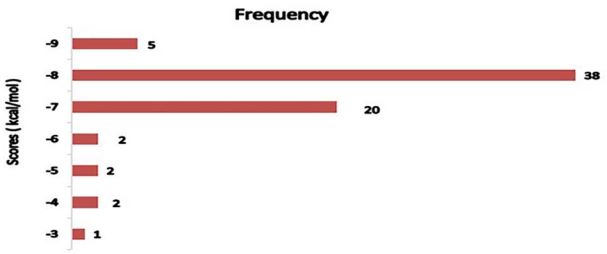


Figure 3. Showing the frequency distribution of 70 phytochemicals neem plant over the range of docking scores.

obtain more accurate scoring of our ligand poses we generated a receptor grid to the MCM7 protein based on the previously obtained binding site residues of the protein. The grid box with a dimension X = 88.89, Y = 38.46 and Z = 43.41 in angstrom (Å) has been identified and used for molecular docking simulation.

Molecular docking analysis

Molecular docking can help to determine the best intermolecular framework formed between a macromolecule and a drug like a small molecular candidate. Initially, a molecular docking study was carried out to screen and determine the best intermolecular interaction between the desired protein and phytochemical compounds. PyRx tools AutoDock Vina wizard has been utilized for carrying out molecular docking between 70 phytochemical compounds and desire protein. The binding affinities found after molecular docking of the phytochemicals compound have shown a distributed range between −3.1 and −9.0 and kcal/mol depicted in Figure 3 and Supplementary Table S1.

Based on the binding affinity top five percentage (%) of 70 phytochemical (total 4) compounds has been chosen, which has a better binding affinity than fluoroquinolone antibiotic ciprofloxacin. In this study, ciprofloxacin has been considered as a control ligand due to previously reported inhibitory activity against MCM7 protein with IC_{50} = 632 µM, which shows a binding affinity of −7.4 kcal/mol. The top four compounds with the highest binding affinity and the binding affinity of ciprofloxacin (control) have shown in Table 1.

Interpretations of protein–ligands interaction

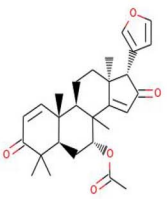
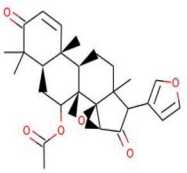
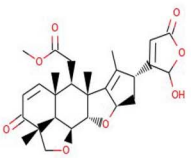
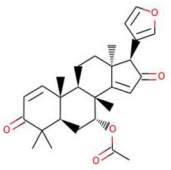
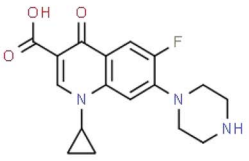
The interaction formed between the selected four ligands with the desired protein has been observed by using the BIOVIA Discovery Studio Visualizer tool. For the compound 156225 (CHEMSPIDER ID) it has been found to form several conventional and carbon–hydrogen bonds with the desire MCM7 protein.

Two conventional hydrogen bonds have been found to form at the position of LYS351 (2.45 Å) and LYS352 (2.55 Å). Four carbon–hydrogen bonds with GLU334 (3.78 Å), ALA338 (3.80 Å) and PRO342 (3.18 Å) have also been observed during the interaction of the compound 156225 (CHEMSPIDER ID). Alkyl and Pi-Alkyl bond has found to form at the position TYR345 (4.64 Å), HIS536 (4.75 Å), and ALA337 (5.05 Å), PRO547 (4.67 Å), respectively showed in Figure 4 and Table 2.

In the case of the compound 122801 (PubChem CID) it has been observed to formed two conventional hydrogen bonds and three carbon–hydrogen bonds with the desired protein. Two conventional hydrogen bonds have noticed at the position of GLY346 (2.01 Å) and LYS351 (2.34 Å), where three carbon–hydrogen bonds have found to form at the position of TYR345 (4.55 Å), ALA338 (3.43 Å) and HIS536 (3.58 Å). One pi–pi T-shaped bond at HIS536 (4.65 Å) and two Pi-Alkyl bonds has also formed at the position TYR345 (4.65 Å), and PRO547 (4.31 Å) showed in Figure 5 and Table 2.

The interaction study of the compound 105377-74-0 (CAS ID) was determined that found one conventional hydrogen bond at the position of ARG532 (2.75 Å) and two carbon–hydrogen bonds

Table 1. List of compound identity, chemical name, two-dimensional (2D) structure of selected best four ligands, and ciprofloxacin (control) with their binding affinity

No	Compound ID	Chemical formula	2D Structure	Score (Kcal/mol)
1	CHEMSPIDER: 156225	(5alpha,7alpha,8eta,13alpha ,17alpha)-17-(3-Furyl)-4,4,8-trimethyl-3,16-dioxoandrosta-1,14-dien-7-yl acetate		-8.8
2	PubChem CID:122801	Epoxyazadiradione		-8.4
3	CASID:105377-74-0	isomargosinolide		-8.2
4	PubChem CID:12308716	17-Epiazadiradione		-8.1
5	PubChem CID:2764	Ciprofloxacin		-7.4

at ALA338 (3.61 Å) and HIS536 (3.45 Å) residual position. It also formed two Alkyl interactions with PRO 342, TYR 345 and PRO 547, where pi-Alkyl bonds have found to form at ALA338 (4.50 Å) and HIS536 (4.46 Å) position depicted in Figure 6 and Table 2.

For the compound 12308716 (PubChem CID) the conventional hydrogen bonds were found to form only for the LYS351 (2.55 Å) position, where pi–Alkyl and Alkyl bonds have predominantly formed. At the position of ARG532, TYR345, HIS536 and PRO547 both pi–Alkyl and Alkyl bonds has observed, where distance for the pi–Alkyl bond was 4.19 Å, 4.10 Å, 5.23 Å, 4.84 Å, and distance for Alkyl bonds has observed 4.73 Å, 4.56 Å, 4.47 Å, 5.23 Å, respectively showed in Figure 7 and Table 2.

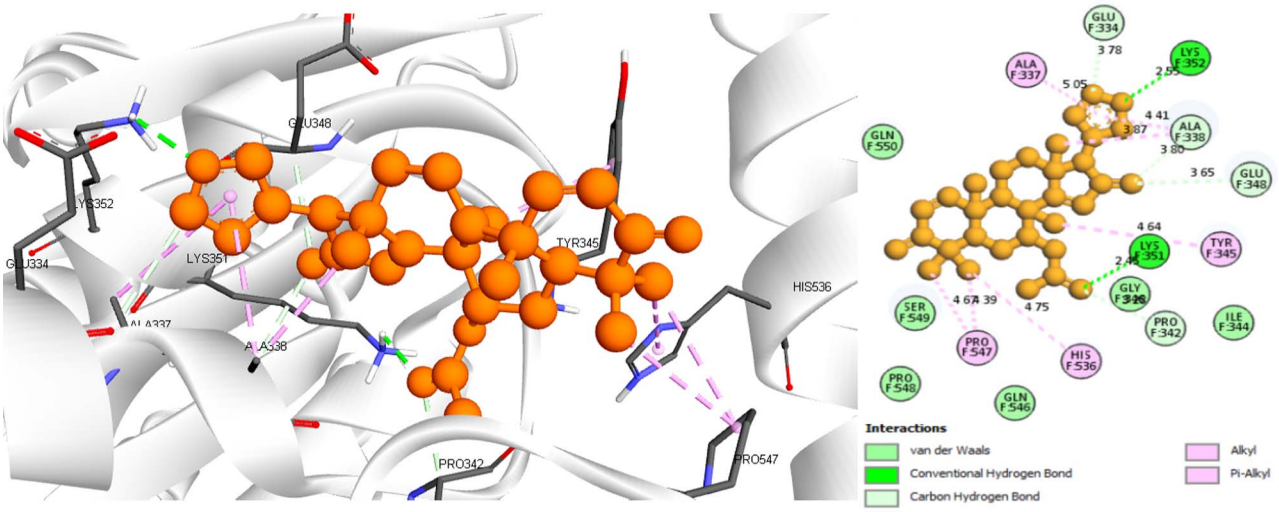


Figure 4. Depicted the interaction between the compound 156225 (CHEMSPIDER ID) and MCM7 protein. Left side representing 3D and the right side representing 2D complex protein–ligand interaction.

