Received: 2 August 2018 | Accepted: 5 November 2018

DOI: 10.1002/jcb.28147

RESEARCH ARTICLE

WILEY Journal of Cellular Biochemistry

Virtual screening of National Cancer Institute database for claudin-4 inhibitors: Synthesis, biological evaluation, and molecular dynamics studies

Majji Rambabu | **Sivaraman Jayanthi**

Department of Biotechnology, Computational Drug Design Lab, School of Bio Sciences and Technology, Vellore Institute of Technology, Vellore, India

Correspondence
Sivaraman Jayanthi, Computational Drug Design Lab, School of Bio Sciences and Technology, Vellore Institute of Technology, Vellore 632014, India.
Email: jayanthi.s@vit.ac.in

Funding information
DST-SERB, India., Grant/Award Number: SB/YS/LS-299/2013

Abstract

Claudin-4 (CLDN4) is a vital member of tight-junction proteins that is often overexpressed in cancer and other malignancies. The three-dimensional structure of human CLDN4 was constructed based on homology modeling approach. A total of 265 242 molecules from the National Cancer Institute (NCI) database has been utilized as a dataset for this study. In the present work, structure-based virtual screening is performed with the NCI database using Glide. By molecular docking, 10 candidate molecules with high scoring functions, which binds to the active site of CLDN4 were identified. Subsequently, molecular dynamics simulations of membrane protein were used for optimization of the top-three lead compounds (NCI110039, NCI344682, and NCI661251) with CLDN4 in a dynamic system. The lead molecule from NCI database NCI11039 (purpurogallin carboxylic acid) was synthesized and cytotoxic properties were evaluated with A549, MCF7 cell lines. Our docking and dynamics simulations predicted that ARG31, ASN142, ASP146, and ARG158 as critically important residues involved in the CLDN4 activity. Finally, three lead candidates from the NCI database were identified as potent CLDN4 inhibitors. Cytotoxicity assays had proved that purpurogallin carboxylic acid had an inhibitory effect towards breast (MCF7) and lung (A549) cancer cell lines. Computational insights and in vitro (cytotoxicity) studies reported in this study are expected to be helpful for the development of novel anticancer agents.

KEYWORDS

claudin-4, homology modeling, molecular dynamics, virtual screening

1 | INTRODUCTION

Claudins (CLDNs) were found to be vital and adequate to form tight-junction (TJ) strands and accounts for some of the specific variability of different barriers.[1,2] CLDNs were one of the important class of cellular adhesion molecules, which had a key role in cell polarity, epithelial cell sheets arrangement, and in paracellular transport. Expression of CLDNs in normal cells is specific and its altered expression is distinguished in different cancer types.[3,4] Usually, tumor development is associated with disruption of TJs and downregulation of CLDNs. They have been accounted for few cancers and it is related to a poor prognosis or metastatic sickness.[5-7] In comparison to other strong tumors, the peritoneal spread is the most widely known technique for ovarian malignancy. Tumor cells subside the ovary into the peritoneal liquid where it scatters all over the abdominal cavity. Thus, it attaches to the mesothelial cell covering which leads to metastatic developments.[8] Nearly 70% of the ovarian growth tissues

expressed claudin-4 (CLDN4) were distinctively expressed across ovarian tumor subtypes with the minimum expression in pure cell subtype.[9] The function of CLDN4 is managed by phosphorylation via kinases as well as forced or knockdown expression. For instance, phosphorylation of CLDN4 by cAMP-dependent protein kinase [10] or protein kinase C (PKC)[11] increases paracellular permeability in ovarian disease cells through mislocalization of CLDNs. In ovarian cancer cells, PKCα enactment results in the mislocalization of CLDN4 with the diminished TJ barrier uprightness in human pancreatic cancer cells.[12] CLDN4 advances the creation of factors that fortify angiogenesis in both in vitro and in vivo processes which recommend its aggressive angiogenic part in ovarian cancer.[13] CLDN4 in ovarian cancer depends on the hypothesis that ovarian malignancies start from typical ovarian surface epithelial cells which do not express CLDN4. In this manner, the overexpression of CLDN4 has been appeared to promote the progression of the ovarian disease. Constrained expression of CLDN4 in ovarian epithelial cells increases invasive behavior by instigating the activation of matrix metalloproteinase.[14] The delocalization of CLDN proteins from the cell membrane is regular among the changed cells and in the ovarian growth which connects with tumor cell migration and invasion.[3,4] The mechanism of the increased CLDN4 expression in the ovarian carcinoma is thought to be the consequence of epigenetic adjustments of the CLDN promoter areas. These areas in the tumor cells bring results in increased cell survival, motility, and invasion.[14-16] Ovarian tumors of various subtypes like mucinous, serous, undifferentiated, clear cell, and endometrioid carcinomas had been found exceptional in the CLDN4 disease state, yet normal ovarian surface epithelium does not express.[13,14,17-19] CLDN4 had a prognostic part in the ovarian growth.[13,18] Epithelial ovarian carcinoma remains gynecologic malignancy with the most noteworthy mortality rate.[20] Even though the mechanism of CLDN4 is yet to be completely elucidated.[21]

In this investigation, we aim to identify potential inhibitors for human CLDN4 with the National Cancer Institute (NCI) database. The curated NCI database is subjected to virtual screening for the CLDN4 target. Further, high throughput virtual screening (HTVS) had been performed to screen the ligand molecules. The docked complexes were screened based on the postdocking analysis which includes binding energy, hydrogen bond (H-bond) interactions, and binding mode. In addition, the pharmacological properties of these inhibitors were validated using the absorption, distribution, metabolism, and excretion (ADME) properties and Lipinski's rule of five. Moreover, the performed molecular dynamic simulations were done for apo and CLDN4 with lead molecules from the NCI database. Virtual screening and dynamics help us to find the top-three molecules. Subsequently, NCI110039 (purpurogallin carboxylic acid) has been synthesized and in vitro studies were conducted. The in silico studies were validated through biological evaluation. This rational approach was adopted for screening the potential inhibitors against CLDN4, which would be helpful in decrypting the binding mechanism that leads to efficient inhibitory action CLDN4.

2 | MATERIALS AND METHODS

2.1 | Data set preparation

The NCI database is used for this study.[22] In Schrodinger suite, the database is curated by visualizing the whole database in Canvas. The improper molecules were excluded from this criterion, such as molecular weight (high and low) of less than 150 and more than 500, number of aromatic rings, and structure similarity. As per the above criteria, the database comprises of 177 236 molecules and exported into the SDF format in maestro were used in ligand preparation.

2.2 | Ligand structure preparation

Curated NCI compounds were used in this study. Preparation of ligands for virtual screening and docking had been done using LigPrep 3.4 module of the Schrodinger suite 2015-2. All ligands were prepared by utilizing the force field called OPLS-2005. The ionization state was fixed to produce every single conceivable state. The molecular weight was set to be less than 150 and greater than 500. Ligand original state using "EpiK," tautomers were generated for the ligands. Utilizing the stereoisomers column maintains the same chiralities and produce 32 at most for each ligand was allowed. Moreover, generated low-energy ring conformations were provided one for each ligand in the output file of ligprep.

2.3 | Homology modeling

Homology modeling of the CLDN4 was performed by using the Schrodinger 10.2 software. The human CLDN4 was retrieved from Swiss-Prot (http://www.uniprot.org/) (accession number: O14493) consists of 209 amino acids. The crystal structure of claudin-15 was used as template (CLDN15) (*Mus musculus*) (PDB ID: 4P79).[23] The template and target sequence were aligned by using ClustalW. Subsequently, secondary structure prediction was carried out for a target and template that helps to find the best model. Further, homology of CLDN4 was based upon the energy-based method. The modeled structure of CLDN4 was validated by PROCHECK in the SAVeS server.

2.4 | Preparation of protein

The modeled structure of CLDN4 is used for protein preparation. Protein Preparation Wizard is available in the Glide. Using import and process results in the elements like assigned bond orders, disulfide bonds, deleted all waters, hydrogens added, and capped termini. Under the refinement, the H-bond assignment enhances to alter the sample water symmetry and possible positions of hydroxyl hydrogen and thiol atoms, tautomers and protonation states of His residues and Chi "flip" assignments chosen through default script allowed in Schrodinger. The range of pH was set to 7.0 and using the OPLS-2005 force field the protein has been minimized.[24,25] For the grid preparation of CLDN4, Glide package had been utilized to signify the properties and shape of the protein by various arrangements and that gives continuously the more exact scoring of the ligand postures.

