



Pharmacophore based Virtual Screening for Identification of Marine Bioactive Compounds as Inhibitors against Macrophage infectivity potentiator (Mip) protein of Chlamydia trachomatis

Journal:	<i>RSC Advances</i>
Manuscript ID	RA-ART-11-2015-024999.R2
Article Type:	Paper
Date Submitted by the Author:	17-Jan-2016
Complete List of Authors:	Vijayan, Ramachandran; Bharathidasan University, Marine Science; Jawaharlal Nehru University, Centre for Computational Biology and Bioinformatics, School of Computational and Integrative Sciences Padmanaban, Elavarasi; Bharathidasan University, Marine Science Ponnusamy, Kalaiarasan; Jawaharlal Nehru University, National Centre for Applied Genetics, School of Life Sciences Subbarao, Naidu; Jawaharlal Nehru University, Centre for Computational Biology and Bioinformatics, School of Computational and Integrative Sciences Manoharan, Natesan; Bharathidasan University, Marine Science
Subject area & keyword:	Drug discovery < Chemical biology & medicinal



Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Pharmacophore based Virtual Screening for Identification of Marine Bioactive Compounds as Inhibitors against Macrophage infectivity potentiator (Mip) protein of *Chlamydia trachomatis*

Ramachandran Vijayan,^{a,b} Padmanaban Elavarasi^a, Ponnusamy Kalaiarasan^c, Naidu Subbarao^b and Natesan Manoharan^a.

Macrophage infectivity potentiator' (Mip) is the virulence factor from *Chlamydia trachomatis* that is primarily responsible for causing Sexual transmitted diseases (STD) and Blindness. Mip possess peptidyl-prolyl-*cis/trans*-isomerase (PPIase) activity that can be inhibited by FK506 or rapamycin. Substitution of aspartate-142 position replaced to leucine-142 and tyrosine-185 position replaced to alanine-185 in the catalytic site that strongly reduces the PPIase activity of Mip proteins as reported earlier. As there are no experimentally determined structures available for the *Chlamydia trachomatis* Mip and the number of reported cases of Blindness and STD is increasing. Due to the fact, it is very important to design of new inhibitors against Mip. So, we have modeled the protein structure by homology modeling Further, pharmacophore model and molecular docking simulations were employed to discover Mip inhibitors from a Universal Natural Product Database (UNPD) of 229358 natural compounds. The docking experiment revealed two potential lead compounds, from which Granaticin from *strpetomyces violaceoruber* and Plumarella from Marine coral species were emerged as top candidates. Along with molecular dynamics (MD) simulations will be carried out to analyze the conformational changes behind the molecular mechanism of Mip (native and mutants) for the screened novel lead compounds.

Key words: *Chlamydia*, *Legionella*, *in-silico* mutagenesis, rapamycin, drug designing.

1 Introduction

Chlamydia is a gram-negative pathogen causing several diseases to mankind.^{1,2} *Chlamydia trachomatis* is the most common sexually transmitted bacterial pathogen, with 100 million cases per year world wide.³ It also affects 400 million people for the major cause of

^a Department of Marine Science, School of Marine Sciences, Bharathidasan University, Trichy-620024, India.

^b Centre for Computational Biology and Bioinformatics, School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi-110067, India.

^c National centre for Applied Human Genetics, School of Life Sciences, Jawaharlal Nehru University, New Delhi-110067, India.

ARTICLE

RSC Advances

preventable blindness.⁴ Besides, genital *C. trachomatis* infection has been identified as a cofactor that is necessary for the transmission of human immunodeficiency virus (HIV).⁵ *Chlamydia pneumoniae* infection is responsible for pneumonia fever and respiratory infections in humans.⁶ These human pathogens are able to evade the host immune system. Homologous proteins are also found in other intra cellular pathogens such as *Legionella pneumophila* MIP⁷, *Chlamydia pneumoniae*⁸ and *Trypanosoma cruzi*.⁹ Macrophage infectivity potentiator (MIP) protein exhibit peptidyl-prolyl *cis/trans* isomerase (PPIase) activity⁷ that can be inhibited by Rapamycin and FK506.¹⁰ Latterly, the 3D structures of Mip proteins from *L. pneumophila* and *E. coli* in complex with the PPIase inhibitors rapamycin¹¹ and FK506,¹² respectively have been reported. Also the mutation of *Legionella pneumophila* Mip protein on catalytic residues at Aspartate-142 position replaced to Leucine-142 and Tyrosine-185 position replaced to Alanine-185 which strongly reduces PPIase activity.¹³ In order to design a drug to treat effectively for chlamydial infections, We constructed a *in silico* mutagenesis^{14,15} model for both important catalytic residues (Aspartate-170 into Leucine-170) and (Tyrosine-213 into Alanine213) of *C. Mip*¹⁶, validated the stability of the mutated model. The Universal Natural Products Database (UNPD)¹⁷ was created to be a comprehensive resource of natural products for virtual screening. UNPD consists of 229358 molecules filtered based on Lipinsky and ADME properties. Further, We have built a pharmacophore model based on the available structural data in protein data bank comprised of major interactions between the native inhibitors and Mip active-site residues. Using our pharmacophore model was able to rank the active molecules and the unfit compounds have been strained. Further we have docked to the known inhibitor rapamycin along with the hit compounds with native and mutants of *C. trachomatis* Mip and to determine the binding affinity. Interestingly, 2 compounds were selected for Molecular Dynamic studies to analyze the stability of the native and mutant model in complex with hit compounds.

2 Methods

2.1 Homology Modeling

The primary sequence of the *Chlamydia trachomatis* Mip protein was retrieved from the UniProt database (Accession number: P26623) (<http://www.uniprot.org/uniprot>). A sequence similarity search for the protein against other sequences with existing structural information in Protein Data Bank (www.rcsb.org/pdb) was performed using the PSI-BLAST [www.ncbi.nlm.nih.gov/blast]. Crystal structure of *Legionella pneumophila* Mip (PDB ID: 1FD9 with 2.4 Å resolution) was selected as template, having 35% sequence identity with target. To analyze conservation of the protein sequence, *Chlamydia trachomatis* Mip and *Legionella pneumophila* Mip were aligned by Clustal W.¹⁸ Gaps were

inserted into the sequences to obtain an optimal alignment, as represented in Fig. 1. The 3D structure of *Chlamydia trachomatis* Mip was modeled using the Modeller¹⁹, and visualized using Discovery Studio²⁰ and PyMOL software²¹. The three-dimensional homology model of the protein is obtained by probability density function with the variable target function procedure in the Cartesian space using the methods of conjugate gradients and molecular dynamics with simulated annealing.¹⁹ Twenty homology models were built and energy minimized with CHARMM force field²², with the 500 steps of steepest descent and 1,000 steps of conjugate gradient with a root mean square (rms) gradient of the potential energy of 0.001kcal/mol Å at each step. The stereochemical quality of the final refined model was evaluated using PROCHECK.²³ The mutant (D170L and Y213A) structure of Cht Mip was built by induced point mutation at the position of 170 and 213 of *C. trachomatis* Mip protein using SPDB viewer package.¹³ This structure was energetically optimized by applying the all atom OPLS force field available in SPDB viewer package.²⁴

2.2 Predicting stability change on mutated single amino acid based on support vector machine (I-Mutant 2.0)

The mutations occurring in the protein-coding region may lead to the deleterious consequences and might disturb its 3D structure. Protein structural stability of the mutants were assessed using I-Mutant 2.0 server²⁵ which is a support vector machine (SVM) based tool for automatic prediction of protein stability changes upon single-point mutations; predictions are performed for both sequence and structure of proteins using I-Mutant 2.0 server.²⁵ The output of this program shows the predicted free energy change value ($\Delta\Delta G$) that is calculated from the unfolding Gibbs free energy value of the native type (kcal/mol). Positive $\Delta\Delta G$ values infer that the mutated protein possesses high stability and vice versa (http://gpcr.biocomp.unibo.it/cgi/predictors/I-Mutant_2.0/I-Mutant_2.0.cgi).