Table 2. List of bonding interactions between selected four phytochemical with MCM7 protein

ID	Residues	Distance (Å)	Bond category	Bond type
CHEMSPIDER:156225	LYS351	2.45	Hydrogen	Conventional hydrogen
	LYS352	2.54	Hydrogen	Conventional hydrogen
	ALA338	3.79	Hydrogen	Carbon hydrogen
	PRO342	3.27	Hydrogen	Carbon hydrogen
	GLU348	3.64	Hydrogen	Carbon hydrogen
	GLU334	3.02	Hydrogen	Carbon hydrogen
	TYR345	4.63	Hydrophobic	Pi-Alkyl
	ALA337	5.04	Hydrophobic	Pi-Alkyl
	ALA338	4.41	Hydrophobic	Pi-Alkyl
CID:122801	GLY346	2.00	Hydrogen	Conventional hydrogen
	LYS351	2.34	Hydrogen	Conventional hydrogen
	ALA338	3.42	Hydrogen	Carbon hydrogen
	TYR345	3.55	Hydrogen	Carbon hydrogen
	HIS536	3.58	Hydrogen	Carbon hydrogen
	HIS536	4.55	Hydrophobic	Pi–Pi T-shaped
	PRO547	4.30	Hydrophobic	Pi-Alkyl
CASID:105377-74-0	ARG532	2.74	Hydrogen	Conventional hydrogen
	HIS536	3.45	Hydrogen	Carbon hydrogen
	PRO342	5.10	Hydrophobic	Alkyl
	TYR345	4.95	Hydrophobic	Pi-Alkyl
CID:12308716	LYS351	2.55	Hydrogen	Conventional hydrogen
	PRO547	4.92	Hydrophobic	Alkyl
	ARG532	4.72	Hydrophobic	Alkyl
	TYR345	4.10	Hydrophobic	Pi-Alkyl
	TYR345	4.55	Hydrophobic	Pi-Alkyl
	HIS536	5.06	Hydrophobic	Pi-Alkyl
	HIS536	5.22	Hydrophobic	Pi-Alkyl
	HIS536	4.51	Hydrophobic	Pi-Alkyl

PK properties

PK is the combination of Greek words pharmakon (drug) and kinetikos (movement), that focuses mainly on dynamic movements of a small molecular candidate into the body and observe the ADME properties of a drug-like compound [31]. PK also can define as a type of xenobiotic regulation process that is required during preclinical studies and the drug development process. It utilizes different mathematical equations to generate a model and provide information regarding the time course of ADME of foreign chemicals (xenobiotics) in the body. PK properties also help us to understand and predict biological action like the toxic or therapeutic effect of a compound. Therefore, to understand the PK properties of the selected four drug-like compounds the SwissADME server has been used in the study. From the server, the ADME properties like lipophilicity (dissolve in fats, oils and nonpolar solvents), water solubility and drug-likeness

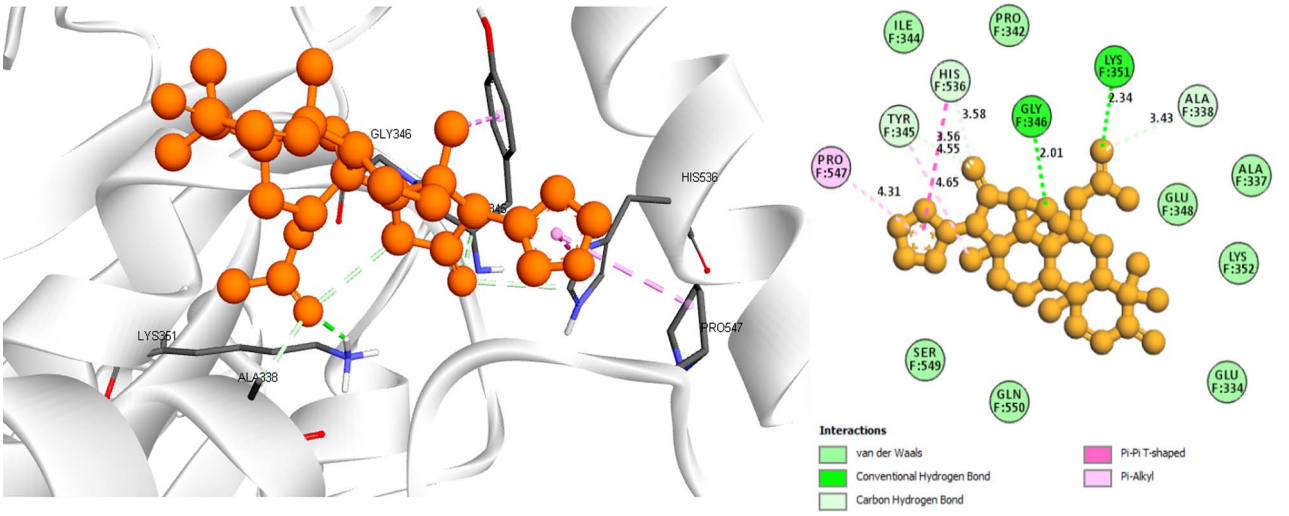


Figure 5. Depicted the interaction between the compound 122801 (PubChem CID) and MCM7 protein. Left side representing 3D and the right side representing 2D complex protein–ligand interaction.