2.5 | Virtual screening workflow

Numerous strategies have been available for virtual screening.[26,27] Structure-based virtual screening is valuable for identifying novel small molecules which bind to protein targets and predicts binding affinity. For docking, the issues included in making scoring functions have been recently highlighted.[28,29] The compounds were initially filtered based on the Lipinski's rule of five that sets the criteria for a drug-like properties.[30] The source of the ligands submitted as ligprep output file, combined all input files and redistributed for sub jobs and set to identify unique compounds by generating a unique property for each input compound. In the filtering tab, QikProp (ADME properties) and prefilter by the Lipinski's rule were performed. In the preparation tab all parameters were set as default such as prepare ligands, target pH, remove high-energy tautomer states, specified 2D properties, and specified as unspecified stereocenters retain up to four stereoisomers, generate low-energy ring conformations one per ligand. Under receptor provided the grid of the protein. In the docking tab, Epik state penalties for docking were used to write interaction score for residue within 12 Å of grid center. The scaling factor is 0.80, partial charge cutoff 0.15 and performed Glide standard precision (SP) for all the ligands.

2.6 | Ligand docking

Grid Based Ligand Docking with Energetics (Glide) v6.7 (Small-Molecule Drug Discovery Suite, 2015-2) package in Schrodinger is utilized for ligand dockings with an OPLS-2005 force field. When it is confirmed that ligands and protein are right structures to dock, the receptor was generated with a grid sizing $80 \times 80 \times 80$ Å through coordinates $X = -24.05$, $Y = 1.65$, $Z = 2.92$ and the small molecules to be docked. It was chosen without exceeding 10 Å. In ligand docking, the protein is flexible, permits the rotation of thiol and hydroxyl group of TYR33, TYR67, TYR139, SER157. All the ligands were docked into the dynamic binding site by utilizing "extra precision" algorithm of the Glide.

2.7 | Free energy calculation

The Prime molecular mechanics generalized born surface area (MM/GBSA) was performed to figure out the binding energy. This method is utilized to anticipate the free binding energy for the set of ligands to the protein. Binding energy is characterized as the entirety of binding and straining energy of the ligand. The accompanying equation calculates the binding energy.[31]

$$\Delta G_{\text{bind}} = \Delta E_{\text{MM}} + \Delta G_{\text{solv}} + \Delta G_{\text{SA}}.$$

2.8 | ADME screening

The QikProp v4.4 program incorporated inside Schrodinger was utilized to acquire the ADME properties of the NCI ligands. The program was processed with neutralized and permitted to compute almost 35 physicochemical properties for all compounds. The predicted descriptors such as H-bond acceptors, H-bond donors, molecular weight, logP (octanol/water), humoral absorption were a couple of imperative properties. It additionally assesses the blood-barrier crossing and central nervous system activity of the compound along with the drug-likeness of the molecules in view of Lipinski's rule of five[32] and Jorgensen rule of three.[33,34] These two had ended up as being solid standards to determine the drug-likeness of a synthetic compound with a specific biological activity or pharmacological properties, that would make it feasible orally active for a drug.

2.9 | MD simulations for membrane protein

The Desmond v4.2 module [35-37] was utilized to ponder the thermodynamic stability of the protein in apo form and protein-ligand complex. Set up a lipid membrane dipalmitoylphosphatidylcholine for protein to cover the four transmembrane domains of the protein. TIP3P water model[38] was utilized to simulate water molecules by utilizing an OPLS-2005 force field. Orthorhombic periodic boundary conditions were fixed up to indicate the shape and size of repeating unit buffered at 10 Å distances. To neutralize the system electrically, the appropriate three

chloride counterions were added to maintain the system charge and place haphazard in the solvated system. Moreover, the solvated system is minimized and relaxed by certain OPLS-2005 force field. The default Desmond minimization and equilibration procedure were followed. The dynamic simulations were accomplished with the periodic conditions to ensemble the amount of substance, volume and temperature (NPT).[39] The procedure was succeeded by running 50-nanoseconds NPT production simulation. Thus, obtained every 5-picoseconds intervals and saved the configurations. SHAKE has been applied to all hydrogen atoms and bond lengths using default settings. The whole Desmond simulations were performed on the Intel (R) Xeon (R) CPU E5–2650v core processor CPU @ 2.60GHz with 16GB RAM. Schrodinger software was incorporated and runs under Ubuntu Linux 14.0 operating system.

2.10 | Synthesis of purpurogallin carboxylic acid

NCI110039 compound[40] chemical name is purpurogallin carboxylic acid and both the compounds structure remain same. Aqueous solutions of pyrogallol (10 mmol, 1.26 in 10 mL of water) and potassium iodate (10 mmol, 2.14 g in 10 mL of water) were added simultaneously with stirring and ice-cooling to the slurry of sodium gallate obtained by adding gallic acid (12 mmol, 2.04 g) to a solution of sodium hydrogen carbonate (12 mmol, 1.04 g) in water (20 mL) so that addition of pyrogallol and potassium iodate mixture was added slowly in 45 minutes. After the reaction mixture was stirred for 30 minutes. The reaction mixture was acidified by adding 10 N-hydrochloric acid (15 mL), and the reaction mixture set aside overnight. After completion of the reaction, the mixture was filtered and the filtrate was extracted three times with chloroform. All the extracts of the organic layer were combined, the solvent was removed by using vacuum distillation until precipitate (5 g). The crude product was recrystallized with tetrahydrofuran/ethanol a small orange microcrystal has been formed at melting above 320°. The molecular formula of the synthesized compound is $C_{12}H_8O_7$ and molecular weight is 264.0270[41] (Supporting Information Figure S1).

2.11 | Nuclear magnetic resonance analysis

Nuclear magnetic resonance (NMR) spectroscopy is a common technique used to find elements of the chemical components particularly proton, carbon, phosphorous, and sulfur. In this study, proton and carbon NMR was used to confirm the compound purpurogallin carboxylic acid using dimethyl sulfoxide (DMSO) as a solvent. After the analysis, the data has been plotted using the Topspin software.

2.12 | Analysis of cell viability

Cancer cells were maintained in Dulbecco's modified Eagle's medium, added with 10% fetal bovine serum. The cells were incubated in a humidified incubator at 37°C with 5% CO_2 for growth. CLDN4 has been involved in many cancers including breast and lung. We have chosen two major cell lines MCF7 (breast cancer) A549 (lung cancer) to evaluate the inhibitor (purpurogallin carboxylic acid) efficiency. Cell viability of A549 and MCF7 after purpurogallin carboxylic acid treatment was estimated by the 3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide (MTT) assay. Briefly, A549, MCF7 cells were seeded in a 96-well plate at a density of 1×10^4 cells/well. Purpurogallin carboxylic acid is added on next day by replacing the medium with fresh medium with various concentration. After 24 hours of incubation, 10 μL of MTT (10 mg/mL) was added into in each well. After keeping for another 4 hours in 37°C, the supernatant is removed and 100 μL of DMSO is added in each well. Microplate reader at a wavelength of 570 nm is used to read the absorbance of each well. The control group which had untreated cells was measured as 100% viability of cells. Results are concluded as a percentage of viable cells when related with the control group. The mitochondrial-dependent reduction of MTT to formazan used to measure cell respiration as an indicator of cell viability.[42]

3 | RESULTS AND DISCUSSION

CLDN4 protein has four transmembrane domains, namely TM1, TM2, TM3, and TM4, two extracellular loops and one cytosolic loop. Extracellular loops are named as ECL1 and ECL2. ECL1 has more residues compared with ECL2[43] (Figure 1).

3.1 | Validation of the modeled structure

A few strategies to assess the geometrical and structural consistency of the homology model was carried out and the similarity between target and template was found to be 45%. Using PROCHECK, the physicochemical and structural elements of the CLDN4 was validated. The Ramachandran plot reveals that CLDN4 has a splendid geometrical consistency. In the Ramachandran plot, residues for the favorable region inside is 91.2%, allowed region is 6.6%, generously allowed region 1.6%, and in the disallowed region only one residue (LEU176) is found.