2.3 Database Screening:

The 2D chemical structures of compounds were obtained from the Universal Natural Products Database (UNPD)¹⁷ and converted into 3D-MOL2 file with the program OpenBABEL2.3.1²⁶. Universal Natural database (UNPD) consisting of 2,29, 358 compounds that is commercially available for experimental testing. As, however it is better to filter all the compounds do not exhibit drug-like properties were eliminating from the database prior to pharmacophore filtering. Therefore, Lipinski's rule of five and ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties were applied by using Discover Studio (DS) program [20] to filter the false hits.

2.4 Pharmacophore model generation:

LigandScout2.03²⁷ software is used to develop pharmacophores from

RSC Advances

protein-ligand complexes based on the interactions between *Legionella pneumophila* Mip in complex with rapamycin (2VCD) and *E. coli* Mip in complex with FK506 (1Q6I) obtained from Protein Data Bank (PDB). This software has an advanced alignment algorithm discover pharmacophoric points/features such as hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), ring aromatic (RA), hydrophobic groups (HYP) and hydrogen bonding vectors on ligand to generate the pharmacophore model of potential inhibitor against *Chlamydia trachomatis* Mip. We have developed a pharmacophore model based on the crucial interactions between the native inhibitors (FK506 and rapamycin) and active site residues of Mip as control. The Compounds were further energy minimized for 100 steps with Swiss-PDB viewer with steepest descent and conjugated gradient algorithms.²⁴

2.5 Virtual Screening:

The hit compounds having specific pharmacophoric features were subjected to molecular docking studies using GOLD (Genetic Optimization for Ligand Docking).²⁸ Hit compounds were docked into the active site to confirm the correct binding mode and to determine the binding affinity. Cambridge Crystallographic Data Center (United Kingdom), GOLD 5.1 program²⁸ uses a genetic algorithm for docking ligands into protein binding sites to explore the full range of ligand conformational flexibility with partial flexibility of protein. Docking procedure consisted of three interrelated components; a) identification of binding site b) a search algorithm to effectively sample the search space (the set of possible ligand positions and conformations on the protein surface) and c) a scoring function. The GOLD fitness score[28] is measured based on the formation of protein-ligand hydrogen bond energy (external H- bond), protein-ligand vanderwals (vdw) energy (external vdw, ligand internal vdw energy (internal vdw) and ligand torsional strain energy (internal torsion). Protein coordinates from the representative structure of the *Chlamydia trachomatis* Mip and mutant structures obtained from Homology modeling were used to define the active site. The active site was defined with a 10Å radius around rapamycin. Ten poses were kept for each ligand, Default settings for GOLD docking were adopted.²⁸ Protein-ligand interactions were analyzed using Discover Studio.²⁰

2.6 Molecular Dynamics Simulations:

The simulations were performed using GROMACS 4.5.3 package^{28,29}, with the GROMOS 96 force field. The protein atom was certain to 1.5 nm far from the wall of box dimensions along with periodic boundary conditions and solvated by simple point charge³⁰ (spc) of water molecules. NaCl counter ions were included for the requirement of the electro-neutrality condition. Energy was minimized using the steepest descent method. In order to maintain the system in a stable environment(300 k, 1 bar), Berendsen temperature coupling³¹ and

Parrinello-Rahman pressure coupling³² were employed and the coupling constants were set to 0.1 and 2.0 ps for temperature and pressure. For the measurement of electrostatic and Van der Waals interactions, the partial mesh Ewald (PME) algorithm³³ was performed, cut-off distance for the small-range VdW (rvdw) was fixed to 1.4 nm and coulomb cut-off (rcoulomb) and neighbour list (rlist) were set to 0.9 nm. LINCS algorithm³⁴, were employed to measure all the constrained bond lengths and the time step was set to 0.002 ps. The complexes in a medium were equilibrated for 100 ps in NPT and NVT ensembles, respectively. Finally, a 20 ns molecular dynamics simulation was carried out for both native and mutant complexes of *Chlamydia trachomatis* Mip. All trajectories were stored every 2 ps for further analysis.

2.7 Structural Analysis of Molecular Dynamics Simulations

Structural properties of the native and mutant complexes of *Chlamydia trachomatis* Mip were calculated from the trajectory files with the built-in modules of GROMACS 4.5.3. Analysis of MD simulations such as root mean-square deviation (RMSD) and root-mean square fluctuation (RMSF) and radius of gyration were analyzed through the use of `g_rmsd`, `g_rmsf` and `g_gyrate` respectively, with the built-in functions of GROMACS. `g_hbond` utility was used to calculate the number of hydrogen bond observed within the protein during the simulation. The formation of hydrogen bonds which were determined based on the donor-acceptor distance were smaller than 3.6 Å and of donor-hydrogen-acceptor angle larger than 90°. SASA of Mip was calculated by using `g_sas` GROMACS, respectively. Graphing Advanced Computation and Exploration (GRACE) program were used to create the various plots for three-dimensional backbone RMSD, RMSF of carbon-alpha, gyration of backbone and SASA analysis.

3 Results and Discussion

3.1 Homology Modeling:

Experimentally determined structure for *Chlamydia trachomatis* Mip has not been available in Protein Data Bank (PDB). Therefore, we have predicted the protein structure of *Chlamydia trachomatis* Mip by homology modeling, a promising tool for providing the insights into the interaction between the drug and receptor. The amino acid sequence of the Cht Mip consist of 243 amino acid residues was retrieved from UniProt(www.uniprot.org). Sequence alignment of the *Chlamydia trachomatis*(Cht) Mip with *Legionella pneumophila* Mip was performed with Clustal W to align conserved key residue as shown in Figure 1. The X-ray structure of *Legionella pneumophila* Mip at 2.4 Å resolution (PDB ID: 1FD9) was used as suitable structural template to built the 3D structure by MODELLER.¹⁹ The input parameters of MODELER were set to generate 20 models by applying the default model building routine 'model' with fast refinement. This procedure is to select the best model from several model structures. Further, the differences among the

ARTICLE

RSC Advances

models are used to evaluate the quality of the protein structure modeling. Energy minimization was performed using the valence force field and the Swiss-PDB Viewer with steepest descent and conjugated gradient algorithms.²⁴ Energy minimization and relaxation of the loop regions was performed using 300 iterations in a simple minimization method. Again the steepest descent was carried out until the energy showed stability in the sequential repetition. The overall conformation of the Cht Mip model was very similar to the template, and even the amino acids involved in the Rapamycin binding were well conserved. Since it has two domains namely N-terminal and C-terminal. N-terminal domain is mainly responsible for dimerisation and C-terminal has peptidyl prolyl *cis-trans* isomerase activity and contribute significantly toward the binding of the ligands. Thus, the 3D model of Cht Mip was considered to be reasonable.

3.2 Protein Structure Validation

The stereo-chemical quality of the *Chlamydia trachomatis* Mip were evaluated by using PROCHECK program to determine the overall stability of the structure. Ramachandran plot showed the distribution of the main chain torsion angles, the phi/psi angles of 82.6% residues fell in the most favored regions, 15.2% of the residues lay in the additional allowed regions, and 2.2% residues fell in the generously allowed regions; no residues lay in the disallowed conformations (Fig. 2). Structural superposition of the modeled protein with respect to the template was calculated using pyMOL and found to have RMSD value of 0.064 Angstrom. The overall conformations of the *Chlamydia trachomatis* Mip model were very similar to the template (Fig. 3). Thus, the predicted three-dimensional model of *C. trachomatis* Mip was considered to be reliable.

