BHARATHIDASAN UNIVERSITY

TIRUCHIRAPPALLI – 620 024

Phone No.: 0431-2407092, Fax : 0431-2407045, Email: office@bdu.ac.in

Website : www.bdu.ac.in

(Re-Accredited with "A" Grade by NAAC)

Dr.C. Thiruchelvam, Ph.D.,
REGISTRAR

Ref.No. 35560/E4/2011 dt. 09.08.2016.

To
The Secretary
University Grants Commission
Bahadurshah Zafar Marg
New Delhi – 110 002.

Sir,

Sub: UGC – Project "Identification of …… virtual screening". – Forwarding of Statement of Expenditure and Utilisation Certificate – Reg.
Ref: UGC Sanction Lr.No.F.No.41 -965/2012(SR) dt.26.07.2012.

****

With reference to the above, I am forwarding herewith the Statement of Expenditure and Utilisation Certificate for the period from 01.07.2012 to 31.12.2015, in respect of the UGC Project entitled "Identification of …… virtual screening" received from Dr. S. Parthasarathy, Associate Professor & Head, Department of Bioinformatics, Bharathidasan University, Tiruchirappalli – 620 024, for your kind consideration.

Yours faithfully,

9/8/16

REGISTRAR

Copy to:

3/14

Dr. S. Parthasarathy, Associate Professor & Head, Department of Bioinformatics, Bharathidasan University, Tiruchirappalli – 620 024

Identification of new inhibitors of penicillin binding protein 2B (PBP2B) of the resistant strains of *Streptococcus pneumonia* using structure based virtual screening

UGC F.NO. 41-965/2012 (SR)

(1July 2012 to 31December 2015)

Final Report of work done of UGC- Major Research Project

Submitted to

UNIVERSITY GRANTS COMMISSION
BAHADURSHAH ZAFAR MARG
NEW DELHI – 110 002

Submitted by

Dr. S. PARTHASARATHY
ASSOCIATE PROFESSOR & HEAD
DEPARTMENT OF BIOINFORMATICS
SCHOOL OF LIFE SCIENCES
BHARATHIDASAN UNIVERSITY
TIRUCHIRAPPALLI – 620 024
TAMIL NADU, INDIA

UNIVERSITY GRANTS COMMISSION
BAHADUR SHAH ZAFAR MARG
NEW DELHI – 110 002

ज्ञान-विज्ञान विमुक्तये

Executive summary of the final report of work done of UGC- Major Research Project

1. Title of the Project	:	Identification of new inhibitors of penicillin binding protein 2B (PBP2B) of the resistant strains of *Streptococcus pneumonia* using structure based virtual screening
2. NAME AND ADDRESS OF THE PRINCIPAL INVESTIGATOR	:	Dr. S. Parthasarathy Associate Professor & Head Department of Bioinformatics School of Life Sciences Bharathidasan University Tiruchirappalli – 620 024, Tamil Nadu.
3. NAME AND ADDRESS OF THE INSTITUTION	:	Department of Bioinformatics School of Life Sciences Bharathidasan University Tiruchirappalli – 620 024, Tamil Nadu.
4. UGC APPROVAL LETTER NO. AND DATE	:	F. No. 41-965/2012 (SR) dated 26-07-2012
5. DATE OF IMPLEMENTATION	:	01-07-2012
6. TENURE OF THE PROJECT	:	03 Years
7. TOTAL GRANT ALLOCATED	:	As 1st installment: Rs. 7,92,800/- As 2nd installment: Rs.2,14,141/-
8. TOTAL GRANT RECEIVED	:	Rs. 10,06,941/-
9. FINAL EXPENDITURE	:	Rs. 10,66,735/- (Includes 10% amount released from the Bharathidasan University)

10. TITLE OF THE PROJECT : Identification of new inhibitors of penicillin binding protein 2B (PBP2B) of the resistant strains of *Streptococcus pneumonia* using structure based virtual screening

11. OBJECTIVES OF THE PROJECT :

- To predict 3D structures of PBP2B of the mutant and resistant strains G54, Hungary 19A-6 and SP195 and validate the structures.
- To collect β-lactam antibiotics like compounds and develop pharmacophore model and build QSAR model and ligand database.
- To prepare set of ligands and target proteins for structure-based virtual screening.
- To dock PBP2B of the strains 5204, G54, Hungary 19A-6 and SP195 with the selected ligands through Glide docking.
- To verify the docked results, redock PBP2B using Induced-Fit docking with the top ranked ligands.
- To valuate biological activities of the screened compounds, build QSAR model with IFD score and IC_{50} values of compounds.

12. WHETHER OBJECTIVES WERE ACHIEVED: (GIVE DETAILS) YES – Detailed report enclosed in **Annexure – 1**

13. ACHIEVEMENTS FROM THE PROJECT :

The increase in penicillin resistant of *Streptococcus pneumoniae* governed by rapid mutations and interactions at molecular level spurred our interest in applying *in silico* approaches. By hierarchical virtual screening, the ZINC database comprising 1,677,620 compounds was screened to identify potential inhibitor against the PBP2B of highly resistant strain 5204. We found 25 promising diverse compounds which showed good affinity against the resistant 5204-PBP2B by using novel *in silico* paradigm, such as, automated virtual screening, customized docking scoring and free energy affinity checks. The important part is that when compounds were screened for novelty by using SciFinder, all were not posing any reported antimicrobial activity for penicillin binding protein 2B. The ligand 5-[(6-hydroxy- 1,2,3,4-tetrahydroisoquinolin-1-yl)methyl]benzene-1,2,3-triol with ID: ZINC59376795 is found to be the most promising inhibitor against the mutated resistant 5204-PBP2B, where it interacts

with active site residues Ser386 and Gly617 of the wild sensitive R6-PBP2B and share glide score of -6.731 kcal/mol. Significantly, the same ligand ZINC59376795 binds with the mutated residues Asn630 and Asn660 of the resistant 5204-PBP2B producing the glide score of -8.060 kcal/mol.

The standard MD, RMSD and protein-ligand contacts analysis was carried out through MD simulations at 5 ns. Asn630 and Asn660 are the mutated residues involved in hydrogen bond and water bridges with the ligand of the 5204-PBP2B. Further, Ser386, Thr616 and Gly617 are the catalytic motifs show more hydrogen bond interactions, which are the active site residues of the sensitive R6-PBP2B. From the results of Glide docking as well as MD simulations, the binding of ligand ZINC59376795 is comparatively better against resistant strain 5204-PBP2B forming more hydrogen contacts and other interactions. This present study indicates that all the 25 compounds are promising inhibitors but the top scoring ligand 5-[(6- hydroxy-1,2,3,4-tetrahydroisoquinolin-1-yl)methyl]benzene- 1,2,3-triol with ID: ZINC59376795 may be identified as the potential inhibitor against the sensitive R6-PBP2B and resistant 5204-PBP2B based on the analysis of molecular docking, free energy calculations and molecular dynamics simulations.

The top five best hit compounds (ZINC59376795, ZINC60175365, ZINC36922620, ZINC39550705 and ZINC36953975) obtained from our high throughput virtual screening (HTVS) analysis with resistant 5204-PBP2B and sensitive R6-PBP2B proteins. Penicillin G is used as a reference drug molecule throughout this study and it forms stable complex with sensitive R6-PBP2B protein. Here, similar stability is observed for the mutant resistant 5204-PBP2B protein structure with the top scoring compound ZINC592376795 which implies that this compound may act as an effective potential inhibitor against *Streptococcus pneumoniae*. The compound ZINC59376795 forms a total of five hydrogen bonds with resistant 5204-PBP2B protein of which three are with mutated residues of transpeptidase (TP) domain (ASN630 and ASN660). Similar to ZINC59376795, the other four compounds including penicillin G also form hydrogen bonds with mutated residue Asn660. The stability analysis of the complexes of wild and mutant forms evaluated through Density functional

theory analysis (DFT), prime-MM/GBSA binding free energy, Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), hydrogen bond formation and Principal Component Analysis (PCA) for a trajectory period of 16 ns and further MD simulations of top scoring compound ZINC59376795 with resistant 5204-PBP2B and sensitive R6-PBP2B confirmed the stability for 50 ns. Thus the compounds ZINC59376795, ZINC60175365, ZINC36922620, ZINC39550705 and ZINC36953975 are found to be a promising gateway for the further development of anti *pneumococcus* therapeutics.

14. SUMMARY OF THE FINDINGS (IN 500 WORDS):

- In this study increase in penicillin resistant of *Streptococcus pneumoniae* governed by rapid mutations and interactions at molecular level spurred our interest in applying *in silico* approaches
- In the study comparison of the sequence alignment of the transpeptidase (TP) domain of the PBP2B of sensitive R6 and resistant 5204 strains of *Streptococcus pneumoniae.* with a block of five mutations can be identified at 565-569 residues
- Virtual screening of 1,677,620 compounds from the ZINC database was screened to identify potential inhibitor against the PBP2B of highly resistant strain 5204 which yielded 25 promising diverse compounds which showed good affinity against the resistant 5204-PBP2B
- Top promising twenty five best hits that bind efficiently with both mutant and wild type proteins
- The ligand 5-[(6-hydroxy- 1,2,3,4-tetrahydroisoquinolin-1-yl)methyl]benzene-1,2,3-triol with ID: ZINC59376795 is found to be the most promising inhibitor against the mutated resistant 5204-PBP2B, where it interacts with active site residues Ser386 and Gly617 of the wild sensitive R6-PBP2B and share glide score of -6.731 kcal/mol. Significantly, the same ligand ZINC59376795 binds with the mutated residues Asn630 and Asn660 of the resistant 5204-PBP2B producing the glide score of -8.060 kcal/mol.

- Molecular docking and dynamics simulations carried out for the top five best hits and the reference drug molecule penicillin G available as reference with 5204-PBP2B and sensitive R6-PBP2B proteins
- The top hit compound (ZINC59376795) form maximum of five hydrogen bonds with resistant 5204-PBP2B protein and other compounds form an average of three hydrogen bonds with this protein
- All the five compounds and penicillin G have hydrogen bond interactions with the mutated residue Asn660 in the resistant 5204-PBP2B. In the case of sensitive R6-PBP2B, Asn445, which is a key active residue, forms hydrogen bond interactions with most of these compounds
- DFT analysis of the five best potential compounds showed the minimal HOMO-LUMO Gap (HLG) with the average energy difference of ~ 0.16 eV which reveals the molecular reactivity and stability of these novel compounds
- A better rigidity and stability is observed during the RMSD analysis of wild type sensitive R6-PBP2B with reference drug molecule penicillin G. A similar finding is observed while considering the RMSD analysis of the mutant structure resistant 5204-PBP2B with the top scoring compound ZINC592376795 which implies that this compound may act as an effective potential inhibitor against *Streptococcus pneumoniae*
- The trace values for the resistant 5204-PBP2B complex, the compounds ZINC59376795 and penicillin G are found to be 74.9631 and 110.192 nm^2, respectively. Penicillin G has higher trace value while compared with five other compounds
- Porcupine plot of PCA movements of the compound ZINC59376795 with resistant 5204-PBP2B complex conformation changed to open state when binding with the mutated residues of TP domain and resistant 5204-PBP2B complex with penicillin G, the conformation changed to close state when binding with one mutated residue.

- Further MD simulation carried out for 50 ns for the best compound ZINC59376795 with resistant 5204-PBP2B confirmed the formation of stable complexes even after 16 ns
- The strains 5204 and G54 in which the mutations occur till the C-terminus end of the TP domain are highly resistant
- The other two partially mutated strains Hungary19A- 6 and SP195 in which the mutations are observed only till the middle of the TP domain are intermediate resistant
- We have generated the 3D structures of PBP2B of the other mutant strains, G54, Hungary19A-6 and SP195 using homology modeling with 2WAF as the template
- It is observed that PBP2B of R6 (2WAF) having 97% identity to the mutant strains, G54, Hungary19A-6 and SP195 and hence R6 (2WAF) protein was chosen as the template
- Three dimensional structures of other mutant strains and to study the molecular interactions with top scoring compounds (ZINC59376795, ZINC60175365, ZINC36922620, ZINC39550705 and ZINC36953975) through *in silico* approach
- Based on the glide docking results, it was observed that all five compounds interacts with penicillin binding protein 2B of the resistance strains G54, Hungary 19A-6 and SP195 of *Streptococcus pneumoniae*.

15. CONTRIBUTION TO THE SOCIETY :

Streptococcus pneumoniae is one of the important virulent pathogen and its virulence is governed by rapid mutations at the molecular level. The conventional antibiotics are not effective and the resistance of this organism poses a serious threat to the society. We identified potential inhibitors ZINC59376795, ZINC60175365, ZINC36922620, ZINC39550705 and ZINC36953975 against the resistance strain 5204-PBP2B G54, Hungary 19A-6 and SP195 of *Streptococcus pneumoniae*. These compounds could be used as potential drug candidates against *Streptococcus pneumoniae* and can be further validated for its drug efficacy using experimental approaches.

16. WHETHER ANY Ph.D. ENROLLED/ PRODUCED OUT OF THE PROJECT : Yes – 01 (Please see **Annexure – 2**)

17. NO. OF PUBLICATIONS OUT OF THE PROJECT (PLEASE ATTACH RE-PRINTS) : Papers published : 03, Poster presented in conferences: 02

Papers published: 03 (Copies Attached in **Annexure – 3**)

1. Suvaithenamudhan S and **Parthasarathy S**. Structure Based Virtual Screening for the Identification of Potential Inhibitors for Penicillin Binding Protein 2B of the Resistant 5204 Strain of *Streptococcus pneumoniae*. *Current Bioinformatics*.2016 11(1), 66-78.

2. Suvaithenamudhan S and **Parthasarathy S**. Molecular dynamics simulation of novel potential inhibitors for penicillin binding protein 2B of the resistant 5204 strain of *Streptococcus pneumoniae. Current ComputerAided Drug Design.* (Communicated), 2016.

3. Bagavathy Shanmugam Karthikeyan, Suvaiyarasan Suvaithenamudhan, Mohammad Abdulkader Akbarsha and **Subbiah Parthasarathy**. Analysis of species selectivity of human, mouse and rat Cytochrome P450 1A and 2B subfamily enzymes using molecular modeling, docking and dynamics simulations. *Cell Biochemistry and Biophysics.*(under revision), 2016.

Poster presented in conferences: 02 (Please see **Annexure – 4**)

1. **Recent Advances in Computational Drug Design held at Indian Institute of Science, Bangalore on 16th Sep-17th Sep 2013**

   Poster Title: On the New Inhibitors for the Resistant 5204 Strain of Penicillin Binding Protein 2B (PBP2B) of *Streptococcus pneumonia* through Structure-Based virtual Screening

2. **2015 NextGen Genomics, Biology, Bioinformatics and Technologies (NGBT) Conference, HICC, Hyderabad on 1st-3rd Oct-2015**

   Poster Title: Structure-Based virtual screening for the identification of potential inhibitors for penicillin binding protein 2B of the resistant 5204 strain *Streptococcus pneumonia.*

(PRINCIPAL INVESTIGATOR)

Dr. S. Parthasarathy
Principal Investigator
Associate Professor and Head
Department of Bioinformatics
Bharathidasan University
Tiruchirappalli - 620 024.
Tamil Nadu, India

(REGISTRAR/PRINCIPAL)

REGISTRAR
BHARATHIDASAN UNIVERSITY
TIRUCHIRAPPALLI - 620 024.

UNIVERSITY GRANTS COMMISSION
University: Bharathidasan University

F. No: **41-965/2012 (SR)** Date : **26.07.2012 & 23.06.2015**
Statement of Actual Expenditure during 01.07.2012 to 31.12.2015

Consolidated Statement of Expenditure incurred during 01.07.2012 to 31.12.2015 in the Department of Bioinformatics, Bharathidasan University under **UGC Major Research Project - "Identification of .... virtual screening"**

S. No	Items	Total Grant Approved by the UGC in Rs.	Grant Released		Total Grant Received Rs.	Total Expenditure Rs.	Unspent balance amount in Rs.	Excess Expenditure
			1st Installment	2nd Installment				
A. 1.	**Non-Recurring** Books &Journals	60,000/-	60,000/-	-Nil-	60,000/-	59,818/-	182/-	-Nil-
2.	Equipments	3,50,000/-	3,50,000/-	-Nil-	3,50,000/-	3,50,000/-	-Nil-	-Nil-
B 3	**Recurring** Honorarium to Principal Investigator	-Nil-	-Nil-	-Nil-	-Nil-	-Nil-	-Nil-	-Nil-
4	Project Fellow	4,77,935/-	2,64,000/-	1,66,141/-	4,30,141/-	4,77,485/-	-Nil-	-47,344/-
5	Chemicals/ Consumables/ Glassware's	-Nil-	-Nil-	-Nil-	-Nil-	-Nil-	-Nil-	-Nil-
6	Hiring Services	-Nil-	-Nil-	-Nil-	-Nil-	-Nil-	-Nil-	-Nil-
7	Contingency	60,000/-	30,000/-	24,000/-	54,000/-	59,765/-	-Nil-	-5,765/-
8	Travel	60,000/-	30,000/-	24,000/-	54,000/-	32,456/-	21,544/-	-Nil-
9	Overhead charges	58,800/-	58,800/-	-Nil-	58,800/-	58,800/-	-Nil-	-Nil-
10	**Total**	**10,66,735/-**	**7,92,800/-**	**2,14,141/-**	**10,06,941/-**	**10,38,324/-**	**21,726/-**	**53,109/-**

1. Amount Received : Rs. 10,66,735/-
2. Total Expenditure on 30th November, 2015 : Rs. 10,06,941/-
3. Unspent Balance Amount : Rs. 21,726/-
4. Excess Expenditure to be reimbursed : Rs. 31,383/-

G.L. & ASSOCIATES (FRN-013540S)
CHARTERED ACCOUNTANTS
No.7, Williams Road, Cantonment,
Trichy - 620 001. Ph: +91-431-2413113

CERTIFICATE

1. Certified that the grant has been utilized for the purpose for which it was sanctioned and in accordance with the terms and conditions attached to the grant.
2. If as a result of check or audit objection, some irregularity is noticed at a later stage, action will be taken to refund, adjust or regularize the objected amount.

PRINCIPAL INVESTIGATOR
Dr. S. Parthasarathy
Principal Investigator
Associate Professor and Head
Department of Bioinformatics
Bharathidasan University
Tiruchirappalli - 620 024.
Tamil Nadu, India

FINANCE OFFICER
25/7/16
Finance Officer
BHARATHIDASAN UNIVERSITY,
TIRUCHIRAPPALLI - 620 024.

REGISTRAR
9/8/16
REGISTRAR
BHARATHIDASAN UNIVERSITY
TIRUCHIRAPPALLI 620 024
7/14

UNIVERSITY GRANTS COMMISSION
BAHADUR SHAH ZAFAR MARG
NEW DELHI – 110 002

Utilization certificate

It is certified that the amount of **Rs.10,38,324/-** (Ten Lakhs thirty eighty thousand three hundred and twenty four only) has been utilized as on **31.12.2015** from the first and second installment of **Rs. 10,06,941/-** (Ten Lakhs six thousand nine hundred and forty one only) received from UGC out of the total grant of **Rs.10,66,735/-** sanctioned to the Department of Bioinformatics, Bharathidasan University, Tiruchirappalli - 24 under the scheme of support for Major Research Project entitled **"Identification of new inhibitors of penicillin binding protein 2B (PBP2B) of the resistant strains of *Streptococcus pneumonia* using structure based virtual screening"** vide UGC letter No. **F. 41-965/2012(SR)** dated: **23.06.2015.** A sum of **Rs.10,38,324/-** has been utilized for the purpose for which it was sanctioned and in accordance with the terms and conditions laid down by the University Grants Commission. The excess expenditure incurred is Rs.53,109/- out of which the unspent balance of Rs.21,726/- may be adjusted and the remaining excess expenditure of Rs.31,383/- has to reimbursed out of the remaining (10% of sanctioned) grant Rs.59,794/- by the UGC, New Delhi.

If as a result of check or audit objection some irregularities are noticed at a later stage, action will be taken to refund, adjust or regularize the objected amount.

SIGNATURE OF THE PRINCIPAL INVESTIGATAR
(Seal)
Dr. S. Parthasarathy
Principal Investigator
Associate Professor and Head
Department of Bioinformatics
Bharathidasan University
Tiruchirappalli - 620 024.
Tamil Nadu, India

9/8/16

SIGNATURE OF THE REGISTRAR/PRINCIPAL
(Seal)

REGISTRAR
BHARATHIDASAN UNIVERSITY
TIRUCHIRAPPALLI - 620 024

12/1/16

25/7/16.

STATUTORY AUDITOR
(Seal)
Finance Officer,
BHARATHIDASAN UNIVERSITY,
TIRUCHIRAPPALLI - 620 024.

12/1/17

LATHA GOVINDARAJULU
M.No. 219371
CHARTERED ACCOUNTANT

G.L. & ASSOCIATES (FRN 013540S)
CHARTERED ACCOUNTANTS
No.7, Williams Road, Cantonment,
Trichy - 620 001. Ph: +91-431-2413113

Latha Govindarajalu B.Sc., F.C.A
Proprietor

G.L. & ASSOCIATES
CHARTERED ACCOUNTANTS

Utilization certificate

It is certified that the amount of **Rs.10,38,324/-** (Ten Lakhs thirty eighty thousand three hundred and twenty four only) has been utilized as on **31.12.2015** from the first and second installment of **Rs. 10,06,941/-** (Ten Lakhs six thousand nine hundred and forty one only) received from UGC out of the total grant of **Rs.10,66,735/-** sanctioned to the Department of Bioinformatics, Bharathidasan University, Tiruchirappalli - 24 under the scheme of support for Major Research Project entitled **"Identification of new inhibitors of penicillin binding protein 2B (PBP2B) of the resistant strains of *Streptococcus pneumonia* using structure based virtual screening"** vide UGC letter No. **F. 41-965/2012(SR)** dated: **23.06.2015.** A sum of **Rs.10,38,324/-** has been utilized for the purpose for which it was sanctioned and in accordance with the terms and conditions laid down by the University Grants Commission. The excess expenditure incurred is Rs.53,109/- out of which the unspent balance of Rs.21,726/- may be adjusted and the remaining excess expenditure of Rs.31,383/- has to reimbursed out of the remaining (10% of sanctioned) grant Rs.59,794/- by the UGC, New Delhi.

If as a result of check or audit objection some irregularities are noticed at a later stage, action will be taken to refund, adjust or regularize the objected amount.

Place : Trichy

Date: 12.01.2017

LATHA GOVINDARAJALU M.No. 219871 CHARTERED ACCOUNTANTS

Chartered Accountant

No. 7, Williams Road, Cantonment,Trichy - 620 001. Ph.. : 0431-2413113 Mobile : 99432 - 15202
email : calathag@gmail.com

Department of Library & Information Science
Bharathidasan University Library
Tiruchirappalli – 620 024

Dr. S. Srinivasa Ragavan
Professor & Head

Phone:0431-2407037
email: bdulib@gmail.com

Librarian

01-03-2017

Certificate

This is to certificate that a copy of the final report of the Major Research Project submitted to UGC by Dr. S Parthasarathy , Head, Department of Bioinformatics entitled **Identification of new inhibitors ..... based Virtual screening** is submitted to the University Library and the same is placed for the reference in the Institutional Repository of University Library.

Dr. S Srinivasa Ragavan
Professor and Head
Dept. of Library and Information Science
University Librarian
Bharathidasan University
Tiruchirappalli-620 024

BHARATHIDASAN UNIVERSITY LIBRARY * TIRUCHIRAPPALLI 620 024 *

DEPARTMENT OF BIOINFORMATICS
SCHOOL OF LIFE SCIENCES
BHARATHIDASAN UNIVERSITY
TIRUCHIRAPPALLI - 620 024
TAMIL NADU, INDIA

Dr. S. PARTHASARATHY
Associate Professor and Head

Phone: +91 431 2407071 Extn. 655 (Off)
Mobile: +91 94435 33095
Fax : +91 431 2407045
Email : bdupartha@gmail.com
partha@bdu.ac.in

16 February 2017

The Director
University informatics center
Bharathidasan University
Tiruchirappalli - 620024

Sir,

Sub: UGC–Major Research Project-Final report – to place in our University website – reg.

I have completed a UGC Major Research Project entitled **"Identification of new inhibitors of penicillin binding protein 2B (PBP2B) of the resistant strains of *Streptococcus pneumonia* using structure based virtual screening"**. As per UGC- Major Research Project norms, the final report of the work done in the project has to be placed on the website of our University. So I request to place our report in our University website.

Thanking You,

Sincerely Yours,

(S. Parthasarathy)

Received the report will be uploaded in the University website shortly

for Dr.S.PARTHASARATHY
DIRECTOR
UNIVERSITY INFORMATICS CENTRE
BHARATHIDASAN UNIVERSITY
TIRUCHIRAPPALLI - 620 024

Annexure – 1

Annexure -1

a) Title of the Project

Identification of new inhibitors of penicillin binding protein 2B (PBP2B) of the resistant strains of *streptococcus pneumonia* using structure based virtual screening.

b) Brief description on the state of the art of the research topic

Streptococcus pneumoniae is a human pathogen which causes diseases through their virulence factors and multi-drug resistance. Infection by *S. pneumoniae* has been classically treated using β-lactam antibiotics but the extensive use of antibiotics leads to antibiotic resistance due increase of mutation in its target Penicillin binding proteins (PBPs). Among the primary resistant determinants of PBPs, PBP1A and PBP2X were already analyzed computationally in order to design novel inhibitors but so far no studies are carried out in PBP2B.