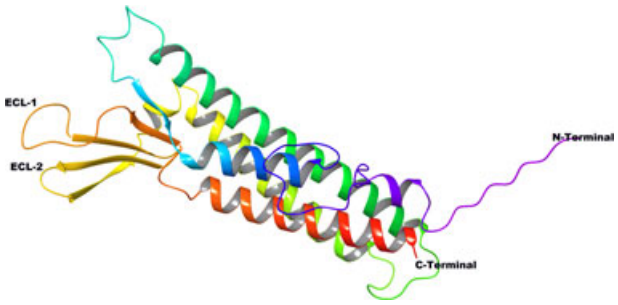

FIGURE 1 Structure of claudin-4 with four transmembrane domains, two extracellular loops with N- and C-terminals. ECL1, extracellular loop1; ECL2, extracellular loop2

The pictorial assessment disclosed that LEU176 is far from the dynamic site region which does not exist in 5 Å of the dynamic site. Furthermore, stereochemical elements of the CLDN4 were confirmed by verifying 3D web server. An ideal 3D-1D profile for each of the 20 amino acids ought to be in the scope of 0 to 0.2. In the 3D server, the values were lesser than zero, which are considered as an erroneous for the model. The 3D plot of CLDN4 demonstrates that average of all the residues is 0.16, which is generally near to 0.20. To assess the reliability of the modeled structure of CLDN4, we calculated the root mean square deviation (RMSD) by superimposing it on the template using the Schrodinger. The CLDN4 model shows 3.0 Å RMSD and an outstanding correspondence with the CLDN15 (*Mus musculus*) experimental structure (Supporting Information Figure S2).

3.2 | Virtual screening

The NCI database screened against the target binding site of CLDN4 protein was utilized for virtual screening workflow (Schrodinger, New York). The Glide HTVS was done with default docking algorithm by selected constraints for every grid in the OPLS-2005 force field. Using Glide HTVS, around 9294 molecules were selected based on GlideScore. After HTVS, the molecules having more than four binding energies had been taken as a separate dataset to perform SP and XP dock.

3.3 | Molecular docking

Molecular docking was implemented to find the depth of molecular interactions to calculate the binding energy between ligand compounds and the CLDN4. The structure of CLDN4 and the best compounds of SP were docked into the active site of CLDN4. The results of docking with compounds NCI110039, NCI344682, and NCI661251 are shown in Table 1. The docking findings confirmed that high binding energy as well as good molecular interactions with important residues in the CLDN4 catalytic site. The ligand NCI110039 docked with CLDN4 as a result −9.3 kcal/mol of binding energy and H-bonds with ARG31, ASN142, ASP146, and ARG158 and hydrophobic interactions with TYR67, LEU71, LEU77, TRP138, and VAL152 (Figure 2).[44] Moreover, three pi-pi interactions with ARG31 and ASP146. The ligand NCI344682 docked with CLDN4 as a result −9.0 kcal/mol of binding energy[45] and H-bonds with ARG31, ASP76, ASN142, ASP146, and ARG158 and hydrophobic interactions with TYR67, LEU71, LEU77, ILE143, and VAL152 (Figure 3). The NCI661251 ligand with CLDN4 target docking was performed as a result −9.6 kcal/mol of binding energy and H-bonds with ARG31, TYR67, ASP76, ASN142, and ASP146 and hydrophobic interactions with TYR67, LEU71, LEU77, and ILE143 (Supporting Information Table S1 and Figure 4).

3.4 | Binding free energy analysis

We have performed binding free energy studies to validate molecular docking energy predictions. The binding energy of NCI110039 is −24.341 kcal/mol, NCI344682 is −33.947, and NCI661251 is −56.212 as shown in Table 2. The MM/GBSA binding energies are approximate free energies of binding, a more negative energy value indicates stronger binding.

TABLE 1 Docking energies and interaction profile of claudin-4 with NCI110039, NCI344682, and NCI661251

Sl no.	Ligand names	Glide docking (XP) energies, kcal/mol	Type of interactions and residues involved	
			H-bonds	Hydrophobic interactions and pi-pi stackings
1	NCI110039	−9.3	ARG31, ASN142, ASP146, ARG158	TYR67, LEU71, LEU77, TRP138, VAL152 (Hphobic), ARG31, ASP146 (pi-pi)
2	NCI344682	−9.0	ARG31, ASP76, ASN142, ASP146, ARG158	TYR67, LEU71, LEU77, ILE143, VAL152
3	NCI661251	−9.6	ARG31, TYR67, ASP76, ASN142, ASP146	TYR67, LEU71, LEU77, ILE143

Abbreviations: H-bonds, hydrogen bonds; NCI, National Cancer Institute.

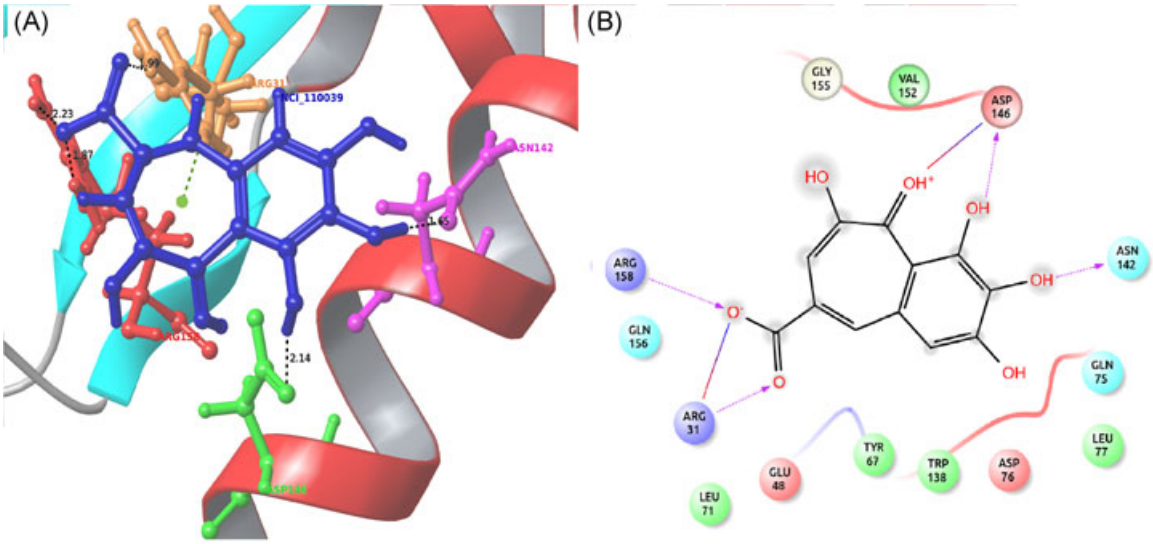

FIGURE 2 Docking snapshot of claudin-4 with NCI110039. A, Showing hydrogen bond formed by NCI110039 with claudin-4 residues. B, 2D representation snapshot showing various interactions formed by NCI110039 at catalytical active site. NCI, National Cancer Institute

3.5 | Predicted ADME properties

ADME properties analyzed with QikProp with appropriate and pharmaceutically noteworthy properties for the short-listed three NCI molecules and the important properties comprise of octanol/water partition coefficient, molecular weight, Madin-Darby canine kidney (MDCK) cell permeability, brain partition coefficient, human oral absorption, aqueous solubility allowing to the Lipinski's rule of five were ended up being a strong criterion for evaluating drug-likeness. The above-mentioned properties from QikProp should be in the particular range such as octanol/water partition coefficient (−2-6.5), molecular weight (150-500), cell permeability (>500 is good, <25 is poor), brain/blood partition (−3.0-1.2), human oral absorption (≤25% is low, ≥ is high), solubility (−6.5-0.5). The above-mentioned NCI110039, NCI344682, and NCI661251 structures are well permissible range (Table 3).