3.3 Protein structural stability of the mutant's analysis

In this mutational study, sequence of *C. trachomatis* Mip (native) at 170th position ASP was replaced by LEU and at 213th position TYR was replaced by ALA to predict protein stability changes through I-mutant server [15]. Loss of stability by the mutant protein with negative Gibbs free energy value of -1.70 and -1.52 at pH 7.0 and 37°C was observed (Table 1) (<http://gpcr.biocomp.unibo.it/cgi/predictors/I-Mutant.2.0/I-Mutant.2.0.cgi>).

Mutation	Stability	RI	DDG	pH	T
D170L	Decrease	3	-1.70	7.0	37
Y185A	Decrease	8	-1.52	7.0	37

Table 1: Protein structural stability of mutated amino acids of *Chlamydia trachomatis* Mip. Note: RI: Reliability Index; T: Temperature in Celsius degrees; pH: $-\log[H^+]$; DDG: $DG(\text{NewProtein})-DG(\text{WildType})$ in Kcal/mol; DDG<0: Decrease Stability; DDG>0: Increase Stability

3.4 Pharmacophore Modeling

The UNPD (229358) molecules has been used for screening were filtered based on Lipinsky rule and Adsorption, Digestion, Metabolism and Excretion (ADME) properties and the resulting molecules were 87,920. LigandScout2.03²⁷ was used to generate pharmacophore model from protein-ligand complexes based on the interactions between *Legionella pneumophila* Mip in complex with rapamycin (2VCD) and *E. coli* Mip in complex with FK506 (1Q6I) obtained from Protein Data Bank (PDB). LigandScout is a program for the automatic detection of relevant interaction points between a ligand and a protein. Its algorithms performed a stepwise interpretation of the ligand molecules: planar ring detection, assignment of functional group patterns, determination of the hybridization state, and finally the assignment of Kekule pattern. Mip is mainly mediated by rapamycin's pipercoline moiety is attached in a hydrophobic pocket of the binding site. The active site is found between the helix and the inner side of the sheet and shows mainly hydrophobic character. Phe180, Leu181, Ile187, Tyr213, and Thr215 residues are formed at the base of the pocket. Asp179, Ile187 and Tyr213 are the major contributors in hydrogen bonding interaction with compounds and are responsible for the main conformational differences. Utilizing available structural information, we developed a pharmacophore model including crucial interactions between the inhibitor molecules and active-site residues of Mip as restraints. The final pharmacophore model was constructed by combining the 9 features as 1 hydrogen bond acceptors (HBAs) (green color), 5 hydrogen bond donors (HBDs) (red color) and 3 aromatic rings (yellow color) using the consensus score (Fig. 4). The pharmacophore of the above mentioned groups have been merged in order to get the common pharmacophore of Mip inhibitors. We used our pharmacophore model (Fig. 4) to filter and rank UNPD compounds (87,920) and the most active compounds bind in a similar fashion at the enzyme's active site. The new pharmacophore model was employed to superimpose with the Universal Natural product database (UNPD) database compounds (87,920) resulting common chemical features containing (150) compounds were chosen based on ranking and the unfit compounds have been strained (Fig. 5).

3.5 Virtual Screening

Before starting the docking, we tested whether the docking program GOLD²⁸ was able to reproduce reasonable results on Mip proteins by docking rapamycin back into the crystal structure of *Legionella* Mip. The binding mode found in the crystal structure could be reproduced (RMSD: 1.2 Angstrom) (Fig. 6) and GOLD was used to dock the top 150 compounds after pharmacophore filtering. The crystal structure of *Legionella* Mip as well as our homology models of *Chlamydia trachomatis* Mip and mutant structures of *Chlamydia trachomatis* Mip were used as receptors for the docking. Further, we have docked with the hit compounds (150) with native and mutants of *C.*

RSC Advances

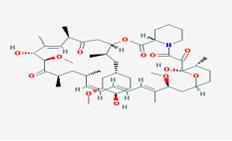
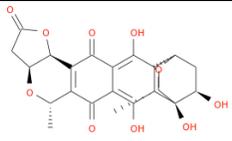
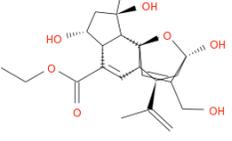
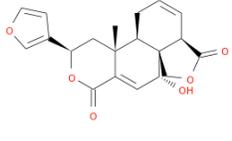
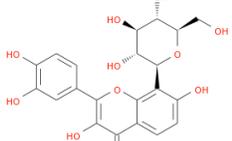
trachomatis Mip and to determine the binding affinity. Interestingly, 8 compounds were selected based on the scoring and hydrogen bonding interactions (Table 2). Out of 8 compounds, 6 compounds were plant species (Table 2). 2 top scoring compounds (UNPD131087 and UNPD175753) from marine sources were chosen for Molecular Dynamic studies to analyze the stability of the native and mutant model of *Chlamydia trachomatis* Mip (Fig. 5). Due to extreme conditions of pH, salinity, temperature and stress synthesis of secondary metabolites are higher in marine species than terrestrial plants. At extreme condition only marine species can able to synthesize active compounds. Therefore, compounds from the marine sources are advantageous.

3.6 Molecular Docking

The native Cht Mip has the binding affinity of 85.75 with compound-1 (UNPD131087) (Table 2; Fig. 7). However, Both D170L and Y185A

mutant showed that the size of the cavity and binding affinity with compound-1 (UNPD131087) were significantly decreased (Table 3). D170, I187 and Y213 make single hydrogen bond with compound-1 (UNPD131087). Six residues are forming electrostatic interactions and 9 residues are forming van der Waals interactions with native Cht Mip-compound-1 (UNPD131087) complex (Table 2; Fig. 7). However, D170L mutant showed less number of interactions compared to native Cht Mip complex with compound-1 (UNPD131087). In D170L, At the 170th position is responsible for the major hydrogen bond interaction is lost and forming van der Waals interaction due to the mutagenesis. Y213A makes two hydrogen bond interactions and I187 form single hydrogen

Table 2: The chemical structure of the compounds and the binding affinity as calculated by using GOLD score for normal and mutated (D170L and Y213A) *C. trachomatis* Mip.

S.No	UNPD-ID	STRUCTURE	CHEMICAL NAME	M. FORMULA	M. WEIGHT	Cht Mip GOLD SCORE	D170L Cht Mip GOLD SCORE	Y213A Cht Mip GOLD SCORE
1.			RAPAMYCIN	C51H79O13	914.17186	37.7	34.52	30.13
2.	131087		GRANATACIN	C22H20O10	444.388	85.75	83.21	79.54
3.	175753		ETHYL ESTER OF PLUMARELLIC ACID	C22H30O7	406.469	85.60	82.47	78.86
4.	144852		6beta-hydroxy-7,8-dehydrobaccharicunaetin	C20H20O6	356.369	83.26	80.15	75.62
5.	166580		8-C-Glucopyranosyl-3,3',4',7,-tetrahydroxyflavone	C21H20O11	448.377	82.58	79.94	74.30

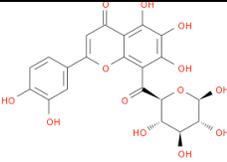
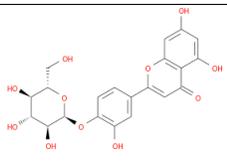
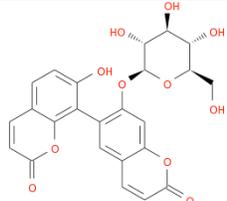
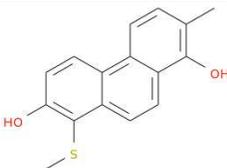
6.	191618		6-hydroxyluteolin 8-glucuronide	C21H18O13	478.36	80.30	76.97	72.55
7.	192068		Luteolin-4'-O-beta-glucosid	C21H20O11	448.377	79.38	75.83	70.67
8.	205547		7-(beta-D-glucopyranosyloxy)-7'-hydroxy-(6,8'-bi-2H-1-benzopyran)-2,2'-dione	C24H20O11	484.409	76.47	74.48	67.89
9.	161908		2,8-Dihydroxy-7-methyl-1-methylthiophenanthrene, 'Micrandrol C'	C16H14O2S	270.346	75.21	73.59	67.15

Table 3: Solvent accessibility surface area (SAS) for normal and mutated *C. trachomatis* Mip.