Experimentally determined three dimensional (3D) structures of PBP2B of both wild sensitive R6 strain, namely, R6-PBP2B (with PDB ID: 2WAF) and mutated resistant 5204 strain, namely, 5204-PBP2B (with PDB ID: 2WAE) are available in Protein Data Bank (PDB). In order to identify new inhibitors for the pneumococcal disease the highly resistant strains 5204 considered for the study, which is having 44 mutations present in the transpeptidase (TP) domain of resistant 5204-PBP2B sequence with respect to the type (mutation free) sensitive R6-PBP2B strain.

c) Definition of the problem

Virulence factors and multidrug-resistantance are the important properties of *streptococcus pneumoniae*. The macromolecular targets of β-lactam antibiotics are PBPs. The 3D structures of penicillin binding protein 2B (PBP2B) of highly resistant 5204-PBP2B strain is available in protein Data Bank (PDB) with PDB ID 2WAE (resistant). PBP2B has the TP domain with residues ranging from 313-680. It is the main target with three catalytic motifs Ser386-Val387-Val388-Lys389 (carries active site Serine), Ser443-Ser444-Asn445 and Lys615-Thr616-Gly617. The β-lactam antibiotics form the stable complex with the active site serine. The use of *in silico* methods complements drug discovery process which delivers new drug candidates more quickly at a lower cost. In this present study, the drug resistant due to versatility of PBP2B TP domain is focused and the application of *in silico* approaches like virtual screening, molecular docking and dynamics facilitates several clues to the problem. Significantly, the influence of structural features, entropic

and enthaplic factors are substantial for the binding of ligand with the target which ultimately scrutinizes plethora of ligands from the chemical databases and effectively facilitates in drug discovery process. Here many *in silico* analysis, such as, virtual screening, molecular docking, free energy calculations and molecular dynamics (MD) simulations have been carried out using Schrodinger suite software.

d) Scope of research work

S. pneumoniae is an important pathogenic organism to human. For the preliminary analysis, sequence alignment of the transpeptidase (TP) domain of the PBP2B of sensitive R6 and resistant 5204 strains, by analyzing the structural changes in the ligand binding site between sensitive R6-PBP2B and resistant 5204-PBP2B structures by superimposition. The study performed a hierarchical virtual screening of 16,77,620 compounds from ZINC database targeting the PBP2B of the mutated resistant 5204 strain of *S. pneumoniae* to identify potential inhibitors. Further, top five best hits compounds and the reference drug molecule penicillin G used for molecular docking (Glide XP docking), binding free energy calculation (prime-MM/GBSA) and molecular dynamics simulations (GROMACS) with mutated resistant 5204-PBP2B and wild sensitive R6-PBP2B to study the structural and dynamic behavior of potential inhibitors targeting TP domain.

e) Methodology

The objective of the work is the virtual screening, molecular docking, binding free energy calculations and molecular dynamics simulations targeting β-lactam antibiotic resistant 5204-PBP2B and sensitive R6-PBP2B of S. *pneumoniae.*

1) Structure based virtual screening for the identification of potential inhibitors for penicillin binding protein 2B of the resistant 5204 strain of *S. pneumoniae*

i. Computational protein structure preparation and receptor site identification

The X-ray crystal structures of the PBP2B of both the mutated resistant strain 5204 (PDB ID: 2WAE) and the wild sensitive strain R6 of *S. pneumoniae* (PDB ID: 2WAF) were downloaded from Protein Data Bank (PDB). It is well known that X-ray crystallographic structures may contain drawbacks, such as missing hydrogen atoms, amino acids or loops and lacking proper protonations, which was addressed. We have used Schrödinger Protein Preparation Wizard to rectify the assignment of bond orders, addition of hydrogen and remove water molecules. Further,

protonation of the protein at the biological pH is done by Epik module of Schrödinger. Assignment of the hydrogen bonding network and minimization were carried out using the OPLS_2005 force field method (Schrödinger, LLC, New York, 2013), where the minimization was set to terminate when the RMSD reaches a maximum cutoff of 0.30Å in order to reach least possible energy. The refined and optimized protein structures were processed using SiteMap module of Schrödinger to specify the location of receptor site using options to report maximum 5 sites that requires at least 15 site points using more restrictive definition of hydrophobicity with standard grid cropping site maps at 4Å from the nearest site point. The SiteMap module generated possible potential receptor site in the receptors (both wild sensitive R6-PBP2B and mutated resistant 5204-PBP2B) with quantitative site-score values for each site which considers size, enclosure, exposure, contacts, hydrophobicity, hydrophilicity and Hbond donor acceptor balance. The receptor grid box for carrying out molecular docking was set to active site residues which are Ser386, Val387, Val388, Lys389, Ser443, Ser444, Asn445, Lys615, Thr616 and Gly617 as cited in the literature were used as a centroid for generating Glide receptor grid generation (with the automatic setting of enclosing box) using van der Waals scaling setup of 0.8 and partial charge cutoff of 0.25 with centroid of grid as selected residues.

ii. Ligand datasets

The collection of virtual commercial compounds is available in ZINC database that contains over 21 million purchasable compounds for which virtual screening can be performed against potential target. In ZINC database, PubChem was chosen from the catalogs option and the 16,77,968 compounds with pH of mid range (6-8) were retrieved in SMILES format. The SMILES format (.smi) was converted into three dimensional structure file in Maestro format (.mae) by using the "Structure File Converter" tool of Schrödinger Maestro which results in 16,77,620 compounds. All 16,77,620 compounds were used in our study for virtual screening against resistant 5204-PBP2B using virtual screening protocol.

iii. Virtual screening workflow

The Virtual Screening Workflow (VSW) module of Schrödinger is used to run an entire sequence of jobs for screening large collections of compounds against one or more targets. The workflow includes ligand preparation using LigPrep, where options were set to generate possible ionization states of the molecules at biological pH using Epik, skipping duplicates, generating at most 32 possible stereoisomers with ring conformation to generate at most one least energy

structure. The workflow had filtering options to skip molecules that had reactive functional groups and generate druggability property calculation by QikProp and receptor grids generated earlier were browsed in the receptor grid tab. The docking was carried out using all the three precision modes of glide HTVS, SP and XP, where options were to retain the top 50% and then docking again with option to retain 70% of top ligands in SP and XP modes. The top compounds were processed for druggabilty check and cheminformatics clustering for chemical diversity.

iv. ADMET prediction and clustering

The compounds were tested for druggability by using QikProp properties calculated during the virtual screening so that we have to check whether ADMET (Adsorption, Distribution, Metabolism, Excretion and Toxicity) parameters are well within the range of acceptability or not. QikProp also provides ranges for comparison of a given compound's properties with those of 95% of known drugs. Important properties, which influence the ADME including molecular weight, solvent accessible volume, solvent accessible surface area (SASA), number of hydrogen bond donors and acceptors, number of atoms in rings, predicted water/gas partition coefficient (QPlogPw) and percentage human oral absorption were calculated for the obtained hits and based on the ranges of the 95% known drug properties we selected the most probable drug like compounds for final selection. In order to explore the structural diversity of compounds from the collection of hits and to obtain the centroids of each cluster, the linear binary fingerprints of all the molecules were calculated by which the hierarchical clustering was carried out using cheminformatics tool of Schrödinger, namely, Canvas using tanimoto similarity algorithm. Cluster representatives were further analyzed for their interaction patterns against both the target site of mutated resistant 5204-PBP2B and wild sensitive R6-PBP2B.

v. Binding energy evaluations

Binding energies of the docked complexes obtained from Glide XP docking were calculated using eMBrAcE MacroModel, version 10.2, Schrödinger, LLC, New York, 2013. Here, the OPLS (Optimized Potentials for Liquid Simulations) 2005 force field is used with no solvation with constant dielectric electrostatic treatment having dielectric constant value 1. Further, the minimization is performed using optimal method with maximum iterations of 2500, where convergence is set to gradient and convergence threshold is 0.05, where energy difference mode for

association energy is used. The energy difference is then calculated using the following Eq. 1,

$$\Delta E = E_{\text{complex}} - (E_{\text{protein}} + E_{\text{ligand}}), \quad (1)$$

where full effects of relaxation and salvation were included in this mode. The values of E_{protein} and E_{ligand} are the lowest energies found for any conformer of the protein and ligand respectively. ΔE is calculated for each complex conformation saved in the output structure file using the energy of the current complex conformation for E_{complex}.

vi. Molecular dynamics simulations

Both for mutated resistant 5204-PBP2B structure 2WAE and the wild sensitive R6-PBP2B structure 2WAF, we used the ligand-receptor complex structures obtained from molecular docking and were prepared with the Schrödinger Protein Preparation Wizard utility and explicit solvated molecular dynamics (MD) was carried out using Desmond. As Desmond is now integrated with Maestro modeling environment of Schrödinger, we performed the molecular dynamics simulations using Desmond instead of using GROMACS (GROningen MAchine for Chemical Simulations) or AMBER (Assisted Model Building with Energy Refinement) or VMD (Visual Molecular Dynamics). The prepared complex systems were solvated using the system builder using TIP4P aqueous solvation of orthorhombic solvent box with a cell size was approximately 51.7Å x 70.1Å x 53.8Å and a solvent buffer extending 10Å beyond the protein in all directions using OPLS 2005 force field. The systems were neutralized with counter ions which entailed adding different numbers of Na^+ (salt positive ion) or Cl^- (salt negative ion) ion to each system to generate the system which is minimized with maximum of 2500 iterations and convergence thresholds of 1 kcal/mol/Å. We used the NPT ensemble with the Nose-Hoover thermostat method set at a reference temperature of 300 K water solvent system. MD simulations were carried out for 5 ns and trajectories were analyzed for RMSD and to the stable residues of interactions against the target site.

2) Molecular dynamics simulation of novel potential inhibitors for penicillin binding protein 2B of the resistant 5204 strain of *S. pneumoniae*

For analyzing the structural and dynamic behavior of top five compound and reference drug molecule penicillin G with protein complexes by calculating the RMSD, RMSF, Hbond and PCA.

i. Binding free energy calculation

The binding free energy calculation was performed through prime-MM/GBSA (Molecular Mechanics/Generalized Born Surface Area). Prime uses a surface GB model employing a Gaussian surface instead of a van der Waals surface for better representation of the solvent accessible surface area. Binding energy (ΔG_{bind}) was calculated by the following equations,

$$\Delta G_{bind} = \Delta E + \Delta G_{solv} + \Delta G_{SA}\ , \quad (1)$$

$$\Delta E = E_{complex} - E_{protein} - E_{ligand}, \quad (2)$$

where, $E_{complex}$, $E_{protein}$, and E_{ligand} are the minimized energies of the protein-inhibitor complex, protein, and inhibitor, respectively.

$$\Delta G_{solv} = G_{solv}(complex) - G_{solv}(protein) - G_{solv}(ligand), \quad (3)$$

where, G_{solv}(complex), G_{solv}(protein), and G_{solv}(ligand) are the solvation free energies of the complex, protein, and inhibitor, respectively.

$$\Delta G_{SA} = G_{SA}(complex) - G_{SA}(protein) - G_{SA}(ligand), \quad (4)$$

where, G_{SA} (complex), G_{SA} (protein), and G_{SA} (ligand) are the surface area energies for the complex, protein, and inhibitor, respectively. The rational criteria for selection of best compounds based on scoring and interaction parameters are shown in XP docking which are further used for MD simulation studies.

ii. Molecular dynamics (MD) simulations of resistant 5204-PBP2B and sensitive R6-PBP2B

The MD simulations of all the systems were carried out using GROMACS 4.5 package. The topology files for the selected proteins were generated using the automated topology builder (ATB) in the framework of GROMOS96 43a1 force field for protein-ligand complex. The topology files for the ligands were generated using PRODRG 2.5 server. The ligand complex obtained from docking was solvated with single point charge (SPC) water model. The solvated system was subjected to 5000 steps of energy minimization employing the steepest descent algorithm. This step was followed by 1 nano second (ns) MD simulation, where the resistant 5204-PBP2B and sensitive R6-PBP2B with ligands complex were position restrained to equilibrate the water and ions under the influence of the solute. The production run was carried out for all the systems for 16 nano seconds (ns) using 2 femto second (fs) time step for the integration of equation of motion in the NPT ensemble at 300 K and at 1 atmospheric pressure, which was controlled using a V-rescale thermostat and Parrinello–Rahman Barostat, respectively. Bond lengths involving hydrogen atoms were constrained by using the Linear Constraint Solver (LINCS) algorithm. The Particle Mesh Ewald (PME) method was used to calculate the electrostatic interaction. The cut off

distances for the long-range electrostatic and van der Waals energy terms were set as 10 Å. The MD simulation coordinates of all the systems were saved at 2 ps interval for further analyses. Post processing and analyses were carried out using GROMACS analysis tools.

iii. Root-Mean-Square Deviation (RMSD)

The RMSD is used to measure the average change in displacement of a selection of atoms for a particular frame with respect to a reference frame. It was calculated for all frames in the trajectory. The RMSD for frame x is given by

$$RMSD_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(r_i'(t_x) - r_i(t_{ref})\right)^2}, \quad (5)$$

where N is the number of atoms in the atom selection t_{ref} is the reference time, (typically the first frame is used as the reference and it is regarded as time $t = 0$); and r' is the position of the selected atoms in frame x is recorded at time t_x. The procedure is repeated for every frame in the simulation trajectory.

iv. Root-Mean-Square Fluctuation (RMSF)

The RMSF is useful for characterizing local changes along the protein chain. The RMSF for residue i is given by

$$RMSF_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T} < \left(r_i'(t) - r_i(t_{ref})\right)^2 >}, \quad (6)$$

where T is the trajectory time over which the RMSF is calculated, t_{ref} is the reference time, r_i is the position of residue i; r' is the position of atoms in residue i after superposition on the reference, and the angle brackets indicate that the average of the square distance is taken over the selection of atoms in the residue.

v. Hydrogen Bond

The hydrogen bonds between protein and ligand were analyzed using the g_hbond utility in the GROMACS. The distance criterion for the hydrogen bonds is d ≤ 3.5 Å between donor and acceptor. The angle between donor and acceptor is greater than 30°.

vi. Principal Component Analysis

Principal component analysis (PCA) was performed for all the trajectories. The GROMACS inbuilt tools g_covar and g_anaeig were used for performing PCA analysis. The trajectory of an MD simulation was utilized to identify the motions of the sensitive R6-PBP2B and

resistant 5204-PBP2B models. We used principal component analysis to extract the principal modes involved in the motion of the protein molecule. A covariance matrix was assembled using a simple linear transformation in Cartesian coordinate space. A vectorial depiction of every single component of the motion indicates the direction of motion. For this, a set of eigenvectors was derived through the diagonalization of the covariance matrix. Each eigenvector has a corresponding eigenvalue that describes the energetic contribution of each component to the motion. The protein regions that are responsible for the most significant collective motions can be acknowledged through PCA.

f) Original contribution

Top hit compound 5-[(6-hydroxy-1,2,3,4-tetrahydroisoquinolin-1-yl)methyl]benzene-1,2,3-triol (ZINC59376795) is identified as the potential inhibitor against resistant 5204-PBP2B protein of *Streptococcus pneumoniae.* ZINC60175365, ZINC36922620, ZINC39550705 and ZINC36953975 are also found to be promising inhibitors gateway for the further development of anti pneumococcus therapeutics.

g) Conclusion

In this study, first virtual screening of 1,677,620 compounds was carried out to identify potential inhibitors against the Penicillin Binding Protein 2B (PBP2B) of the resistant 5204 strain of *Streptococcus pneumoniae* which resulted in top promising five best hits that bind efficiently with both mutant and wild type proteins. Next, molecular docking and dynamics simulations are carried out for the top five best hits and the reference drug molecule penicillin G with 5204-PBP2B and sensitive R6-PBP2B proteins. The top hit compound 5-[(6-hydroxy-1,2,3,4-tetrahydroisoquinolin-1-yl)methyl]benzene-1,2,3-triol (ZINC59376795) form maximum of five hydrogen bonds with resistant 5204-PBP2B protein and other compounds form an average of three hydrogen bonds with this protein. All the five compounds and penicillin G have hydrogen bond interactions with the mutated residue Asn660 in the resistant 5204-PBP2B. In the case of sensitive R6-PBP2B, Asn445, which is a key active residue, forms hydrogen bond interactions with most of these compounds.

The stability of the resistant 5204-PBP2B with the top five compounds evaluated through RMSD and RMSF analysis shows that throughout the time period of 16 ns the protein-ligand backbone stability was not affected by these compounds and is stable throughout the simulation period without any significant fluctuation. A better rigidity and stability is observed during the RMSD analysis of wild type sensitive R6-PBP2B with reference drug molecule penicillin G. A

similar finding is observed while considering the RMSD analysis of the mutant structure resistant 5204-PBP2B with the top scoring compound ZINC592376795 which implies that this compound may act as an effective potential inhibitor against *Streptococcus pneumoniae*. Principal component analysis is performed by the trace of diagonalized covariance matrix. The trace values for the resistant 5204-PBP2B complex, the compounds ZINC59376795 and penicillin G are found to be 74.9631and 110.192 nm^2, respectively. Penicillin G has higher trace value while compared with five other compounds. Further, based on the porcupine plot of PCA movements of the top scoring compound resistant 5204-PBP2B-ZINC59376795 complex depicts that the distribution of free form of resistant 5204-PBP2B is large compared to the resistant 5204-PBP2B-penicillin G complex bound form, which confirms that this compound may act as an effective potential inhibitor. Along with this top screening compound, the remaining four compounds presented in this study are also found to be a promising gateway for the further development of anti *pneumococcus* therapeutics.

References

1. Appelbaum PC. Antimicrobial resistance in *Streptococcus pneumoniae*: An overview. Clin Infect Dis. 1992; 15(1): 77-83.
2. Hakenbeck R, Kaminski K, König A, van der Linden M, Paik J, Reichmann P, Zähner D. Penicillin-binding proteins in beta-lactam-resistant *Streptococcus pneumoniae*. Microb Drug Resist. 1999; 5(2): 91-99.
3. Musher DM. A Pathogenetic categorization of clinical syndromes caused by *Streptococcus pneumonia*. In: Tuomanen El., editor. The *Pneumococcus*. Washington: ASM press; 2004; 211-220.
4. Klugman KP. Clinical relevance of antibiotic resistance in *pneumococcal* infections. In: Tuomanen El., editor. The *Pneumococcus*. Washington: ASM press; 2004; 331-338.
5. Zapun A, Contreras-Martel C, Vernet T. Penicillin-binding proteins and beta-lactam resistance. FEMS Microbiol Rev. 2008; 32(2): 361-385.
6. Contreras-Martel C, Dahout-Gonzalez C, Martins Ados S, Kotnik M, Dessen A. PBP active site flexibility as the key mechanism for beta-lactam resistance in pneumococci. J Mol Biol. 2009; 387(4): 899-909.

7. Miguet L, Zervosen A, Gerards T, Pasha FA, Luxen A, Distèche-Nguyen M, Thomas A. Discovery of new inhibitors of resistant *Streptococcus pneumoniae* penicillin binding protein (PBP) 2x by structure-based virtual screening. J Med Chem. 2009; 52(19): 5926-5936.
8. Suvaithenamudhan, S. & Parthasarathy, S. Structure Based Virtual Screening for the Identification of Potential Inhibitors for Penicillin Binding Protein 2B of the Resistant 5204 Strain of Streptococcus Pneumoniae. Current Bioinformatics. 2016; 11(6):66-78.
9. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms forHighly Efficient, Load-Balanced, and Scalable Molecular Simulation. J Chem TheoryComput. 2008; 4(3):435-447.
10. Mobley DL, Dill KA. Binding of small-molecule ligands to proteins: "what you see" is not always "what you get". Structure. 2009; 17(4):489-498.
11. Jakubík J, Randáková A, Doležal V. On homology modeling of the M_2 muscarinic acetylcholine receptor subtype. J Comput Aided Mol Des. 2013; 27(6):525-538.

Annexure – 2

பாரதிதாசன் பல்கலைக்கழகம்
திருச்சிராப்பள்ளி - 620 024

BHARATHIDASAN UNIVERSITY
PALKALAIPERUR
TIRUCHIRAPPALLI-620 024
TAMIL NADU

Off : 0431-2407092 0431-2407099 (P)
Fax : 0431-2407045 Email : office@bdu.ac.in
Grams :"BARD"
Website :www.bdu.ac.in

(Accredited with "A" Grade by NAAC)

Prof. S. ARUCHAMY, M.Sc., Ph.D.,
REGISTRAR i/c

Ref:35560/E4/2011/dated: 03 .10.2012

OFFICIAL MEMORANDUM

Sub: UGC Project "Identification ..screening" – Appointment of Project Fellow Order issued - Reg.

Mr.S.Suvaithenamudhan , is appointed as a "Project Fellow" to serve under the supervision of Dr.S.Parathasarathy, Associate Professor & Head , Department of Bioinformatics , and the Principal Investigator of the UGC Project, subject to the following conditions.

1. The appointment is initially for a period of one year on a fellowship of Rs.14,000/- p.m.+HRA and will be extended further, (for two more years or upto the tenure of the project, whichever is earlier) only if the performance of the appointee is found satisfactory.
2. He shall work under the overall direction and supervision of the Principal Investigator/Co-investigator of the project..
3. The appointment will take effect from the date on which the candidate reports for work to the Principal Investigator.
4. The candidate shall not accept or hold any appointment paid or receive any emoluments, salary, stipend etc., from any other source during the tenure of appointment. Those who are already in receipt of any payment from any other source will not be entitled to this appointment unless they are relinquished in favour of the post of "PROJECT FELLOW" as aforesaid.
5. If it appears at any time to the University that the progress of work or conduct of the candidate has not been satisfactory, the appointment may be suspended or withdrawn.
6. If the appointment is acceptable under the conditions stipulated, the candidate may report to Dr.S.Parathasarathy, Associate Professor & Head , Department of Bioinformatics , Bharathidasan University, Tiruchirappalli – 620 024, as aforesaid. He/She is requested to communicate his/her acceptance under the aforesaid terms and conditions within a week from the date of receipt of this memorandum, failing which, it will be presumed that he is not interested in the offer and the post will be otherwise filled.
7. No traveling or any other allowance will be paid for joining the project.

REGISTRAR

To
Mr. S.Suvaithenamudhan, 16/34, Ellaiyah street Mathalam pettai, Kumbakonam 612 001, Thanjavur .

Copy to: 1. Dr.S.Parathasarathy, Associate Professor & Head , Department of Bioinformatics, Bharathidasan University, Trichy – 620 024.
2.The Finance Officer, Bharathidasan University, Trichy – 620 024.

BHARATHIDASAN UNIVERSITY
TIRUCHIRAPPALLI – 620 024
Phone No.: 0431-2407092, Fax : 0431-2407045, Email: office@bdu.ac.in
Website : www.bdu.ac.in
(Accredited with "A" Grade by NAAC)

REGISTRAR

Ref.No.47432/Ph.D.K1/Bioinformatics/Full Time/January 2013/Date: 26.12.2012

To
Mr.S.Suvaithenamudhan
Research Scholar
Department of Bioinformatics
Bharathidasan University
Tiruchirappalli - 620 024.

Sir,

Ref :Ph.D Programme - provisional registration application No. 6275 dt.30.11.2012.

I am to inform you that you have been provisionally registered for the Ph.D. Degree under the Full Time category. As per the regulations of this University you have to carryout research work under Research Supervisor for a minimum period of two years and a maximum period of Four Years from the date of registration i.e. from 01.01.2013 to 31.12.2016.

Be it informed that you will be governed by the regulations, rules and conditions for the Degree of Doctor of Philosophy of this University operative at the time of your registration. The Regulations framed for the course work for the programme is enclosed along with this letter.

The Subject/Discipline of Research chosen by you is Bioinformatics and the broad topic of your research is "IDENTIFICATION OF NEW INHIBITORS OF CERTAIN PENICILLIN BINDING PROTEINS OF THE RESISTANT STRAINS OF STREPTOCOCCUS PNEUMONIAE USING STRUCTURE-BASED VIRTUAL SCREENING".

The subject/discipline cannot be subsequently changed. You are requested to quote the reference number cited above in all correspondence with the University regarding your registration.

Yours sincerely,

for REGISTRAR

Encl: as above.

P.T.O.

To Mr. S.S
2/1/13

BHARATHIDASAN UNIVERSITY
TIRUCHIRAPPALLI – 620 024
Phone No.: 0431-2407092, Fax : 0431-2407045, Email: office@bdu.ac.in
Website : www.bdu.ac.in
(Accredited with "A" Grade by NAAC

REGISTRAR

Ref.No.47432/Ph.D.1/Bioinformatics/January 2013/Full Time/Confirmation/Date: 17.12.2014

To
Mr. S. Suvaithenamudhan
Research Scholar
Department of Bioinformatics
Bharathidasan University
Tiruchirappalli – 620 024.

Sir,

Sub: Ph.D.Programme – **Confirmation order** – reg.
Ref: Letters received on 04.12.2014 from Dr. S. Parthasarathy.

******

With reference to the above cited communication, the provisional registration for the Ph.D. degree of the candidate whose research topic titled **"IDENTIFICATION OF NEW INHIBITORS OF CERTAIN PENICILLIN BINDING PROTEINS OF THE RESISTANT STRAINS OF STREPTOCOCCUS PNEUMONIAE USING STRUCTURE-BASED VIRTUAL SCREENING"** is hereby confirmed on due consideration of the progress reports and recommendations of Doctoral Committee. He is also permitted to proceed with his research work further.

Yours Sincerely,
17.12
for Registrar

Copy to: Dr. S. Parthasarathy, Associate Professor and Head, Department of Bioinformatics, Bharathidasan University, Tiruchirappalli – 620 024.

All Communications are to be addressed to the Registrar only
Please quote our reference in all your replies

18.12.2014

BHARATHIDASAN UNIVERSITY

(Re-accredited with "A" Grade by NAAC)

Palkalaiperur, Tiruchirappalli - 620 024.
Tamil Nadu, India

Off : 4031 - 2407016, 2407060
Telex : 0455 - 253
Fax : 0431 - 2407444
Grams : "BARD"
Email : coe@bdu.ac.in
Website : www.bdu.ac.in

Dr. R. THIRUMURUGAN Ph.D.,
CONTROLLER OF EXAMINATIONS (i/c)

Ref:No.15371 / Ph.D/CE/Bioinformatics/2370 Date: 18.10.2016

SUBMISSION CERTIFICATE

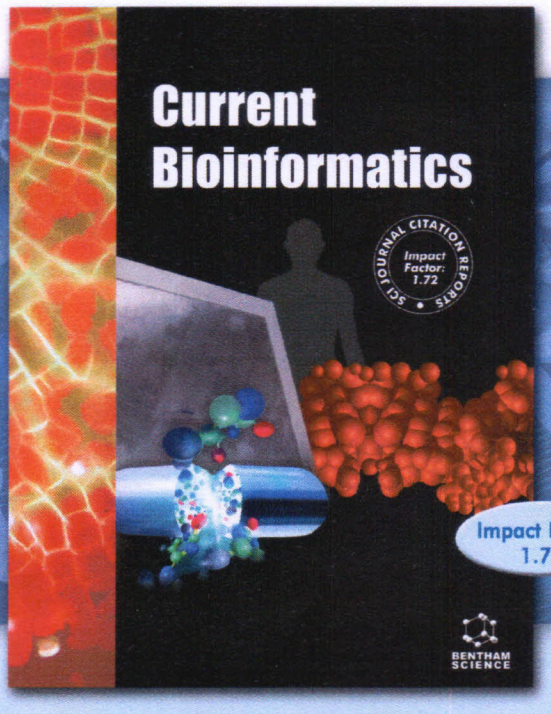
This is to certify that **S. SUVAITHENAMUDHAN** has submitted his Ph.D. Thesis entitled "IDENTIFICATION OF NEW INHIBITORS OF PENICILLIN BINDING PROTEIN 2B OF THE RESISTANT STRAINS OF STREPTOCOCCUS PNEUMONIAE USING STRUCTURE BASED VIRTUAL SCREENING" under the guidance of **Dr. S. PARTHASARATHY,** Associate Professor and Head, Department of Bioinformatics, Bharathidasan University, Tiruchirappalli – 620 024, on **13.10.2016.**

This Certificate is issued on his request letter dated **18.10.2016.**

CONTROLLER OF EXAMINATIONS

Annexure – 3

JOURNAL IMPACTING BIOINFORMATICS



Impact Factor 1.72

Editor-in-Chief:
Alessandro Giuliani, Italy

Current Bioinformatics aims to publish all the latest and outstanding developments in bioinformatics. Each issue contains a series of timely, in-depth reviews written by leaders in the field, covering a wide range of the integration of biology with computer and information science.

- Publishing Peer Reviewed Articles Rapidly
- Abstracted in Chemical Abstracts, EMBASE, SCI Expanded, BIOSIS, JCR/Science Edition and more...
- FREE Online Trials for Institutions
- FREE Sample Issues Available on Website

ISSN: 2212-392X (Online)
ISSN: 1574-8936 (Print)

For Subscriptions
Contact: subscriptions@benthamscience.org

For Advertising & Free Online Trials
Contact: marketing@benthamscience.org

www.benthamscience.com

BENTHAM SCIENCE

Publishers of Quality Research

Structure Based Virtual Screening for the Identification of Potential Inhibitors for Penicillin Binding Protein 2B of the Resistant 5204 Strain of *Streptococcus pneumoniae*

Suvaiyarasan Suvaithenaumudhan and Subbiah Parthasarathy*

Department of Bioinformatics, School of Life Sciences, Bharathidasan University, Tiruchirappalli 620024, Tamil Nadu, India

Subbiah Parthasarathy

Abstract: In this paper, we have performed virtual screening of compounds to identify potential inhibitors against the Penicillin Binding Protein 2B (PBP2B) of the resistant 5204 strain of *Streptococcus pneumoniae*. We have considered 1,677,620 compounds from ZINC database for virtual screening workflow of Schrödinger suite software to identify potential inhibitors that are capable of binding to mutated resistant 5204-PBP2B. Initially, we have obtained 1,247 hits and were prioritized based on protein-ligand contacts which resulted in 99 compounds. These 99 compounds were further clustered to obtain 25 structurally diverse compounds of which the top scoring compound 5-[(6-hydroxy-1,2,3,4-tetrahydroisoquinolin-1-yl)methyl]benzene-1,2,3-triol) with ID: ZINC59376795 may be identified as the potential inhibitor. Molecular dynamics simulations were performed for the wild-sensitive R6-PBP2B and mutated-resistant 5204-PBP2B complexes with this top scoring compound ZINC59376795 and the binding patterns, RMSD calculations, protein-ligand contacts analysis provides deeper insights into the interaction patterns of this novel inhibitor against the sensitive-R6-PBP2B and resistant 5204-PBP2B of *S. pneumoniae*.

Keywords: Structure based virtual screening, glide docking, molecular dynamics simulation, penicillin binding protein, *Streptococcus pneumoniae*.

INTRODUCTION

Streptococcus pneumoniae is one of the major human diseases causing pathogens responsible for diseases such as pneumonia, bronchitis, sinusitis, endocarditis, sepsis and meningitis. The mortality due to *Streptococcus pneumoniae* rose to 1.6 million yearly deaths worldwide and it is admitted that the increase in penicillin-resistant by *Streptococcus pneumoniae* is one of the major reasons for the causative [1-6]. In spite of the conventional administration of β-lactam antibiotics, even the high levels are found not to be therapeutic due to the spread of drug-resistance which makes the organism resistant to the treatment [7].