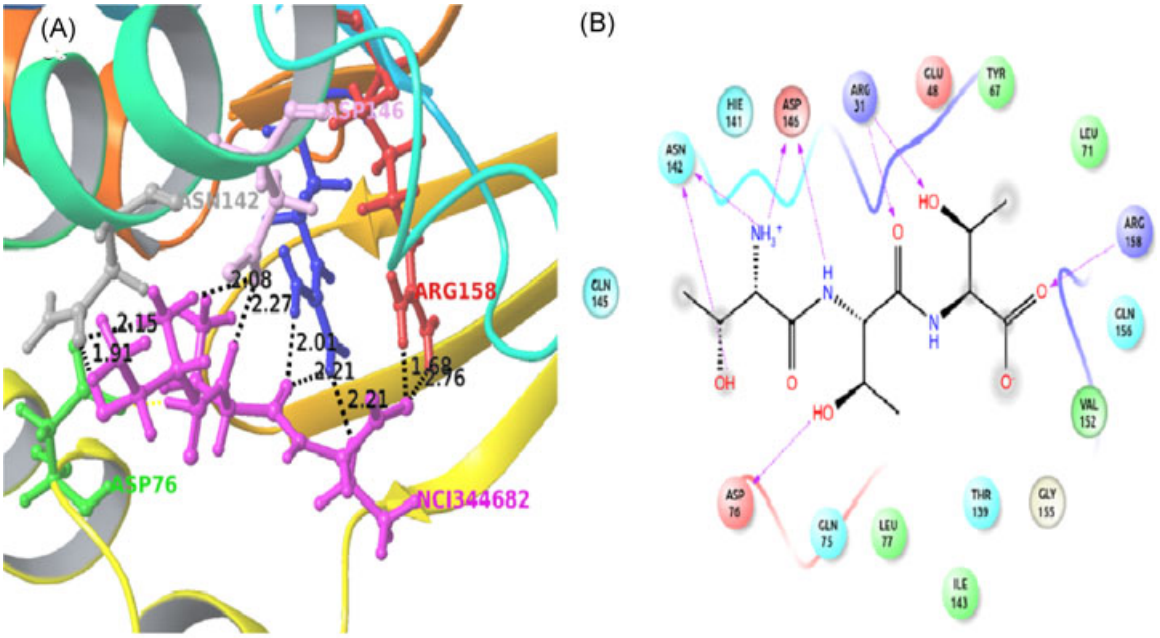

FIGURE 3 Docking snapshot of claudin-4 with NCI344682. A, Hydrogen bond formed by NCI344682 with claudin-4 residues. B, Two-dimensional representation snapshot showing various interactions formed by NCI344682 at the catalytical active site of claudin-4. NCI, National Cancer Institute

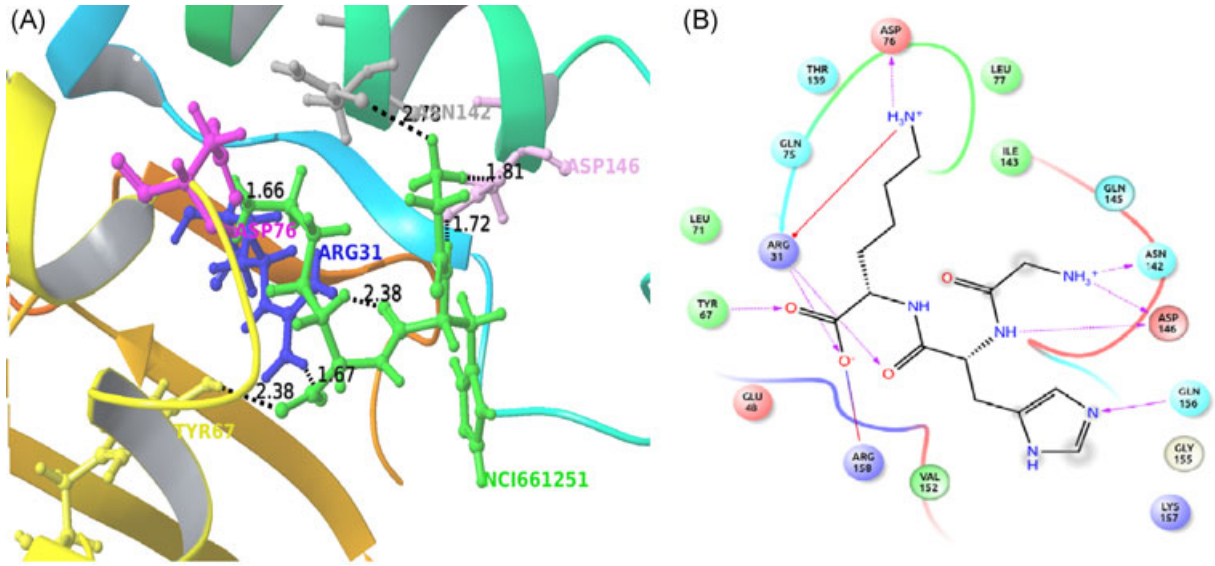

FIGURE 4 Docking snapshot of claudin-4 with NCI661251. A, Hydrogen bond formed by NCI661251 with claudin-4 residues. B, Two-dimensional representation snapshot showing various interactions formed by NCI661251 at the catalytical active site of claudin-4

TABLE 2 Binding free energy calculation for CLDN with NCI110039, NCI344682, and NCI661251

Sl no.	Molecules ID	ΔG, kcal/mol	ΔGvdW, kcal/mol
1	110 039	−24.341	−14.954
2	344 682	−33.947	−22.886
3	661 251	−56.212	−29.712

3.6 | Molecular dynamic simulations

For better understanding, the results of CLDN4 were subjected into protein in apo (no ligand) state first. Later, CLDN4 with the lead compounds like NCI110039, NCI344682, and NCI661251. The statistically critical outcomes and trajectory analysis of the dynamics have been represented (Supporting Information Table S2).

3.7 | MD simulations of CLDN4 membrane protein (apo form)

To inspect the stability and conformational changes of the CLDN4 apo state we observed the root mean square fluctuations (RMSF) (Figure 6A), RMSD (Figure 5A), the radius of gyration (ROG) (Figure 5B), total energy (Figure 5D), and intramolecular H-bonds (Figure 5C). The CLDN4 RMSD backbone was observed to be fluctuating about 4.0 Å on an average for the simulation. However, it is noticed that the convergence after 3-nanosecond of simulated time. The RMSD averaged constant snaps recommended that the structure of the CLDN4 is stable throughout the 50-nanosecond simulation. Using the RMSF diagram (Figure 6A) of CLDN4, it is found that the main peaks of fluctuation between 38 to 40 residues were more than 2.6 Å. At 152 amino acid, it is up to 5.2 Å and residues between 200 to 209 up to 13.8 Å which is the most elevated deviation amid MD simulations. Rest all residues were observed to be entirely steady and under 3.0 Å. ROG is used to find the compactness of the CLDN4.[46] During the CLDN4 apo state simulation measured ROG in the scope of 20.9 to 22.6 Å and with an average of 21.7 Å (Figure 5B). With the ROG information, CLDN4 protein is stable without major peaks present in the structure.[47] CLDN4 holds an average 145 intramolecular H-bonds inside the scope 122 to 171 (Figure 5C). At last, studied the CLDN4's

TABLE 3 ADME properties of NCI lead compounds

Molecules	Molecular weight (mol_MW)	Octanol/water partition coefficient (log Po/W)	Aqueous solubility (QplogS)	Apparent MDCK cell permeability (QPPMDCK)	Brain/blood partition coefficient (QplogBB)	Percent human oral absorption	Rule of five
NCI110039	264.191	−0.467	−1.728	0.79	−2.468	29.866	0
NCI344682	321.33	−3.243	−0.395	0.497	−2.4446	0	0
NCI661257	340.381	−3.866	0.879	−0.079	−2.453	0	0

Abbreviations: ADME, absorption, distribution, metabolism, and excretion; MDCK, Madin-Darby canine kidney; NCI, National Cancer Institute.