PROTEIN	TOTAL (SAS) Å	PER RESIDUE (SAS) Å (NORMAL)	PER RESIDUE (SAS) Å (MUTATION)	% CHANGE
Cht. Mip (NORMAL)	10885.26			
D170L Cht Mip (MUTANT)	10850.32	59.09	33.70	42.9
Y213A Cht. Mip(MUTANT)	10852.14	109.32	50.07	54.2

bond interaction. Five residues were forming electrostatic interactions and 8 residues were forming van der Waals interactions (Table 2; Fig. 7). Similarly, Y213A mutant lost hydrogen bond interaction at 213th position and form electrostatic interactions. D170 and I187 make single hydrogen bond interactions with compound-1. Six residues were able to make electrostatic interactions and 7 residues were forming van der Waals interactions. Solvent accessible surface area was calculated for Cht Mip (without mutation) and that after each of such *in silico* mutation mentioned above (Table 3). The native Cht Mip has the binding affinity of 85.60 with compound-2 (UNPD175753) (Table 2; Fig. 8). However, Both D170L and Y213A mutant showed that the size of the cavity and binding affinity with compound-2 (UNPD175753) were significantly decreased (Table 3). I187 and Y213 make two hydrogen bond with compound-2 (UNPD175753). Four residues are forming electrostatic

interactions and 10 residues are forming van der Waals interactions with native Cht Mip- compound-2 (UNPD175753) complex (Table 2). However, D170L mutant showed less number of interactions compared to native Cht Mip complex with compound-2 (UNPD175753). In D170L, at the 170th position is responsible for the major hydrogen bond interaction is lost and forming van der Waals interaction due to the mutagenesis. In case of Y213A mutant, Y213 makes two hydrogen bond interaction and I187 form two hydrogen bond interaction. Four residues were forming electrostatic interactions and 10 residues were forming van der Waals interactions (Table 2; Fig. 8). Similarly, Y213A mutant lost hydrogen bond interaction at 213th position and form van der Waals interactions. D170 and I187 make single hydrogen bond interactions with compound-1. Six residues were able to make electrostatic interactions and 10 residues were forming van der Waals interactions. Solvent accessible surface area (SAS) was calculated for Cht Mip (without mutation) and that after each of such *in silico* mutation mentioned above (Table 3). Subsequently, we studied the effects of substitution of Asp-170 to Leu- 170 and Tyr-213 to Ala-213 on SAS and binding affinities with rapamycin. The native Cht Mip has the binding affinity of 37.7 with rapamycin (Table 2; Fig. 9). However, the size of the cavity (42.9%) and binding affinity with rapamycin (34.52) for D170L and the size of the cavity (54.2%) and binding affinity with rapamycin (30.13) for Y213A were significantly decreased (Table 3). Tyr-159 make single hydrogen bond with rapamycin. Five residues are

RSC Advances

forming electrostatic interactions and 13 residues are forming van der Waals interactions with native Cht Mip- rapamycin complex (Table 2; Fig. 9). However, D170L mutant showed less number of interactions compared to native Cht Mip complex with rapamycin. In D170L, Tyr-159 make single hydrogen bond with rapamycin. 11 residues are forming electrostatic interactions and 11 residues are forming van der Waals interactions with native Cht Mip- rapamycin complex (Table 2; Fig. 9). At the 170th position is responsible for electrostatic interaction is lost and forming van der Waals interaction due to the mutagenesis. Gly217, Leu225 and Phe227 is responsible for van der Waals interaction formation is lost due to the mutagenesis. However, Y213A mutant showed less number of interactions compared to native Cht Mip complex with rapamycin. In D170L, Asp-170 make single hydrogen bond with rapamycin. Tyr-159 is responsible for hydrogen bond formation is lost. 4 residues are forming electrostatic interactions and 10 residues are forming van der Waals interactions with native Cht Mip- rapamycin complex (Table 2; Fig. 9). At the Val-186 position is responsible for electrostatic interaction is lost and forming van der Waals interaction due to the mutagenesis. Ser171, Gly214 and Leu225 is responsible for van der Waals interaction formation is lost and 2 new residues (Ile-187 and Val-186) makes van der Waals interaction due to the mutagenesis.

3.7 Simulation Study of Native and Mutant Complexes

Molecular dynamics simulations are used to understand better the knowledge on time evolution and the dynamic behavior of proteins conformations.^{28,29,30,31} It helps to get information on motions of individual atoms as a function of time and enable to explain the properties of molecules and the interactions between them.^{32,33,34} We performed MD simulations to study the effect of specific mutations on the structural characteristics of the Cht Mip protein-wild type and the two single mutants (D170L and Y213A) with inhibitors (compound-1 and compound-2) as explained in the Methods section. We have analysed the root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), solvent accessible surface area (SASA), number of hydrogen bonds (NH), and variation of secondary structure pattern between the native and mutant complexes (native, D170L and Y213A with compound-1 and compound-2). Six independent simulations were carried out for the native and mutant complexes for a total of 20 ns simulation time. We found that compound-1 with mutant D170L showed similar significantly deviation pattern compared to native in complex with compound-1 till the end of the simulation (Fig. 5A), whereas native and mutant Y213A complex with compound-1 tend to reach a higher equilibrium compared to native in complex with compound-1 (Fig. 10), Mutant Y213A remained distinguished throughout the simulation resulting in backbone RMSD of 0.1 to 0.4 nm. From 2 ns till the end of simulations, Cht Mip (native) showed RMSD values of 0.3nm-04nm. A small variation in the average

RMSD of native and mutants lead to the conclusion that the mutations could influence the changes in flexibility of dynamic behavior of protein. After 10 ns leads to stable trajectory throughout the simulation. Similarly, We have observed that compound-2 with both mutant D170L and Y213A showed similar significantly pattern till 12 ns and there will minor difference in trajectory leads to stable equilibrium through end of the simulation (Fig. 5A), whereas native and mutant Y213A complex with compound-1 tend to reach a higher equilibrium compared to native in complex with compound-1 (Fig. 10), Mutant Y213A remained distinguished throughout the simulation resulting in backbone RMSD of 0.1 to 0.4 nm. After 10 ns tend to reach stable trajectory in all the complexes throughout the simulation. The higher RMSD obtained for all the complexes was limited to 0.4 nm demonstrates that the simulations produced stable trajectories, and provided an appropriate basis for further investigation.

The radius of gyration is used to calculate the mass weighted root mean square distance of atoms from their centre of mass. The overall Cht Mip structure at various time points during the trajectory can be analyzed for the competence, shape and folding in the plot of Rg (Fig. 10). Throughout the simulation, for both the compound-1 and compound-2 with mutant protein D170L and Y213A exhibited a similar pattern of Rg value, out of which mutant D170L showed a higher deviation with Rg score of 1.25nm nm (Fig. 10). In the simulation of Y213A protein, we found that the aggregate with Rg score of 1.3nm is decreasing in complex with both the compound-1 and compound-2 followed by stabilisation after 10ns observed in all the complexes towards the end of the simulation. (Fig 10, Fig. 11).

We have measured the C-RMSF to observe the overall flexibility of atomic positions in the trajectory for native and mutant complexes. Mutant D170L complex with compound-1 shows significant change in the protein structure conformation (as compared in backbone RMSD) with increase in the C-RMSF being also observed (Fig. 10). This suggests that Y213A mutation causes the binding of compound-1&2 to make the backbone become more flexible for motion. Y213A-compound-2 mutation affects the interacting residues at the maximum of 0.46nm (Fig. 10). Furthermore, the flexibility of mutant Y213A was found to be in consistent with the native Cht Mip. This is due to the restriction of surrounding residues caused by the in the active site of protein due to mutation in Cht Mip. The results suggest that there exists a significant change of structural deviation in the mutant complex Y213A when compared to the native.