Penicillin binding proteins (PBPs) are the key enzymes involved in the extracellular steps of the biosynthesis of peptidoglycan pathway and found to be the main constituent of the bacterial cell wall [8, 9]. Cell wall peptidoglycan has repeating units, such as N-acetyl glucosamine (NAG) and N-acetyl muramic acid (NAM), which are cross-linked by pentapeptidic chains. It is responsible for the maintenance of bacterial shape and internal pressure. PBPs catalyze reactions like polymerization of NAG and NAM and peptide cross-linking [6, 10]. There are six PBPs which are classified into two major categories based on molecular masses. PBP1A, PBP1B, PBP2A, PBP2B and PBP2X are the high molecular mass enzymes, of which PBP1A, PBP1B, PBP2A are Class A bi-functional enzymes showing glycosyltransferase (GT) and transpeptidase (TP) activities and PBP2B, PBP2X are Class B monofuntional enzymes possess only transpeptidase activity. PBP3 is the only low molecular mass PBP enzyme which catalyzes carboxypeptidation reaction [6, 11].

Experimentally determined three dimensional (3D) structures of PBP2B of both wild sensitive R6 strain, namely, R6-PBP2B (with PDB ID: 2WAF) and mutated resistant 5204 strain, namely, 5204-PBP2B (with PDB ID: 2WAE) are available [6] in Protein Data Bank (PDB) [12]. PBP2B has the TP domain [6] with residues ranging from 313-680 as shown in Fig. (**1A**). It is the main target with three catalytic motifs Ser386-Val387-Val388-Lys389 (carries active site Serine), Ser443-Ser444-Asn445 and Lys615-Thr616-Gly617 (c.f. Fig. **1A**). The β -lactam antibiotics form the stable complex with the active site serine [6].

The use of *in silico* methods complements drug discovery process which delivers new drug candidates more quickly at a lower cost. In this present study, the drug resistant due to versatility of PBP2B TP domain is focused and the application of *in silico* approaches like molecular docking and dynamics facilitates several clues to the problem. Significantly, the influence of structural features, entropic and enthaplic factors are substantial for the binding of ligand with the target which ultimately scrutinizes plethora of ligands from the chemical databases and effectively facilitates in drug discovery process. Here many *in silico* analysis, such as, virtual screening [13], molecular docking,

*Address correspondence to this author at the Department of Bioinformatics, School of Life Sciences, Bharathidasan University, Tiruchirappalli 620024, Tamil Nadu, India; Tel: +91 431 2407071, Ext. 655; Fax: +91 431 2407045; E-mail: bdupartha@gmail.com

2212-392X/16 $58.00+.00

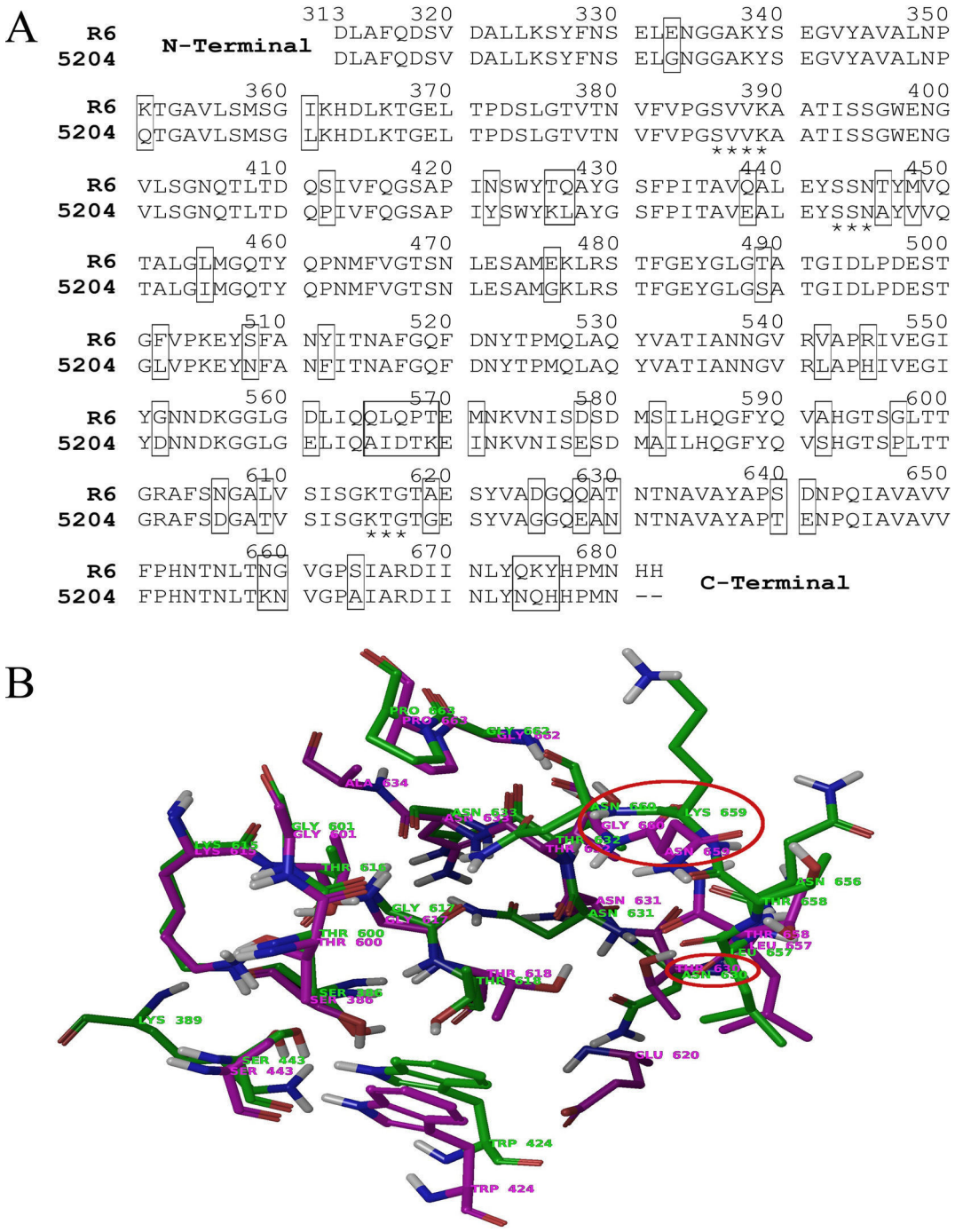


Fig. (1). (**A**) Sequence alignment of the transpeptidase (TP) domain of the PBP2B of sensitive R6 and resistant 5204 strains of *Streptococcus pneumoniae*. The mutated residues are indicated by the boxes and the catalytic motifs are indicated with stars. (**B**) Analyzing the structural changes in the ligand binding site between wild sensitive R6-PBP2B and mutated resistant 5204-PBP2B structures by the superimposition of these structures. Binding site details of sensitive R6-PBP2B and resistant 5204-PBP2B are shown. The residues of R6-PBP2B are indicated in magenta and residues of 5204-PBP2B are indicated in green and the mutated residues are encircled in red colour.

free energy calculations and molecular dynamics (MD) simulations have been carried out using Schrödinger suite [14].

In this paper, we performed a hierarchical virtual screening of the 1,677,620 compounds in ZINC database targeting the PBP2B of the mutated resistant 5204 strain of *Streptococcus pneumoniae*. Further, top scoring 25 virtual screened compounds were analysed using molecular docking with wild sensitive R6-PBP2B and mutated resistant 5204-PBP2B. The binding energies between the sensitive R6-PBP2B and the resistant 5204-PBP2B complexes with the top scoring 25 virtual screened compounds were also evaluated. The comparisons made between these two strains exemplify the nature of intermolecular interactions of the

complex. Molecular dynamics simulations are the versatile computational method in studying dynamics of biological macromolecules, stability, structural conformations with respect to the time scale [15]. Molecular dynamics were performed for the sensitive R6-PBP2B and resistant 5204-PBP2B complexes with the top scoring compound, which may be identified as the novel potential inhibitor. This effective combined computational strategy, such as, virtual screening, molecular modeling and molecular dynamics methods give insights into the interaction patterns of the novel potential inhibitors against the sensitive R6-PBP2B and resistant 5204-PBP2B of *S. pneumoniae.*

MATERIALS AND METHODS

The entire computational study was carried out on a personal computer with 4GB RAM on the 64-bit CentOS Linux operating system using Schrödinger suite software [14] with graphical user interface Maestro 9.4 [16].

Computational Protein Structure Preparation and Receptor Site Identification

The X-ray crystal structures of the PBP2B of both the mutated resistant strain 5204 (PDB ID: 2WAE) and the wild sensitive strain R6 of *S. pneumoniae* (PDB ID: 2WAF) [6] were downloaded from Protein Data Bank (PDB) [12]. It is well known that X-ray crystallographic structures may contain drawbacks, such as missing hydrogen atoms, amino acids or loops and lacking proper protonations, which should be addressed. We have used Schrödinger Protein Preparation Wizard [17, 18] to rectify the assignment of bond orders, addition of hydrogen and remove water molecules. Further, protonation of the protein at the biological pH is done by Epik module [19] of Schrödinger suite. Assignment of the hydrogen bonding network and minimization were carried out using the OPLS_2005 force field method (Schrödinger, LLC, New York, 2013), where the minimization was set to terminate when the RMSD reaches a maximum cutoff of 0.30Å in order to reach least possible energy.

The refined and optimized protein structures were processed using SiteMap module of Schrödinger suite to specify the location of receptor site using options to report maximum 5 sites that requires at least 15 site points using more restrictive definition of hydrophobicity with standard grid cropping site maps at 4Å from the nearest site point. The SiteMap module generated possible potential receptor site in the receptors (both wild sensitive R6-PBP2B and mutated resistant 5204-PBP2B) with quantitative site-score values for each site which considers size, enclosure, exposure, contacts, hydrophobicity, hydrophilicity and Hbond donor acceptor balance. The receptor grid box for carrying out molecular docking was set to active site residues which are S386, V387, V388, K389, S443, S444, N445, K615, T616 and G617 as cited in the literature [6] were used as a centroid for generating Glide receptor grid generation (with the automatic setting of enclosing box) using van der Waals scaling setup of 0.8 and partial charge cutoff of 0.25 with centroid of grid as selected residues.

Ligand Datasets

The collection of virtual commercial compounds is available in ZINC database [20] that contains over 21 million purchasable compounds for which virtual screening can be performed against potential target. In ZINC database, PubChem was chosen from the catalogs option and 1,677,968 compounds with pH of mid range (6-8) were retrieved in SMILES format. The SMILES format (.smi) was converted into three dimensional structure file in Maestro format (.mae) by using the "Structure File Converter" tool of Schrödinger Maestro which results in 1,677,620 compounds. All these 1,677,620 compounds were used in our study for virtual screening against resistant 5204-PBP2B using virtual screening protocol.

Virtual Screening Workflow

The Virtual Screening Workflow (VSW) module of Schrödinger is used to run an entire sequence of jobs for screening large collections of compounds against one or more targets. The workflow includes ligand preparation using LigPrep module [21], where options were set to generate possible ionization states of the molecules at biological pH using Epik, skipping duplicates, generating at most 32 possible stereoisomers with ring conformation to generate at most one least energy structure. The workflow had filtering options to skip molecules that had reactive functional groups and generate druggability property calculation by QikProp module [22] and receptor grids generated earlier were browsed in the receptor grid tab. The docking was carried out using all the three precision modes of glide HTVS, SP and XP, where options were to retain the top 50% and then docking again with option to retain 70% of top ligands in SP and XP modes. The top compounds were processed for druggability check and cheminformatics clustering for chemical diversity.

ADMET Prediction and Clustering

The compounds were tested for druggability by using the properties calculated during the virtual screening so that we have to check whether ADMET (Adsorption, Distribution, Metabolism, Excretion and Toxicity) parameters are well within the range of acceptability or not. QikProp module [22] also provides ranges for comparison of a given compound's properties with those of 95% of known drugs. Important properties, which influence the ADME including molecular weight, solvent accessible volume, solvent accessible surface area (SASA), number of hydrogen bond donors and acceptors, number of atoms in rings, predicted water/gas partition coefficient (QPlogPw) and percentage human oral absorption were calculated for the obtained hits and based on the ranges of the 95% known drug properties, we have selected the most probable drug like compounds for final selection. In order to explore the structural diversity of compounds from the collection of hits and to obtain the centroids of each cluster, the linear binary fingerprints of all the molecules were calculated by which the hierarchical clustering was carried out using cheminformatics tool of

Schrödinger suite, namely, Canvas module [23] using tanimoto similarity algorithm. Cluster representatives were further analyzed for their interaction patterns against both the target site of mutated resistant 5204-PBP2B and wild sensitive R6-PBP2B.

Binding Energy Evaluations

Binding energies of the docked complexes obtained from Glide XP docking were calculated using eMBrAcE MacroModel [24], version 10.2, Schrödinger, LLC, New York, 2013. Here, the OPLS (Optimized Potentials for Liquid Simulations) 2005 force field [25] is used with no solvation with constant dielectric electrostatic treatment having dielectric constant value 1. Further, the minimization is performed using optimal method with maximum iterations of 2500, where convergence is set to gradient and convergence threshold is 0.05, where energy difference mode for association energy is used. The energy difference is then calculated using the following Eq. 1,

$$\Delta E = E_{\text{complex}} - (E_{\text{protein}} + E_{\text{ligand}}), \quad (1)$$

where full effects of relaxation and salvation were included in this mode. The values of E_{protein} and E_{ligand} are the lowest energies found for any conformer of the protein and ligand respectively. ΔE is calculated for each complex conformation saved in the output structure file using the energy of the current complex conformation for E_{complex}.

Molecular Dynamics Simulations

For both sensitive and resistant strains, we used the receptor-ligand complexes obtained earlier by molecular docking and explicit solvated molecular dynamics (MD) was carried out using Desmond module [26]. As Desmond module was used to perform molecular dynamic simulations of biological macromolecules [27], we have also performed the molecular dynamics simulations using Desmond module instead of using GROMACS (GROningen MAchine for Chemical Simulations) or AMBER (Assisted Model Building with Energy Refinement) or VMD (Visual Molecular Dynamics). The prepared complexes were solvated using the system builder using TIP4P aqueous solvation of orthorhombic solvent box with a cell size was approximately 51.7Å x 70.1Å x 53.8Å and a solvent buffer extending 10Å beyond the protein in all directions using OPLS 2005 force field. The systems were neutralized with counter ions which entailed adding different numbers of Na^+ (salt positive ion) or Cl^- (salt negative ion) ion to each system to generate the system which is minimized with maximum of 2500 iterations and convergence thresholds of 1 kcal/mol/Å. We used the NPT ensemble with the Nose-Hoover thermostat method set at a reference temperature of 300 K water solvent system. MD simulations were carried out for 5 ns and trajectories were analyzed for RMSD and to the stable residues of interactions against the target site.

RESULTS AND DISCUSSION

Ligand Binding Site Analysis of Wild Sensitive R6-PBP2B and Mutated Resistant 5204-PBP2B

The highly resistant 5204-PBP2B sequence has a total of 58 mutations compared to the sensitive R6-PBP2B sequence [6]. Significantly, the major number of mutations (*i.e.,* 44 mutations) were present in the transpeptidase (TP) domain (313-680 residues) of PBP2B with a block of five mutations can be identified at 565-569 residues as shown in Fig. (**1A**). In order to design drugs for PBP2B, it is very vital to analyze the structural changes in the ligand binding site between sensitive R6-PBP2B and resistant 5204-PBP2B structures and hence both these structures were superimposed by considering the residues around the top scoring ligand (ZINC59376795) within the radius of 4Å and are shown in Fig. (**1B**). The residues of R6-PBP2B are indicated in magenta and residues of 5204-PBP2B are indicated in green colour. From Fig. (**1B**), it is observed that 16 residues, namely, Ser386, Trp424, Ser443, Thr600, Gly601, Lys615, Thr616, Gly617, Thr618, Asn631, Thr632, Asn633, Leu657, Thr658, Gly662 and Pro663, in the receptor active site are similar between sensitive R6-PBP2B and resistant 5204-PBP2B structures. The residues Thr630, Asn659 and Gly660 of the sensitive R6-PBP2B structure are mutated as Asn630, Lys659 and Asn660, respectively, in the mutated resistant 5204-PBP2B structure and are encircled in red colour, as shown in Fig. (**1B**).

SiteMap analysis gives an idea about comparison of receptor cavity size, H-bond acceptor, H-bond donor and hydrophobic regions from contour maps. This gives insights into the active site changes between the sensitive R6-PBP2B and resistant 5204-PBP2B structures and is illustrated in Fig. (**2**). Further, the values of the important SiteMap properties like, amino acid residues, cavity size, site score, enclosure, hydrophobic, hydrophilic and contact for the structures of sensitive R6-PBP2B and resistant 5204-PBP2B are given in Table **1**. The complementary terms of different features that define the affinity of the ligand in the receptor is analyzed and the site points are superimposed to indicate the vibrant differences, as shown in Fig. (**2**). We find that the cavity size of the sensitive R6-PBP2B is 110$Å^3$ and that of resistant 5204-PBP2B is shrinked to 39$Å^3$ (c.f. Table **1**) as indicated by the purple and blue colour site points for the sensitive R6-PBP2B and resistant 5204-PBP2B, which are shown in Fig. (**2A**). Similarly, we find that there is a significant change in the pattern of the hydrophobic regions for the resistant 5204-PBP2B when compared with the sensitive R6-PBP2B. The sensitive R6-PBP2B has hydrophobic and hydrophilic scores of 0.702 and 1.134, respectively, while the resistant 5204-PBP2B structure has less hydrophobic and hydrophilic scores of 0.335 and 0.985, respectively (c.f. Table **1**). The H-bond acceptor regions are indicated as yellow colour solid contours for sensitive R6-PBP2B and green colour solid contours indicate for resistant 5204-PBP2B, as shown in Fig. (**2B**), which provides the details about the difference in patterns but its contribution is little exterior of the active site. In a similar way, the H-bond donor regions are shown in Fig. (**2C**), which indicates its contribution is little interior to the active site. Finally, Fig. (**2D**) shows the hydrophobic regions of resistant 5204-PBP2B and sensitive R6-PBP2B, which have contributions from Trp424 and Leu657. The above SiteMap analysis signifies that the existing active compounds against sensitive R6-PBP2B can have less affinity with the resistant 5204-PBP2B. The three main mutations in the amino acid residues which are involved in the sitemap analysis are i) Thr630 is mutated to Asn630, ii) Asn659 to Lys659 and iii) Gly660 to Asn660 (c.f. Fig. **1B**).

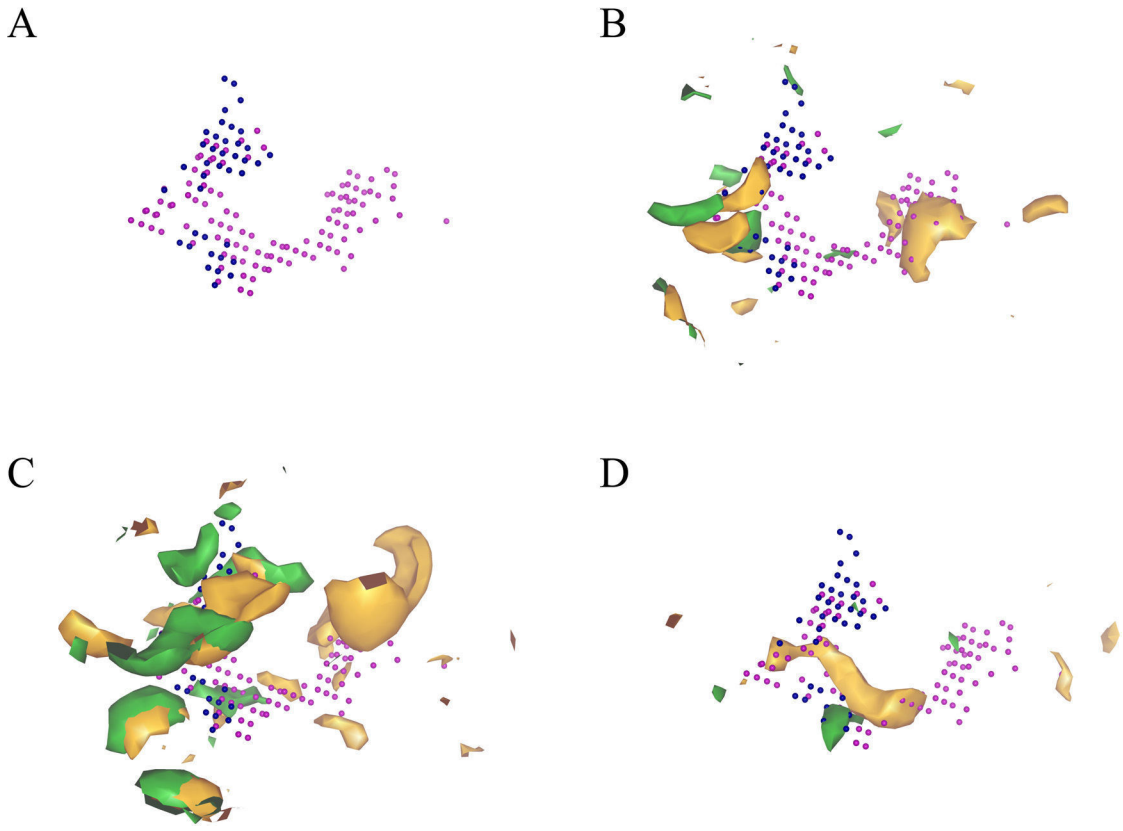


Fig. (2). SiteMap shows structural complementarities of sensitive R6-PBP2B and resistant 5204-PBP2B structures. (**A**) Cavity of the receptor. Purple colour points indicates sensitive R6-PBP2B. Blue colour points indicating resistant 5204-PBP2B. (**B**) H-bond acceptor. Yellow colour solid contours indicates sensitive R6-PBP2B and Green colour solid contours indicates resistant 5204-PBP2B. (**B**) H-bond acceptor (**C**) H-bond donor and (**D**) Hydrophobic.

Hence, we have used the sites of mutations for screening the novel compounds effective against resistant 5204-PBP2B.

Structure Based Virtual Screening

Our objective is to identify a potential inhibitor against resistant 5204-PBP2B through structure based virtual screening [13, 14] and the detailed workflow of our study is shown in Fig. (**3**). First, we carried out virtual screening against the crystal structure 2WAE of the resistant 5204-PBP2B [6] obtained from PDB against 1,677,620 compounds of the ZINC database, using high-throughput virtual screening (HTVS) workflow of Schrödinger [14]. Initially, we check with the orientation and position of the ligand pose complementary checkup against the receptor with less sampling scale of conformers followed by standard precision (SP) with additional larger scale of pose exploration option to refine the torsional thoroughness of the final torsional refinement and sampling. Ultimately, we use extra precision (XP) algorithm that uses extensive scoring by employing penalty, such as, ligand solvent exposed regions, rotatablility, ionization Epik state and rewards, such as, H-bonding, hydrophobic interactions, correlated H-bonding, hydrophobically packed correlated H-bonding, electrostatic interactions, with also reward for low molecular weight that weeds out false positive molecules if they posses bad interaction. This methodology is a perfect balance between

Table 1. SiteMap properties for the structures of sensitive R6-PBP2B and resistant 5204-PBP2B.

Strain Type	Amino Acid Residues	Cavity Size (Å)^3	Site Score	Enclosure	Hydrophobic	Hydrophilic	Contact
Wild Sensitive R6-PBP2B (2WAF)	Gly385, Ser386, Lys389, Ile421, Asn422, Ser423, Trp424, Ser443, Asn445, Phe517, Gln519, Thr600, Lys615, Thr616, Gly617, Thr618, Ala619, Glu620, Thr630, Thr632, Leu657, Thr658, Asn659, Gly660, Pro663	110	1.042	0.744	0.702	1.134	1.005
Mutated Resistant 5204-PBP2B (2WAE)	Trp424, Thr600, Thr616, Thr618, Asn630, Asn631, Thr632, Asn654, Asn656, Leu657, Thr658, Lys659, Asn660	39	0.761	0.752	0.335	0.985	0.976

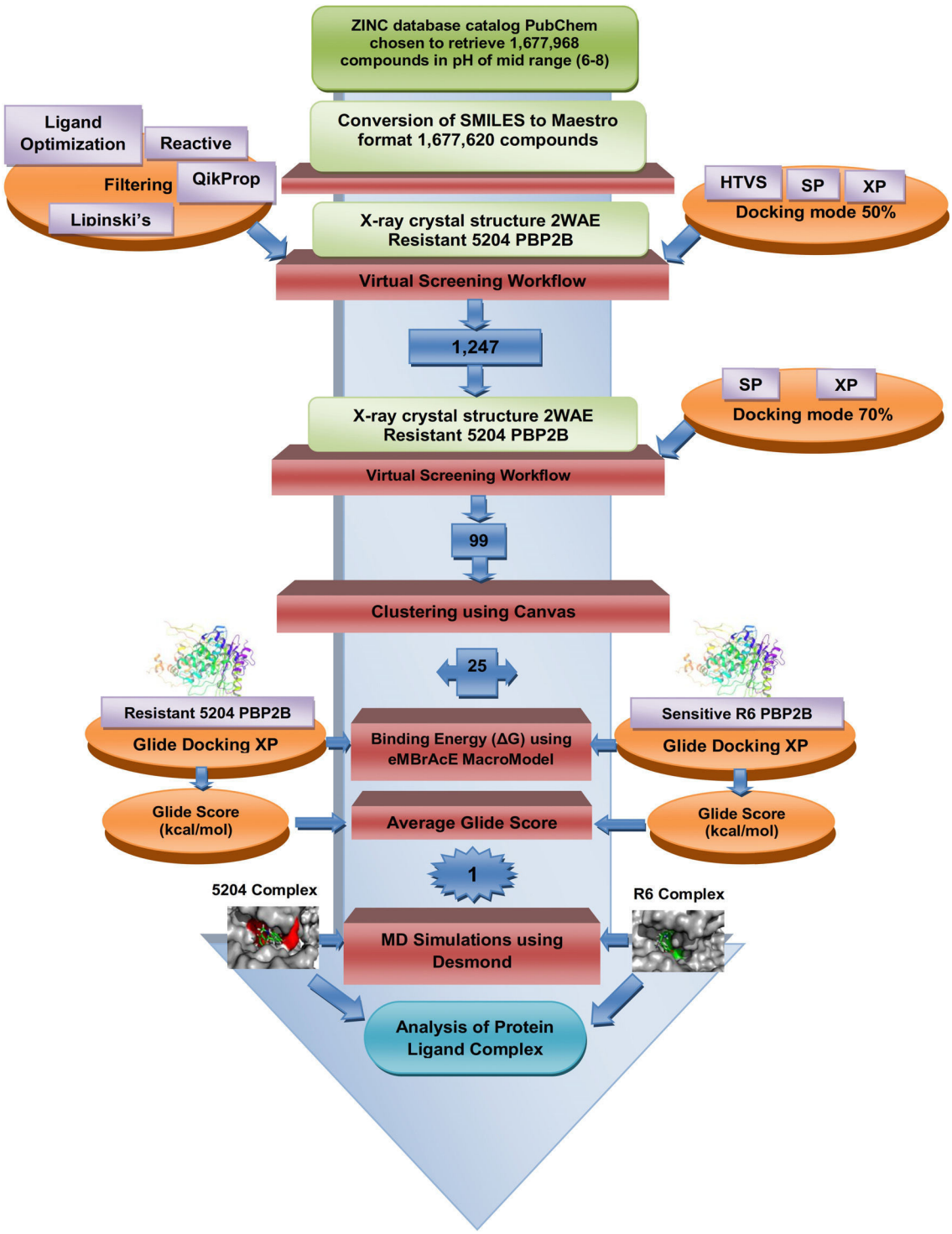


Fig. (3). Flow chart shows the strategies of virtual screening work flow and comparative molecular docking.

accuracy and speed that would be ideal for our study using the appropriate modules of Schrödinger suite software [14].

In the virtual screening workflow, the docking was carried out using all the three precision modes of glide HTVS, SP and XP, with options to retain the top 50% of the compounds. First, we screened with HTVS, followed by SP and finally by XP docking. After XP docking, we got 1,247 hits binding to mutant 5204-PBP2B. Further these 1,247 compounds were screened without using any filters and by using Glide docking at the two accuracy levels SP and XP with options to retain the top 70% of the compounds against the resistant 5204-PBP2B and we have obtained 99 promising ligands (c.f. Fig. **3**).