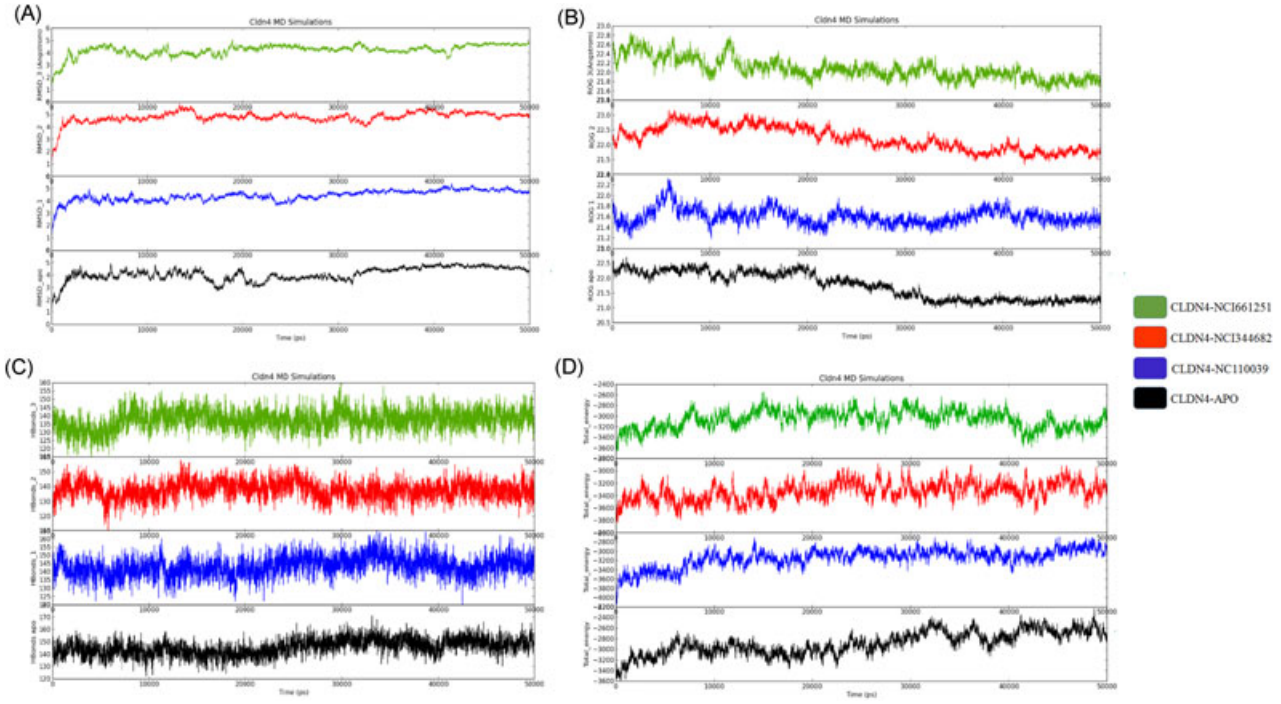

FIGURE 5 RMSD, ROG, intrahydrogen bonds, total energy. A, MD simulation trajectory analysis RMSD of claudin-4 protein in the presence of no ligand and in complex with NCI110039, NCI344682, and NCI661251. B, MD simulation trajectory analysis ROG of claudin-4 protein in the presence of no ligand and in complex with NCI110039, NCI344682, and NCI661251. C, MD simulation trajectory analysis intrahydrogen bonds of claudin-4 protein in the presence of no ligand and in complex with NCI110039, NCI344682, and NCI661251. D, MD simulation trajectory analysis total energy of claudin-4 protein in the presence of no ligand and in complex with NCI110039, NCI344682, and NCI661251. MD, molecular dynamic; NCI, National Cancer Institute; RMSD, root mean square deviation; ROG, radius of gyration

energy in the most alleviated conformation and it was noticed that the average −2889 kcal/mol (Figure 5D). Because every single perception in RMSF, ROG, RMSD, energy contributions, and intramolecular H-bonds collectively, a study can say that the CLDN4 structure is stable throughout the simulation.

3.8 | CLDN4-NCI110039 molecular dynamic simulations

The molecular dynamics of CLDN4-NCI110039 complex was executed to better comprehend the impact of NCI110039 with the CLDN4 binding site. The CLDN4-NCI110039 complex had a binding affinity of −9.3 kcal/mol. The results

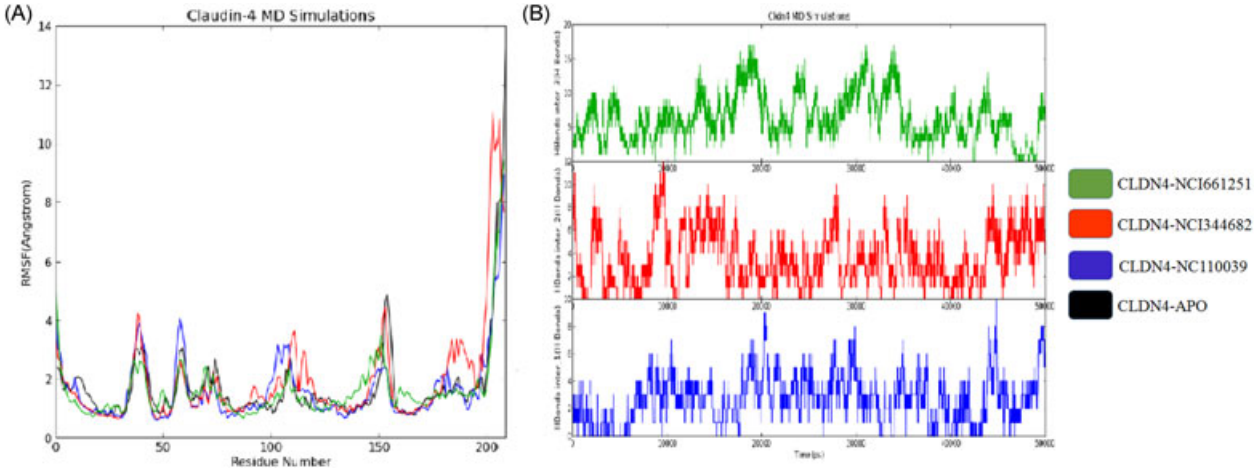

FIGURE 6 RMSF, Interhydrogen bonds. A, MD simulation trajectory analysis RMSF of claudin-4 protein in the presence of no ligand and in complex with NCI110039, NCI344682, and NCI661251. B, MD simulation trajectory analysis interhydrogen bonds of claudin-4 protein in the presence of NCI110039, NCI344682, and NCI661251. CLDN4, claudin-4; MD, molecular dynamic; NCI, National Cancer Institute; RMSF, root mean square fluctuations

in Figure 5A shows that the RMSD is usual for the first 26-nanosecond. Although at the time of binding pocket opens (Figure 5A) ensued a raise up to 5.4 Å for a CLDN4-NCI1110039 complex. The CLDN4 RMSF considered with the presence of NCI110039, which had been noticed the most highly active amino acid motions in CLDN4 native state and minimized (Figure 6A) in the existence of NCI110039. These data are extremely braced to the solid restraining and steady of NCI110039 on CLDN4 with a comparison of CLDN4's apo state residue fluctuations.[48] Fifty-nanoseconds of dynamics were used in the current study of protein-ligand complex, which enables the most stable conformation.[49] The ROG of the CLDN4-NCI110039 indicates CLDN4 had somewhat elaborated (Figure 5B) in the existence of NCI110039 through continuing 21.5 Å as an average within the range of 21.1 to 22.3 Å statistically. The CLDN4 structure as apparent with ROG associating with the RMSD, ROG expansion could be the reason for additional stabilization of the CLDN4-NCI110039. We also considered the intramolecular H-bonds exist through the simulation of the CLDN4 in complex with NCI110039 and found out that it is continuing with an average of 143 intramolecular H-bonds in the range of120 to 165, respectively. The investigation of the energy required for stable conformation of CLDN4-NCI110039 and it has been detected to continue −3146 kcal/mol average energy (Figure 5D) which is very much minimized in comparing to CLDN4's apo state average energy of −2889 kcal/mol, respectively. Finally, approaching the interH-bonds between receptor and ligand throughout the simulation of an average 2.6 within the range of 0 to 10 maximum H-bonds (Figure 6B).