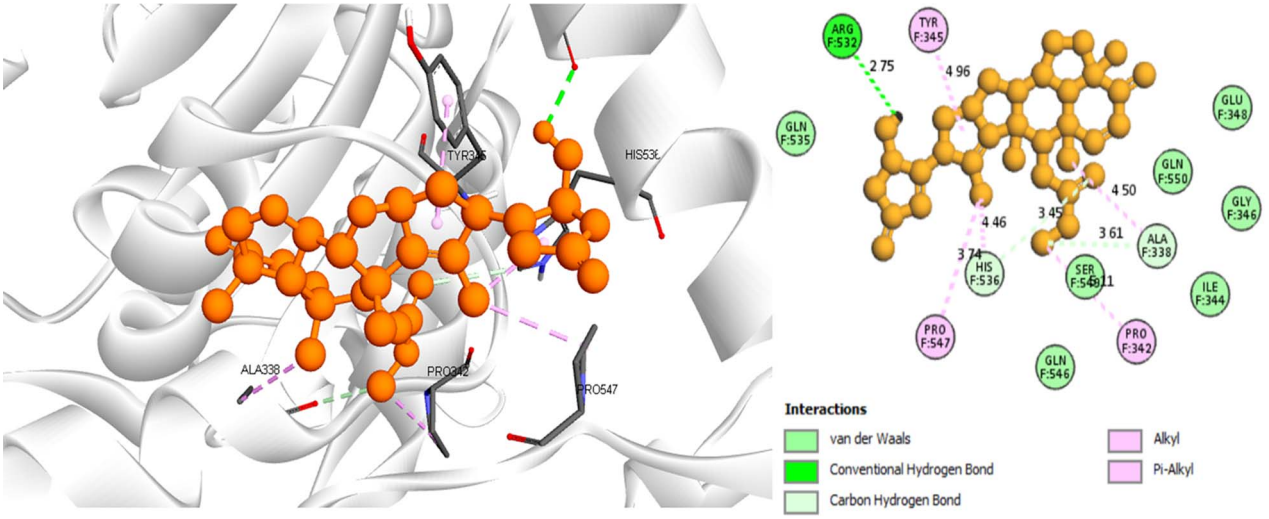


Figure 6. Depicted the interaction between the compound 105377-74-0 (CAS ID) and MCM7 protein. Left side representing 3D and the right side representing 2D complex protein–ligand interaction.

of the selected compounds have been identified and listed in Table 3 and Supplementary Table S2. The PK properties for all the selected compounds were found effective and druggable in the study.

Toxicity prediction

Toxicity prediction is an important step in the modern drug design process that helps to determine the adverse effects of a specific compound on humans, animals, plants or the environment. Conventional drug design approaches use different animal tests to evaluate compounds' toxicity, which is time-consuming, costly and requires ethical considerations [34]. Compared to conventional methods, computer-aided toxicity tests are fast and inexpensive methods that eliminate potentially toxic chemical compounds and reduce a large number of biological experimental tests. For instance, inhibition of human Ether-à-go-go-Related Gene (hERG) potassium ion channels is toxic for the heart and may produce lethal cardiac arrhythmia. Therefore, early-stage identification of putative hERG inhibitors or non-inhibitor may play an important role in reducing cardiotoxicity. To identify adverse effects of selected four compounds an *in silico* toxicity test has been performed by using admetSAR 2.0 web server. The server has identified drug-induced hERG toxicity, AMES toxicity, carcinogenicity, *tetrahymena pyriformis* (TP) toxicity, and honeybee (HB) toxicity listed in Table 4 and Supplementary Table S3.

MD simulations analysis

MD simulation can confirm the stability of protein–ligand complexes in a specific and artificial environment. Therefore, the study performed a 50 ns MD simulation to analyze the steady nature and conformations stability of the protein–ligand complexes. Simulations were performed for the complex docking structure along with one reference antagonist that binds with MCM7 to illustrate the stability of the selected protein–ligands complex. The results of MD simulation have described based on

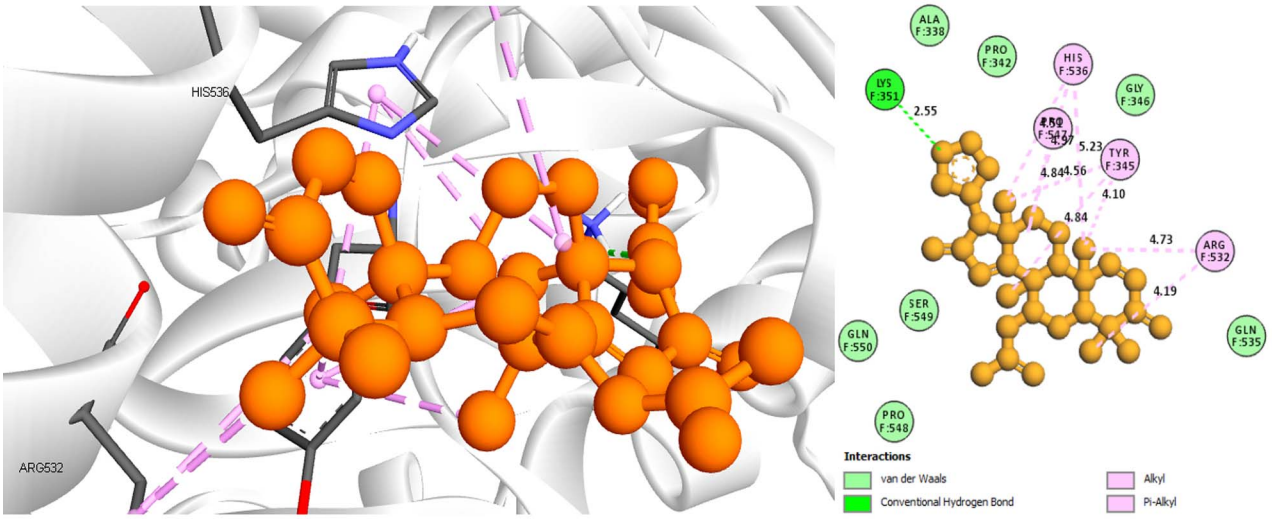


Figure 7. Depicted the interaction between the compound 12308716 (PubChem CID) and MCM7 protein. Left side representing 3D and the right side representing 2D complex protein–ligand interaction.

Table 3. List of pharmacokinetics properties includes physicochemical properties, lipophilicity, water solubility, drug-likeness and medicinal chemistry of selected four compounds

Properties		CHEMSPIDER:156225	CID:122801	CASID:105377-74-0	CID:12308716
Physicochemical properties	MW (g/mol)	450.57	466.57	484.54	450.57
	Heavy atoms	33	34	35	33
	Arom. heavy atoms	5	5	0	5
	Rotatable bonds	3	3	4	3
	H-bond acceptors	5	6	8	5
	H-bond donors	0	0	1	0
Lipophilicity	Log $P_{o/w}$	4.34	3.17	2.86	3.05
Water solubility	Log S (ESOL)	Moderate	Moderate	Soluble	Moderate
Pharmacokinetics	GI absorption	High	High	High	High
Drug-likeness	Lipinski	Yes	Yes	Yes	Yes
Medi. Chemistry	Synth. accessibility	Easy	Easy	Easy	Easy

Table 4. List of the drug-induced hERG inhibition, AMES toxicity, carcinogens, *Tetrahymena pyriformis* (TP) toxicity, honeybee (HB) toxicity, rat acute toxicity (LD_{50} in mol/kg) along with P-glycoprotein inhibitor (PGI) activity of selected four compounds

Compound ID	hERG inhibition	AMES	Carcinogens	Fish toxicity	TP toxicity	HB toxicity	PGI	RAT (LD50)
CHEMSPIDER:156225	No	No	No	Yes	Yes	Yes	Yes	2.770
CID:122801	No	No	No	Yes	Yes	Yes	Yes	3.105
CASID:105377-74-0	No	No	No	Yes	Yes	Yes	Yes	3.817
CID:12308716	No	No	No	Yes	Yes	Yes	Yes	2.770

the RMSD, RMSF, intramolecular hydrogen bonds (Intra HB), and protein–ligand contact analysis (P–L contact).

RMSD analysis

In MD simulation, RMSD also known as the root-mean-square error (RMSE) is frequently used to calculate the mean value change by dislocation of atoms from a particular frame compared to a reference frame and help to determine if the simulation has equilibrated or not [24]. It is mainly the square root of the mean of squared errors that calculate the difference between two sets of values (observed value and estimated value) [36]. The mean or average value change from a particular frame to a reference frame with a range order of 1–3 Å or 0.1–0.3 nm are completely admissible, where a value more than the desired range indicates a large conformational change of the protein. At the end of the MD simulation, if the fluctuations increase or decrease on some thermal average then it should consider that the system has not equilibrated well. The MD simulation with a 50 ns time frame should generate an RMSD that can be calculated by using the equation (Eq: i) and described in detail below.