In order to obtain chemically diverse scaffolds from these 99 compounds, clustering was done using Canvas module [23]. We first generated the linear hashed binary fingerprints

which could be used to compare molecules among each other for clustering the collection of molecules as per the chemical structure similarity. The hierarchical clustering was carried out to obtain the centroid of each group using tanimoto algorithm setup and the resulted dendrogram is shown in Fig. (**4**) with proper ZINC IDs. Finally, the 25 unique structures scaffold representative compounds are exported to be used for further steps and these structures along with their ZINC IDs are shown in Fig. (**5**). In order to check whether the cluster representative compounds are well within the druggabilty range, we processed these 25 compounds using QikProp module [22], where it has more than 50 calculated and predicted properties but for documentation we have tabulated certain important properties in Table **2**, where all the compounds were well within the range of druggability. As shown in Table **2**, all the 25 compounds are falling in the ranges of the known drug properties. The molecular weight of the 25 compounds are in the range of 287 to 450 Daltons and they are in the limits of the drug molecular properties i.e. 130 to 725 Daltons. The total solvent-accessible area (SASA) and volume of the 25 compounds are also well within the range. The percentage human oral absorption is in the range of 43 to 100% and on comparison of the drug molecules oral absorption, less than 25% indicates the poor absorption for the compounds. Out of the 25 hits, 21 compounds are showing greater than 75% oral absorption, which signifies that the compounds can be used for oral administration. Again, these 25 compounds were docked against both sensitive R6-PBP2B and resistant 5204-PBP2B targets, so that we could analyze about the interaction patterns and docking scores. These generated poses were used to calculate the binding free energy in order to consider the desolvation parameter of ligand using eMBrAcE module [24, 25].

In order to address the dual inhibition or resistant adaptability against both sensitive R6-PBP2B and resistant 5204-PBP2B, we took 25 compounds that are showing strong interactions specifically to the mutated residues will have greater affinity towards resistant 5204-PBP2B but to consider the affinity to sensitive R6-PBP2B. In the present work, we want to select the compounds having activity against both sensitive R6-PBP2B and resistant 5204-PBP2B, hence we performed docking into sensitive and resistant PBP2B and selected the compounds having activity for both strains. Further, the resistant 5204-PBP2B has two mutations over the sensitive R6-PBP2B in the binding pocket, which indicates that the compounds of resistant 5204-PBP2B also bind to the sensitive R6-PBP2B. Hence, we have used only half of the glide score ($S_{sen}/2$) of the sensitive R6-PBP2B in calculating the average glide score (S_{ave}). The average glide score (S_{ave}) is calculated by considering half of the glide score ($S_{sen}/2$) of the sensitive R6-PBP2B and the full glide score (S_{res}) of the resistant 5204-PBP2B, as given by the following Eq. 2.

$$S_{ave} = \frac{(S_{sen} / 2) + S_{res}}{2} \quad . \qquad (2)$$

Using this average glide score (S_{ave}), we ranked all the 25 hits obtained from XP docking. The values of the Glide score, binding energy (ΔG) of both sensitive R6-PBP2B and resistant 5204-PBP2B and average glide score (c.f. Eq. 2) are given in Table **3**. From Table 3, the best active ligand may be selected by the lowest average glide score (S_{ave}). The top scoring compound 5-[(6-hydroxy-1,2,3,4-tetrahydroisoquinolin-1-yl)methyl]benzene-1,2,3-triol with ID: ZINC59376795 has been identified as the potential inhibitor.

The docking studies showed that the active ligand (ZINC59376795) with resistant 5204-PBP2B interacts with the two mutated residues of TP domain as shown in Fig. (**6A**) with Glide score of -8.060 kcal/mol as given in Table **4**. The hydrogen bond formed with amino acids along with their bond lengths are given in Table **4** and are shown in Fig. (**6A**). For visual inspection, Fig. (**6B**) shows the docking pose of ligand with the mutated residues (in red) and other residues (in grey) of the TP domain. Here Asn630 and Asn660 are mutated residues and the other two residues Leu657 and Thr600 (c.f. Table **4**) also interact with the ligand ZINC59376795. The docking studies showed that the active ligand (ZINC59376795) with the sensitive R6-PBP2B interacts with two active site residues Ser386 and Gly617, as shown in Fig. (**7A**) of TP domain, with Glide score of -6.731

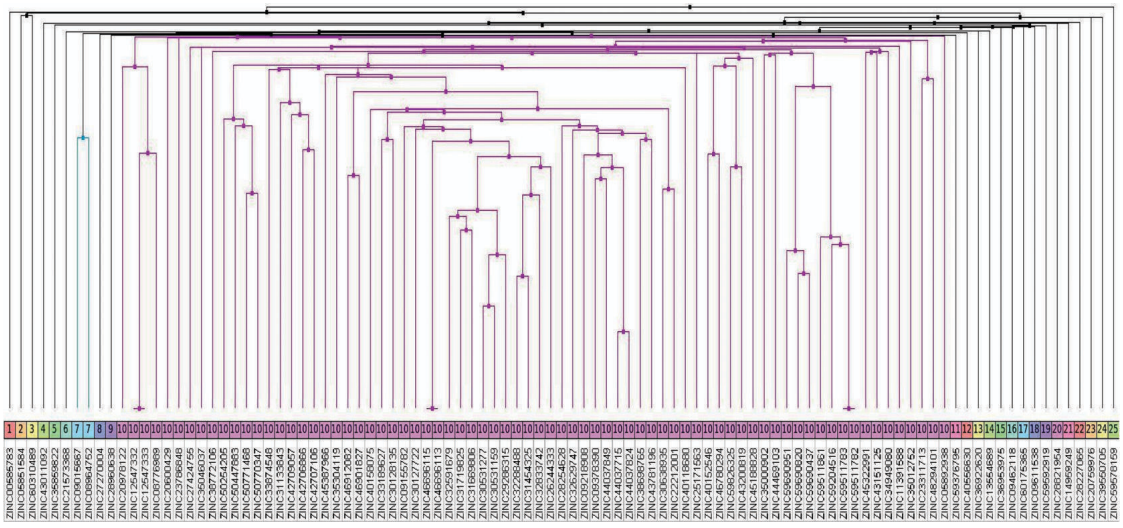


Fig. (4). Dendrogram of clustering 25 active ligands by their binary fingerprints, defining clusters using the same color for members of each cluster. The x-axis is the ZINC ID.

Table 2. QikProp properties of the 25 compounds obtained from clustering using Canvas.

S. No.	Compound ID	Molecular Weight (130-725 Daltons)*	Solvent Accessible Volume (500-2000 Å^3)*	SASA[a] (300-1000 Å^2)*	H-Bond Acceptor Groups (2-20)*	H-Bond Donor Groups (0-6)*	Number of Atoms in Ring	QPlogPw[b] (4-45)*	Percent Human Oral Absorption (>80% High; <25% Poor)*
1	ZINC59376795	287.315	896.28	513.70	4.50	5	6	13.872	54.883
2	ZINC60175365	494.977	1464.46	790.65	7.00	4.5	13	15.797	43.039
3	ZINC36922620	298.771	929.13	535.13	2.75	2	4	8.362	100.000
4	ZINC39550705	302.338	969.35	567.07	8.50	2	4	13.852	52.067
5	ZINC36953975	278.353	943.43	539.63	3.25	2	5	9.046	100.000
6	ZINC20759976	487.644	1573.32	832.29	10.00	0	6	13.955	77.381
7	ZINC09611538	480.581	1473.91	804.25	7.25	1	11	12.597	100.000
8	ZINC59592919	469.926	1392.64	782.98	6.00	2	4	16.381	95.977
9	ZINC35859822	481.609	1509.99	778.75	8.75	0	5	13.246	93.089
10	ZINC21573368	457.528	1426.12	753.63	7.50	0	7	11.467	87.556
11	ZINC59578159	364.441	1178.81	659.36	9.10	3	11	14.546	61.468
12	ZINC40556630	399.448	1287.81	738.42	4.70	5	9	15.69	96.345
13	ZINC29286315	483.968	1411.32	784.16	10.00	2	7	19.381	75.366
14	ZINC27270004	381.451	1181.03	668.01	8.00	2	4	17.442	78.920
15	ZINC05851584	428.508	1244.51	649.92	5.75	3	4	13.896	100.000
16	ZINC28227065	410.472	1248.71	676.88	7.20	0	8	8.793	95.197
17	ZINC60310489	314.774	962.11	555.85	3.50	1	2	7.514	100.000
18	ZINC14959249	424.861	1236.71	697.00	6.20	4	5	15.936	89.595
19	ZINC00585783	396.448	1192.88	648.11	5.50	2	7	11.766	91.057
20	ZINC43011092	427.904	1178.47	625.45	6.45	3	6	13.667	90.798
21	ZINC09462118	498.552	1517.39	837.11	7.75	1	7	13.899	85.937
22	ZINC09015667	497.596	1589.01	864.51	8.25	0	7	11.771	89.657
23	ZINC28821954	488.585	1542.41	841.67	9.70	2	7	19.106	74.910
24	ZINC13554689	433.555	1300.24	669.55	7.70	3	4	14.261	100.000
25	ZINC28960638	450.539	1419.90	768.94	6.00	2	5	13.183	100.000

*Range values obtained from QikProp manual.
[a]**SASA:** Solvent Accessible Surface Area, [b]**QPlogPw:** Predicted water/gas partition coefficient.

kcal/mol (c.f. Table **4**). The hydrogen bond interactions formed with the active site residues are shown in Fig. (**7B**), where the active site residues are indicated in green. From the above results, it is inferred that the mutated resistant 5204-PBP2B has comparatively high affinity with ligand (ZINC59376795) than the wild sensitive R6-PBP2B.

In order to check the novelty of the screened 25 scaffold leader representative compounds, which were obtained *via* cheminformatics clustering, these compounds were searched against SciFinder. All these compounds seem to be novel as they are not reported for any antibacterial activity so far and few molecules were reported with anticancer and antiviral activities and the details are provided as the supplementary data in Table **T1**. Hence, these compounds may be novel and promising against Penicillin Binding Protein 2B of the resistant 5204 strain of *S. pneumoniae*.

Molecular Dynamics Evaluation of the Resistant 5204-PBP2B - Ligand ZINC59376795 Complex

In order to assess stability and interaction pattern of the ligand-receptor complex, explicit molecular dynamics simulations of the sample accessible configuration space of the resistant 5204-PBP2B and ligand ZINC59376795 complex with simulation has been effected at 5 ns using Desmond module [26] of Schrödinger suite. The standard MD, RMSD and protein-ligand contacts analysis were performed. The stability of the complex was evaluated by the RMSD values of the backbone atoms and ligand complex. From Fig. (**8A**), it is inferred that complex get equilibrated around 4 ns and it fluctuates around its average value. It is also verified using GROMACS 4.6.5 [28] that the complex is stable by continuing the MD simulations upto 10 ns and is included in the supplementary data as Fig. (**S1**). Fig. (**8B**) shows the protein-ligand contact analysis. Protein

Table 3. Glide scores, Binding energy (ΔG) and Average glide score of 25 compounds with the sensitive R6-PBP2B and resistant 5204-PBP2B of *Streptococcus pneumonia.*

S. No.	Compound ID	Sensitive R6-PBP2B		Resistant 5204-PBP2B		Average Glide Score (S_{ave}) (kcal/mol)
		Glide Score (S_{sen}) (kcal/mol)	Binding Energy(ΔG) (kcal/mol)	Glide Score (S_{sen}) (kcal/mol)	Binding Energy(ΔG) (kcal/mol)	
1	ZINC59376795	-6.731	-88.12	-8.060	-61.58	-5.713
2	ZINC60175365	-5.369	-71.64	-8.206	-21.02	-5.445
3	ZINC36922620	-6.100	-51.08	-7.810	-35.60	-5.430
4	ZINC39550705	-7.179	-106.87	-7.014	-52.37	-5.302
5	ZINC36953975	-5.923	-55.54	-7.540	-31.34	-5.251
6	ZINC20759976	-5.550	-29.61	-7.672	-46.21	-5.224
7	ZINC09611538	-5.839	-73.49	-7.442	-53.17	-5.180
8	ZINC59592919	-5.152	-87.77	-7.692	-50.92	-5.134
9	ZINC35859822	-6.840	-32.88	-6.821	-34.50	-5.121
10	ZINC21573368	-6.181	-35.68	-7.127	-51.83	-5.109
11	ZINC59578159	-6.163	-42.85	-7.135	-68.32	-5.109
12	ZINC40556630	-7.127	-47.43	-6.534	-38.28	-5.049
13	ZINC29286315	-5.314	-45.75	-7.207	-27.94	-4.932
14	ZINC27270004	-5.714	-22.21	-6.751	-41.05	-4.804
15	ZINC05851584	-6.519	-39.50	-6.289	-66.25	-4.774
16	ZINC28227065	-5.594	-16.92	-6.736	-44.50	-4.767
17	ZINC60310489	-5.465	-33.01	-6.776	-36.73	-4.754
18	ZINC14959249	-5.963	-81.72	-6.392	-58.57	-4.687
19	ZINC00585783	-5.556	-46.88	-6.595	-22.37	-4.687
20	ZINC43011092	-6.204	-28.01	-6.048	-55.74	-4.575
21	ZINC09462118	-5.334	-6.79	-6.355	-83.13	-4.511
22	ZINC09015667	-4.135	-77.39	-6.694	-30.19	-4.381
23	ZINC28821954	-5.087	-54.93	-6.094	-42.79	-4.319
24	ZINC13554689	-4.289	-26.81	-4.641	-31.77	-3.393
25	ZINC28960638	-3.833	-48.98	-3.623	-56.87	-2.770

interactions with the ligand can be monitored throughout the simulation. Protein-ligand contacts are categorized into four types: hydrogen bonds, hydrophobic, ionic and water bridges and shown as stacked bar chart in Fig. (**8B**). In total, 14 residues have interactions (or 'contact') with the ligand ZINC59376795. Among these residues Asn630, Lys659 and Asn660 are the mutated residues of 5204-PBP2B (Fig. **8B**). Trp424 shows more hydrophobic contacts compared to other residues.

Residues Thr616 and Lys659 show hydrogen bond interactions exclusively. Significantly, it is to be noted that the hydrogen bond contacts with residues Thr616 (catalytic motif) and Lys659 (mutated residue) were not observed in glide docking (c.f. Fig. **6A**). Ala603 and Asn654 are the only two residues exhibit the water bridges exclusively. Thr658 is the only one residue involved in ionic contact. The mutated residues of Asn630 and Asn660 and normal residues of Thr600, Asn631 and Asn656 are involved in hydrogen bond as well as contact as water bridges with the ligand. The structure file (5204_2WAE_ZINC59376795.mae) for the resistant 5204-PBP2B - ligand ZINC59376795 complex in Maestro format of Schrödinger is included as supplementary data File **F1**.

Molecular Dynamics Evaluation of Sensitive R6-PBP2B - Ligand ZINC59376795 Complex

MD simulation of the sensitive R6-PBP2B backbone and ligand ZINC59376795 complex has been effected at 5 ns using Desmond module [26] and the results are shown in Fig. (**9**). From Fig. (**9A**), it is evident that the system is well equilibrated within the 5 ns. As mentioned earlier, here also the MD simulations are verified upto 10 ns and the complex is stable, which is included in the supplementary data as Fig. (**S2**). The protein-ligand contacts such as hydrogen bonds, hydrophobic, ionic and water bridges are shown as stacked bar chart in Fig. (**9B**). In this case, 17 residues have interactions with the ligand. Among these 17 residues,

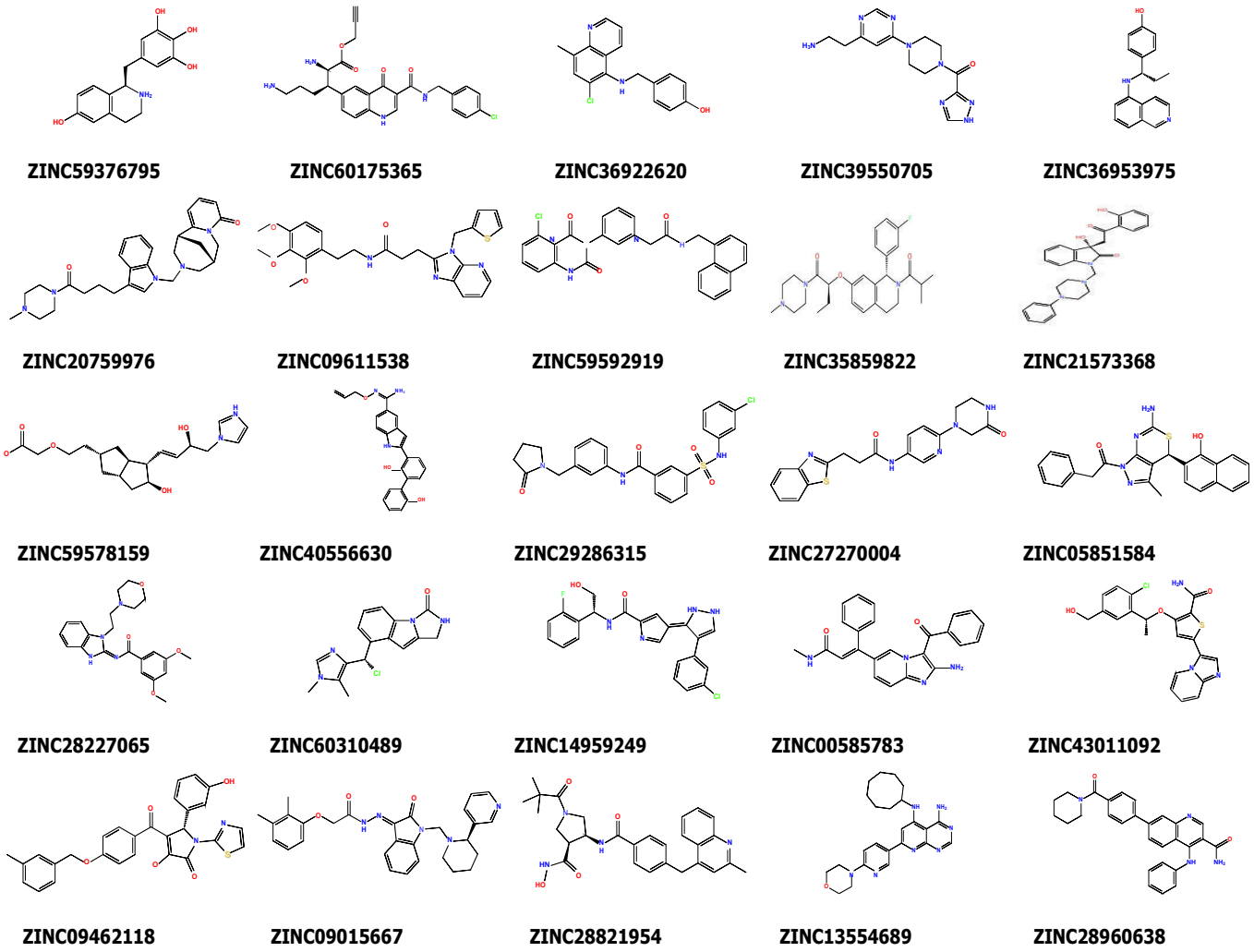


Fig. (5). Structures of compounds of the 25 unique clusters along with their ZINC IDs as listed in Table **3**.

Table 4. Glide scores and hydrogen bond interactions of glide docking for the ligand ZINC59376795 : (5-[(6-hydroxy-1,2,3,4-tetrahydroisoquinolin-1-yl) methyl] benzene-1,2,3-triol for both sensitive R6-PBP2B and resistant 5204-PBP2B.

Strains	Amino Acid	H-Bond Interaction	Bond Length (Å)	Glide Score (kcal/mol)
Sensitive R6-PBP2B (2WAF)	SER 386	O.... H	2.259	-6.731
	GLY 617	O.... H	2.152	
Resistant 5204-PBP2B (2WAE)	THR 600	O.... H	1.763	-8.060
	ASN 630	O.... H	1.714	
	ASN 630	O.... H	1.849	
	LEU 657	O.... H	1.980	
	ASN 660	H.....O	2.048	

Ser386, Lys389, Ser443, Asn445, Lys615, Thr616 and Gly617 are the seven residues of the catalytic motifs. Thr616 and Gly617 are the two residues involves only in hydrogen bond contacts compared to other residues. Lys389, Asn445, Thr618, Glu620, Asn659 and Gly660 are the six residues exhibit water bridges. Ser386, Ser443, Leu657 and Thr658 residues involved in hydrogen bond as well as contact as water bridges with the ligand. The structure file (R6_2WAF_ZINC59376795.mae) for the sensitive R6-PBP2B - ligand ZINC59376795 complex in Maestro format of Schrödinger is included as supplementary data File **F2**.

CONCLUSION

The increase in penicillin resistant of *Streptococcus pneumoniae* governed by rapid mutations and interactions at molecular level spurred our interest in applying *in silico* approaches. By hierarchical virtual screening, the ZINC

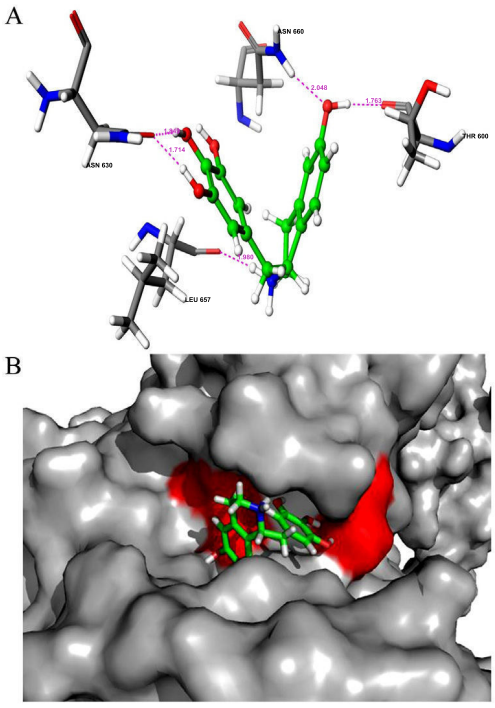


Fig. (6). Interacted residues of penicillin-binding protein 2B (PBP2B) of the mutated resistant 5204 strain structure 2WAE with the ligand ZINC59376795 are shown. (**A**) The ligand is shown as ball and stick model and the dotted pink lines indicate hydrogen bond interactions. (**B**) The binding site shape of compound and the residues of the mutated residues are represented in red.

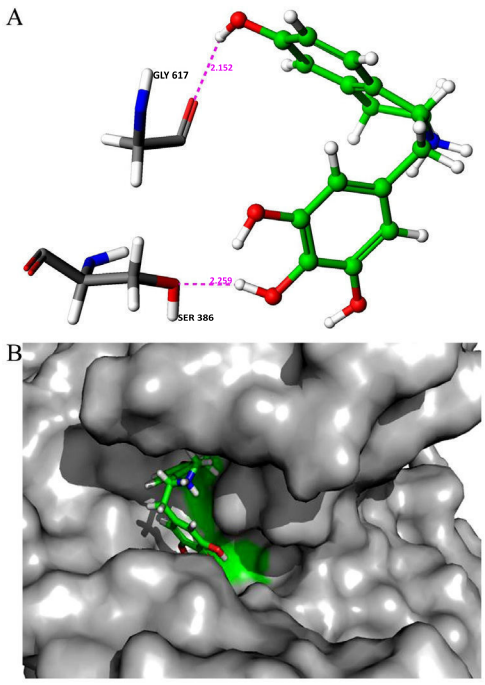


Fig. (7). Interacted residues of penicillin-binding protein 2B (PBP2B) of wild sensitive R6 strain structure 2WAF with the ligand ZINC59376795 are shown. (**A**) The ligand is shown as ball and stick model and the dotted pink lines indicate hydrogen bond interactions. (**B**) The binding site shape of compound and the residues of the catalytic motif residues are represented in green.

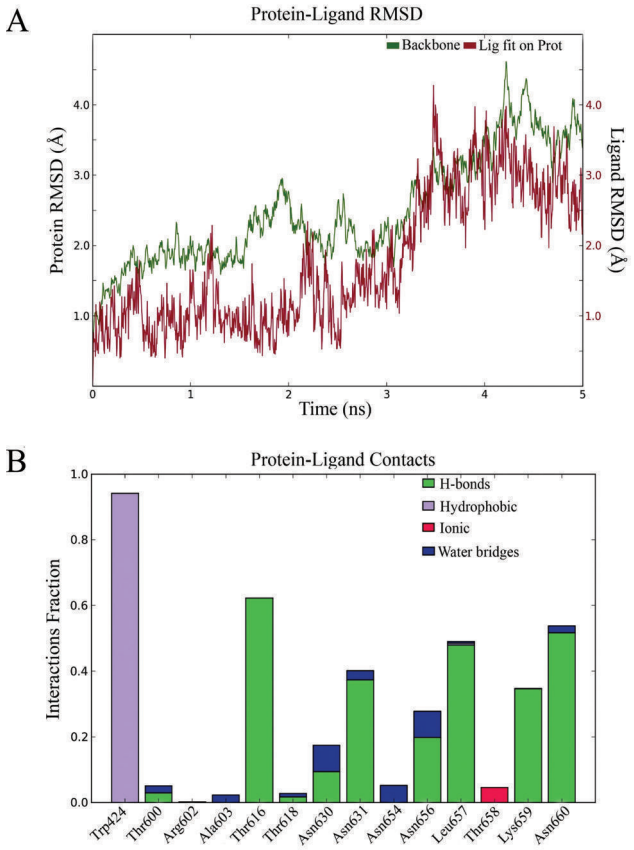


Fig. (8). Molecular dynamic simulations for the resistant 5204-PBP2B. (**A**) Trajectory of protein - ligand RMSD spanning the binding site for 5204-PBP2B backbone is shown in green and the ligand is represented in cherry rose. (**B**) Stacked bar chart of protein - ligand contacts.

database comprising 1,677,620 compounds was screened to identify potential inhibitor against the PBP2B of highly resistant strain 5204. We found 25 promising diverse compounds which showed good affinity against the resistant 5204-PBP2B by using novel *in silico* paradigm, such as, automated virtual screening, customized docking scoring and free energy affinity checks. The important part is that when compounds were screened for novelty by using SciFinder, all were not posing any reported antimicrobial activity for penicillin binding protein 2B. The ligand 5-[(6-hydroxy-1,2,3,4-tetrahydroisoquinolin-1-yl)methyl]benzene-1,2,3-triol with ID: ZINC59376795 is found to be the most promising inhibitor against the mutated resistant 5204-PBP2B, where it interacts with active site residues Ser386 and Gly617 of the wild sensitive R6-PBP2B and share glide score of -6.731 kcal/mol. Significantly, the same ligand ZINC59376795 binds with the mutated residues Asn630 and Asn660 of the resistant 5204-PBP2B producing the glide score of -8.060 kcal/mol.

The standard MD, RMSD and protein-ligand contacts analysis was carried out through MD simulations at 5 ns. Asn630 and Asn660 are the mutated residues involved in hydrogen bond and water bridges with the ligand of the 5204-PBP2B. Further, Ser386, Thr616 and Gly617 are the catalytic motifs show more hydrogen bond interactions,

which are the active site residues of the sensitive R6-PBP2B. From the results of Glide docking as well as MD simulations, the binding of ligand ZINC59376795 is comparatively better against resistant strain 5204-PBP2B forming more hydrogen contacts and other interactions. This present study indicates that all the 25 compounds are promising inhibitors but the top scoring ligand 5-[(6-hydroxy-1,2,3,4-tetrahydroisoquinolin-1-yl)methyl]benzene-1,2,3-triol with ID: ZINC59376795 may be identified as the potential inhibitor against the sensitive R6-PBP2B and resistant 5204-PBP2B based on the analysis of molecular docking, free energy calculations and molecular dynamics simulations. MD simulations may be performed for the top 5 or 10 compounds with the sensitive R6-PBP2B and resistant 5204-PBP2B which will provide deeper insights into the interaction patterns of these novel inhibitors of *Streptococcus pneumoniae*.