3.9 | CLDN4-NCI344682 molecular dynamic simulations

The native state CLDN4 dynamics and CLDN4-NCI110039 complex were performed, the molecular dynamics of CLDN4-NCI344682 was executed to better comprehend the impact of NCI344682 with the CLDN4 binding site. The NCI344682-CLDN4 complex has a binding affinity of −9.0 kcal/mol acquired by utilizing the Glide module of Schrodinger. Subsequently, it is considered RMSD for the CLDN4 complex with NCI344682. The results represented in Figure 5A shows that the RMSD is usual for the first 12-nanosecond. Although, at the time of binding pocket opens (Figure 5A) ensued a raise up to 5.8 Å for the CLDN4 complex. The CLDN4 RMSF considered with the presence of NCI344682 which had been noticed the most highly active amino acid motions in CLDN4 native state and minimized (Figure 6A) in the existence of NCI344682. The ligand molecule NCI344682 on CLDN4 with a comparison of CLDN4's apo state residue fluctuation, which enables the most stable conformation.[49] The ROG of the CLDN4-NCI344682 indicates that CLDN4 had a little elaborated (Figure 5B) in the existence of NCI110039 through continuing an average 22.2 Å within the range of 21.3 to 23.1 Å statistically. A reasonable expansion of CLDN4 can be detected around the middle and initiated after the 22-nanosecond of the simulation. The CLDN4 structure is as apparent as ROG associating with the RMSD, ROG expansion could be the reason for additional stabilization of the CLDN4-NC344682 complex. We also considered the intramolecular H-bonds existing through the molecular dynamics CLDN4-NCI344682 complex and noticed that it continues 137 average intramolecular H-bonds in the range 110 to 158, respectively. The energy required for stable conformation of CLDN4-NCI344682 and it has been detected to continue −3335 kcal/mol average energy (Figure 5D) which is very much minimized when compared with CLDN4's apo state average energy −2889 kcal/mol, respectively. To end, approaching to interH-bonds between receptor and ligand throughout the simulation average was 3.6 within the range of 0 to 12 maximum H-bonds (Figure 6B).

3.10 | CLDN4-NCI661251 molecular dynamic simulations

The CLDN4-NCI661251 complex showed binding affinity of −9.6 kcal/mol which was acquired by utilizing the Glide module of Schrodinger. Subsequently, considered RMSD for the CLDN4 complex with NCI661251. The results in Figure 5A shows that the RMSD at the time of binding pocket opens ensued in a decrease up to 3.2 Å for the CLDN4-NCI661251 complex. The CLDN4 RMSF is considered with the presence of NCI661251, which had been noticed the most highly active amino acid motions in CLDN4 native state and minimized (Figure 6A) in the existence of NCI661251. The data are highly supporting to restraining and the steadying of NCI661251 on CLDN4 with a comparison of CLDN4's apo state residue fluctuations.[49] The ROG of the CLDN4-NCI661251 indicates that CLDN4 had expanded (Figure 5B) in the existence of NCI661251 through continuing 22.0 Å as an average within the range of 21.5 to 22.8 Å statistically. The CLDN4 structure as apparent with ROG associating with the RMSD, ROG expansion could be the reason for additional stabilization of the CLDN4-NCI661251 complex. The intramolecular H-bonds exist throughout the simulation and found out that it is continuing with an average of 137 intramolecular H-bonds in range 115 to 160 respectively, through the simulation representing complex stability. The energy required for stable conformation of CLDN4-NCI661251 and which had been detected to continue with −3066 kcal/mol of average

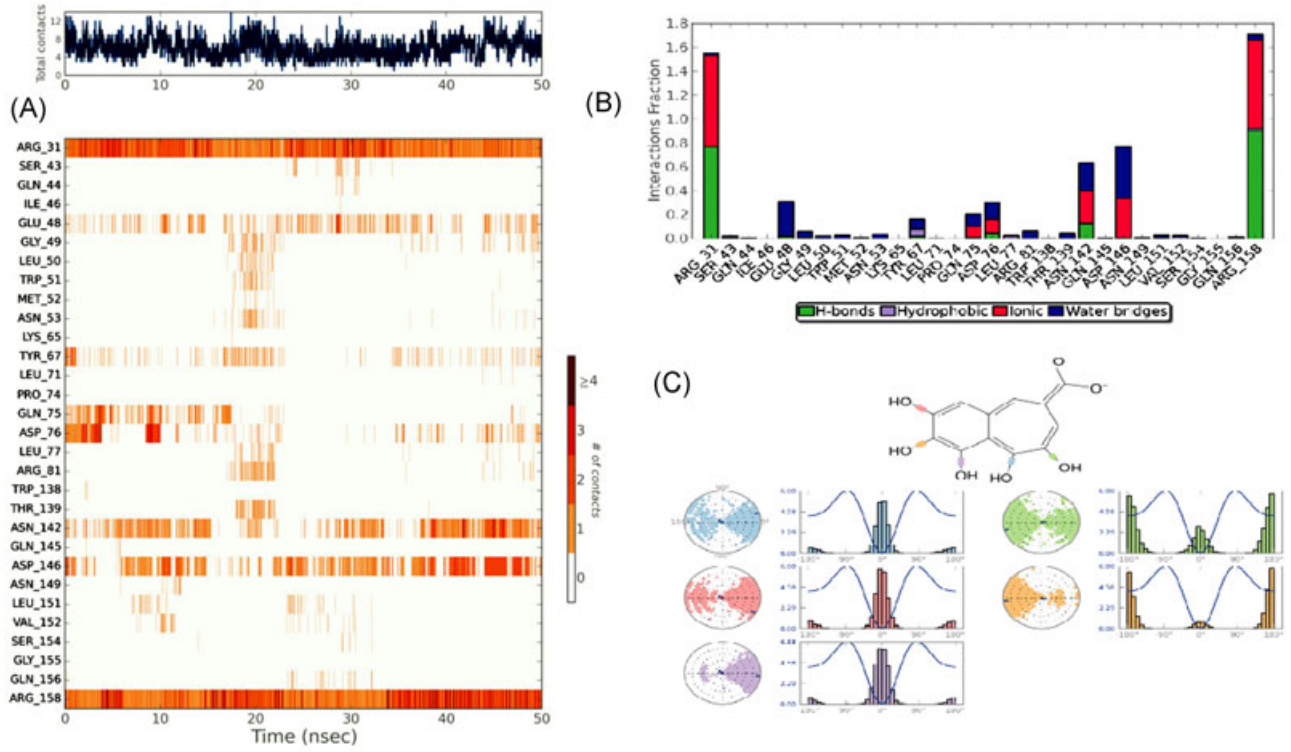

FIGURE 7 Total contacts, interaction diagram, torsional angle. A, Analysis of total contacts formed between claudin-4 residues and NCI110039 during MD simulations. B, Analysis of various interactions involved in stabilizing the claudin-4-NCI110039 complex. C, Analysis of the torsional degree of freedom during MD simulation trajectory for the rotatable bonds present in the NCI110039. MD, molecular dynamic

energy (Figure 5D) and is minimized when compared with CLDN4's apo state average energy −2889 kcal/mol in the apo state. Further approaching interH-bonds between the CLDN4-NCI661251 simulation average is 6.3 within the range starts from 0 to 19 maximum H-bonds (Figure 6B).

3.11 | CLDN4-NCI110039 interaction profile

The interaction profile incorporated inside the Schrodinger's Desmond module for canvassing the point by point intermolecular interactions among CLDN4 and NCI110039 compound. A total of 5 to 10 (Figure 7A) H-bonds were existing among CLDN4 and NCI110039. Regular H-bonds were noticed with amino acids ARG31; ARG158 with over 69% tenancy throughout MD trajectory. Aside from H-bonds, CLDN4-NCI110039 complex shown two hydrophobic contacts with residues like TYR67, LEU77 (Table 4 and Figure 7B).

To analyze and figure out the ligand torsions dynamics, it enables interactions such as H-bonds alongside different interactions amid CLDN4 and NCI110039 complex.[50] The calculated torsional degree of freedom for rotational bonds in the NCI110039. NCI110039 molecule holds five rotatable bonds amid ligand positions 1 to 5, 6 to 13, 7 to 19, 3 to 18, and 2 to 17. These rotational bonds help for binding (Figure 7C).

Through a few computational examinations, it is found that NCI110039 is the potential ligand and synthesized. The synthesized ligand NCI110039 had been confirmed with NMR analysis.