RMSD of protein

Initially, the protein frames of MCM7 have aligned on the reference frame backbone (red color) and based on the selection of the ligand fit protein atom the RMSD has calculated for

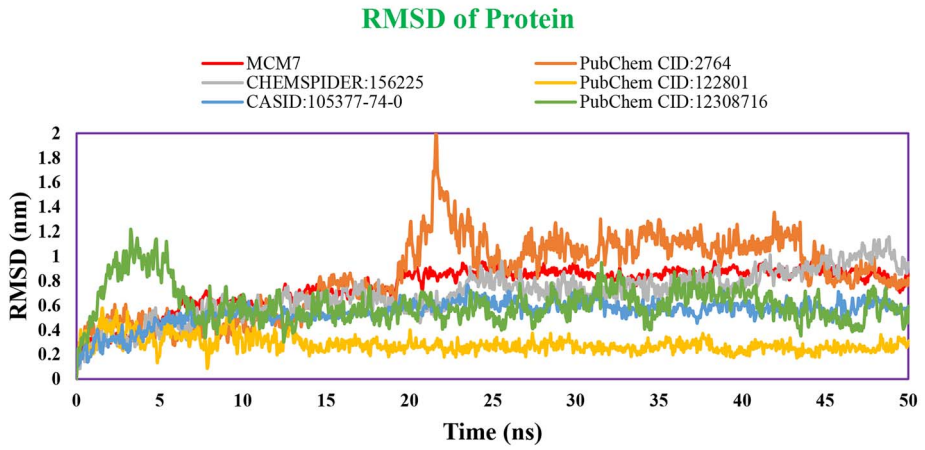


Figure 8. Showing the RMSD values extracted from ligand fit protein (Lig_wrt_Protein) atoms of the complex structure, viz, PubChem CID: 2764 (orange), CHEMSPIDER: 156225 (gray), PubChem CID: 122801 (gold), CASID:105377-74-0 (blue), PubChem CID: 12308716 (green) and MCM7 backbone (red) concerning 50 ns simulation time.

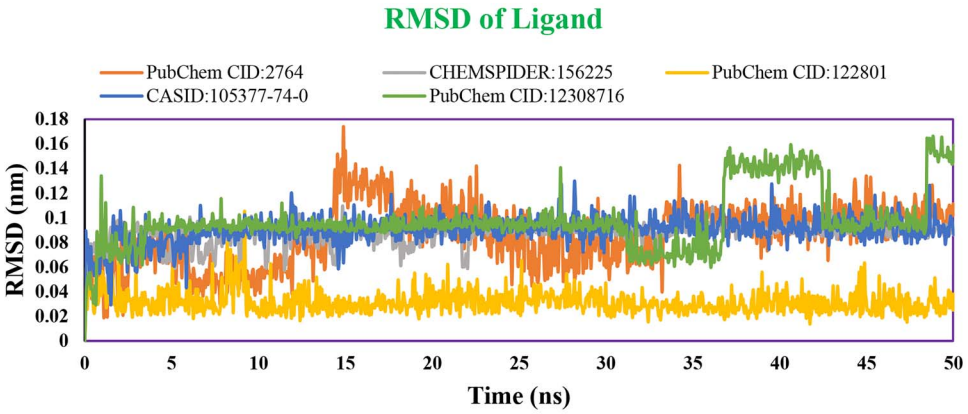


Figure 9. Showing the RMSD values extracted from ligand atoms of the complex structure, viz, PubChem CID: 2764 (orange), CHEMSPIDER: 156225 (gray), PubChem CID: 122801 (gold), CASID:105377-74-0 (blue), and PubChem CID: 12308716 (green) to 50 ns simulation time.

compounds CHEMSPIDER:156225 (gray), CID:122801 (gold), CASID:105377-74-0 (blue), CID:12308716 (green) and control compound CID: 2764 (orange) shown in Figure 8. Monitoring the results of RMSD found the stability of the three compounds except for the compounds CID:122801. The average value change from the MCM7 backbone to CID:12308716, CASID:105377-74-0 and CID:12308716 compounds within a range of 0.1–0.3 nm, where the value changed for the compound CID:122801 has >0.5 nm that was more than the desired range indicate the large conformational change of the protein. At the beginning of 0–8 ns MD simulation, it has found the fluctuations increasing for the compound CID:12308716, which has shown the state of equilibration after 10 ns of MD simulation time. In the case of the other three compounds the fluctuations were optimum until 45 ns, which has a bit increase at the end, but maintain an average range distance between 0.1 and 0.3 nm therefore it should consider that the fluctuations will be optimized during the long simulation run.

RMSD of ligand

Ligands or compounds RMSD were calculated to determine the stability of the selected compounds to the control compounds. In Figure 9 the RMSD of the compounds CHEMSPIDER:156225 (gray), CID:122801 (gold), CASID:105377-74-0 (blue), CID:12308716 (green),and control compound CID: 2764 (orange) has calculated. At first, the complex docking structure has aligned on the protein backbone of the reference and then the RMSD of the compounds has been measured. Herein, the observation of the compound found optimum RMSD for all compounds except compound CID:122801 (>0.4 nm), which distance are significantly larger than the RMSD of the control compound, therefore the compound can diffuse away from its initial binding site.

RMSF analysis

MD simulation not only provides the RMSD value of protein–ligand complex structure, but it also provides information regarding protein heterogeneity and the steady state of macromolecules by generating information about RMSF. The RMSF value is necessary for characterizing a protein that provides an idea about the local changes of protein along with the protein chain [34]. The RMSF for selected natural compounds CHEMSPIDER:156225 (gray), CID:122801 (gold), CASID:105377-74-0 (blue), CID:12308716 (green) and control compound CID: 2764 (orange) in complex with the protein MCM7 has calculated from the Cα by utilizing the equation 2 (Eq: ii) and shown in the Figure 10.

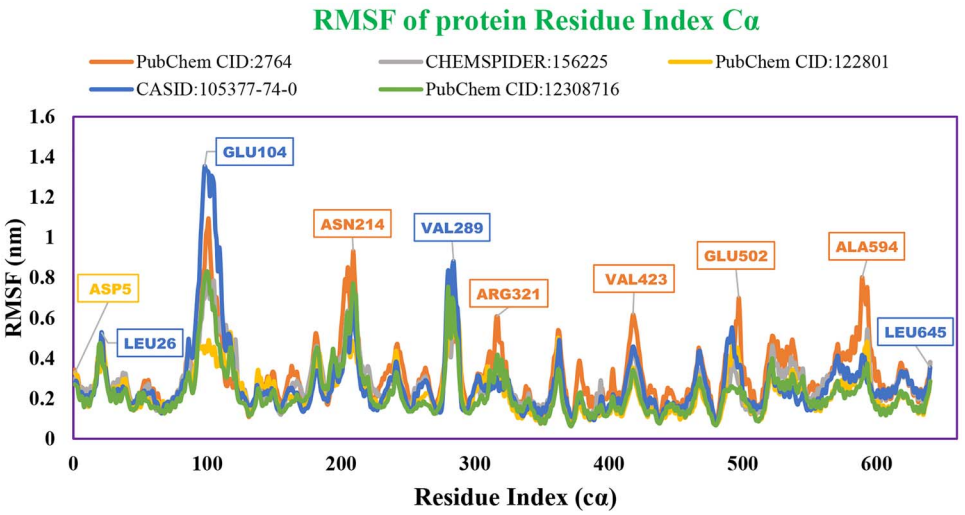


Figure 10. Showing the RMSF values extracted from protein residues Cα atoms of the complex structure, viz, PubChem CID: 2764 (orange), CHEMSPIDER: 156225 (gray), PubChem CID: 122801 (gold), CASID:105377-74-0 (blue), and PubChem CID: 12308716 (green) to 50 ns simulation time.