To measure the compactness of hydrophobic core forming mutant D170L complex with compound-1 and compound-2 indicates higher values of SASA (35 nm) with time when compared to the native protein (Fig. 9, Fig. 11). As in case of mutant Y213A in complex with

ARTICLE

RSC Advances

compound-1 and compound-2 has observed small differences of SASA [32nm]. Hence, the findings are in direct correlation with the changes of the solvent accessible surface area (SASA) that reflects the major exposure was due to the loss of hydrophobic contacts formation between the catalytic residues and compound-1 and compound-2.

The numbers of hydrogen bonds formed between cpd-1 and cpd2 and protein (native and mutant) during the MD simulation were also calculated (Fig.9, Fig. 10). From our analysis, it is well revealed that native complex forms more number of NH bond with native and D170L-cpd1 with an average of 2-3 hydrogen bonds(Fig. 10). While the mutant complex Y213A-compound-1(UNPD131087) exhibited less number of intermolecular hydrogen bonds of an average ~1-2. (Fig. 10). Number of hydrogen bond with native Cht Mip and D170L compound-2(175753) with an average of ~3-4 hydrogen bonds (Fig.10). while the mutant complex Y213A-compound-2 (UNPD175753) exhibited less number of hydrogen bonds with an average of ~1-2. (Fig. 11).

4 Conclusion

In the present work, the 3D structure of native and mutant *Chlamydia trachomatis* Mip was constructed based on the X-ray structure of *Legionella pneumophila* Mip using homology modeling. The model was then validated and further used for docking analysis. We developed a pharmacophore model based on the important features based on rapamycin and FK506. The new pharmacophore model was used to superimpose with the Universal Natural product database (UNPD) database compounds resulting common chemical features containing compounds were chosen. Further the lead molecules were docked into the active site of native and mutant *Chlamydia trachomatis* Mip. Docking analysis showed that native *Chlamydia trachomatis* Mip has the highest GOLD score of 85.75 in terms of compound-1(UNPD131087) binding compared to D142L mutant that has the 83.21 and Y213A that has 79.54, Similarly for compound-2(UNPD175753) has the GOLD score of 85.60 and D142L has 82.47 and Y213A has 78.86 which indicates that the binding affinity is highly affected in the both the mutants due to changes in the conformation of the active site by which resistance has been developed. Molecular Dynamics simulations analysis infers changes in the binding pattern and structural modification in the binding site by means of Rg, SASA, RMSD and RMSF. Hence, this study gives insight into the impact of novel mutation on the activity of this protein, which can be attributed to the drug resistance observed. This study will help shed light on designing novel inhibitors for Mip.

Acknowledgement

We performed Docking Simulation studies using High Performance Computing (HPC) facility [DST -PURSE (Grant No. SR/FT/LS-113/2009)] at Bharathidasan University, Trichy, India

References

- 1 Centers for Disease Control and Prevention. 1997. *Chlamydia trachomatis* genital infections—United States, 1995. *JAMA* **277**, 952-953.
- 2 D. Corsaro and G. Greub, *Clin. Microbiol. Rev.*, 2006, **19**, 283-297.
- 3 D. C. Mabey, A. W. Solomon and A. Foster, *Lancet*, 2003, **362**, 223-229.
- 4 A. G. Lundemose, S. Birkelund, P. M. Larsen, S.J.Fey and G. Christiansen, *Infect. Immun.*, 1990, **58**, 2478-2486.
- 5 R.C. Brunham, J. Bwayo, G. Maitha, I. Maclean, C. Yang, C. Shen, S. Roman, N.J. Nagelkerke and M. Cheang, *J. Infect. Dis.*, 1996, **173**, 950-956.
- 6 J.T. Grayston, M.B. Aldous, A. Easton, S.P. Wang, C.C. Kuo, L.A. Campbell and J. Altman, *J. Infect. Dis.*, 1993, **168**, 1231-1235.
- 7 A. Riboldi-Tunncliffe, B. Konig, S. Jessen, M.S. Weiss, J. Rahfeld, J. Hacker, G. Fischer and R. Hilgenfeld, *Nat. struct. Biol.*, 2001, **8**, 779-783.
- 8 M. V. Kalayoglu, P. Libby and G. I. Byrne, *JAMA.*, 2002, **288**, 2724-2731.
- 9 B. P. Pedro José, M. Cristina Vega, G.R. Elena, F.C. Rafael, M.R. Sandra, F.X. Gomis-Rüth, G. Antonio and C. Miquel, *EMBO Rep.*, 2002, **3**, 88-94.
- 10 A. G. Lundemose, J. E. Kay and J. H. Pearce, *Mol. Microbiol.*, 1993, **7**, 777-783.
- 11 A. Ceymann, M. Horstmann, P. Ehses, K. Schweimer, A.K. Paschke, A M. Steinert and C. Faber, *BMC Struct. Biol.*, 2008, **8**, 17. 7.
- 12 F. A. Saul, J.P. Arié, N.B. Vulliez-le, R. Kahn, J.M. Betton and G.A. Bentley, *J. Mol. Biol.*, 2004, **335**, 595-608.
- 13 E. Wintermeyer, B. Ludwig, M. Steinert, B. Schmidt, G. Fischer and J. Hacker *Infect. Immun.*, 1995, **63**, 4576-4583.
- 14 R. Vijayan, N. Subbarao and B.N. Mallick, *Am. J. Biochem. Biotech.*, 2007, **3**, 216-224.
- 15 R. Vijayan, N. Subbarao and N. Manoharan, *Interdiscip. Sci. Comput. Life Sci.*, 2015, **7**, 1-8.
- 16 R. Vijayan, N. Subbarao and N. Manoharan, *J. Marine. Biosci.*, 2015, **1**, 50-62.
- 17 G. Jiangyong, G. Yuanshen, C. Lirong, G. Yuan, L. Hui-Zhe and X. Xiaojie, *PLOS ONE*, 2013, **8**, e62839.
- 18 H. McWilliam, W. Li, M. Uludag, S. Squizzato, Y.M. Park and N. Buso, *Nucleic. Acids. Res.*, 2013, **41**, 597-600.