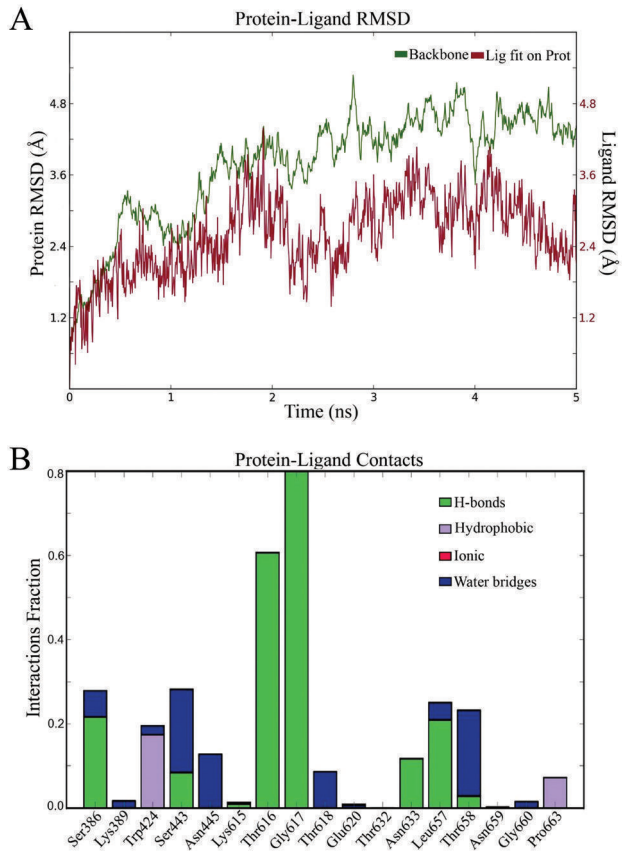


Fig. (9). Molecular dynamic simulations for the sensitive R6-PBP2B. (**A**) Trajectory of protein - ligand RMSD spanning the binding site for the R6-PBP2B backbone is shown in green and the ligand is represented in cherry rose. (**B**) Stacked bar chart of protein - ligand contacts.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This work forms part of the major research project funded by the University Grants Commission (UGC), New Delhi. One of the authors (SS) gratefully acknowledges the support provided by the UGC (MRP) in the form a project fellow. The authors thank Dr. Ravikumar Muttineni and Mr. Vinod Devaraji, Schrödinger, Bangalore for all their suggestions during the work and also for the critical reading of the manuscript.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's web site along with the published article.

REFERENCES

[1] Appelbaum PC. Antimicrobial resistance in Streptococcus pneumoniae: An overview. Clin Infect Dis 1992; 15(1): 77-83.

[2] Hakenbeck R, Kaminski K, König A, *et al*. Penicillin-binding proteins in beta-lactam-resistant Streptococcus pneumoniae. Microb Drug Resist 1999; 5(2): 91-9.

[3] Musher DM. A Pathogenetic categorization of clinical syndromes caused by *Streptococcus pneumonia*. In: Tuomanen El., editor. The *Pneumococcus*. Washington: ASM press; 2004; 211-220.

[4] Klugman KP. Clinical relevance of antibiotic resistance in *pneumococcal* infections. In: Tuomanen El., editor. The *Pneumococcus*. Washington: ASM press; 2004; 331-338.

[5] Zapun A, Contreras-Martel C, Vernet T. Penicillin-binding proteins and beta-lactam resistance. FEMS Microbiol Rev 2008; 32(2): 361-85.

[6] Contreras-Martel C, Dahout-Gonzalez C, Martins Ados S, Kotnik M, Dessen A. PBP active site flexibility as the key mechanism for beta-lactam resistance in pneumococci. J Mol Biol 2009; 387(4): 899-909.

[7] Pallares R, Viladrich PF, Liñares J, Cabellos C, Gudiol F. Impact of antibiotic resistance on chemotherapy for pneumococcal infections. Microb Drug Resist 1998; 4(4): 339-47.

[8] Rogers HJ. Bacterial growth and the cell envelope. Bacteriol Rev 1970; 34(2): 194-214.

[9] Barreteau H, Kovac A, Boniface A, Sova M, Gobec S, Blanot D. Cytoplasmic steps of peptidoglycan biosynthesis. FEMS Microbiol Rev 2008; 32(2): 168-207.

[10] Contreras-Martel C, Job V, Di Guilmi AM, Vernet T, Dideberg O, Dessen A. Crystal structure of penicillin-binding protein 1a (PBP1a) reveals a mutational hotspot implicated in beta-lactam resistance in Streptococcus pneumoniae. J Mol Biol 2006; 355(4): 684-96.

[11] Pagliero E, Chesnel L, Hopkins J, *et al.* Biochemical characterization of Streptococcus pneumoniae penicillin-binding protein 2b and its implication in beta-lactam resistance. Antimicrob Agents Chemother 2004; 48(5): 1848-55.

[12] Berman HM, Westbrook J, Feng Z, *et al*. The Protein Data Bank. Nucleic Acids Res 2000; 28(1): 235-42.

[13] Miguet L, Zervosen A, Gerards T, *et al*. Discovery of new inhibitors of resistant Streptococcus pneumoniae penicillin binding protein (PBP) 2x by structure-based virtual screening. J Med Chem 2009; 52(19): 5926-36.

[14] Schrödinger Suite, Schrödinger, LLC, New York, NY, 2013.

[15] Alonso H, Bliznyuk AA, Gready JE. Combining docking and molecular dynamic simulations in drug design. Med Res Rev 2006; 26(5): 531-68.

[16] Maestro, version 9.4, Schrödinger, LLC, New York, NY, 2013.

[17] Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des 2013; 27(3): 221-34.

[18] Olsson MHM, Søndergard CR, Rostkowski M, Jensen JH. PROPKA3. Consistent Treatment of Internal and Surface Residues in Empirical pK_a predictions. J Chem Theor Comput 2011; 7: 525-37.

[19] Shelley JC, Cholleti A, Frye LL, Greenwood JR, Timlin MR, Uchiyama M. Epik: a software program for pKa prediction and protonation state generation for drug-like molecules J Comput Aided Mol Des 2007; 21: 681-91.

[20] ZINC Database http://zinc.docking.org

[21] LigPrep version 2.6, Schrödinger, LLC, New York, NY, 2013.

[22] QikProp, version 3.8, Schrödinger, LLC, New York, NY, 2013.

[23] Duan J, Dixon SL, Lowrie JF, Sherman W. Analysis and comparison of 2D fingerprints: insights into database screening performance using eight fingerprint methods. J Mol Graph Model 2010; 29(2): 157-70.

[24] MacroModel, version 10.2, Schrödinger, LLC, New York, NY, 2013.

[25] The OPLS_2005 parameters are described in Banks JL, Beard HS, Cao Y, Cho AE, Damm W, Farid R, Felts AK, Halgren TA, Mainz DT, Maple JR, Murphy R, Philipp DM, Repasky MP, Zhang LY, Berne BJ, Friesner RA, Gallicchio E, Levy RM. Integrated Modeling Program, Applied Chemical Theory (IMPACT); J. Comp. Chem. 2005; 26(16): 1752-1780. (The OPLS_2005 parameters are located in $SCHRODINGER/mmshare-vversion/data/f14/).

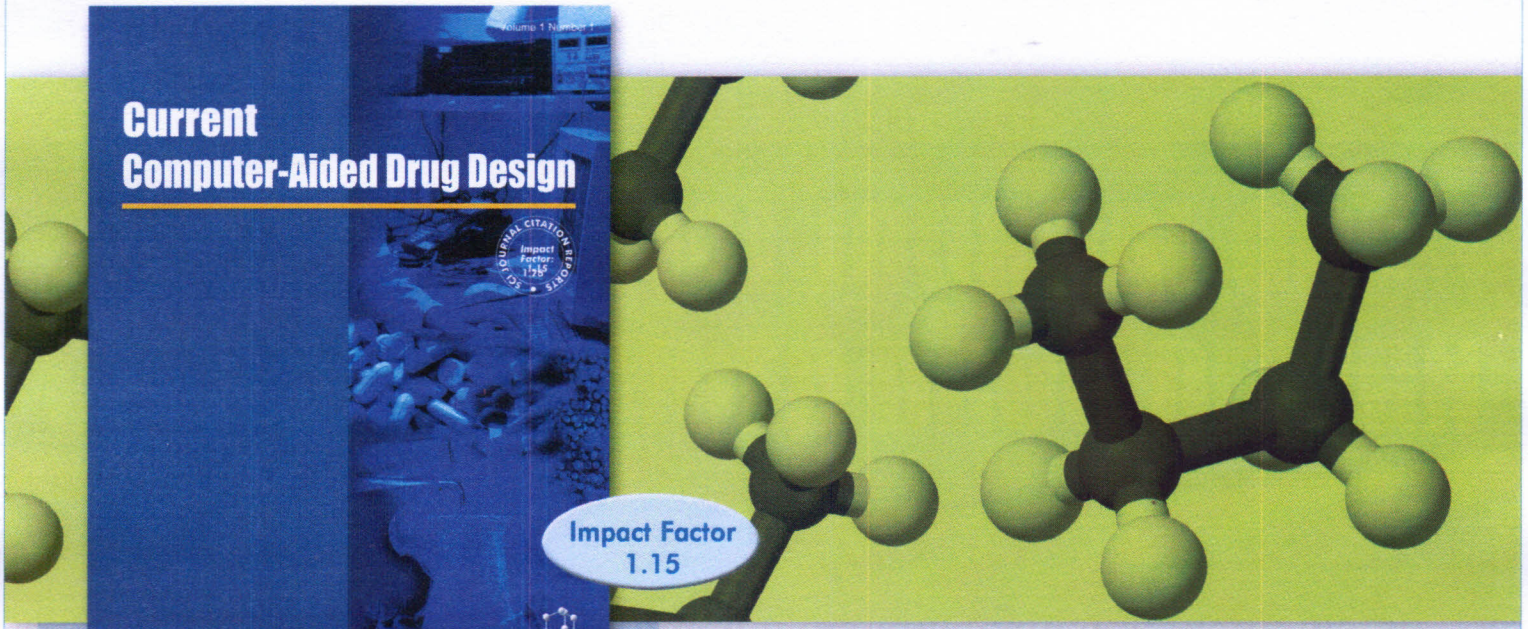
[26] Bowers KJ, Chow E, Xu H, *et al.* Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters; Proceedings of the ACM/IEEE Conference on Supercomputing (SC06) Tampa, FL2006.

[27] Osguthorpe DJ, Sherman W, Hagler AT. Exploring protein flexibility: incorporating structural ensembles from crystal structures and simulation into virtual screening protocols. J Phys Chem B 2012; 116(23): 6952-9.

[28] Hess B, Kutzner C, van der Spoel D, Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J Chem Theory Comp 2008; 4(3): 435-47.

Received: March 25, 2015 Revised: April 23, 2015 Accepted: June 20, 2015

AN ESSENTIAL JOURNAL ON COMPUTER-AIDED DRUG DESIGN



Editor-in-Chief:
Subhash C. Basak, USA

Current Computer-Aided Drug Design aims to publish all the latest developments in drug design based on computational techniques. It is an essential journal for all medicinal chemists who wish to be kept informed and up-to-date with all the latest and important developments in computer-aided methodologies and their applications in drug discovery.

- Publishing Peer Reviewed Articles Rapidly
- Abstracted in Chemical Abstracts, EMBASE, MEDLINE, PubMed and more .....
- FREE Online Trials for Institutions
- FREE Sample Issues Available on Website

For Subscriptions
Contact: subscriptions@benthamscience.org

For Advertising & Free Online Trials
Contact: marketing@benthamscience.org

www.benthamscience.com

BENTHAM SCIENCE

Publishers of Quality Research

Current Computer Aided-Drug Design

ISSN (Print): 1573-4099
ISSN (Online): 1875-6697

Subhash C. Basak
Departments of Chemistry, Biochemistry & Molecular Biology University of Minnesota Duluth
Duluth, MN 55811
USA

Back (/node/582/molecular-dynamics-simulations-of-novel-potential-inhibitors-for-penicillin-binding-protein-2b-of-the-resistant-5

Molecular Dynamics Simulations of Novel Potential Inhibitors for Penicillin Binding Protein 2B of the Resistant 5204 Strain of Streptococcus pneumoniae

(E-pub Abstract Ahead of Print)

Author(s): Subbiah Parthasarathy, Suvaiyarasan Suvaithenamudhan.

Abstract:

Background: Top five best hit compounds (ZINC59376795, ZINC60175365, ZINC36922620, ZINC39550705 and ZINC36953975) were obtained through our high throughput virtual screening (HTVS) analysis with resistant 5204-PBP2B (5204 Penicillin Binding Protein 2B) and sensitive R6-PBP2B (R6 Penicillin Binding Protein 2B) proteins of Streptococcus pneumoniae. Objective: To gain insight in molecular docking and dynamics simulations of these top five best hit compounds with both resistant 5204-PBP2B and sensitive R6-PBP2B targets. Method: We have employed Glide XP docking and molecular dynamics simulations of these five best hit compounds with 5204-PBP2B and R6-PBP2B targets. The stability analysis has been carried out through DFT, prime-MM/GBSA binding free energy, RMSD, RMSF and Principal Component Analysis. Results: The reference drug, penicillin G forms stable complex with sensitive R6-PBP2B protein. Similar stability is observed for the mutant resistant 5204-PBP2B with the top scoring compound ZINC592376795 which implies that this compound may act as an effective potential inhibitor. The compound ZINC59376795 forms a total of five hydrogen bonds with resistant 5204-PBP2B protein of which three are with mutated residues. Similarly, the other four compounds including penicillin G also form hydrogen bonds with mutated residue. The MD simulations and stability analysis of the complexes of wild and mutant forms are evaluated for a trajectory period of 16ns and further MD simulations of ZINC59376795 with resistant 5204-PBP2B and sensitive R6-PBP2B confirmed the stability for 50 ns. Conclusion: These results suggest that the top five best hit compounds are found to be a promising gateway for the further development of anti pneumococcus therapeutics.

Keywords: Glide docking; Molecular dynamics simulation; Penicillin G; penicillin binding protein 2B; Streptococcus pneumoniae; resistant 5204 strain; sensitive R6 strain

Full-Text Inquiry Rights & Permissions Print Export

Other

Article Details

VOLUME: 13
Year: 2017
(E-pub Abstract Ahead of Print)
DOI: 10.2174/1573409913666170301120421
Price: $95

(/terms/termandcondition.html?1)

Molecular dynamics simulations of novel potential inhibitors for penicillin binding protein 2B of the resistant 5204 strain of *Streptococcus pneumoniae*

Suvaiyarasan Suvaithenamudhan and Subbiah Parthasarathy*

*Corresponding author

Department of Bioinformatics, School of Life Sciences, Bharathidasan University, Tiruchirappalli – 620 024, Tamil Nadu, India

Email: bdupartha@gmail.com, partha@bdu.ac.in

Phone: +91 431 2407071

Fax: +91 431 2407045

Mobile: + 91 9443533095

Abstract

Our present study aims to gain insight in molecular docking and dynamics simulations of the top five best hit compounds (ZINC59376795, ZINC60175365, ZINC36922620, ZINC39550705 and ZINC36953975) obtained from our high throughput virtual screening (HTVS) analysis with resistant 5204-PBP2B (5204 Penicillin Binding Protein 2B) and sensitive R6-PBP2B (R6 Penicillin Binding Protein 2B) proteins. Penicillin G is used as a reference drug molecule throughout this study and it forms stable complex with sensitive R6-PBP2B protein. Here, similar stability is observed for the mutant resistant 5204-PBP2B protein structure with the top scoring compound ZINC592376795 which implies that this compound may act as an effective potential inhibitor against *Streptococcus pneumoniae*. The compound ZINC59376795 forms a total of five hydrogen bonds with resistant 5204-PBP2B protein of which three are with mutated residues of transpeptidase (TP) domain. Similar to ZINC59376795, the other four compounds including penicillin G also form hydrogen bonds with mutated residue. The stability analysis of the complexes of wild and mutant forms evaluated through Density functional theory analysis (DFT), prime-MM/GBSA binding free energy, Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), hydrogen bond formation and Principal Component Analysis (PCA) for a trajectory period of 16ns and further MD simulations of top scoring compound ZINC59376795 with resistant 5204-PBP2B and sensitive R6-PBP2B confirmed the stability for 50 ns. Thus the top five best hit compounds are found to be a promising gateway for the further development of anti *pneumococcus* therapeutics.

Keywords: Glide docking, Molecular dynamics simulation, Penicillin G, penicillin binding protein 2B**,** *Streptococcus pneumonia,* resistant 5204 strain, sensitive R6 strain.

1. Introduction

Streptococcus pneumoniae is a gram-positive human pathogenic bacterium. It serves as the common cause of pneumonia, bacteremia, bronchitis, sinusitis, endocarditis, otitis media, sepsis, and meningitis [1, 2]. For the treatment of pneumococcal infections, β-lactams (Penicillins) have been effectively used as potent inhibitors for the penicillin binding proteins (PBPs), because of their effectiveness, low cost, ease of use, and minimal side effects [3]. β-lactams antibiotics form a stable acyl enzyme complex with penicillin binding proteins (PBPs) in the bacterial cell membrane, thus inhibiting the final stages of peptidoglycan biosynthesis [4, 5]. The macromolecular targets of β-lactam antibiotics are PBPs. The 3D structures of penicillin binding protein 2B (PBP2B) of both wild-type R6 (sensitive) and mutant 5204 (resistant) strains are available in protein Data Bank (PDB) [6] with their PDB IDs 2WAF (sensitive) and 2WAE (resistant), respectively. PBP2B of the strain 5204 is a major drug resistance target and its modification is essential for the development of high levels of resistance to penicillin. The 58 mutations observed in full length 5204-PBP2B are located mostly on the transpeptidase (TP) domain and display flexibility in their active site region [7].

We have already obtained five best potential compounds through hierarchical virtual screening of 1,677,620 compounds in ZINC database targeting the PBP2B of the mutated resistant 5204 strain of *Streptococcus pneumonia* [8]. These five compounds were filtered from top scoring 25 hits by carrying out molecular docking with wild type (sensitive) R6-PBP2B and mutated (resistant) 5204-PBP2B protein structures. Our present study aims to gain insight to interaction of the top five best hit compounds (ZINC59376795, ZINC60175365, ZINC36922620, ZINC39550705 and ZINC36953975) and the reference drug molecule penicillin G with resistant 5204-PBP2B and sensitive R6-PBP2B using molecular docking and dynamics simulation investigation. The stability of these compounds with the wild type and mutant

structures are evaluated through Density functional theory (DFT) analysis, prime-MM/GBSA binding free energy, Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF) analysis, Principal Component Analysis (PCA) and hydrogen bond formation for a trajectory period of 16ns. Results observed from the above studies show that those top five compounds were found to serve as potential inhibitors for the development of anti *pneumococcus* therapeutics. The workflow of our entire study is shown as Fig. **1**.

2. Materials and methods

2.1 Hardware and Software

The entire molecular docking and molecular dynamics simulations were performed using high performance computing (HPC) cluster operated with CentOS V6.5 Linux operating system using Schrödinger suite 2015-3 with graphical user interface Maestro 10.3 [9] and GROMACS 4.5 package [10]. The hardware specifications of HPC cluster master node are IBM X3550M4 1U Rack server with dual Intel Xeon E5-2670 V2 10c 2.5 GHz processors with 128 GB of Memory.

2.2 Virtual screening and Glide XP docking

In this study, the five best compounds namely ZINC59376795, ZINC60175365, ZINC36922620, ZINC39550705, ZINC36953975 which were obtained [8] through the high-throughput virtual screening were chosen for molecular docking and dynamics studies. These five compounds were docked against both resistant 5204-PBP2B and sensitive R6-PBP2B using Glide XP docking methodology [11, 12]. In addition, the antibiotic penicillin G drug was also taken into our study to understand interaction profiles, structural behavior and flexibility between the penicillin G and top five best hit compounds.

2.3 Density functional theory analysis

Density functional theory (DFT) calculations were performed to predict the molecular reactivity and stability of the potential compounds identified through virtual screening. The

energy difference between highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) is known as HOMO-LUMO gap [13-15] was calculated and tabulated. It was done by exporting the best binding poses to Maestro 10.3 of Schrödinger 2015-3 version and geometry was optimized in the Jaguar panel using Becke's three-parameter exchange potential [16] and Lee-Yang-Parr correlation functional (B3LYP) theory with 6- 31G* basis set [17, 18].

2.4 Binding free energy calculation

The binding free energy calculation was performed through prime-MM/GBSA (Molecular Mechanics/Generalized Born Surface Area) [19, 20]. Prime uses a surface GB model employing a Gaussian surface instead of a vander Waals surface for better representation of the solvent accessible surface area [21]. Binding energy (ΔG_{bind}) was calculated by the following equations,

$$\Delta G_{bind} = \Delta E + \Delta G_{solv} + \Delta G_{SA}, \quad (1)$$

$$\Delta E = E_{complex} - E_{protein} - E_{ligand}, \quad (2)$$

where, $E_{complex}$, $E_{protein}$, and E_{ligand} are the minimized energies of the protein-inhibitor complex, protein, and inhibitor, respectively [22].

$$\Delta G_{solv} = G_{solv}(complex) - G_{solv}(protein) - G_{solv}(ligand), \quad (3)$$

where, G_{solv}(complex), G_{solv}(protein), and G_{solv}(ligand) are the solvation free energies of the complex, protein, and inhibitor, respectively.

$$\Delta G_{SA} = G_{SA}(complex) - G_{SA}(protein) - G_{SA}(ligand), \quad (4)$$

where, G_{SA} (complex), G_{SA} (protein), and G_{SA} (ligand) are the surface area energies for the complex, protein, and inhibitor, respectively. The rational criteria for selection of best compounds based on scoring and interaction parameters are shown in XP docking which are further used for MD simulation studies.

2.5 Molecular dynamics (MD) simulations of resistant 5204-PBP2B and sensitive R6-PBP2B

The MD simulations of all the systems were carried out using GROMACS 4.5 package [10]. The topology files for the selected proteins were generated using the automated topology builder (ATB) in the framework of GROMOS96 53a6 force field for protein-ligand complex [23, 24]. The topology files for the ligands were generated using PRODRG 2.5 server [25]. The ligand complex obtained from docking was solvated with single point charge (SPC) water model [26]. The solvated system was subjected to 5000 steps of energy minimization employing the steepest descent algorithm. This step was followed by 1 nano second (ns) MD simulation, where the resistant 5204-PBP2B and sensitive R6-PBP2B with ligands complex were position restrained to equilibrate the water and ions under the influence of the solute. The production run was carried out for all the systems for 16 nano seconds (ns) using 2 femto second (fs) time step for the integration of equation of motion in the NPT ensemble at 300 K and at 1 atmospheric pressure, which was controlled using a V-rescale thermostat and Parrinello–Rahman Barostat, respectively. Bond lengths involving hydrogen atoms were constrained by using the Linear Constraint Solver (LINCS) algorithm [27]. The Particle Mesh Ewald (PME) method was used to calculate the electrostatic interaction [28, 29]. The cutoff distances for the long-range electrostatic and van der Waals energy terms were set as 10 Å. The MD simulation coordinates of all the systems were saved at 2 ps interval for further analyses. Post processing and analyses were carried out using GROMACS analysis tools.

2.6 Trajectory analysis

Each trajectory produced after MD simulations was further analyzed and plotted using GnuPlot tool. The MD trajectories were analyzed using g_rms, g_rmsf, g_covar, g_anaeig and g_hbond utilities of GROMACS package to obtain the Root Mean Square Deviation (RMSD),

Root Mean Square Fluctuation (RMSF), Principal Component Analysis (PCA) and number of hydrogen bonds respectively. All the graphs were generated using the GnuPlot tool [30].

2.6.1 Root-Mean-Square Deviation (RMSD)

The RMSD is used to measure the average change in displacement of a selection of atoms for a particular frame with respect to a reference frame. It was calculated for all frames in the trajectory. The RMSD for frame x is given by

$$RMSD_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(r_i'(t_x) - r_i\left(t_{ref}\right)\right)^2}, \quad (5)$$

where N is the number of atoms in the atom selection t_{ref} is the reference time, (typically the first frame is used as the reference and it is regarded as time $t = 0$); and r' is the position of the selected atoms in frame x is recorded at time t_x. The procedure is repeated for every frame in the simulation trajectory.

2.6.2 Root-Mean-Square Fluctuation (RMSF)

The RMSF is useful for characterizing local changes along the protein chain. The RMSF for residue i is given by

$$RMSF_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T} < \left(r_i'(t) - r_i\left(t_{ref}\right)\right)^2 >}, \quad (6)$$

where T is the trajectory time over which the RMSF is calculated, t_{ref} is the reference time, r_i is the position of residue i; r' is the position of atoms in residue i after superposition on the reference, and the angle brackets indicate that the average of the square distance is taken over the selection of atoms in the residue.

2.6.3 Hydrogen Bond

The hydrogen bonds between protein and ligand were analyzed using the g_hbond analysis in the GROMACS. The distance criterion for the hydrogen bonds is d $\leq$ 3.5 Å between donor and acceptor. The angle between donor and acceptor is greater than 30°.

2.6.4 Principal Component Analysis

Principal component analysis (PCA) was performed for all the trajectories. The GROMACS inbuilt tools g_covar and g_anaeig were used for performing PCA analysis. The trajectory of a MD simulation was utilized to identify the motions of the sensitive R6-PBP2B and resistant 5204-PBP2B models. We used principal component analysis to extract the principal modes involved in the motion of the protein molecule [31]. A covariance matrix was assembled using a simple linear transformation in Cartesian coordinate space. A vectorial depiction of every single component of the motion indicates the direction of motion. For this, a set of eigenvectors was derived through the diagonalization of the covariance matrix. Each eigenvector has a corresponding eigenvalue that describes the energetic contribution of each component to the motion [32]. The protein regions that are responsible for the most significant collective motions can be acknowledged through PCA.

3. Results and discussion

3.1 Docking and interaction analysis

Glide XP docking results (interaction patterns, docking score and binding free energy) of five best hit compounds (ZINC59376795, ZINC60175365, ZINC36922620, ZINC39550705, ZINC36953975) and penicillin as the reference drug molecule with both resistant 5204-PBP2B and sensitive R6-PBP2B targets are tabulated in Table **1**. Figures **2** and **3** explain the confirmation of the ligands in the binding pocket of target. From Table **1**, it is inferred that the best hit compound 5-[(6-hydroxy-1,2,3,4-tetrahydroisoquinolin-1-yl)methyl]benzene-1,2,3-triol with ID: ZINC59376795 shows higher affinity towards both the sensitive and resistant protein and the reference drug molecule penicillin exhibited least affinity compared with other compounds.

3.2 XP Docking pose analysis and binding energy evaluation of resistant 5204-PBP2B

To determine the binding affinity between resistant 5204-PBP2B and drug complexes we have calculated MM-GBSA binding free energy of the complexes. It is observed from the results that the top hit ZINC59376795 compound is found to bind efficiently with binding energy (ΔG_{bind}) of -49.25 kcal/mol and formed five hydrogen bonds (Table **1**). The compound ZINC59376795 forms five hydrogen bonds of which three are with mutated residues of TP domain (ASN630 forms two hydrogen bonds and ASN660 forms a single hydrogen bond) which are encircled in red colour in Fig. **2A** and also with other residues THR600 and LEU657. The top hit ZINC59376795 compound forms five hydrogen bonds while compared with four other compounds.

Apart from the top scoring compound, the remaining four compounds, ZINC60175365, ZINC36922620, ZINC39550705, ZINC36953975 and penicillin G with resistant 5204-PBP2B protein (Fig. **2B-F**) forms an average of three hydrogen bonds. ZINC60175365, ZINC36922620, ZINC36953975 and penicillin G form hydrogen bonds with the mutated residue ASN660 (encircled in red colour in Fig. **2A-C, 2E-F**) and with other residues (Table **1**). The compound ZINC39550705 exhibits hydrogen bonding interactions with residues ASN654 and ASN656 of TP domain which are neither mutated residue nor active site residues (Fig. **2D**). Though penicillin G has a low binding free energy of -25.65 kcal/mol, it shows less efficient binding with resistant 5204-PBP2B protein in contrast to other compounds as inferred from the docking results (Table **1**).

3.3 XP Docking pose analysis and binding energy evaluation of sensitive R6-PBP2B

In sensitive R6-PBP2B protein, the top hit compound ZINC59376795 forms hydrogen bonds with active site residues SER386 and GLY617 (encircled in green color in Fig. **3A**) of TP domain with bond lengths of 2.259 and 2.152 Å, respectively (Table **1**). The compounds ZINC60175365, ZINC36922620, ZINC39550705 and ZINC36953975 form hydrogen bonds

with Asn445 (encircled in green color in Fig. **3B-E**) which is at the active site and also with other residues (Table **1**). Sensitive R6-PBP2B shows that the reference drug molecule penicillin G forms interactions with active site residues SER386 and THR616 (encircled in green color in Fig. **3F**). All the five compounds have interaction with the active site residues of sensitive R6-PBP2B and hence it should work well in mutant resistant 5204-PBP2B.