Analysis of the RMSF graph found the main peaks of fluctuations between 100 and 300 residues maximum for CASID:105377-74-0 with a fluctuation of 1.4 nm in GLU104 residue. The compound has also shown a second round of maximum fluctuations at VAL289 residue with a range of nearly 1 nm, which has shown optimum fluctuation for the rest of the AA residue. The control PubChem CID: 2764 ligand showed the most and maximum fluctuations within 100–600 the residual position with a range of 0.8–1 nm. The remaining compounds were found to be quite stable and fluctuating below 0.5 nm. However, the fluctuations of the compound CID:122801 have always been low compared to control compounds with a maximum of 1 nm between GLU104 residue and did not maintain an average minimum distance for further residues indicating less flexibility of the compound. The RMSF of CHEMSPIDER: 156225-protein-complex and CASID: 105377-74-0-protein complex had the same flow as the Apo and control one, whereas CID: 122801—protein complex RMSF values were higher for all AA residues in the protein complex Indicating the protein complex had a more flexible nature than other complexes structure.

Protein–ligand contacts

Protein in complex with the selected ligands and their intermolecular interactions has been analyzed through 'Simulation Interactions Diagram (SID)' during the 50 ns simulation run. These interactions (or 'contacts') of the compound's CID: 2764, CHEMSPIDER: 156225, CID: 122801, CASID:105377-74-0 and CID: 12308716 has been described based on hydrogen, hydrophobic, ionic and water bridges bonds and shown in Figure 11. The stacked bar charts for the compound CHEMSPIDER: 156225 is showing an interaction fraction value (IFV) of a maximum of 1.0 at the residue LYS 351, which makes contact by using hydrogen and hydrophobic bonds suggesting the specific interaction has maintained over 100% of the simulation time. For the compounds, CID: 122801 the IFV found a maximum 1.0 (ALA 338) that has formed by hydrophobic and water bridge bond, and IFV also found 0.85 at GLY 346 produced by hydrogen and water bridge bond indicating maintaining 100% and 85% interaction, respectively during the simulation time. The compound formed the minimum hydrogen bonding (only two at ALA 338 and GLY 346) with the desired protein, therefore the stability of the compound should be hindered. Compare to the two ligands and control compound CID: 2764, it has found that the compounds CASID:105377-74-0 minted most frequent contact with the protein residue, where maximum IFV 0.85 found at GLY346 position. In the case of compound CID: 12308716 the IFV observed very low until ALA 337 position but maintained an optimized IFV value-form ALA338 to PRO553 position. The compound has also been found to form multiple hydrogens, hydrophobic and water bridge bonding interactions at the position PHE 551 with the IFV value of 0.5. Interactions between the ligand and protein that occur more than 30.0% of the MD simulation time for the selected compound have also been presented in 2D format shown in Figure 11 (left). In the case of compound CID: 2764 found to form multiple interactions with the same residue within the same atom of the ligand due to the sidechain having more than one H-bond donor features.

Hydrogen bond analysis

Hydrogen (H) bonds play an important role to stabilize the ligand with the desired protein and help to influence the drug specificity, acceleration of metabolism and adsorption [24, 34]. Therefore, the number of H-bond produced during protein–ligand complex interaction has been observed during the 50 ns simulation run and shown in Figure 12. Observation of H-bond reveals that all the compound's maintenance optimizes H-bonding until 15 ns simulation runs except the compound CID: 122801. Therefore, the compound will not be able to show the desired effect on the drug specificity, metabolism, and adsorption.

Discussion

CADD has become one of the main and indispensable tools to modern drug design, as it can greatly reduce the expense, time and labor involvement in the process of drug discovery [37]. CADD has accelerated the invention of drugs by enabling scientists to narrow down their biological and synthetic research efforts. Besides, the drug candidate previously identified through CADD approaches like molecular docking, ADMET and MD simulation has shown the highest biological efficacy [25]. Understanding the hypothesis of how ligands bind, interact and inhibit a specific protein can identify drug candidates against a specific