RSC Advances

- 19 N. Eswar, B. John, N. Mirkovic, A. Fiser, V.A. Ilyin, and U. Pieper, *Nucleic. Acids. Res.*, 2003, **31**, 3375-3380.
- 20 The PyMOL molecular graphics system, version 1.5.0.4 Schrödinger, LLC.
- 21 Accelrys Software Inc. Discovery studio modeling environment, release 4.0. Accelrys Software Inc., San Diego, 2013.
- 22 A.D. MacKerell, *J. Comp. Chem.*, 2004, **25**, 1584-1604.
- 23 R.A. Laskowski, J.A. Rullmann, M.W. Macarthur, R. Kaptein and J.M. Thornton, *J. Biomol. NMR*, 1996, **8**, 477-486.
- 24 N. Guex, and M.C. Peitsch, *Electrophoresis*, 1997, **18**, 2714-2723.
- 25 E. Capriotti, P. Fariselli and R. Casadio, *Nucl. Acids. Res.*, 2005, **33**, W306–W310
- 26 I.V. Tetko, J. Gasteiger, R. Todeschini, A. Mauri, D. Livingstone, P. Ertl, V.A. Palyulin, E.V. Radchenko, N.S. Zefirov, A.S. Makarenko, V.Y. Tanchuk and V.V. Prokopenko, *J. Comput. Aid. Mol. Des.*, 2005, **19**, 453-63.
- 27 G. Wolber and T. Langer, *J. Chem. Inf. Model*, 2005, **45**, 160-169.
- 28 E. Lindahl, B. Hess, D. van der Spoel. *Mol. Model. annual*. 2001, **7**, 306–317.
- 29 S. Lin, W. Gao, Z. Tian, C. Yang, L. Lu, J.L. Mergny, C.H. Leung and D.L. Ma, *Chem. Sci.*, 2015, **6**, 4284-4290.
- 30 E. Puziukynaitė, L. Wang, K.S. Schanze, J.M. Papanikolas and J.R. Reynolds, *Polym. Chem.*, 2014, **5**, 2363-2369.
- 31 H.J. Zhong, L. Lu, K.H. Leung, C.C.L. Wong, C. Peng, S.C. Yan, D.L. Ma, Z. Cai, H.M.D Wang and C.H. Leung, *Chem. Sci.*, 2015, **6**, 5400-5408.
- 32 N. Eshuis, B.J. van Weerdenbug, M.C. Feiters, F.P. Rutjes, S.S. Wikmenga and M. Tessari, *Angew. Chem. Int. Ed. Engl.*, 2015, **54**, 1481-1484.
- 33 D.L. Ma, D.S. Chan and C.H. Leung, *Acc. Chem. Res.*, 2014, **47**, 3614-3631
- 34 D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof and A.E. Mark, *J. Comput. Chem.* **26**, 1701–1718.
- 35 H.J.C. Berendsen, J.P.M. Postma, W.F. Gunsteren, J. Hermans *Intermolecular Forces.*, 1981, 331–342.
- 36 H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola and J.R. Haak *T. J. Chem. Phy.*, 1984, **81**, 3684–3690.
- 37 M. Parrinello and A. Rahman, *J. Appl. Phy.*, 1981, **52**, 7182–7190.
- 38 T. Darden, D. York and L. Pedersen, *T. J. Chem. Phy.*, 1993, **98**: 10089–10092.
- 39 J.Y. Xie, G.H. Ding and M. Karttunen, *Biochem. Biophys. Acta.*, 2014, **1838**, 994–1002.

```

CHT      MKNILSWMLMFAVALPIVGCNNGGGSQTSATEKSMVEDSALTDNQKLSRTFGHLLSRQLS
1FD9     -----ATDATSLATDKDKLSYSIGADLGKNFK
          :      *  **::***  ::*  *:::..

CHT      RTEDFSLLDLVEVIKGMQSEIDGQSAPLTDTE-----YEKQMAEVQKASFEAKCSENLA
1FD9     -NQGIDVNP EAMAKGMQDAMSGAQLALTEQQMKDVLNKFQKDLMAKRTAEFNKKADENKV
          .:  ::::  :  ***.  :.*  .  **::  :  ::::  :.*:*  *..**  .

CHT      SAEKFLKENKEKAGVIELEPNKLQYRVVKEGTGRVLSGKPTALLHYTGSGFIDGKVFDSSE
1FD9     KGEAFLTENKNKPGVVVL-PSGLQYKVINSGNGVKPGKSDTVTVEYTGRLIDGTVFDSTE
          ..*  **.***:*  **.*  *  **::***.***  .  .  *  .:***  :***.***:*

CHT      KNKEPILLPLTKVIPGFSQGMQMKEGEVRVLYIHPDLAYGTA---GQLPPNSLLIFEVK
1FD9     KTGKPFATFQVSQVIPGWTEALQLMPAGSTWEIYVPSGLAYGPRSVGGPIGPNETLIFKIH
          *  .  :*  :  ::***:***:*  *  *..  :::  ****  *  :  **  ***:::

CHT      LIEANDDNVSVTE
1FD9     LISVKKSS-----
          **..:..

```

Figure 1: Pairwise alignment of *Chlamydia trachomatis* Mip and the template (*Legionella pneumophila* Mip)

75x50mm (300 x 300 DPI)

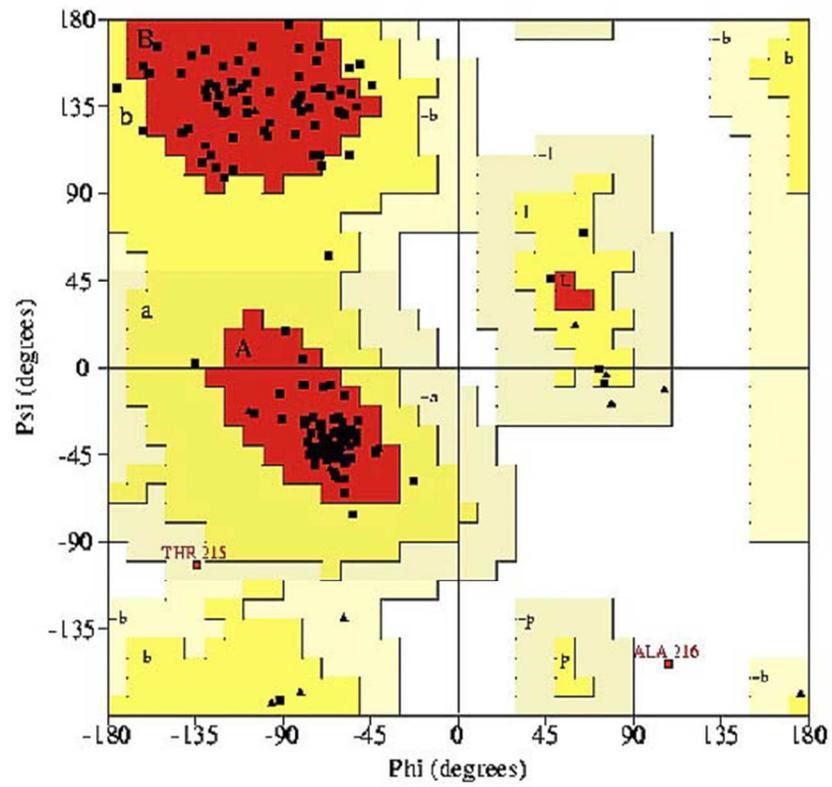


Figure 2: Ramachandran plot for *Chlamydia trachomatis* Mip.

54x49mm (300 x 300 DPI)

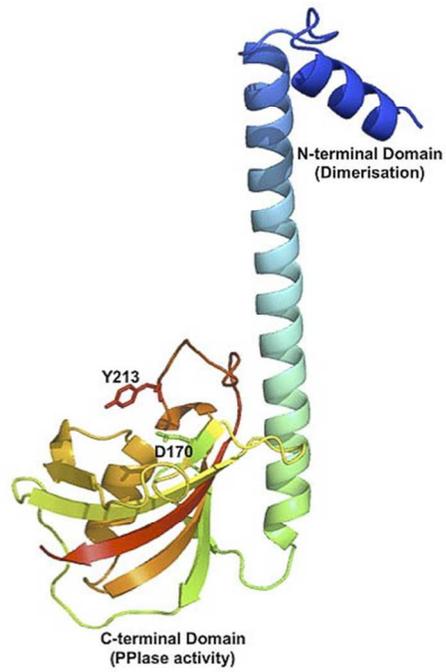


Figure 3: Three-dimensional structure of *Chlamydia trachomatis* Mip. Active site residues are represented in sticks.