3.4 Density functional theory analysis

HOMO and LUMO of potential molecules are important to establish the interactions between drug and receptor [15]. Hence, the HOMO and LUMO of five potential compounds were analyzed using DFT calculations. The HOMO-LUMO gap which can explain the chemical reactivity and stability of drug molecules was calculated for five compounds and tabulated in Table **2**. The contour plots of HOMO and LUMO are given Fig **4**. For ZINC59376795, the contribution of hydroxyl groups is significant and the same involved in hydrogen bond formation with receptor which is observed from docking analysis. The calculated HOMO-LUMO gap is high for ZINC59376795 (0.17990 eV) and it is the lowest in the case of ZINC60175365 (0.14293 eV). The calculated HOMO-LUMO gaps of drug molecules lies in the range from 0.142 to 0.179 eV and it indicates the significant reactivity of the five potential compounds. The HOMO-LUMO gap values are inversely proportional to binding free energies of the drugs with receptor. These results are close in agreement with the MM-GBSA calculations.

3.5 Molecular dynamics simulations

To compare the structural behavior and flexibility of the wild-sensitive R6-PBP2B and mutated-resistant 5204-PBP2B, all the five lead compounds are studied using MD simulations for 16 ns for each complex. Figure **5** shows the variation of RMSD from the starting conformation. It is observed that throughout the simulation period (16ns), the protein backbone of the resistant 5204-PBP2B (2WAE) is flexible than its wild type R6-PBP2B (2WAF) with their maximum RMSD values at 0.50 and 0.40 nm, respectively. From the RMSD analysis of the

trajectories of both protein-ligand complexes (resistant 5204-PBP2B (Fig.**6A-F**) and sensitive R6-PBP2B (Fig.**6G-L**)), it is observed that the fluctuations are more in sensitive R6-PBP2B than resistant 5204-PBP2B.

From Table **3**, it is evident that for the mutant protein, resistant 5204-PBP2B, the top two compounds ZINC59376795 and ZINC60175365 have lower average RMSD values of 0.33 and 0.36 nm, respectively (Fig. **6A** and **B**) and these compounds are more stable inside the protein than the remaining three compounds. The compounds ZINC36922620, ZINC39550705, ZINC36953975 and penicillin G with resistant 5204-PBP2B have average RMSD values at 0.44, 0.39, 0.49 and 0.45 nm, respectively (Fig. **6C-F**). The above three compounds and penicillin G have higher average RMSD values and hence it is found that the top two compounds ZINC59376795 and ZINC60175365 are more stable. Investigation of RMSD of protein backbone of resistant 5204-PBP2B complex with penicillin G reveals a steady increase in the RMSD values till 16 ns, where it reaches the maximum at 0.69 nm (Fig. **6F**). The strain 5204 is highly resistant to penicillin G, as the value Minimum Inhibitory Concentration (MIC) is found to be 6 μg/ml as reported by Pagliero E *et al*., in 2004 [33]. The present study also adopts the above resistant 5204-PBP2B complex with penicillin G and investigates this as a reference drug molecule throughout our analysis.

From Table **3**, it is observed that forthe sensitive R6-PBP2B complex, the top two compounds 5-[(6-hydroxy-1,2,3,4-tetrahydroisoquinolin-1-yl)methyl]benzene-1,2,3-triol (ZINC59376795) and prop-2-ynyl 2,6-diamino-3-[3-[(4- chlorophenyl) methylcarbamoyl]-4-oxo-1H-quinolin-6-yl]hexanoate (ZINC60175365) have average RMSD values of 0.54 and 0.56 nm, respectively (Fig. **6G** and **H**). From the trajectories plotted in Fig. **6I-K**, it is found that the compounds ZINC36922620, ZINC39550705 and ZINC36953975 have average RMSD values at 0.45, 0.46 and 0.41 nm, respectively. The penicillin G has a lower average RMSD value of 0.35 nm and hence it is stable with sensitive R6-PBP2B. The RMSD analysis of penicillin G with

R6-PBP2B (Fig. **6L**) showed that there is no significant increase in RMSD values of the backbone. This result is in contrast to the results obtained for the resistant 5204-PBP2B where a steady increase in RMSD is observed (Fig. **6F**) as expected.

Comparison of RMSD values of the resistant 5204-PBP2B (Fig. **6A-E**) and sensitive R6-PBP2B complex (Fig. **6G-K**) shows that R6-PBP2B has more fluctuations and hence resistant 5204-PBP2B is stable than R6-PBP2B. Further MD simulation for 50 ns was carried out for the best potential compound ZINC59376795 to verify the stability of ZINC59376795 with 5204-PBP2B and R6-PBP2B complex beyond 16 ns. The results showed in supplementary Fig. **S1A** shows that the compound ZINC59376795 form stable complexes even at 50 ns with the resistant 5204-PBP2B and with sensitive R6-PBP2B, the compound ZINC59376795 form complexes with more fluctuations (supplementary Fig. **S1B**).

3.5.1 Root Mean Square Fluctuation

In order to calculate the residual mobility of each residue, the RMSF values are computed for the wild-type and mutant protein complexes. The calculated RMSF values are plotted in Fig. **7**. The high RMSF value indicates more flexibility whereas the low RMSF value indicates limited movements during simulation in relation to the residues average position. The RMSF of the residues are shown in Fig. **7A** and **B** clearly depicting different flexibility in resistant 5204-PBP2B and sensitive R6-PBP2B complexes.

In resistant 5204-PBP2B (Fig.**7A**), it is observed that reference drug molecule penicillin G induces the flexibility of the protein residues than the remaining five compounds average position. It is observed that the residues 110-180 in resistant 5204-PBP2B show more fluctuation. This same sequence segment in sensitive R6-PBP2B (Fig. **7B**) is found to be more flexible than resistant 5204-PBP2B. In sensitive R6-PBP2B the RMSF values of the residues 110-180 ranges upto a maximum value of 1.3 nm.

In resistant 5204-PBP2B, the residues LYS210, GLN458 and GLY594 has high fluctuation with penicillin G, whereas the same residues show less fluctuation with the remaining five compounds. In sensitive R6-PBP2B, the residues GLU247, GLY417 and GLY467 has fluctuation with ZINC59376795.

3.5.2 Hydrogen bond analysis

To determine the number of hydrogen bonds with the TP binding site, g_hbond utility of GROMACS is used. From Table **3** and Fig.**8**, it is observed that except the top ZINC59376795 compound all the other four compounds form approximately equal number of hydrogen bonds with resistant 5204-PBP2B. Whereas in sensitive R6-PBP2B, ZINC60175365 forms an average of ~4 hydrogen bonds followed by ZINC39550705 with ~3 hydrogen bonds and the remaining three compounds form an average of one hydrogen bond throughout the period of 16 ns. Penicillin G forms an average of two hydrogen bonds with both resistant 5204-PBP2B and sensitive R6-PBP2B protein (Table **3** and Fig. **8**). The ZINC59376795, ZINC60175365 and ZINC39550705 with resistant 5204-PBP2B and sensitive R6-PBP2B complexes exhibited stable and strong hydrogen bonds throughout the simulation time period.

3.5.3 Principal component analysis

Principal component analysis (PCA) was performed on all the twelve trajectories of resistant 5204-PBP2B (Fig. **9A-F**) and sensitive R6-PBP2B (Fig. **9G-L**) forms to monitor the overall strenuous motions of the protein. Diagonal covariance matrices were built over the backbone of the protein for each trajectory. The eigenvalues obtained through the diagonalization of the covariance matrix elucidates the atomic contribution on motion. Similarly, the eigenvectors explain a collective motion accomplished by the particles [34]. Here, the overall flexibility is calculated by the trace of diagonalized covariance matrix. The trace values for the resistant 5204-PBP2B complex, the compounds ZINC59376795, ZINC60175365,

ZINC36922620, ZINC39550705, ZINC36953975 and penicillin G are found to be 74.9631, 85.7122, 82.2791, 58.5064, 69.4204 and 110.192 nm^2, respectively (Fig. **9A-F**). Penicillin G has higher trace value while compared with 5 other compounds. The trace values for the sensitive R6-PBP2B complex, the compounds ZINC59376795, ZINC60175365, ZINC36922620, ZINC39550705, ZINC36953975 and penicillin G are found to be 136.399, 97.9876, 117.533, 88.214, 70.6205 and 88.7796 nm^2, respectively (Fig. **9G-L**). Among these trace values, sensitive R6-PBP2B complexes show high values suggesting an overall escalation in the flexibility than the resistant 5204-PBP2B complexes. Whereas resistant 5204-PBP2B complexes exhibit the lowest value confirming the decrease in flexibility in the collective motion of the protein. From these projections, it is observed that clusters of resistant 5204-PBP2B were well defined and was more stable compared to the sensitive R6-PBP2B.

Porcupine plot of PCA movements of resistant 5204-PBP2B (Fig. **9A-F**) and sensitive R6-PBP2B (Fig. **9G-L**) show the confirmation changes from ‘‘open’’ state to ‘‘close’’ state. From Fig. **9A** it is found that, the compound 5-[(6-hydroxy-1,2,3,4-tetrahydroisoquinolin-1-yl)methyl]benzene-1,2,3-triol (ZINC59376795) with resistant 5204-PBP2B complex conformation changed to open state when binding with the mutated residues of TP domain. For resistant 5204-PBP2B complex with penicillin G, the conformation changed to close state when binding with one mutated residue (Fig. **9F**). Based on the porcupine plot of PCA movements of ZINC59376795-resistant 5204-PBP2B complex depicts that the distribution of free form of resistant 5204-PBP2B is large compared to the penicillin G-resistant 5204-PBP2B bound form.

4. Conclusion

In the present study, molecular docking and dynamics simulations were carried out for the top five best hits and the reference drug molecule penicillin G available as reference with 5204-PBP2B and sensitive R6-PBP2B proteins. The top hit compound 5-[(6-hydroxy-1,2,3,4-tetrahydroisoquinolin-1-yl)methyl]benzene-1,2,3-triol (ZINC59376795) formed maximum of

five hydrogen bonds with resistant 5204-PBP2B protein and other compounds formed an average of three hydrogen bonds with this protein. All the five compounds and penicillin G have hydrogen bond interactions with the mutated residue ASN660 in the resistant 5204-PBP2B. In the case of sensitive R6-PBP2B, ASN445, which is a key active residue, forms hydrogen bond interactions with most of these compounds. Also DFT calculations show the HOMO-LUMO gap of the five best potential compounds in the range from 0.142 - 0.179 eV and it signifies the reactivity of these novel compounds.

The stability of the resistant 5204-PBP2B with the top five compounds evaluated through RMSD and RMSF analysis showed that throughout the time period of 16 ns the protein-ligand backbone stability was not affected by these compounds and was stable throughout the simulation period without any significant fluctuation. A better rigidity and stability was observed during the RMSD analysis of wild type sensitive R6-PBP2B with reference drug molecule penicillin G (Fig. **6L**). A similar finding was observed while considering the RMSD analysis of the mutant structure resistant 5204-PBP2B with the top scoring compound ZINC592376795 (Fig. **6A**) which implies that this compound may act as an effective potential inhibitor against *Streptococcus pneumoniae*. Principal component analysis was performed by the trace of diagonalized covariance matrix. The trace values for the resistant 5204-PBP2B complex, the compounds ZINC59376795 and penicillin G were found to be 74.9631 and 110.192 nm^2, respectively. Penicillin G had higher trace value while compared with five other compounds. Further, based on the porcupine plot of PCA movements of the top scoring compound resistant 5204-PBP2B-ZINC59376795 complex depicts that the distribution of free form of resistant 5204-PBP2B was large compared to the resistant 5204-PBP2B-penicillin G complex bound form, which confirmed that this compound may act as an effective potential inhibitor. Further MD simulation carried out for 50 ns for the best compound ZINC59376795 with resistant 5204-PBP2B confirmed the formation of stable complexes even after 16 ns. Along

372 with this top screening compound, the remaining four compounds presented in this study are also
373 found to be a promising gateway for the further development of anti *pneumococcus* therapeutics.
374

Acknowledgments

This work forms part of the major research project (F. No. 41-965/2012 (SR)) funded by the University Grants Commission (UGC), New Delhi. One of the authors (SS) gratefully acknowledges the support provided by the UGC (MRP) in the form a project fellow. Molecular dynamics studies were performed using High Performance Computing (HPC) facility [DST-PURSE (Grant no. SR/FT/LS-113/2009)] at University Informatics Centre, Bharathidasan University, Tiruchirappalli.

References

1. Musher, DM. in *The Pneumococcus*, ed. E. I. Tuomanen, ASM Press, Washington, DC, **2004**, ch.14, pp. 211-220.

2. Klugman, KP. in *The Pneumococcus*, ed. E. I. Tuomanen, ASM Press, Washington, DC, **2004**, ch. 20, pp. 331-338.

3. Frère, JM.; Joris, B. Penicillin-sensitive enzymes in peptidoglycan biosynthesis. *Crit. Rev. Microbiol*. **1985**, 11(4), 299-396.

4. Bergmann, C.; Chi, F.; Rachid, S.; Hakenbeck, R. Mechanisms for penicillin resistance in *Streptococcus pneumoniae*: penicillin-binding proteins, gene transfer, and cell wall metabolism. *The Pneumococcus*, ed. E. I. Tuomanen, ASM Press, Washington, DC, **2004**, ch. 21, pp. 339-349.

5. Barreteau, H.; Kovac, A.; Boniface, A.; Sova, M.; Gobec, S.; Blanot, D. Cytoplasmic steps of peptidoglycan biosynthesis. *FEMS Microbiol Rev.*, **2008**, 32(2), 168-207.

6. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The protein data bank. *Nucleic Acids Res.*, **2000**, 28(1), 235-242.

7. Contreras-Martel, C.; Dahout-Gonzalez, C.; Martins Ados, S.; Kotnik, M.; Dessen, A. PBP active site flexibility as the key mechanism for beta-lactam resistance in pneumococci. *J. Mol. Biol.*, **2009**, 387(4), 899-909.

8. Suvaithenamudhan, S.; Parthasarathy, S. Structure based virtual screening for the identification of potential inhibitors for penicillin binding protein 2B of the resistant 5204 strain of *streptococcus pneumoniae*. *Current Bioinformatics*, **2016**, 11(1), 66-78.

9. Maestro, version 10.3, Schrödinger, LLC, New York, NY, **2015**.

10. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory. Comput.*, **2008**, 4(3), 435-447.

11. Glide, version 6.8, Schrödinger, LLC, New York, NY, **2015**.

12. Friesner, R.A.; Banks, J.L.; Murphy, R.B.; Halgren, T.A.; Klicic, J.J.; Mainz, D. T.; Repasky, M.P.; Knoll, E.H.; Shelley, M.; Perry, J.K.; Shaw, D.E.; Francis, P.; Shenkin, P.S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.*, **2004**, 47(7), 1739-1749.

13. Matysiak, J. Evaluation of electronic, lipophilic and membrane affinity effects on antiproliferative activity of 5-substituted-2-(2,4-dihydroxyphenyl)-1,3,4-thiadiazoles against various human cancer cells. *Eur. J. Med. Chem.*, **2007,** 42(7), 940–947.

14. Zhan, C.G.; Nichols, J.A.; Dixon, D.A. Ionization potential, electron affinity, electronegativity, hardness, and electron excitation energy: molecular properties from density functional theory orbital energies. *J. Phys. Chem. A.*, **2003**, 107(20), 4184–4195.

15. Zheng, Y.; Zheng, M.; Ling, X.; Liu, Y.; Xue, Y.; An, L.; Gu, N.; Ji, M. Design, synthesis, quantum chemical studies and biological activity evaluation of pyrazole-benzimidazole derivatives as potent Aurora A/B kinase inhibitors. *Bioorg. Med. Chem.Lett.*, **2013**, 23(12), 3523-3530.

16. Becke, A.D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.*, **1993**, 98(7):5648–5652.

17. Gill, P.M.W.; Johnson, B.G.; Pople, J.A.; Frisch, M.J. The performance of the Becke—Lee—Yang—Parr (B—LYP) density functional theory with various basis sets. *Chem. Phys. Lett.*, **1992**, 197(4), 499–505.

18. Stephens, P.J.; Devlin, F.J.; Chabalowski, C.F.; Frisch, M.J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.*, **1994**, 98(45), 11623–11627.

19. Prime, version 4.1, Schrödinger, LLC, New York, NY, **2015**.

20. Mobley, D.L.; Dill, K.A. Binding of small-molecule ligands to proteins: "what you see" is not always "what you get". *Structure*, **2009**, 17(4), 489-498.

21. Jakubík, J.; Randáková, A.; Doležal, V. On homology modeling of the M_2 muscarinic acetylcholine receptor subtype. *J. Comput. Aided. Mol.*, **2013**, 27(6), 525-538.

22. Das, D.; Koh, Y.; Tojo, Y.; Ghosh, A.K.; Mitsuya, H. Prediction of potency of protease inhibitors using free energy simulations with polarizable quantum mechanics based ligand charges and a hybrid water model. *J. Chem. Inf. Model.*, **2009**, 49(12), 2851-2862.

23. Oostenbrink, C.; Villa, A.; Mark, A.E.; van Gunsteren, W.F. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force field parameter sets 53A5 and 53A6. *J. Comput. Chem.*, **2004**, 25(13), 1656-1676.

24. Van Gunsteren, W.F.; Billeter, S.R.; Eising, A.A.; Hünenberger, P.H.; Krüger, P.; Mark, A.E.; Scott, W.R.P.; Tironi, I.G. Biomolecular simulation: The GROMOS96 manual and user guide. *Verlag der Fachvereine, Zürich*, 1996, pp.1-1024.

25. Schüttelkopf, A.W.; van Aalten, D.M. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.*, **2004**, 60(8), 1355-1363.

26. Berendsen, H.J.C.; Postma, J.P.M.; van Gunsteren, W.F.; Hermans, J. Interaction models for water in relation to protein hydration. In Pullman, B. (Ed.), Intermol. Forces, **1981**, pp. 331-342.

27. Hess, B.; Bekker, H.; Berendsen, H.J.; Fraaije, J.G. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.*, **1997**, 18(12), 1463–1472.

28. Essmann, U.; Perera, L.; Berkowitz, M.L.; Darden, T.; Lee, H.; Pedersen, L.G. A smooth particle mesh Ewald method. *J. Chem. Phys.*, **1995**, 103(19), 8577-8593.

29. Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald-an N.Log (N) method for Ewald sums in large systems. *J. Chem. Phys.*, **1993**, 98(12), 10089–10092.

30. T. Williams and C. Kelley, available from: http://www.gnuplot.info/

31. Amadei, A.; Linssen, A.B.; Berendsen, H.J. Essential dynamics of proteins. *Proteins*, **1993**, 17(4), 412-425.

32. Mesentean, S.; Fischer, S.; Smith, J.C. Analyzing large-scale structural change in proteins: comparison of principal component projection and Sammon mapping. *Proteins*, **2006**, 64(1), 210-218.

33. Pagliero, E.; Chesnel, L.; Hopkins, J.; Croizé, J.; Dideberg, O.; Vernet, T.; Di Guilmi, A.M. Biochemical characterization of *Streptococcus pneumoniae* penicillin-binding protein 2b and its implication in beta-lactam resistance. *Antimicrob. Agents Chemother.*,**2004**, 48(5), 1848-1855.

34. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.E.; Berendsen, H.J. GROMACS: fast, flexible, and free. *J. Comput. Chem.*, **2005**, 26(16), 1701-1718.

List of Tables

List of Figures

Figure 9. 2D projections of backbone motion and displacement of residues along eigenvectors 1and 2 for the resistant 5204-PBP2B (A-F) and sensitive R6-PBP2B (G-L) structures. Porcupine plot of resistant 5204-PBP2B (red colour) and Sensitive R6-PBP2B (green colour).

Supplementary Figure

Figure S1**.** Backbone RMSD values of ligand ZINC59376795from both the types resistant 5204-PBP2B (A) and sensitive R6-PBP2B (B) of protein-ligand complexes were generated against the initial structures of protein-ligand complexes during 50 ns of molecular dynamics (MD) simulation period. Red colour indicates resistant 5204-PBP2B (A); the sensitive R6-PBP2B (B) is shown in blue and green colour indicates ligands.

Table 1

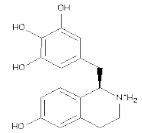
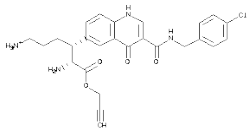
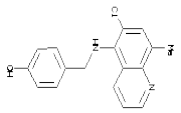
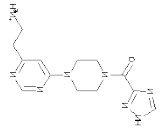
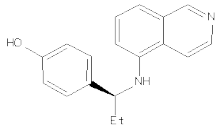
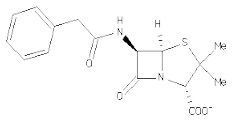
Compound ID 2D Structures	Resistant 5204-PBP2B (Mutant)					Sensitive R6- PBP2B (Wild)			
	IUPAC Name	Docking Score	Interacting residues	Bond length (Å)	MM-GBSA ΔG_{bind} kcal/mol	Docking Score	Interacting residues	Bond length (Å)	MM-GBSA ΔG_{bind} kcal/mol
ZINC59376795 	5-[(6-hydroxy-1,2,3,4-tetrahydroisoquinolin-1-yl)methyl]benzene-1,2,3-triol	-8.060	THR600 ASN630 ASN630 LEU657 ASN660	1.763 1.849 1.714 1.980 2.048	-49.25	-6.731	SER386 GLY617	2.259 2.152	-60.06
ZINC60175365 	prop-2-ynyl 2,6-diamino-3-[3-[(4- chlorophenyl) methylcarbamoyl]-4-oxo-1H-quinolin-6-yl]hexanoate	-8.206	ASN631 ASN656 LEU657 ASN660	2.329 2.440 2.095 2.483	-68.23	-5.369	ASN445 THR599 THR616 GLU620	1.833 2.157 2.190 1.814	-81.55
ZINC36922620 	4-[[(6-chloro-8-methylquinolin-5-yl)amino]methyl]phenol	-7.810	LEU657 THR600 ASN660	1.943 1.781 2.054	-54.82	-6.100	ASN445 GLY660	2.145 2.473	-53.62
ZINC39550705 	[4-[6-(2-aminoethyl)pyrimidin-4-yl]piperazin-1-yl]-(1H-1,2,4-triazol-5-yl)methanone	-7.014	ASN654 ASN656	2.007 2.075	-51.43	-7.179	ASN445 GLU620 THR658	2.195 1.643 2.073	-54.99
ZINC36953975 	4-[1-(isoquinolin-5- ylamino) propyl]phenol	-7.540	THR600 LEU657 ASN660	1.776 2.124 2.048	-50.65	-5.923	ASN445	2.231	-48.04
Penicillin G 	(2S,5R,6R)-3,3-dimethyl-7-oxo-6-[(2-phenylacetyl)amino]-4-thia-1-azabicyclo [3.2.0]heptane-2-carboxylic acid	-4.261	ASN660 ASN660	2.00 1.73	-25.65	-5.126	THR616 SER386	1.900 2.050	-52.56

Table 2

Sl.No	Compound ID	HOMO (eV)	LUMO (eV)	*HLG (eV)
1	ZINC59376795	-0.31234	-0.13244	0.17990
2	ZINC60175365	-0.28502	-0.14209	0.14293
3	ZINC36922620	-0.20278	-0.05162	0.15116
4	ZINC39550705	-0.33448	-0.17610	0.15838
5	ZINC36953975	-0.21922	-0.04862	0.17060

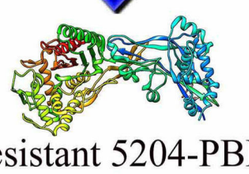
* HOMO-LUMO gap energy (HLG)

Table 3

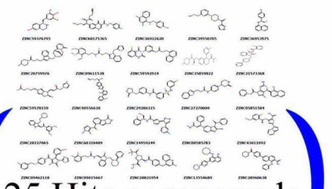
Complexes with Resistant	Resistant 5204-PBP2B		Complexes with Sensitive	Sensitive R6-PBP2B		RMSD Difference Between 5204 &R6 *(a-b)*
	Average (RMSD) (nm) *a*	Average number of hydrogen bond interactions		Average (RMSD) (nm) *b*	Average number of hydrogen bond interaction	
ZINC59376795	0.33	0.54	ZINC59376795	0.54	0.86	-0.21
ZINC60175365	0.36	1.43	ZINC60175365	0.56	3.41	-0.2
ZINC36922620	0.44	1.12	ZINC36922620	0.45	0.51	-0.01
ZINC39550705	0.39	1.41	ZINC39550705	0.46	2.35	-0.07
ZINC36953975	0.49	1.57	ZINC36953975	0.41	0.86	0.08
Penicillin G	0.45	1.47	Penicillin G	0.35	1.51	0.1

ZINC database retrieve 1,677,620 compounds in pH of mid range (6-8)
Filtering (Ligand Optimization, Reactive, QikProp, Lipinski's)
X-ray crystal structure 2WAE Resistant 5204 PBP2B
Virtual Screening Docking mode (HTVS, SP, XP) c.f. Ref. 8

ZINC Database
1,677,620 compounds

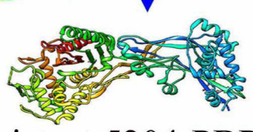


Resistant 5204-PBP2B

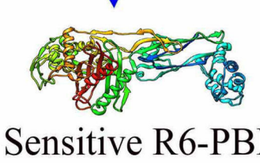


25 Hits compounds

25 Hits compounds
Again, these 25 compounds Glide Docking XP against both Resistant 5204-PBP2B & Sensitive R6-PBP2B targets

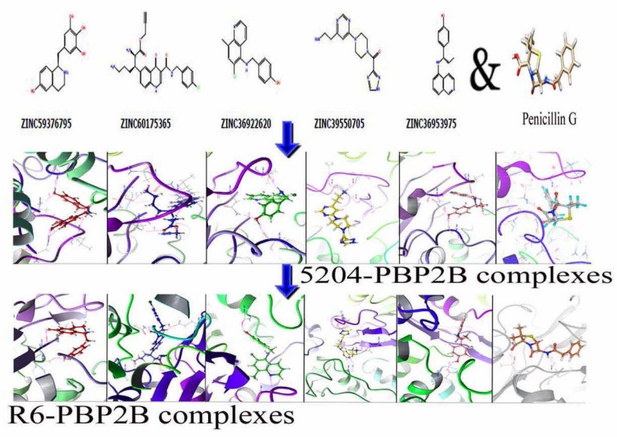


Resistant 5204-PBP2B

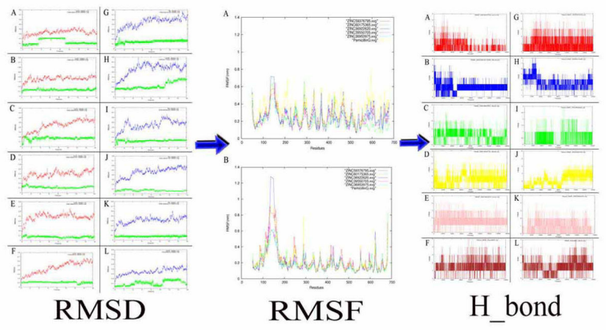


Sensitive R6-PBP2B

Top 5 best hits and reference drug Penicillin G
Molecular docking stidy of Resistant 5204-PBP2B & Sensitive R6-PBP2B
prime-MM/GBSA binding free energy



Simulations stidy of Resistant 5204-PBP2B & Sensitive R6-PBP2B complexes
Analysis of RMSD, RMSF, H_bond, PCA