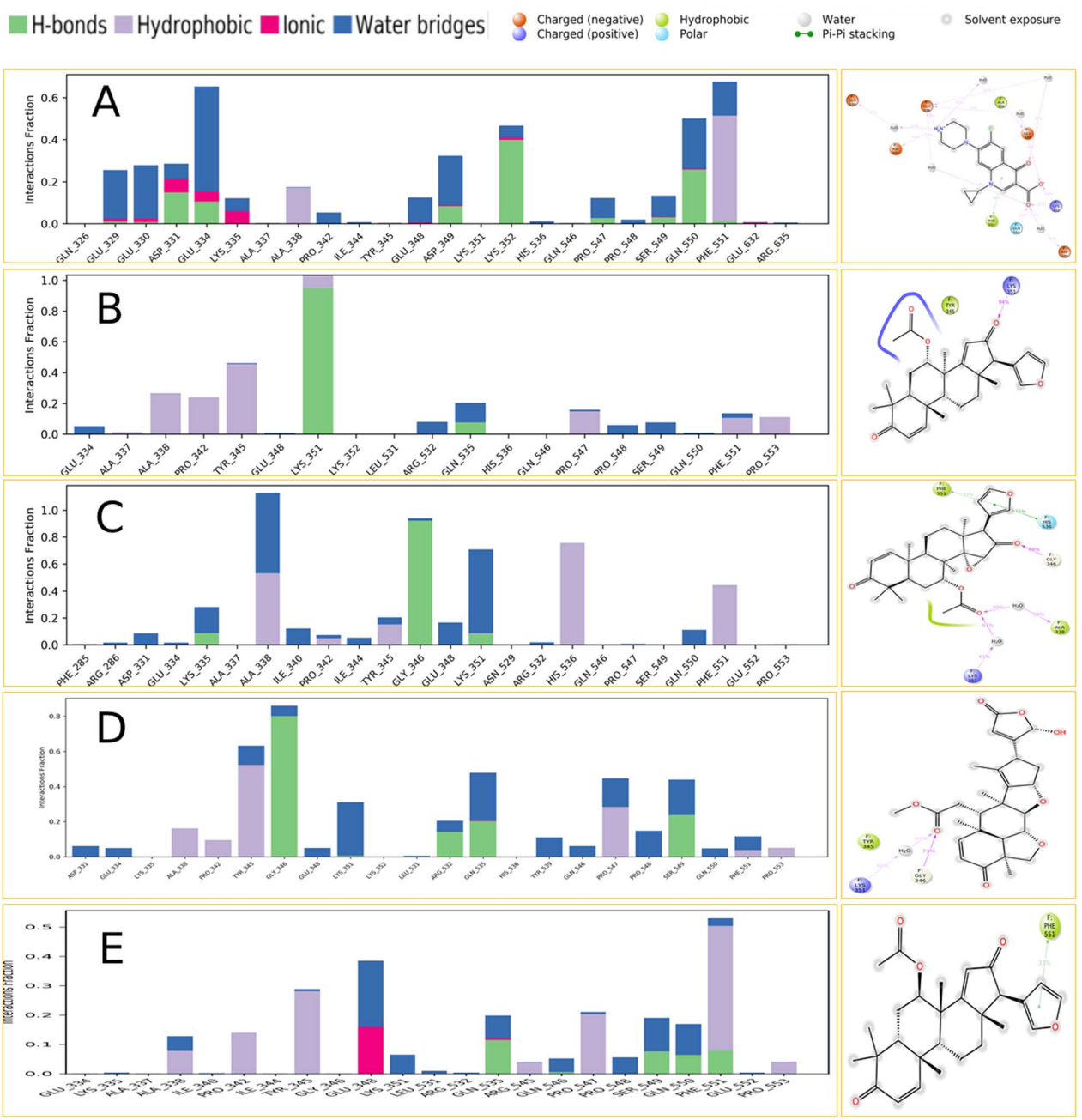


Figure 11. Showing the protein–ligand contacts (left) and ligand–protein contact (right) of the (A). PubChem CID: 2764, (B). CHEMSPIDER: 156225, (C). PubChem CID: 122801, (D). CASID:105377-74-0, and (E). PubChem CID: 12308716 compounds to 50 ns simulation time. The right side of the figure depicted the 2D interactions of the ligand–protein complex that took place more than 30.0% of the total 50 ns MD simulation time.

disease. CADD approach has made it easier to understand the behavior and binding mode of the ligand with a specific target molecule, whether molecular docking can predict the predominant binding modes between a ligand and protein, and MD simulation can reveal the mechanism of complex protein–ligand interaction thus ranks small molecule as drug candidate against a specific disease.

The study analyzed and screened 70 natural phytochemicals targeting the MCM7 protein to fight against human cancer. Initially, the molecular docking approach has been performed to screen the phytochemical, and the best four compounds have been chosen according to their highest binding affinity. Higher binding energy corresponds to lower binding affinity has been observed to the compounds CHEMSPIDER: 156225, CID: 122801, CAS ID: 105377-74-0, CID: 12308716 with binding score − 8.8, −8.4, −8.2 and − 8.0 kcal/mol, respectively. Interpretation of binding interaction demonstrated strong hydrogen and hydrophobic bonding between the protein and ligand.

PK mainly focuses on observing the metabolite kinetics of small molecular candidates in the body. PK of a drug is a temporary and time-course evaluation of drug components in whole blood, tissue and target organs [32]. The activity of a drug mostly depends on ADME that is related to pharmacokinetic properties. During the drug design process, the pharmacokinetic parameters must be optimized so that they can pass standard clinical trials required to be a promising drug candidate [24]. Molecular

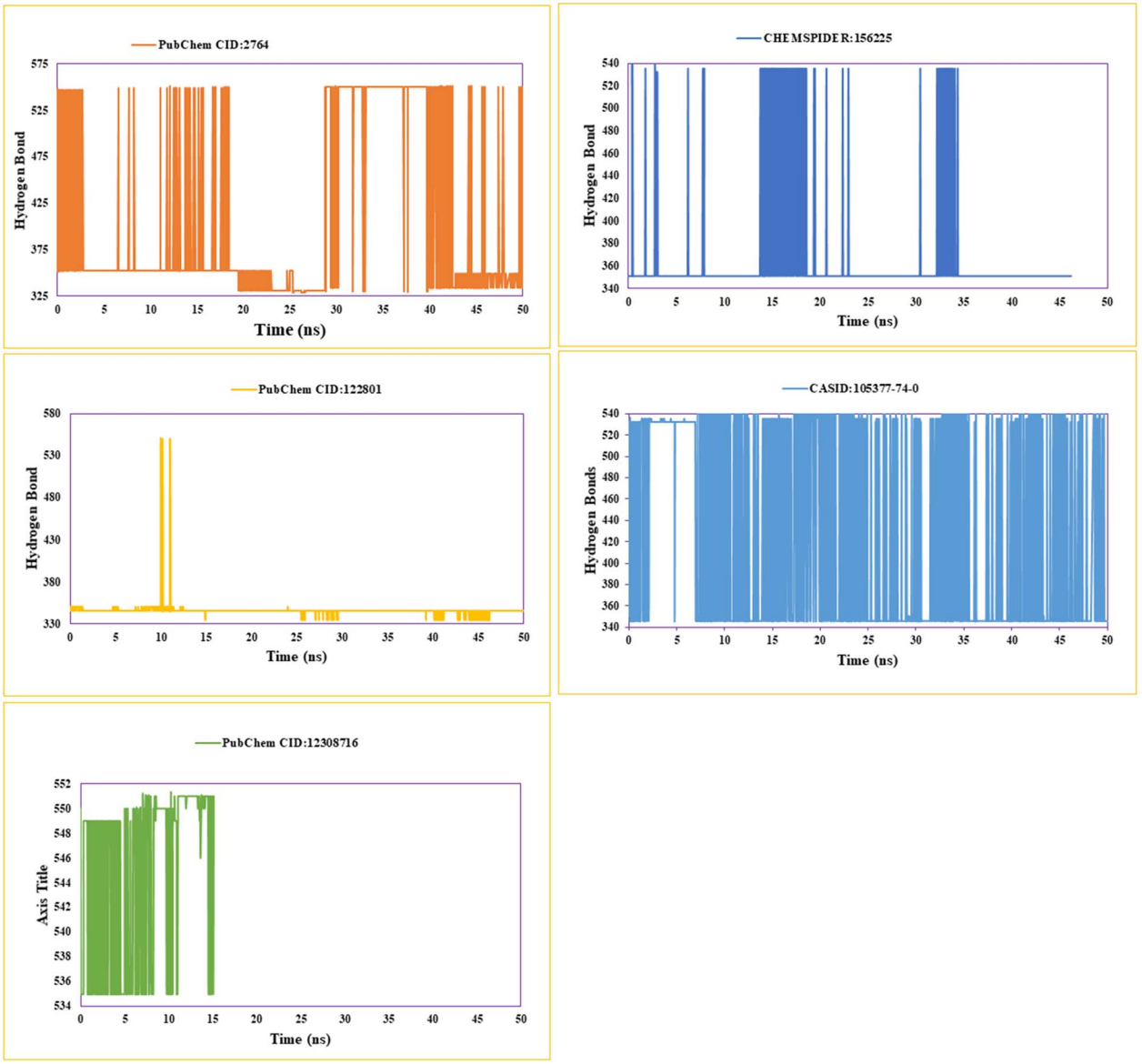


Figure 12. Showing the hydrogen bonds formed during the protein–ligand interaction. Herein, present the hydrogen bond of PubChem CID: 2764 (orange), CHEMSPIDER: 156225 (blue), PubChem CID: 122801 (gold), CASID:105377-74-0 (accent blue), and PubChem CID: 12308716 (green) compounds to 50 ns simulation time.

weight and polar surface topological area (TPSA) have been included under ADME properties that affect the drug molecule's permeability across the biological barrier. Permeability of a drug candidate may decrease due to greater molecular weight and TPSA related to permeability is improved by lower ones. LogP is expressed in the sense of lipophilicity and is referred to as the logarithm of the inorganic and aqueous phase partition coefficient of the target molecule. Lipophilicity influences the drug molecule's absorption within the body. Lower absorption and vice versa are correlated with higher LogP. The LogS value affects the solubility of the candidate molecule and it is often chosen to have the lowest value [32]. The number of donors and acceptors of hydrogen bonds beyond the appropriate range again affects the capacity of a drug molecule to cross the bilayer of the membrane. The overwhelming amount of rotatable bonds is associated with oral bioavailability and the appropriate range is reported to be within 10. In this study, all selected compounds implemented the typical PK properties and further evaluation has performed.