75x54mm (300 x 300 DPI)

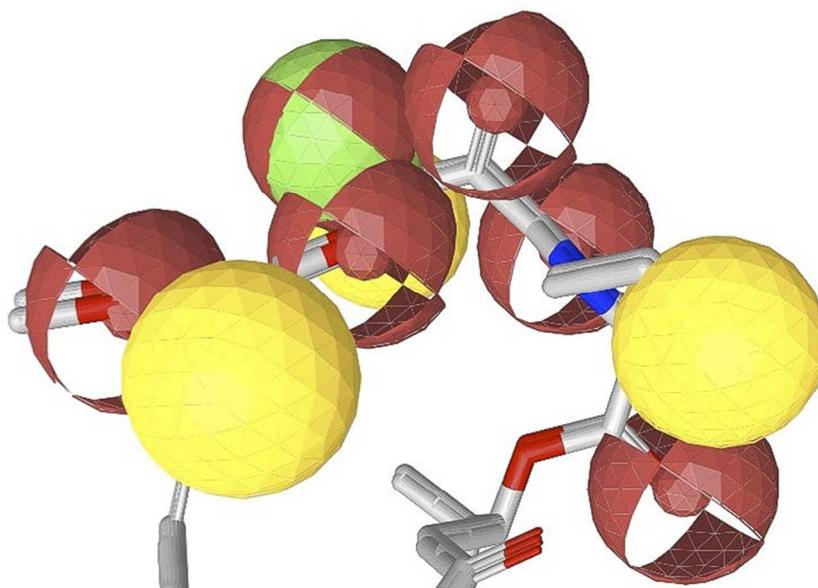


Figure 4: Pharmacophore model generated based on rapamycin and FK506. Color: hydrogen bond acceptors(HBAs) (green), hydrogen bond donors(HBDs) (red) and aromatic rings (yellow color).

75x59mm (300 x 300 DPI)

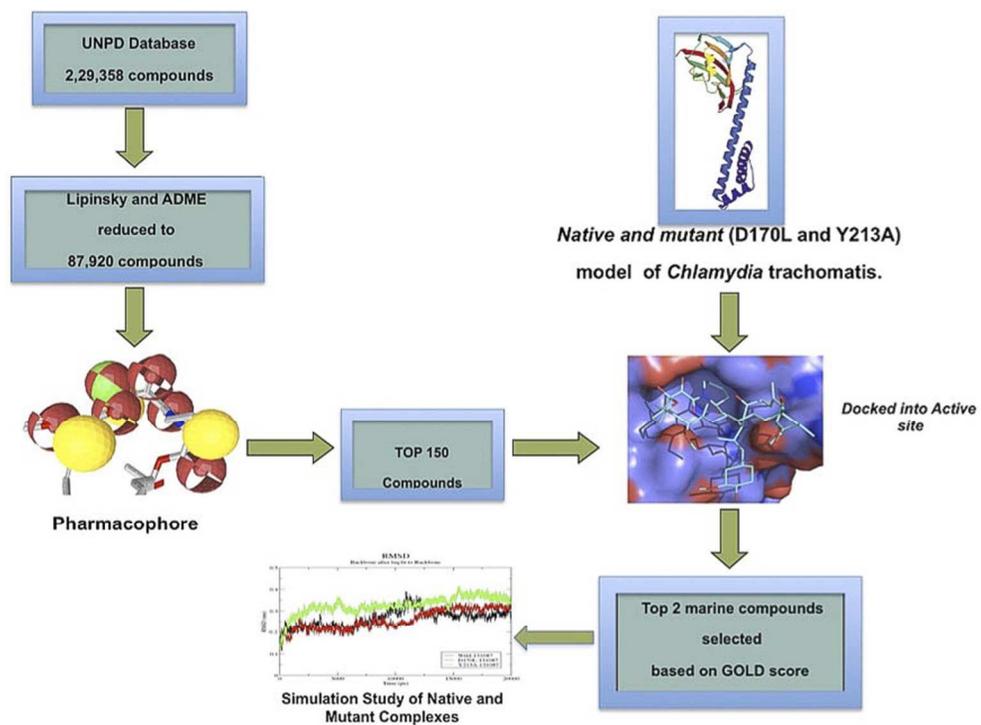


Figure 5: Virtual screening flowchart.

75x61mm (300 x 300 DPI)

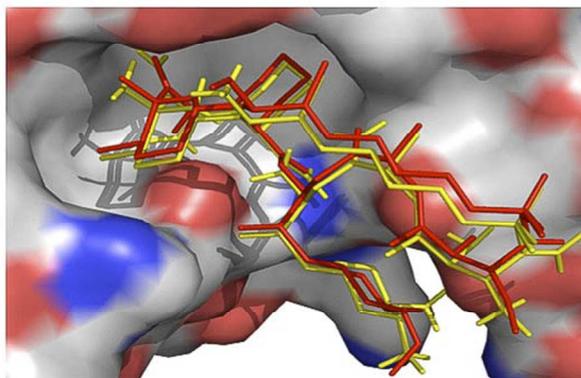


Figure 6: Binding mode reproduced for *Legionella pneumophila* Mip with rapamycin (RMSD 1.2Å). color: yellow: rapamycin(native) from pdb; red: docked conformation of rapamycin.

75x40mm (300 x 300 DPI)

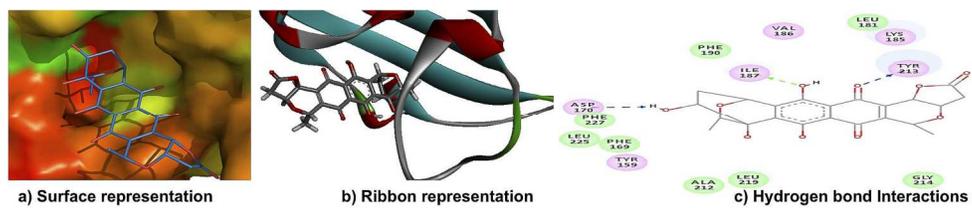


Fig 7A. Docked conformation of Cht Mip (native) with compound-1(UNPD131087)

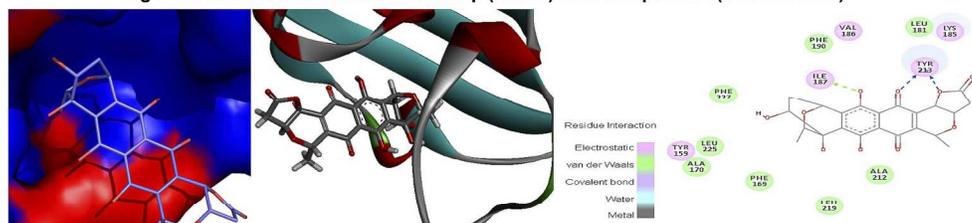


Fig 7B. Docked conformation of D170L Cht Mip (mutant) with compound-1(UNPD131087)