Top scoring compound
ZINC592376795

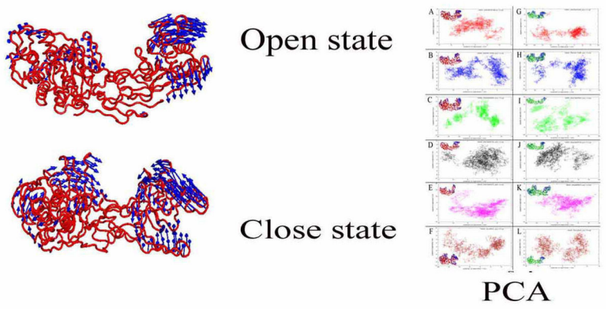
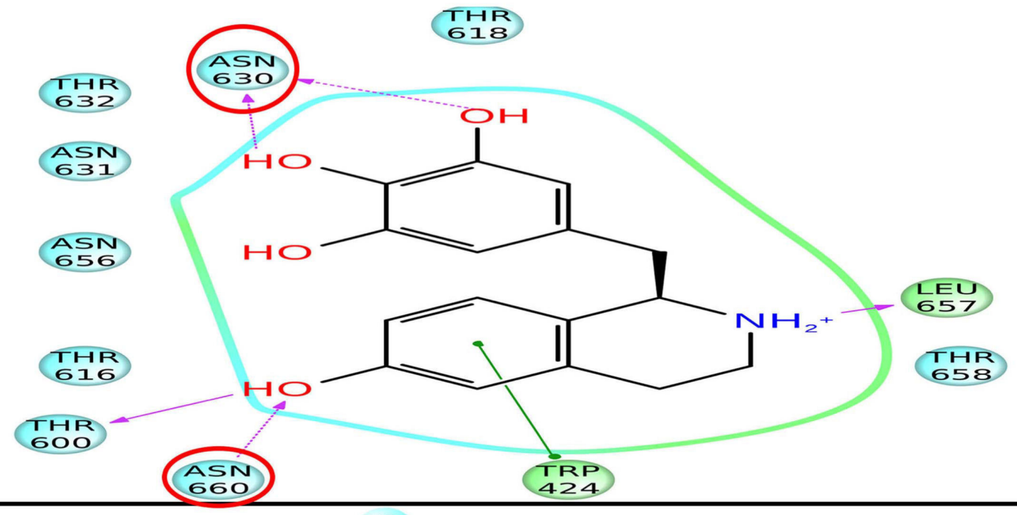
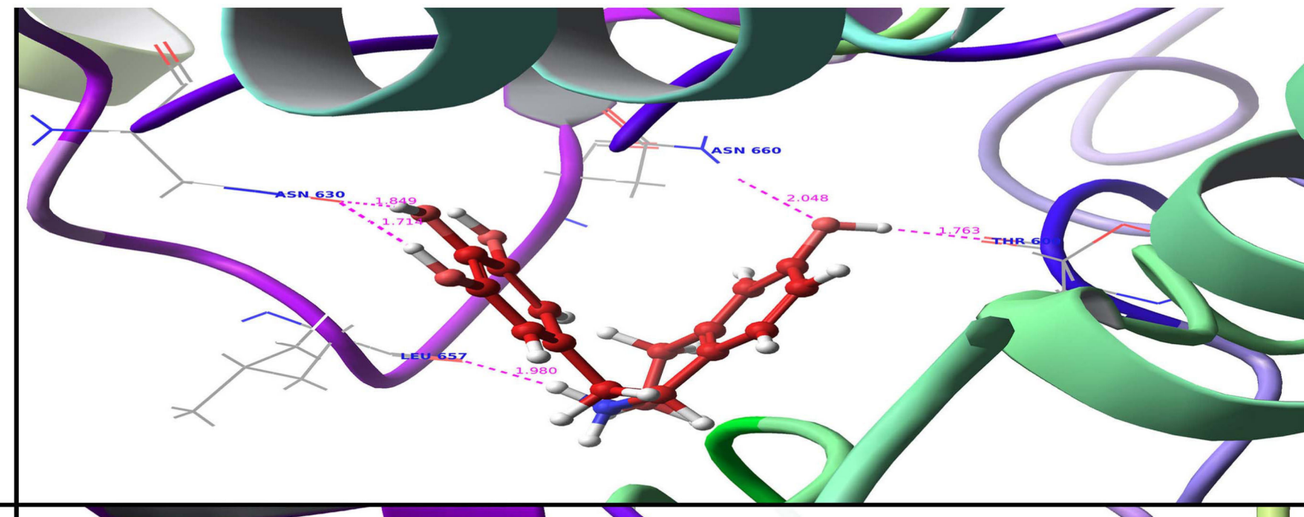


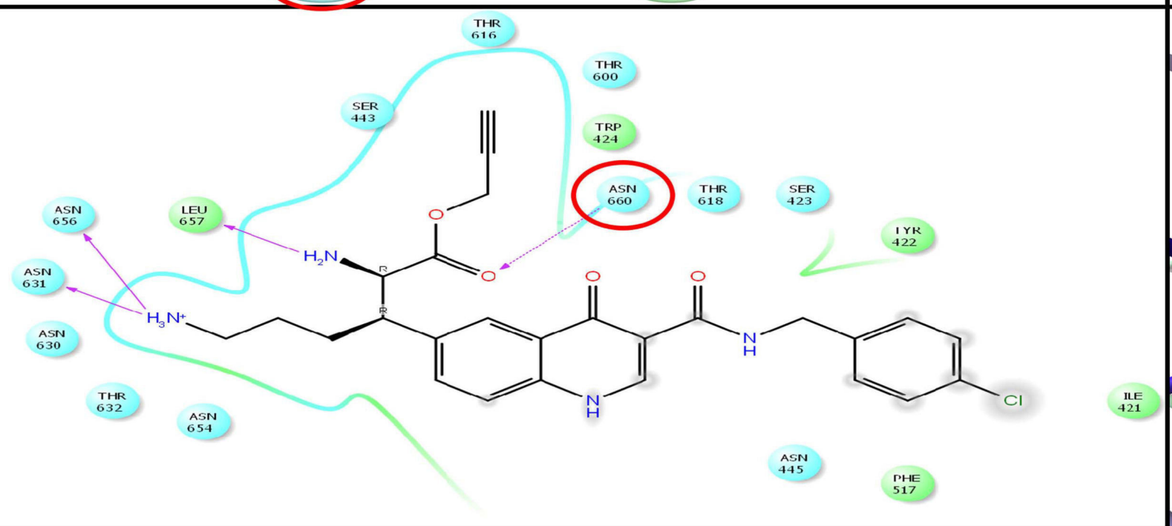
Figure 1

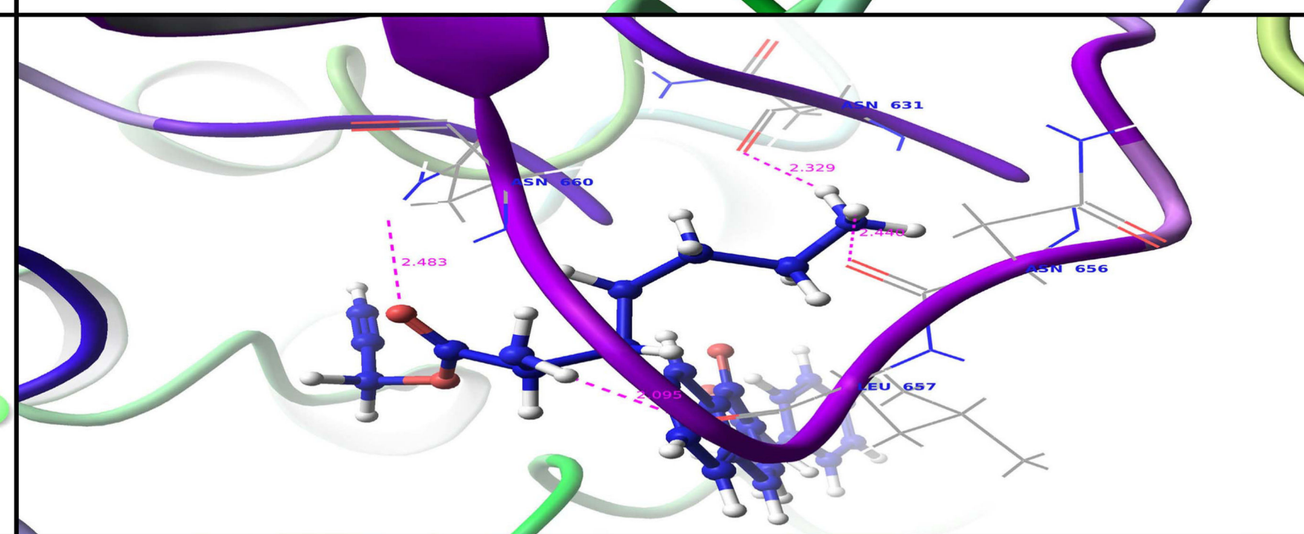
A



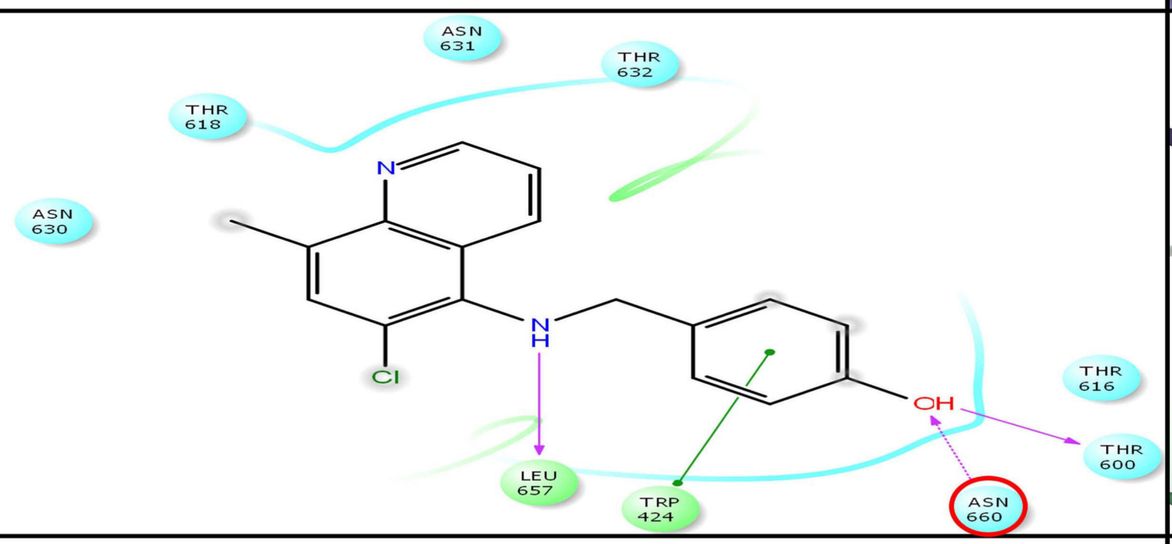


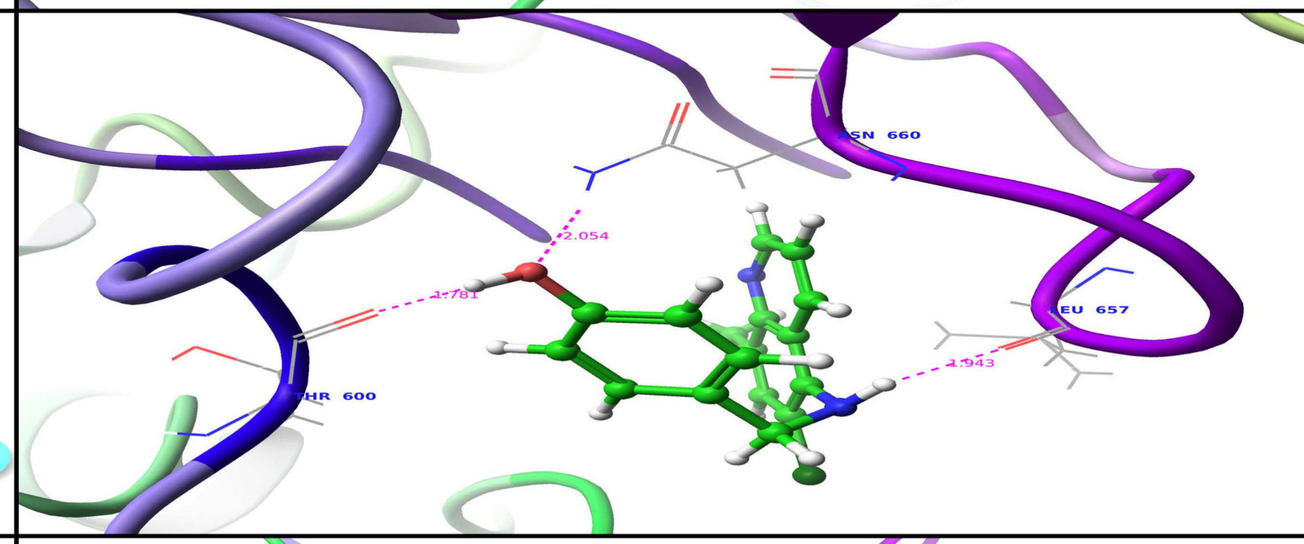
B



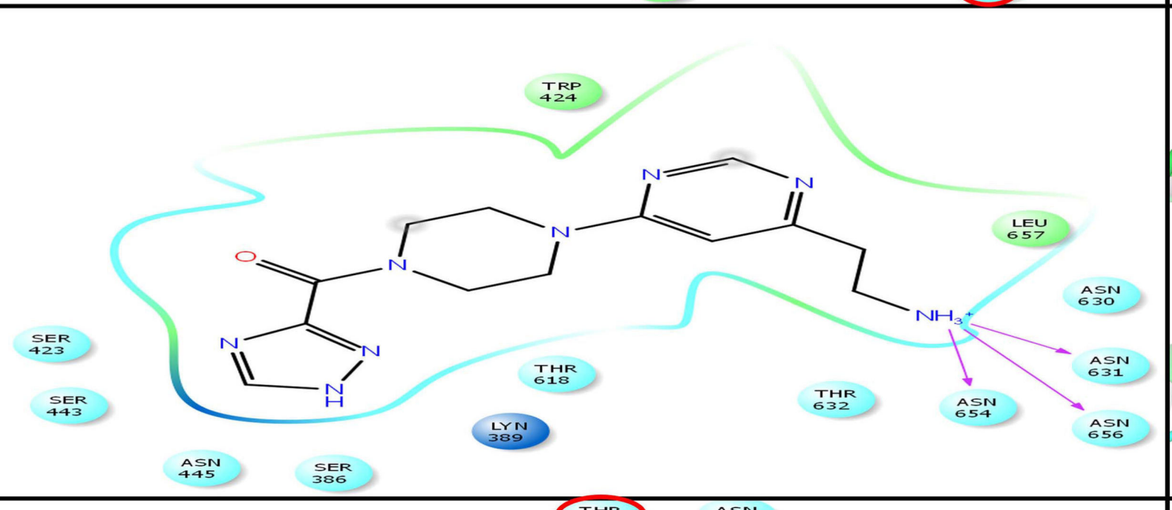


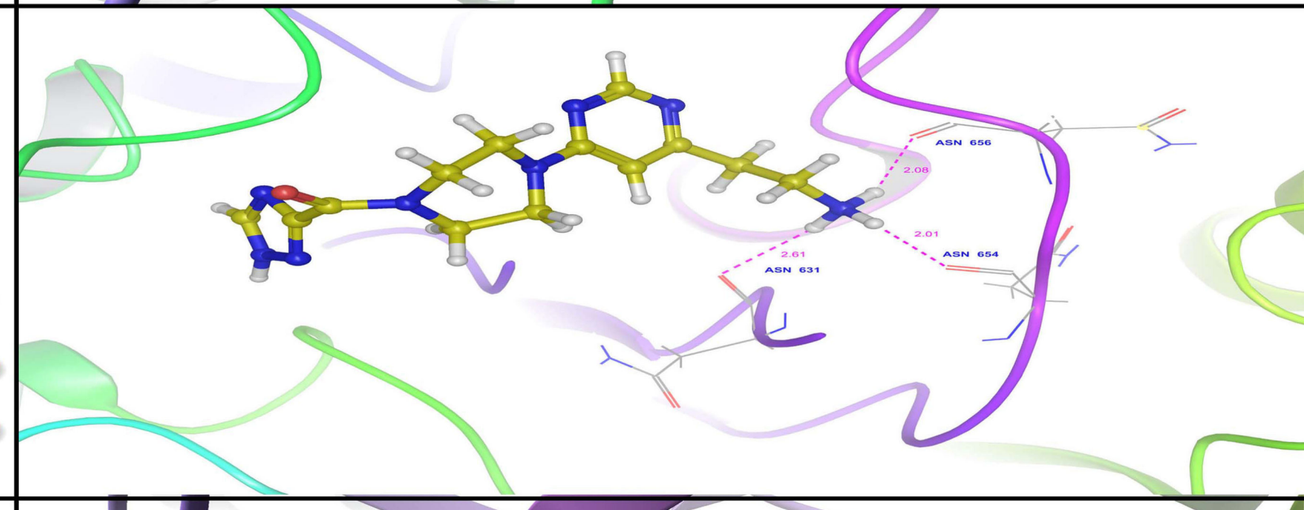
C



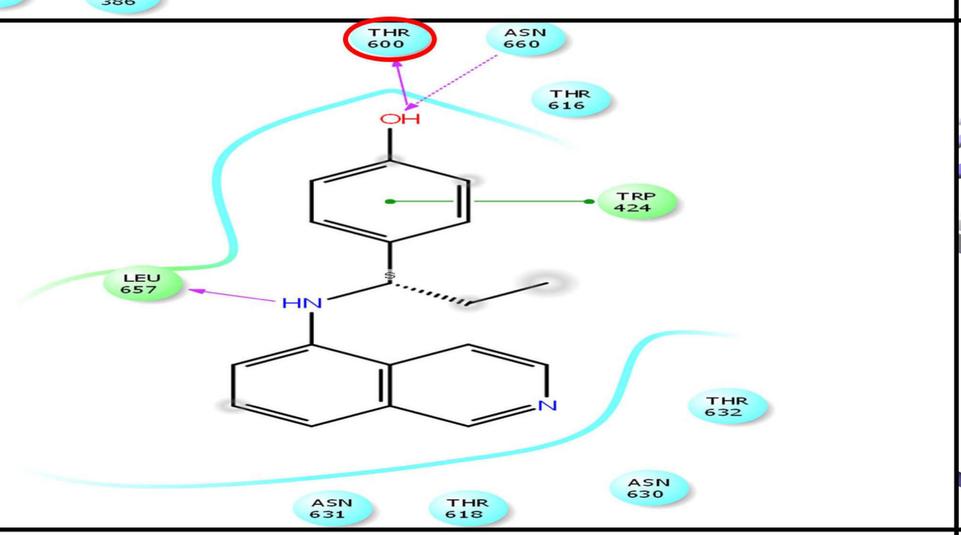


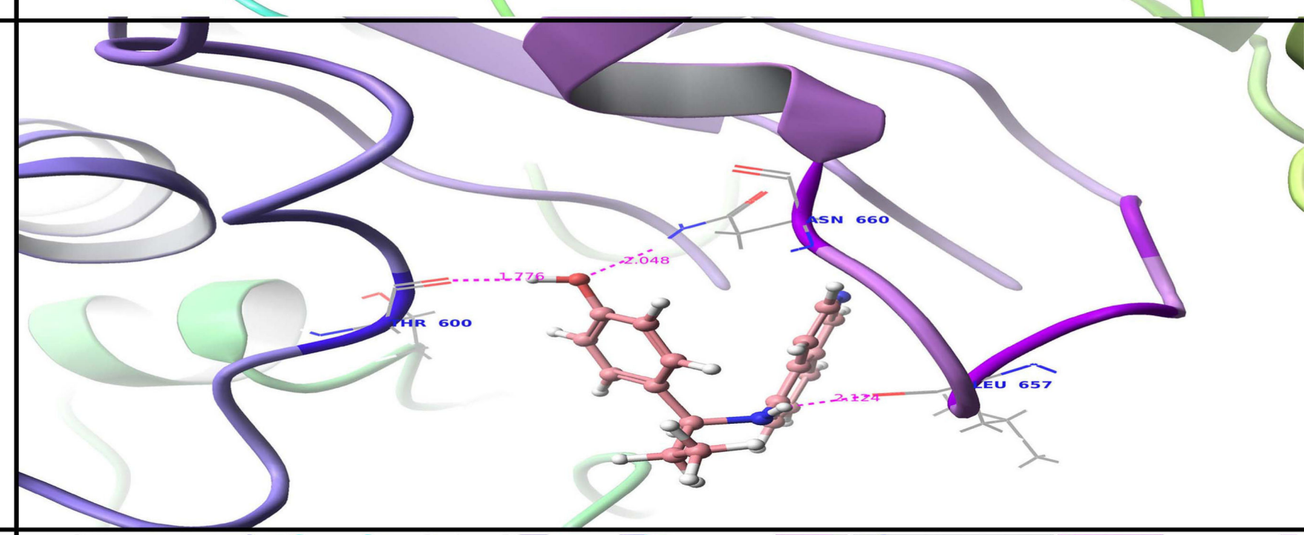
D



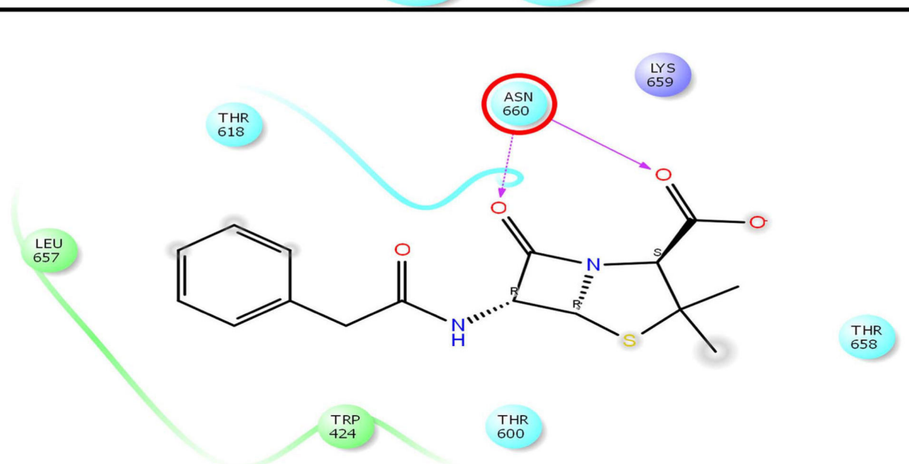


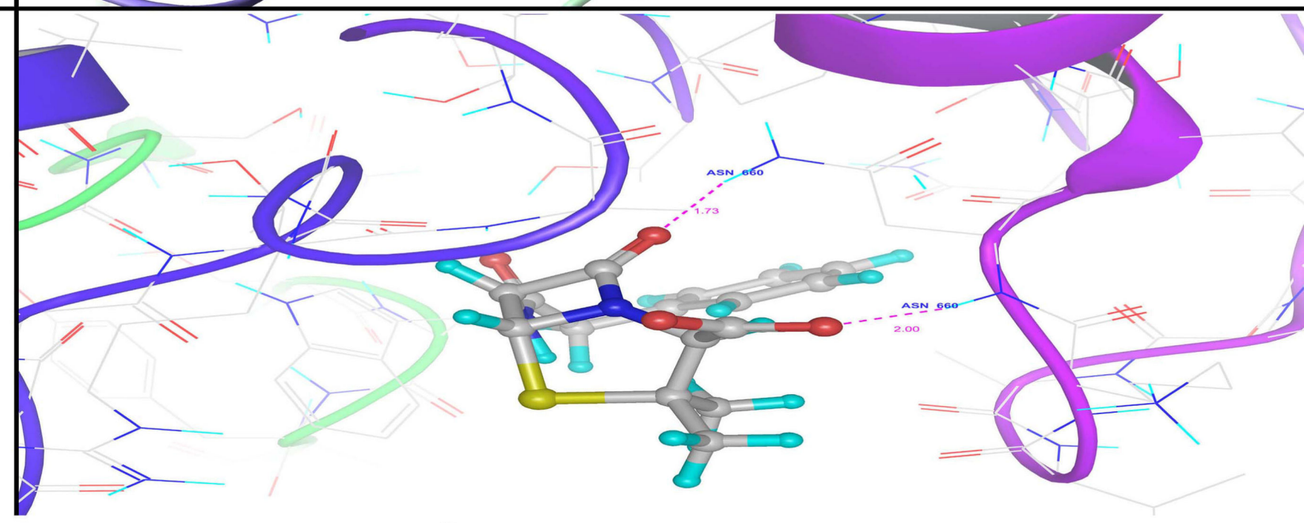
E





F





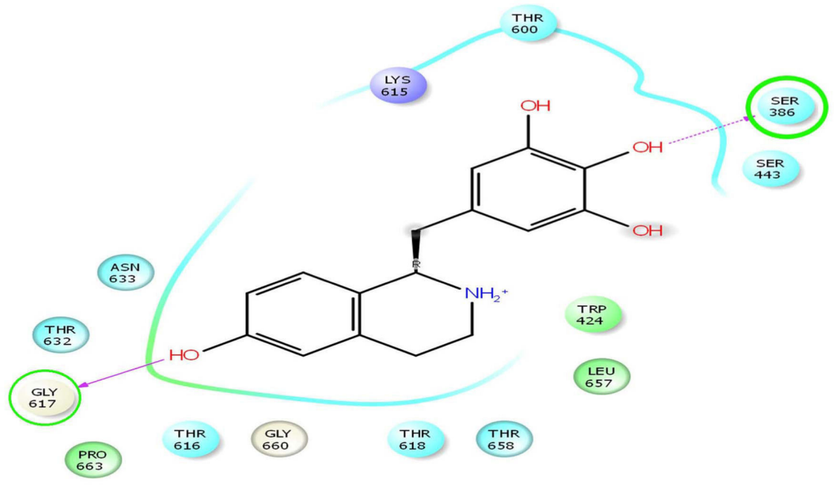
Charged (negative)
Charged (positive)
Polar
Hydrophobic
Glycine

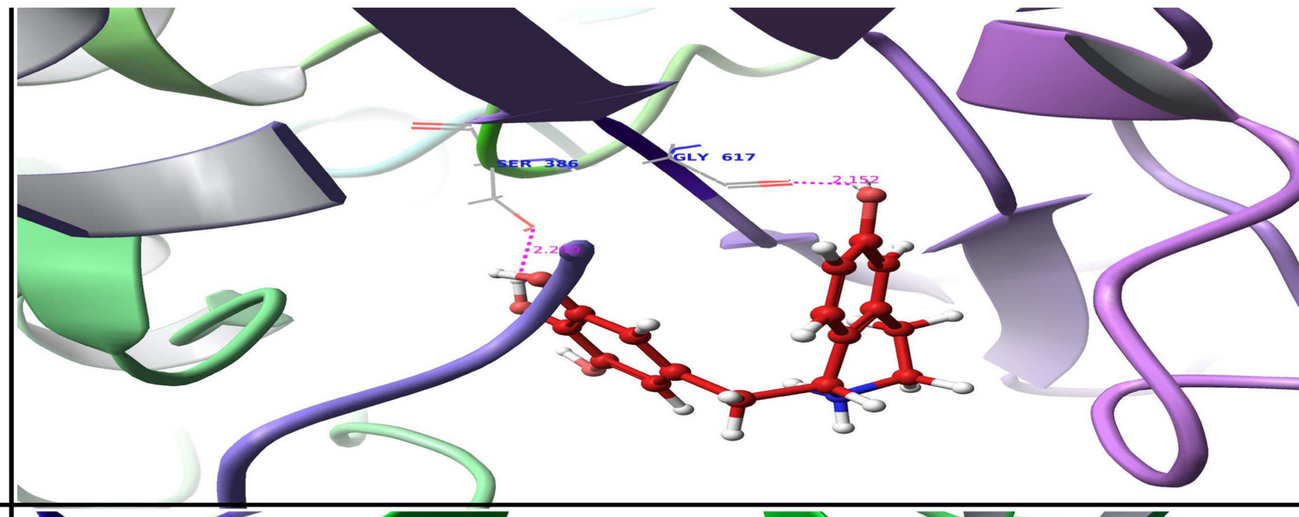
Metal
Water
Hydration site
Displaced hydration site
π–π stacking

π–cation
H–bond (backbone)
H–bond (side chain)
Metal coordination
Solvent exposure