Toxicity can refer to a state of a compound being toxic and that can damage an organism [38]. It has found 20% of failures in the late drug development occur due to toxicity. Toxicity testing of a compound requires animal trials, which is a complex, costly and time-consuming procedure. *In silico* toxicity analysis that requires no animal experiments, low time and inexpensive can be represented as an alternative method for rationalizing preclinical drug development [39]. The best four phytochemicals identified in the study have been evaluated based on *in silico* methods to identify their toxicity. Data collected from *in silico* toxicity test servers determined the noncarcinogenic properties of the compounds. Ames tests have also been employed to evaluate the compound potential genotoxicity by evaluating the capability of reverse mutations, which has been found negative for all the compounds. The toxicity prediction test also found that the compounds are inhibitors of p-glycoprotein and noninhibitor of hERG. The LD_{50} mainly provides information regarding the immediate or acute toxicity of compounds found optimum in the study.

MD simulation analyzes the physical movements of atoms that have become an essential tool for CADD [24, 34]. MD simulation can confirm the stability of the desired drug candidate to a targeted protein. Therefore, MD simulation has implicated the four selected compounds and found the lowest RMSD and RMSF value except for the compound CID:122801. The compound with unexpected RMSD and RMSF values has further been evaluated based on hydrogen bond interaction, protein–ligand contact, Rg and SASA values (Supplementary Figures S1 and S2), which has also indicated a diverse result, therefore the compound has omitted.

Conclusion

Overexpression of DNA replication licensing complex component MCM7 is responsible for multiple human malignancies including hepatocellular carcinoma, esophageal squamous cell carcinoma, head and neck squamous cell carcinoma, and prostate carcinoma [40]. To date, no effective compounds have been identified targeting the protein that can fight against human malignancies. Therefore, the study has been designed to identify natural and effective compounds that can inhibit the activity of the protein resulting in the hindrance of cancer development. The study has utilized comprehensive *in silico* approaches including molecular docking, ADMET and MD simulation, which has identified three compounds as potential drug candidate namely (5α,7α,8ξ,13α,17α)-17-(3-Furyl)-4,4,8-trimethyl-3,16-dioxoandrosta-1,14-dien-7-yl acetate (CHEMSPIDER:156225), isomargosinolide (CASID:105377-74-0) and 17-Epiazadiradione (PubChem CID:12308716) that can inhibit the activity of MCM7 protein. Although *in vitro* and *in vivo* study is required to further confirm the activity of the compounds against desire protein.

Additional information

There is no additional information available for this article.

Key Points

- Minichromosome Maintenance Complex Component 7 (MCM7) is a family member of MCM protein family that plays a key role in DNA replication and G1/S cell-cycle progression. Upregulation of the MCM7 protein is responsible for proliferation of several types of cancer cell.
- Effective, natural and low toxic compounds that able to bind with the protein can inhibit the activity of the protein and hinder undesired cell proliferation, but to date no effective compounds has been identified to targeting the MCM7 protein that can inhibit the uncontrol cell proliferation.
- Previously extracts from neem plant (*Azadirachta indica*) have shown anticancer activities by inhibiting cancer cell growth of different cancer cell. Therefore, compounds from neem plant can inhibit the activity of MCM7 protein.
- To identify natural compounds as MCM7 inhibitor an *in silico* drug design approaches including molecular docking, pharmacoinformatics, toxicity and molecular dynamics simulation has been utilized that determine three compounds and will be able to inhibit the activity of MCM7 protein.

Supplementary Data

Supplementary data are available online at *Briefings in Bioinformatics*.

Acknowledgment

The authors would like to thank to the BioSol Center (Biological Solution Centre, (www.biosolcentre.org) for providing support on molecular dynamics simulation and other necessary equipment required in the study. The research organization known as BioSol Center has also provided suggestions that were very helpful to revise the manuscript.

Funding

No research grants from funding agencies of the public, commercial or nonprofit sectors was received for this study.

References

1. Samad A, Haque F, Nain Z, *et al*. Computational assessment of MCM2 transcriptional expression and identification of the prognostic biomarker for human breast cancer. *Heliyon* 2020;**6**(10). doi: 10.1016/j.heliyon.2020.e05087.
2. Nieduszynski CA, Blow JJ, Donaldson AD. The requirement of yeast replication origins for pre-replication complex proteins is modulated by transcription. *Nucleic Acids Res* 2005;**33**(8):2410–20, Apr. doi: 10.1093/nar/gki539.
3. M. A. Mou *et al*., "Validation of CSN1S1 transcriptional expression, promoter methylation, and prognostic power in breast cancer using independent datasets," *Biochem Biophys Rep*, vol. **24**, p. 100867, Dec. 2020, doi: 10.1016/j.bbrep.2020.100867.
4. Nishitani H, Lygerou Z. Control of DNA replication licensing in a cell cycle. *Genes Cells* 2002;**7**(6) John Wiley & Sons, Ltd:523–34Jun. 01. doi: 10.1046/j.1365-2443.2002.00544.x.
5. Rahman F, Mahmood TB, Amin A, *et al*. A multi-omics approach to reveal the key evidence of GDF10 as a novel therapeutic biomarker for breast cancer. *Inform Med Unlocked* 2020;**21**. doi: 10.1016/j.imu.2020.100463.
6. Tognetti S, Riera A, Speck C. Switch on the engine: how the eukaryotic replicative helicase MCM2–7 becomes activated. *Chromosoma* 2015;**124**(1):13–26. doi: 10.1007/s00412-014-0489-2.
7. Xu X, Wang JT, Li M, *et al*. Timeless suppresses the accumulation of aberrant CDC45 • MCM2-7 • GINS replicative helicase complexes on human chromatin. *J Biol Chem* 2016;**291**(43):22544–58Oct. doi: 10.1074/jbc.M116.719963.
8. Powell SK, MacAlpine HK, Prinz JA, *et al*. Dynamic loading and redistribution of the Mcm2-7 helicase complex through the cell cycle. *EMBO J* 2015;**34**(4):531–43Feb. doi: 10.15252/embj.201488307.
9. Bruck I, Kaplan D. Dbf4-Cdc7 phosphorylation of Mcm2 is required for cell growth. *J Biol Chem* 2009;**284**(42): 28823–31Oct. doi: 10.1074/jbc.M109.039123.
10. Nguyen VQ, Co C, Li JJ. Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. *Nature* 2001;**411**(6841):1068–73Jun. doi: 10.1038/35082600.
11. Huang TH, *et al*. Epidermal growth factor receptor potentiates MCM7-mediated dna replication through tyrosine phosphorylation of lyn kinase in human cancers. *Cancer Cell* Jun. 2013;**23**(6):796–810. doi: 10.1016/j.ccr.2013.04.027.