TABLE 4 Interaction profile of NCI110039 with claudin-4 complex post-MD simulation comparison to pre-MD simulations (ie, molecular docking interactions)

S.no	Ligand name	Type of interactions	Pre-MD	Post-MD
		H-bond	ARG31, ASN142, ASP146, ARG158	ARG31, ARG158
1	NCI110039	Hydrophobic	TYR67, LEU71, LEU77, TRP138, VAL152	TYR67, LEU77, VAL152
		Pi-Pi	ARG31, ASP146	N/A

Abbreviations: MD, molecular dynamic; NCI, National Cancer Institute.

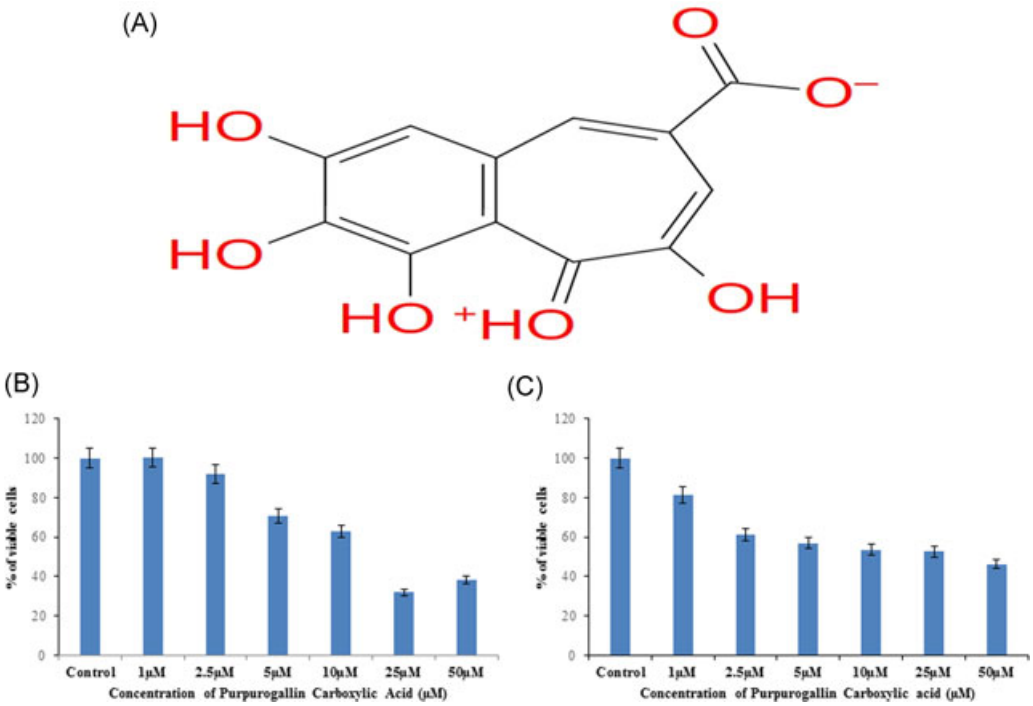

FIGURE 8 Structure of purpurogallin carboxylic acid (PCA), MCF7 cells viability, A549 cells viability. A, Structure of synthesized molecule PCA. B, Purpurogallin carboxylic acid effect on MCF7 cell line. C, Purpurogallin carboxylic acid effect on A549 cell line

3.12 | 1 H Nuclear magnetic resonance

Ten grams yield 38%; orange crystals, melting point 320°C; ^{1}H NMR spectrum (400 MHz, DMSO-d6) had shown the peaks obtained at δ 6.706-6.758 (q, 20 Hz, 1 H), δ 6.904 (s, OH), δ 7.062-7.085 (d, 8 Hz, 1 H), δ 7.331-7.360 (d, 12 Hz, 1 H), δ 9.392 (s, 2 OH), δ 10.586 (s, OH), δ 15.321 (s, OH); ^{13}C NMR spectrum (400 MHz, DMSO-d6): ppm 110.7, 115.3, 116.9, 124.0, 133.5, 134.8, 135.2, 152.0, 152.2, 155.2, 182.7. Calculated for $C_{12}H_8O_7$ 264.0270; found 264.1320.

3.13 | Biological activity of purpurogallin carboxylic acid

The cytotoxic effect of purpurogallin carboxylic acid (Figure 8A) was determined by MTT assay through in vitro on MCF7 and A549 cell lines. Cytotoxicity of purpurogallin carboxylic acid was measured with different concentrations. Results showed compare with the control, 1 μM had not shown any cytotoxicity effect. With increasing concentration of purpurogallin carboxylic acid, cytotoxicity also significantly increased till 25 μM (31.9%). After increasing the concentration (50 μM) it had not shown any cytotoxic effect (38%). It indicates that the 25 μM is the optimal concentration for cytotoxicity of MCF7 cell line (Figure 8B). On A549 cell line with 1 and 2.5 μM cytotoxicity significantly increased (81.2% and 61%). With 5, 10, and 25 μM it is shown a moderate cytotoxicity (56.9%, 53.4%, and 52.5%). At 50 μM concentration cytotoxicity had significantly increased (46.3%) (Figure 8C).

4 | CONCLUSION

In the present study, homology modeling was carried out to build an appropriate structure of CLDN4. The structure-based virtual screening was carried out to bring new and potent inhibitors for CLDN4 protein. Further, docking studies with compounds from the NCI database elucidated the binding mechanism of CLDN4 and active residues like ARG31, ASN142, ASP146, and ARG158 were identified. Subsequently, MD simulations have been performed for CLDN4-NCI110039, CLDN4-NCI344682, and CLDN4-NCI661251 complexes that provide insight into interacting residues and protein stability. Moreover, molecular dynamic simulations results showed that the trajectory of the CLDN4 complex with NCI110039 is stable and insignificant conformational change was observed in CLDN4 because of NCI110039 binding. The stability NCI110039 was determined by executing MM/GBSA and MD simulation studies. In addition, preliminary in vitro studies were also performed and the cytotoxicity assays showed that NCI110039 has an inhibitory effect in the breast (MCF7) and lung (A549) cancer cell lines. From the overall analysis through molecular docking, dynamic simulations and cytotoxic studies, the current study

reveals that NCI110039 could be a promising lead for CLDN4 inhibitory action. Further CLDN4 knockout studies are required in future to substantiate the mechanism of action of NCI110039 to serve as a potent anticancer agent.

ACKNOWLEDGMENTS

The authors are thankful for the financial support from the Department of Science and Technology, Science and Engineering Research Board (DST-SERB), DST No: SB/YS/LS-299/2013. The authors would like to thank the management of Vellore Institute of Technology, Vellore for providing necessary support for carrying out this study work.

ORCID

Sivaraman Jayanthi http://orcid.org/0000-0001-6722-1658

Uncited reference

50.