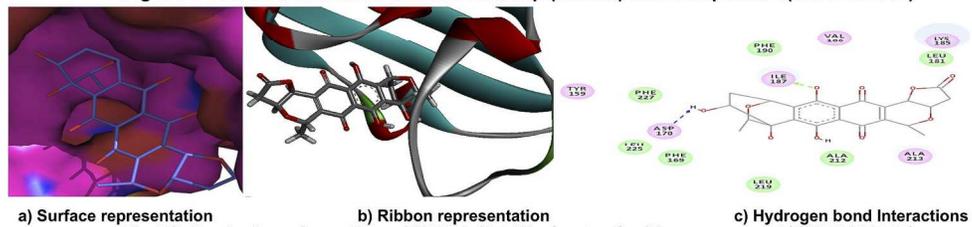
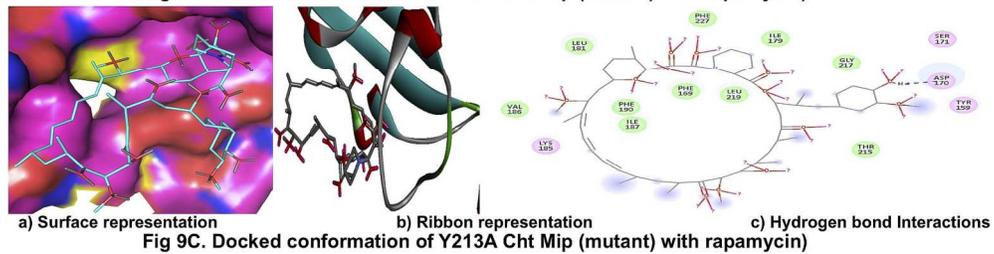
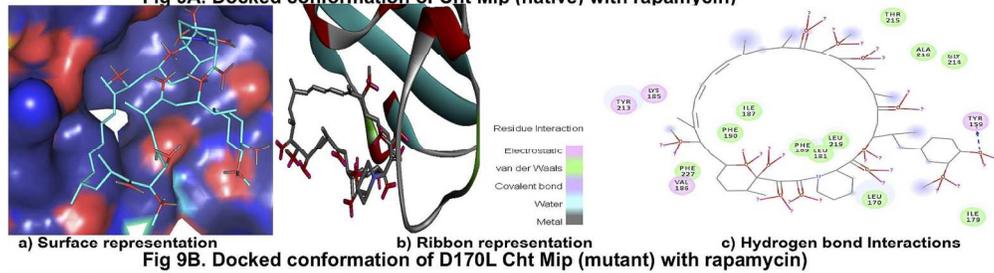
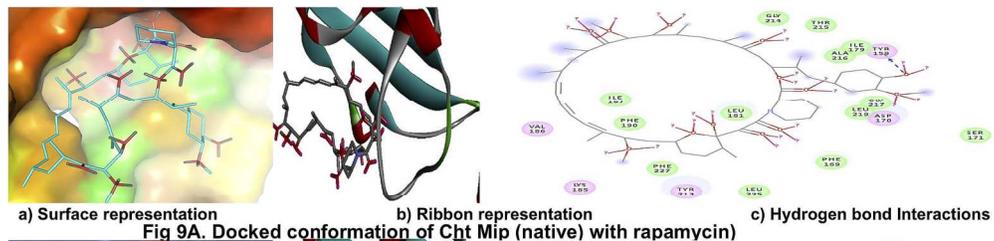


Fig 7C. Docked conformation of Y213A Cht Mip (mutant) with compound-1(UNPD131087)

173x128mm (300 x 300 DPI)



170x128mm (300 x 300 DPI)

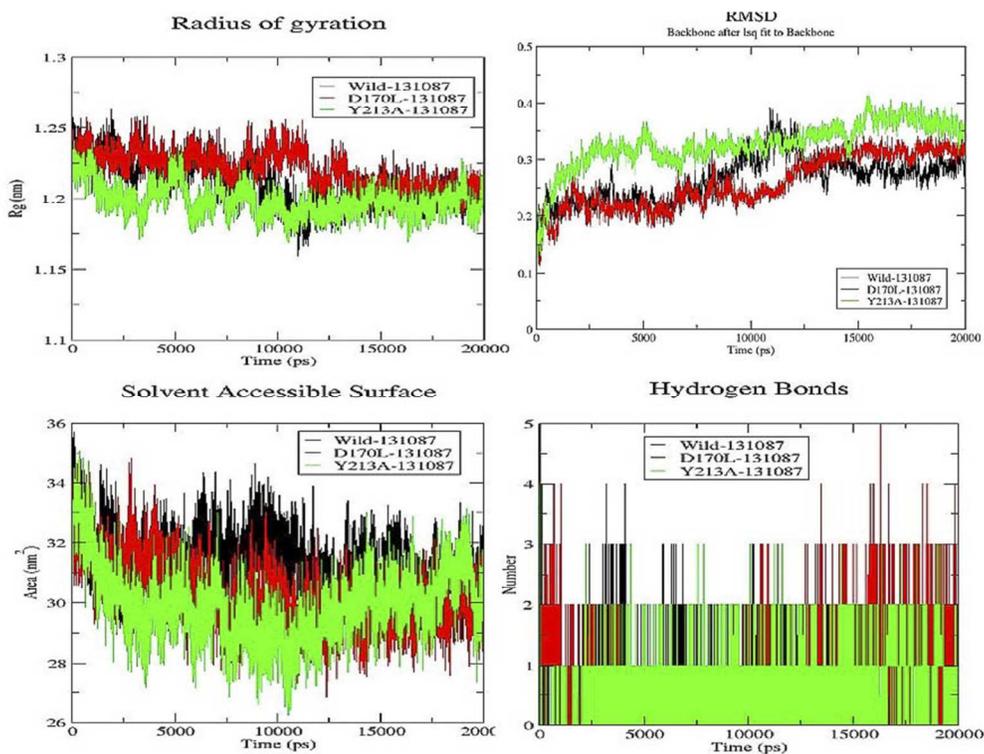


Figure 10: Analysis of Rg, RMSD, SASA, and Hydrogen bond interactions of native and mutant Cht Mip-compound-131087 complex at 20000 ps. (a) Rg of the protein backbone over the entire simulation. The ordinate is Rg (nm), and the abscissa is residue. b) Time evolution of backbone RMSDs of the native and mutant structures. c) The ordinate is SASA (nm²), and the abscissa is time (ns). The symbol coding scheme is as follows: native (green colour), mutant D170L (red colour), and Y213A (blue colour).

75x72mm (300 x 300 DPI)

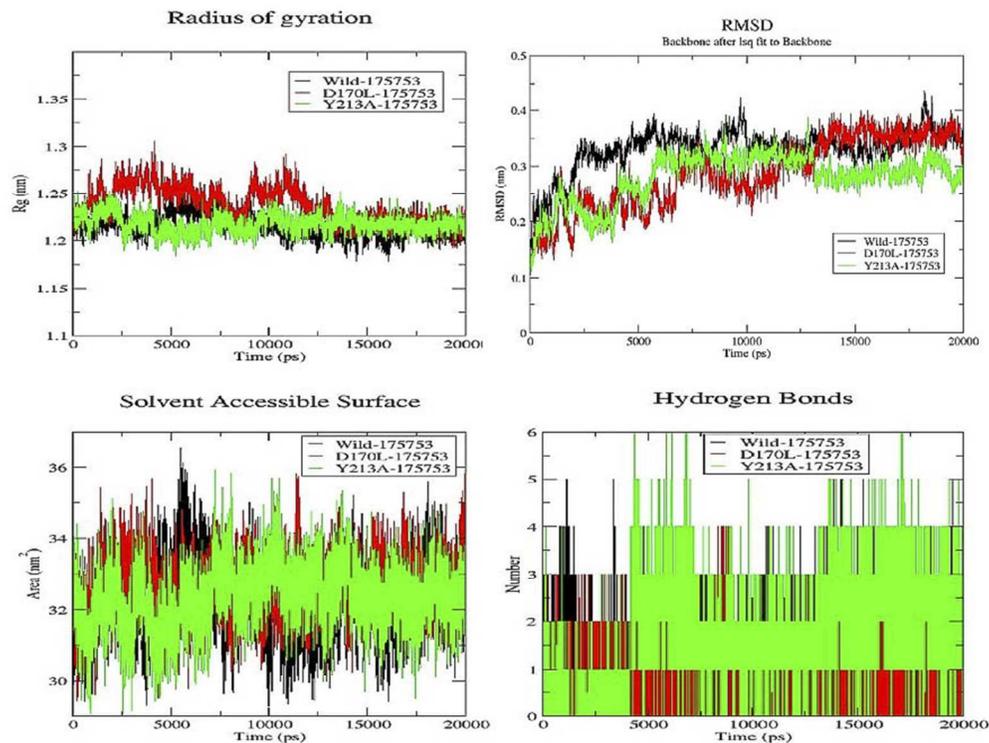


Figure 11: Analysis of Rg, RMSD, SASA, and Hydrogen bond interactions of native and mutant Cht Mip-compound-175753 complex at 20000 ps. (a) Rg of the protein backbone over the entire simulation. The ordinate is Rg (nm), and the abscissa is residue. (b) Time evolution of backbone RMSDs of the native and mutant structures. (c) The ordinate is SASA (nm²), and the abscissa is time (ns). The symbol coding scheme is as follows: native (green colour), mutant D170L (red colour), and Y213A (blue colour).

75x72mm (300 x 300 DPI)