12. Tabassum A, *et al*. Transporter associated with antigen processing 1 2 (TAP1) as a potential biomarker for breast, lung, liver and ovarian cancer using health informatics 4Aug. 2020. doi: 10.20944/preprints202008.0322.v1 August 14, 2020, preprint: not peer reviewed.
13. Rafi JH, *et al*. High expression of bone morphogenetic protein 1 (BMP1) is associated with a poor survival rate in human gastric cancer, a dataset approaches. *Genomics* 2021;**113**(1P2):1141–54Jan. doi: 10.1016/j.ygeno.2020.11.012.
14. Rahman SMM, Atikullah M, Islam MN, *et al*. Anti-inflammatory, antinociceptive and antidiarrhoeal activities of methanol and ethyl acetate extract of Hemigraphis alternata leaves in mice. *Clinical phytoscience* 2019;**5**(1). doi: 10.1186/s40816-019-0110-6.
15. Sharma V, Sarkar IN. Bioinformatics opportunities for identification and study of medicinal plants. *Brief Bioinform* 2013;**14**(2):238–50Mar. doi: 10.1093/bib/bbs021.
16. F. Ahammad and F. A. A. Fuad, "The in silico identification of potent natural bioactive anti-dengue agents by targeting the human hexokinase 2 enzyme," in *Proceedings of 5th International Electronic Conference on Medicinal Chemistry*, Oct. 2020, p. 6342, doi: 10.3390/ecmc2019-06342.
17. Karpiński TM, Adamczak A. Anticancer activity of bacterial proteins and peptides. *Pharmaceutics* 2018;**10**(2) MDPI AG, Jun. 01. doi: 10.3390/pharmaceutics10020054.
18. Fridlender M, Kapulnik Y, Koltai H. Plant derived substances with anti-cancer activity: from folklore to practice. *Front Plant Sci* 2015;**6**OCTOBER. Frontiers Research Foundation, Oct. 01. doi: 10.3389/fpls.2015.00799.
19. Patridge E, Gareiss P, Kinch MS, *et al*. An analysis of FDA-approved drugs: natural products and their derivatives. *Drug Discov Today* 2016;**21**(2) Elsevier Ltd:204–7Feb. 01. doi: 10.1016/j.drudis.2015.01.009.
20. Alzohairy MA. Therapeutics role of azadirachta indica (neem) and their active constituents in diseases prevention and treatment. *Evid Based Complement Alternat Med*vol. 2016. Hindawi Limited 2016. doi: 10.1155/2016/7382506.
21. Moga MA, Bălan A, Anastasiu CV, *et al*. An overview on the anticancer activity of azadirachta indica (neem) in gynecological cancers. *Int J Mol Sci* Dec. 2018;**19**(12). doi: 10.3390/ijms19123898.
22. Fang J, Liu C, Wang Q, *et al*. In silico polypharmacology of natural products. *Brief Bioinform* May 2017;**19**(6):1153–71. doi: 10.1093/bib/bbx045.
23. Ahammad F, Abd Rashid TRT, Mohamed M, *et al*. Contemporary strategies and current trends in designing antiviral drugs against dengue fever via targeting host-based approaches. *Microorganisms* 2019;**7**(9). doi: 10.3390/microorganisms7090296.
24. Opo FADM, Rahman MM, Ahammad F, *et al*. Structure based pharmacophore modeling, virtual screening, molecular docking and ADMET approaches for identification of natural anti-cancer agents targeting XIAP protein. *Sci Rep* 2021;**11**(1):4049Dec. doi: 10.1038/s41598-021-83626-x.
25. Mahmud S, *et al*. Virtual screening and molecular dynamics simulation study of plant-derived compounds to identify potential inhibitors of main protease from SARS-CoV-2. *Brief Bioinform* 2021;**2021**(00):1–13Feb. doi: 10.1093/bib/bbaa428.
26. Rzechorzek NJ, Hardwick SW, Jatikusumo VA, *et al*. CryoEM structures of human CMG-ATPγS-DNA and CMG-AND-1 complexes. *Nucleic Acids Res* 2020;**48**(12):6980–95Jul. doi: 10.1093/nar/gkaa429.
27. Mohanraj K, *et al*. IMPPAT: a curated database of Indian Medicinal Plants, Phytochemistry and Therapeutics. *Sci Rep* 2018;**8**(1):4329Dec. doi: 10.1038/s41598-018-22631-z.
28. Ko J, Murga LF, Wei Y, *et al*. Prediction of active sites for protein structures from computed chemical properties. *Bioinformatics* 2005;**21**(SUPPL. 1):i258–65Jun. doi: 10.1093/bioinformatics/bti1039.
29. Dallakyan S, Olson AJ. Small-molecule library screening by docking with PyRx. *Methods Mol Biol* 2015;**1263**:243–50. doi: 10.1007/978-1-4939-2269-7_19.
30. Mohammad T, Mathur Y, Hassan MI. InstaDock: a single-click graphical user interface for molecular docking-based virtual high-throughput screening. *Brief Bioinform*Oct. 2020. doi: 10.1093/bib/bbaa279.
31. Hsiao Y, Su B-H, Tseng YJ. Current development of integrated web servers for preclinical safety and pharmacokinetics assessments in drug development. *Brief Bioinform*Aug. 2020. doi: 10.1093/bib/bbaa160.
32. 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**. doi: 10.1038/srep42717.
33. Yang H, *et al*. admetSAR 2.0: web-service for prediction and optimization of chemical ADMET properties. *Bioinformatics* 2019;**35**(6):1067–9Mar. doi: 10.1093/bioinformatics/bty707.
34. Bharadwaj S, Dubey A, Yadava U, *et al*. Exploration of natural compounds with anti-SARS-CoV-2 activity *via* inhibition of SARS-CoV-2 Mpro. *Brief Bioinform* 2021;**2021**(00):1–17Jan. doi: 10.1093/bib/bbaa382.
35. Han Y, Wang Z, Ren J, *et al*. Potential inhibitors for the novel coronavirus (SARS-CoV-2). *Brief Bioinform* 2020; **2020**(00):1–7Sep. doi: 10.1093/bib/bbaa209.
36. Samad A, Ahammad F, Nain Z, *et al*. Designing a multi-epitope vaccine against SARS-CoV-2: an immunoinformatics approach. *J Biomol Struct Dyn* 2020;1–17Jul. doi: 10.1080/07391102.2020.1792347.
37. Madhavi Sastry G, Adzhigirey M, Day T, *et al*. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des* 2013;**27**(3):221–34Mar. doi: 10.1007/s10822-013-9644-8.
38. Bhuiyan MA, Quayum ST, Ahammad F, *et al*. Discovery of potential immune epitopes and peptide vaccine design - a prophylactic strategy against Rift Valley fever virus. *F1000Research* 2020;**9**:999Aug. doi: 10.12688/f1000research.24975.1.
39. Zhou W, Wang Y, Lu A, *et al*. Systems pharmacology in small molecular drug discovery. *Int J Mol Sci* 2016;**17**(2) MDPI AG, Feb. 01. doi: 10.3390/ijms17020246.
40. Qiu YT, Wang WJ, Zhang B, *et al*. MCM7 amplification and overexpression promote cell proliferation, colony formation and migration in esophageal squamous cell carcinoma by activating the AKT1/mTOR signaling pathway. *Oncol Rep* 2017;**37**(6):3590–6Jun. doi: 10.3892/or.2017.5614.