REFERENCES

1. Furuse M, Sasaki H, Fujimoto K, Tsukita S. A single gene product, claudin-1 or -2, reconstitutes tight junction strands and recruits occludin in fibroblasts. *J Cell Biol*. 1998;143:391-401.
2. Tsukita S, Furuse M. Pores in the wall: claudins constitute tight junction strands containing aqueous pores. *J Cell Biol*. 2000;149:13-16.
3. Morin PJ. Claudin proteins in human cancer: promising new targets for diagnosis and therapy. *Cancer Res*. 2005;65:9603-9606.
4. Singh AB, Sharma A, Dhawan P. Claudin family of proteins and cancer: an overview. *J Oncol*. 2010;2010:541957-541968.
5. Ohtani S, Terashima M, Satoh J, et al. Expression of tight-junction-associated proteins in human gastric cancer: down-regulation of claudin-4 correlates with tumor aggressiveness and survival. *Gastric Cancer*. 2009;12:43-51.
6. Sung CO, Han SY, Kim SH. Low expression of claudin-4 is associated with poor prognosis in esophageal squamous cell carcinoma. *Ann Surg Oncol*. 2011;18:273-281.
7. Ueda J, Semba S, Chiba H, et al. Heterogeneous expression of claudin-4 in human colorectal cancer: decreased claudin-4 expression at the invasive front correlates cancer invasion and metastasis. *Pathobiology*. 2007;74:32-41.
8. Shield K, Ackland ML, Ahmed N, Rice GE. Multicellular spheroids in ovarian cancer metastases: biology and pathology. *Gynecol Oncol*. 2009;113:143-148.
9. Boyle P, Levin B. *World Cancer Report 2008*. France: International Agency for Research on Cancer; 2008.
10. D'Souza T, Agarwal R, Morin PJ. Phosphorylation of claudin-3 at threonine 192 by cAMP-dependent protein kinase regulates tight junction barrier function in ovarian cancer cells. *J Biol Chem*. 2005;280:26233-26240.
11. D'souza T, Indig FE, Morin PJ. Phosphorylation of claudin-4 by PKCepsilon regulates tight junction barrier function in ovarian cancer cells. *Exp Cell Res*. 2007;313:3364-3375.
12. Kyuno D, Kojima T, Ito T, et al. Protein kinase Cα inhibitor enhances the sensitivity of human pancreatic cancer HPAC cells to *Clostridium perfringens* enterotoxin via claudin-4. *Cell Tissue Res*. 2011;346:369-381.
13. Li J, Chigurupati S, Agarwal R, et al. Possible angiogenic roles for claudin-4 in ovarian cancer. *Cancer BiolTher*. 2009;8:1806-1814.
14. Agarwal R, D'souza T, Morin PJ. Claudin-3 and claudin-4 expression in ovarian epithelial cells enhances invasion and is associated with increased matrix metalloproteinase-2 activity. *Cancer Res*. 2005;65:7378-7385.
15. Honda H, Pazin MJ, Ji H, Wernyj RP, Morin PJ. Crucial roles of sp1 and epigenetic modifications in the regulation of the cldn4 promoter in ovarian cancer cells. *J Biol Chem*. 2006;281:21433-21444.
16. Kwon MJ, Kim SS, Choi YL, et al. Derepression of cldn3 and cldn4 during ovarian tumorigenesis is associated with loss of repressive histone modifications. *Carcinogenesis*. 2010;31:974-983.
17. Rangel LB, Agarwal R, D'souza T, et al. Tight junction proteins claudin-3 and claudin-4 are frequently overexpressed in ovarian cancer but not in ovarian cystadenomas. *Clin Cancer Res*. 2003;9:2567-2575.
18. Boylan KLM, Misemer B, Derycke M, et al. Claudin 4 is differentially expressed between ovarian cancer subtypes and plays a role in spheroid formation. *Int J Mol Sci*. 2011;12:1334-1358.
19. Bellone S, Tassi R, Betti M, et al. Mammaglobin b (scgb2a1) is a novel tumor rejection antigen and highly differentially expressed gene in all major histological types of ovarian cancer. *Br J Cancer*. 2013;109:462-471.
20. Patonai A, Erdélyi-Belle B, Korompay A, et al. Claudins and tricellulin in fibrolamellar hepatocellular carcinoma. *VirchowsArchiv*. 2011;458:679-688.
21. Aravindakshan J, Chen X, Sairam MR. Differential expression of claudin family proteins in mouse ovarian serous papillary epithelial adenoma in aging FSH receptor-deficient mutants. *Neoplasia*. 2006;8:984-994.
22. NCI, https://cactus.nci.nih.gov/download/nci/index.html. Accessed May 2012.
23. Suzuki H, Nishizawa T, Tani K, et al. Crystal structure of a claudin provides insight into the architecture of tight junctions. *Science*. 2014;344:304-307.
24. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and testing of the OPLS all-tom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc*. 1996;118:11225-11236.
25. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. Evaluation and reparameterization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B*. 2001;105:6474-6487.
26. Jorgensen WL. The many roles of computation in drug discovery. *Science*. 2004;303:1813-1818.
27. Song CM, Lim SJ, Tong JC. Recent advances in computer-aided drug design. *Brief Bioinform*. 2009;10:579-591.

28. Leach AR, Shoichet BK, Peishoff CE. Prediction of protein-ligand interactions. Docking and scoring: successes and gaps. *J Med Chem*. 2006;49:5851-5855.
29. Tirado-Rives J, Jorgensen WL. Contribution of conformer focusing to the uncertainty in predicting free energies for protein-ligand binding. *J Med Chem*. 2006;49:5880-5884.
30. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*. 2001;46:3-26.
31. Mariadasse R, Biswal J, Jayaprakash P, et al. Mechanical insights of oxythiamine compound as potent inhibitor for human transketolase-like protein 1 (TKTL1 protein). *J Recept Signal Transduct Res*. 2016;36:233-242.
32. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*. 1997;23:3-25.
33. Jorgensen WL, Duffy EM. Prediction of drug solubility from monte carlo simulations. *Bioorg Med Chem Lett*. 2000;10:1155-1158.
34. Jorgensen WL, Duffy EM. Prediction of drug solubility from structure. *Adv Drug Deliv Rev*. 2002;54:355-366.
35. Bowers KJ, Chow E, Xu H et al. *Scalable algorithms for molecular dynamics simulations on commodity clusters*. In SC 2006 conference: proceedings of the ACM/IEEE; Heraklion, Greece, 2006.
36. Schrödinger Release 2014-2: Desmond Molecular Dynamics System, version 3.8, D. E. Shaw Research, New York, NY, 2014. Maestro-Desmond Interoperability Tools, version 3.8, Schrödinger, New York, NY, 2014.
37. Shivakumar D, Williams J, Wu Y, Damm W, Shelley J, Sherman W. Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *J Chem Theory Comput*. 2010;6:1509-1519.
38. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1993;79:926-935.
39. Shinoda W, Mikami M. Rigid-body dynamics in the isothermal-isobaric ensemble: a test on the accuracy and computational efficiency. *J Comput Chem*. 2003;24:920-930.
40. Jhoo JW. Anti-inflammatory effects of purpurogallin carboxylic acid, an oxidation product of gallic acid in fermented tea. *Korean J Food Sci Technol*. 2008;40:707-711.
41. Crow WD, Haworth RD. 293. Purpurogallin. Part VII. Constitution of purpurogallincarboxylic acid. *J Chem Soc (Resumed)*. 1951;131:1325-1327.
42. Reddy LVK, Sen D. DADLE enhances viability and anti-inflammatory effect of human Mscs subjected to 'serum free'apoptotic condition in part via the Dor/pi3k/akt pathway. *Life Sci*. 2017;191:195-204.
43. Bhavaniprasad V, Febin prabhu dass J, Jayanthi S. Activation mechanism of claudin-4 by ephrin type-A receptor 2: a molecular dynamics approach. *Mol Biosyst*. 2013;9:2627-2634.
44. Venkatesan A, Rambabu M, Jayanthi S, Febin Prabhu Dass J. Pharmacophore feature prediction and molecular docking approach to identify novel anti-HCV protease inhibitors. *J Cell Biochem*. 2018;119:960-966.
45. Pandey B, Sharma P. Structural insights into impact of Y134F mutation and discovery of novel fungicidal compounds against CYP51 in *Puccinia triticina*. *J Cell Biochem*. 2018;119:2588-2603.
46. Lobanov MY, Bogatyreva NS, Galzitskaya OV. Radius of gyration as an indicator of protein structure compactness. *J Mol Biol*. 2008;42:623-628.
47. Purohit R. Role of ELA region in auto-activation of mutant KIT receptor: a molecular dynamics simulation insight. *J Biomol Struct Dyn*. 2014;32:1033-1046.
48. Kumar A, Purohit R. Computational investigation of pathogenic nsSNPs in CEP63 protein. *Gene*. 2012;503:75-82.
49. DuBay KH, Geissler PL. Calculation of proteins' total side-chain torsional entropy and its influence on protein–ligand interactions. *J Mol Biol*. 2009;391:484-497.
50. Reddy SVG, Reddy KT, Kumari VV, Basha SH. Molecular docking and dynamic simulation studies evidenced plausible immunotherapeutic anticancer property by Withaferin A targeting indoleamine 2, 3-dioxygenase. *J Biomol Struct Dyn*. 2015;33:2695-2709.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Rambabu M, Jayanthi S. Virtual screening of National Cancer Institute database for claudin-4 inhibitors: Synthesis, biological evaluation, and molecular dynamics studies. *J Cell Biochem*. 2018;1-13. https://doi.org/10.1002/jcb.28147