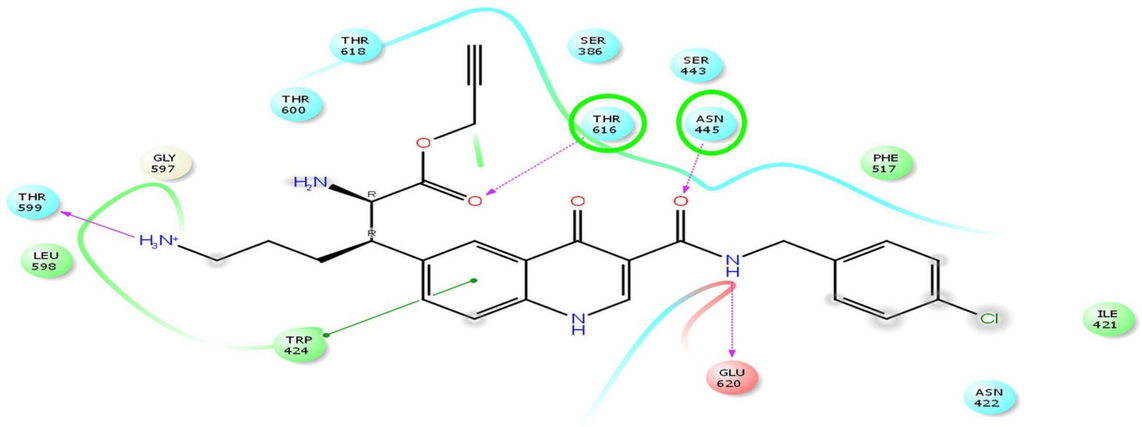
Figure. 2

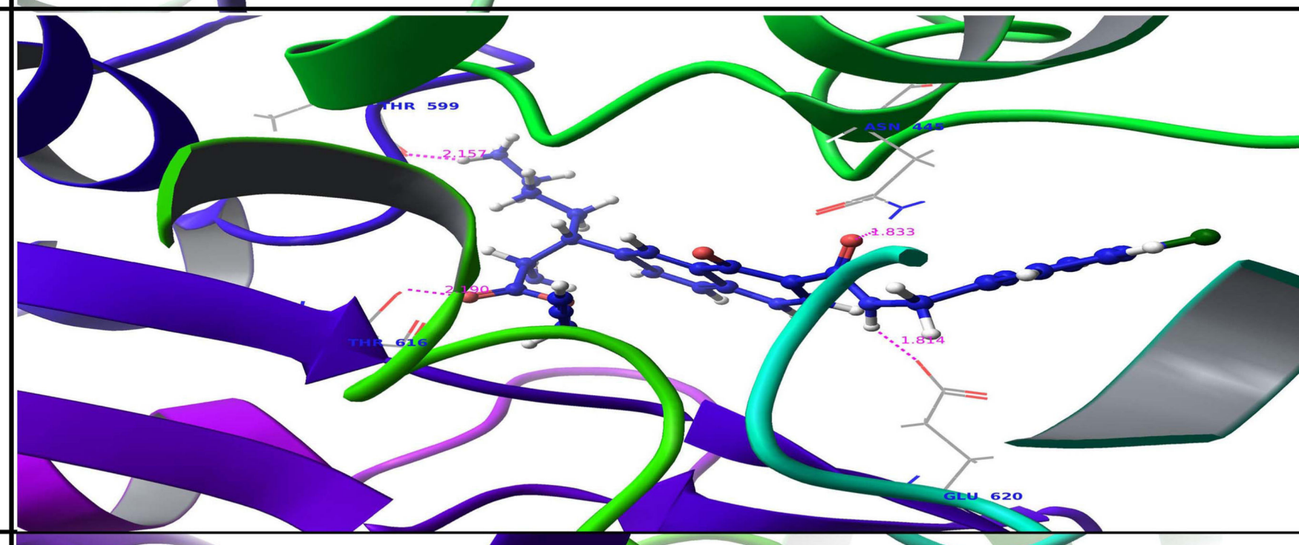
A



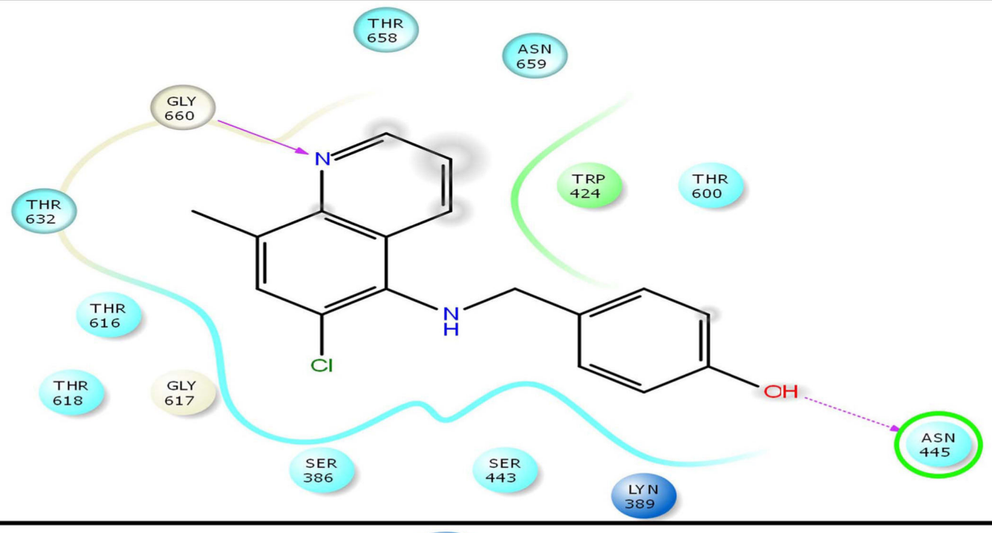


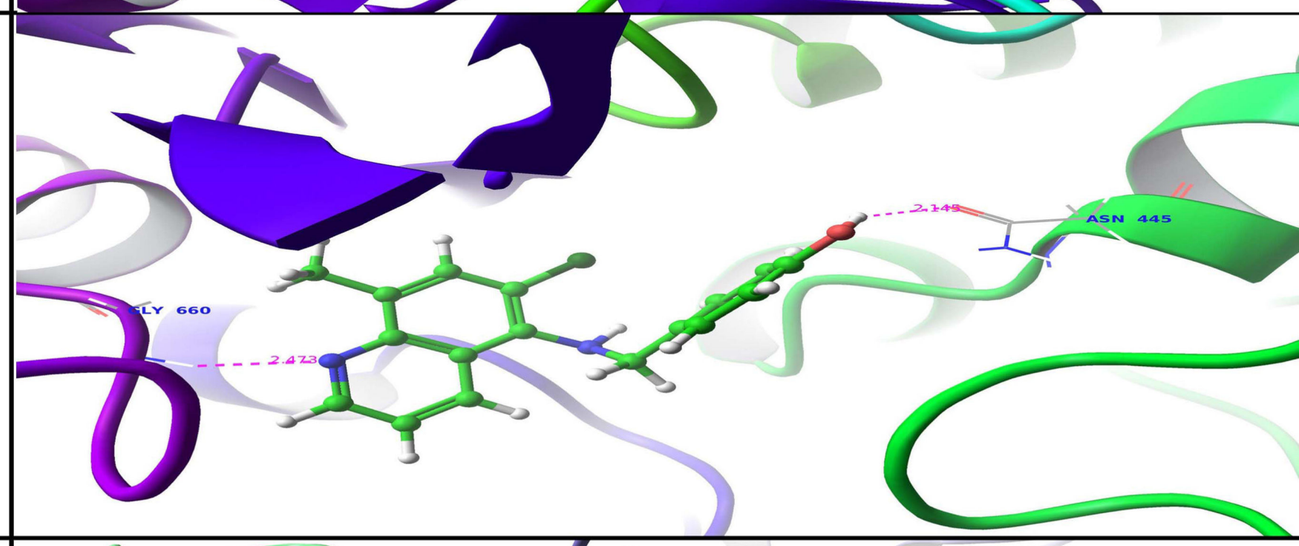
B



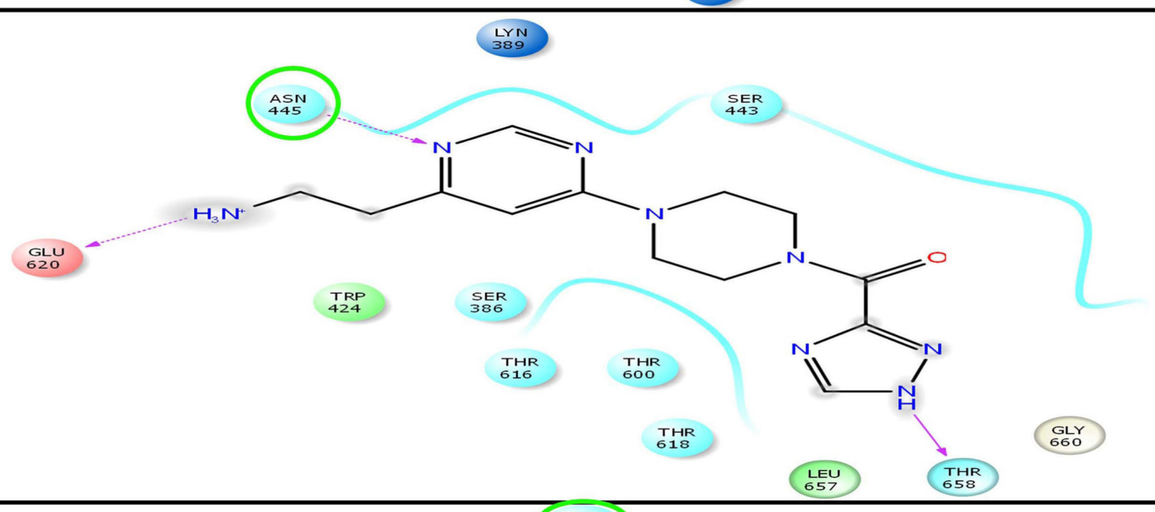


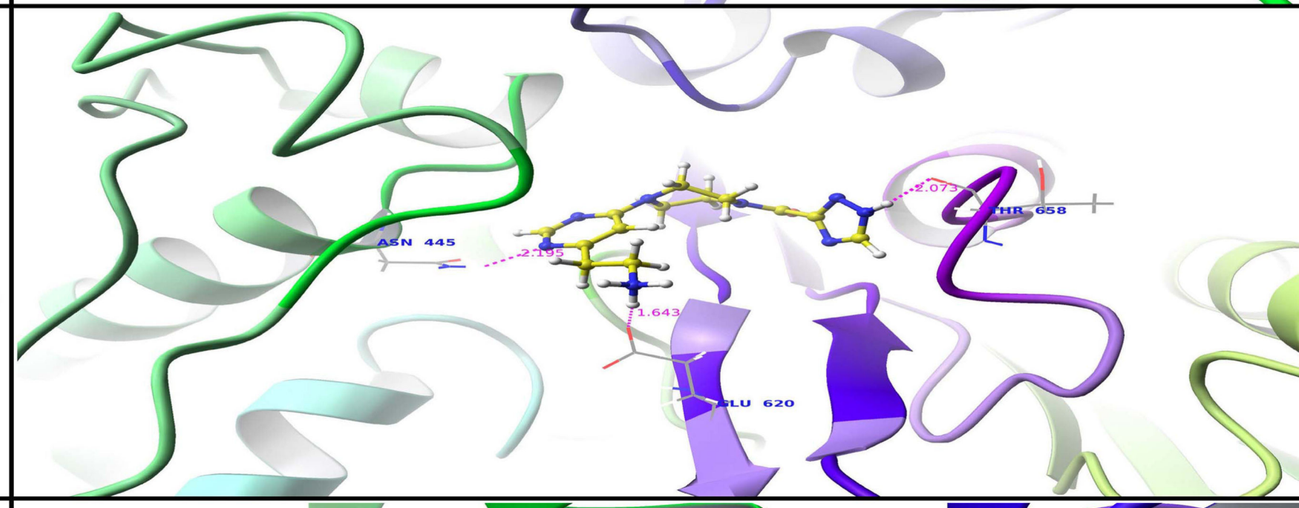
C



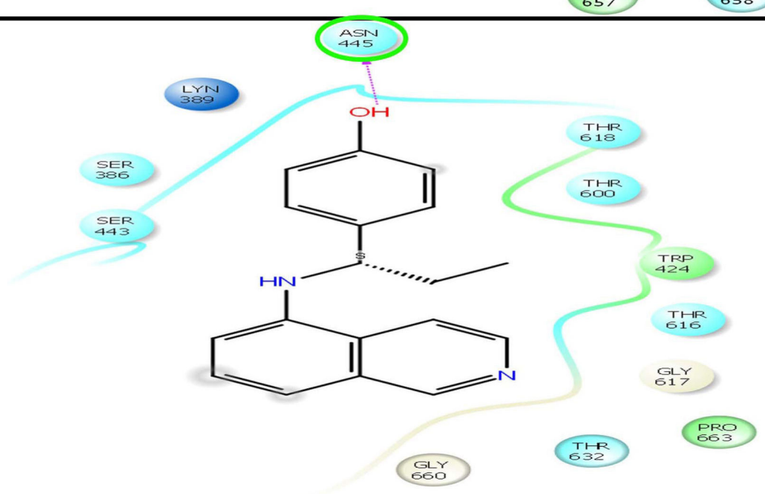


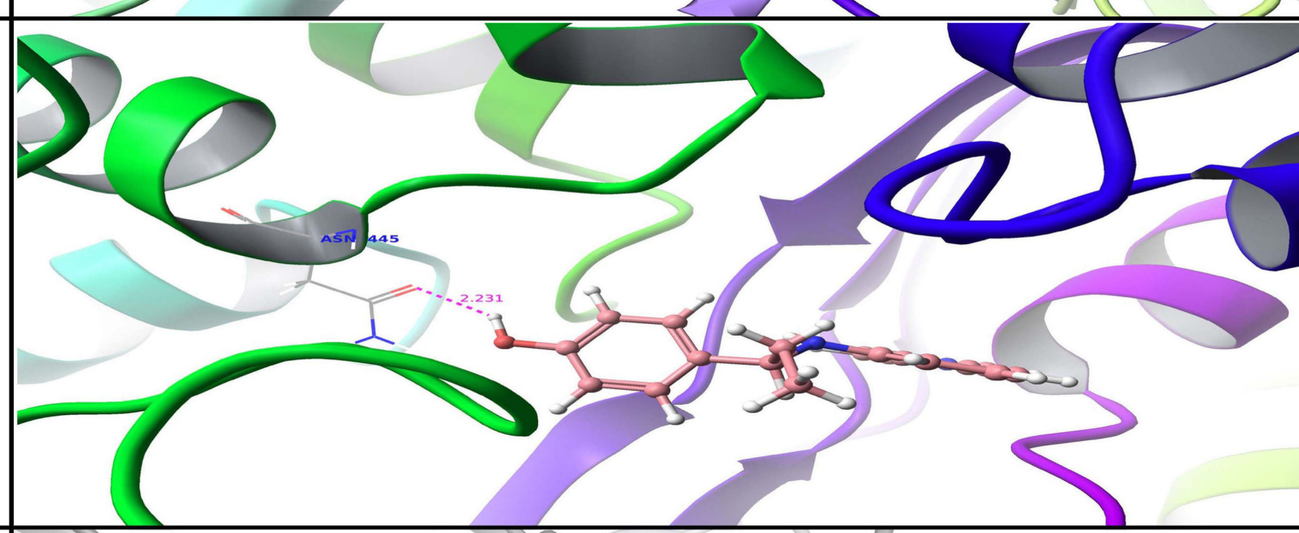
D





E





F

Charged (negative)
Charged (positive)
Polar
Hydrophobic
Glycine

Metal
Water
Hydration site
Displaced hydration site
π–π stacking

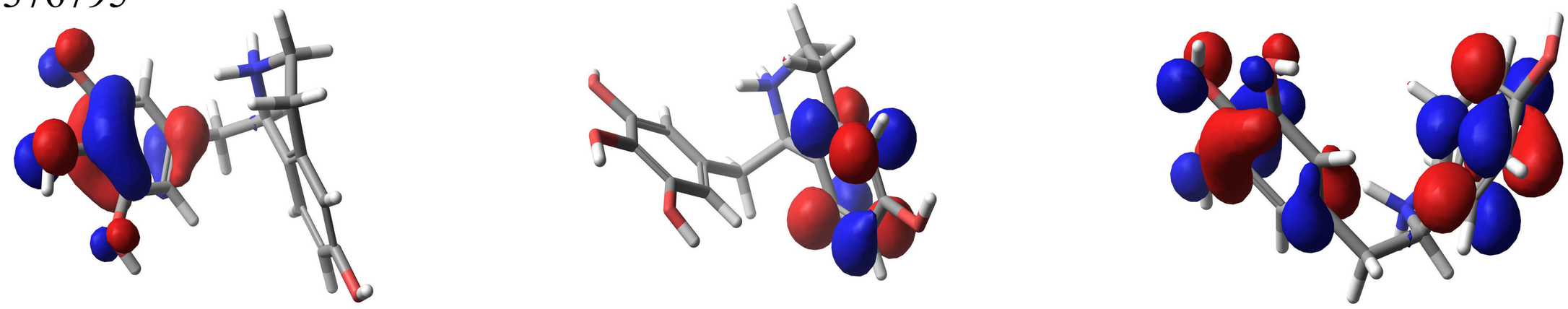
π–cation
H–bond (backbone)
H–bond (side chain)
Metal coordination
Solvent exposure

Figure. 3

HOMO → LUMO → HOMO + LUMO

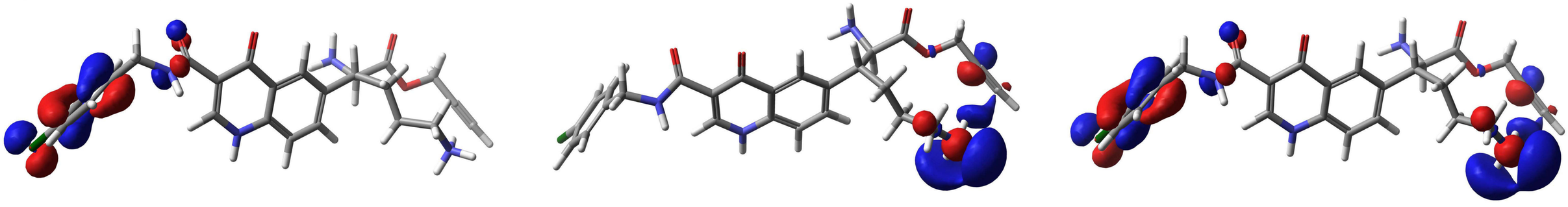
ZINC59376795

A)



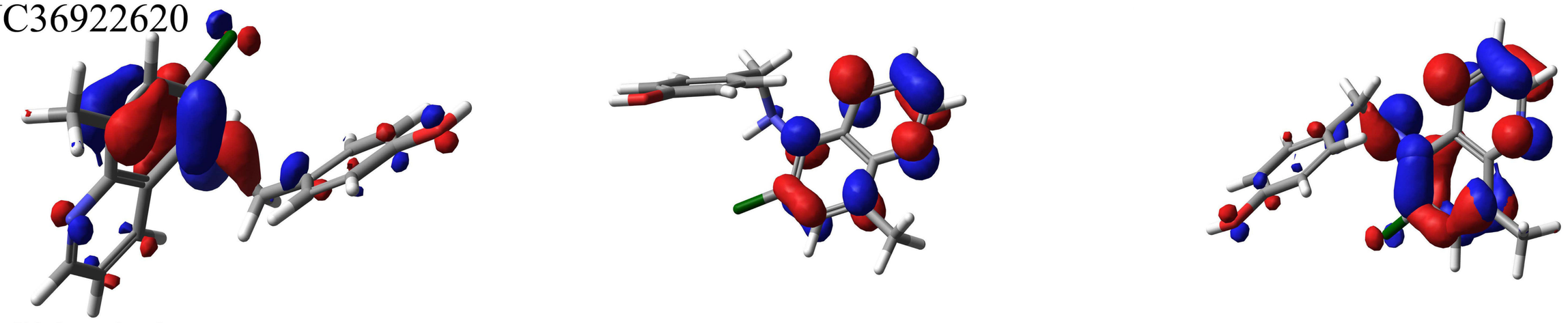
ZINC60175365

B)



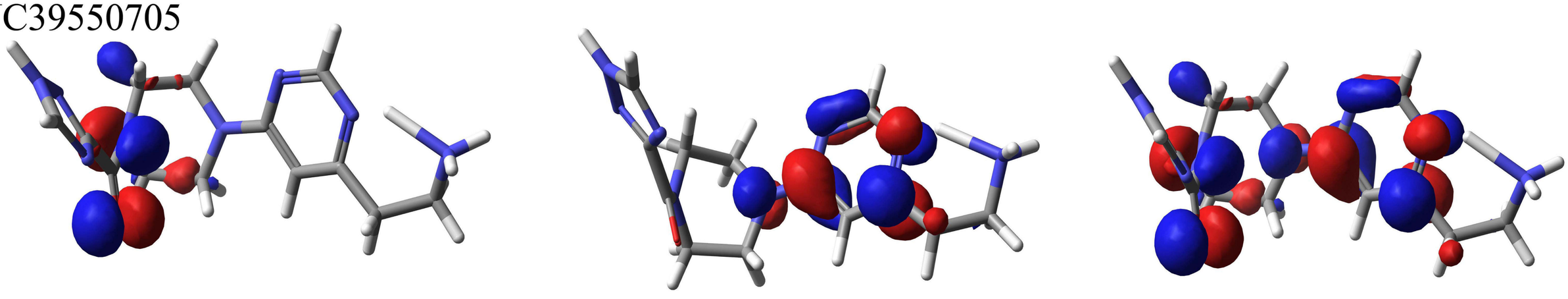
ZINC36922620

C)



ZINC39550705

D)



ZINC36953975

E)

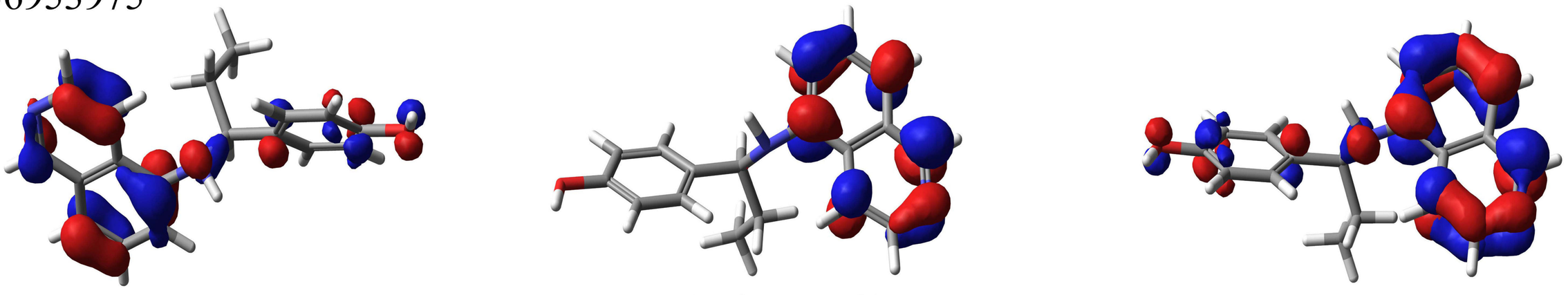
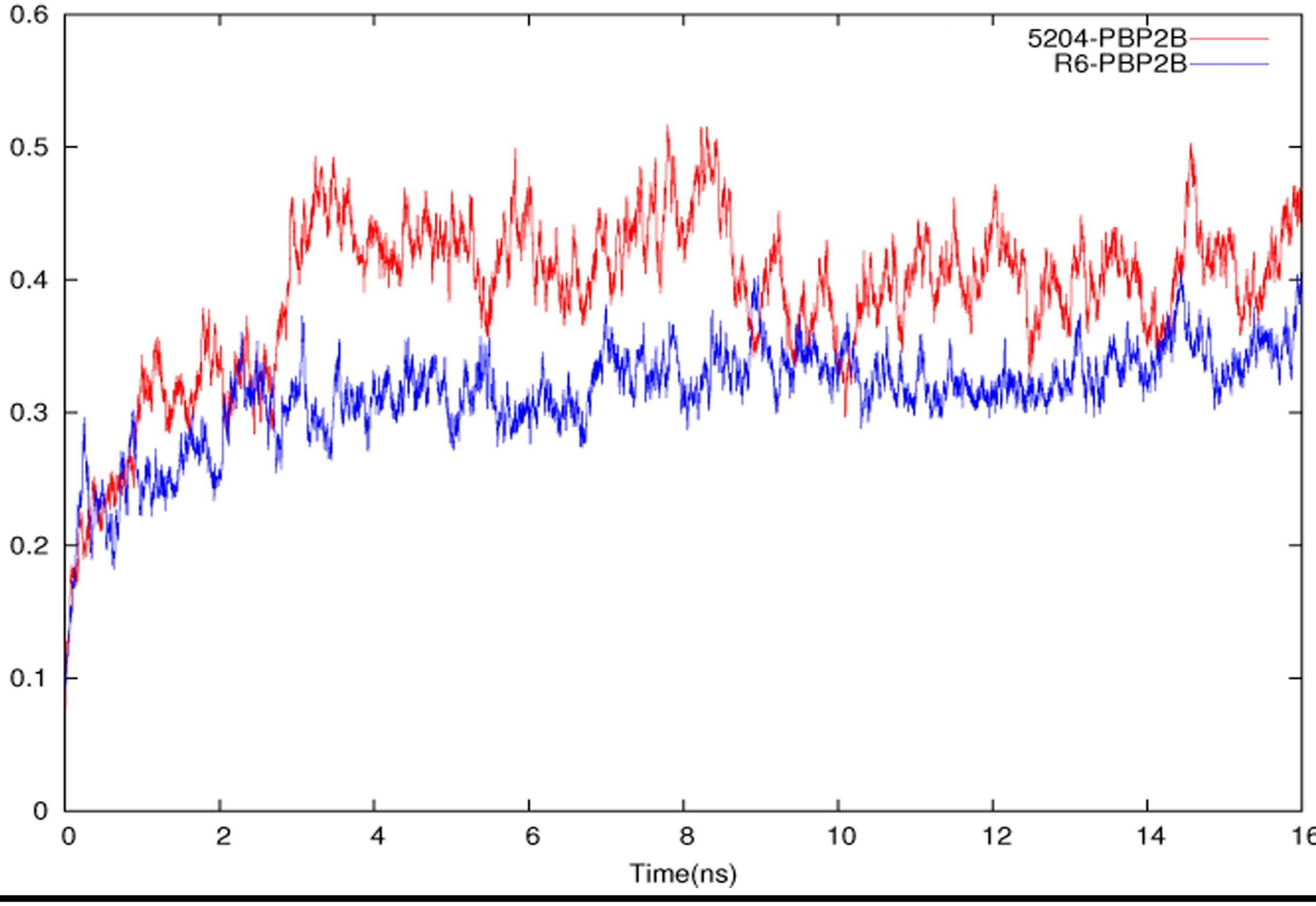


Figure. 4

A


5204-PBP2B
R6-PBP2B
RMSD(nm)
0.6
0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
12
14
16
Time(ns)

B

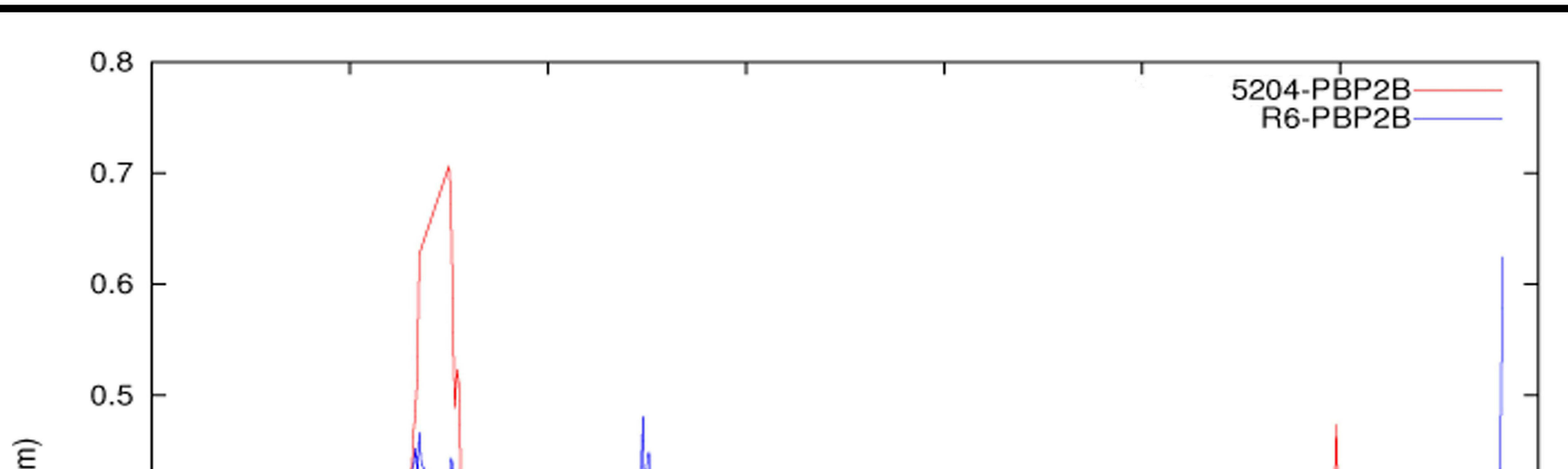
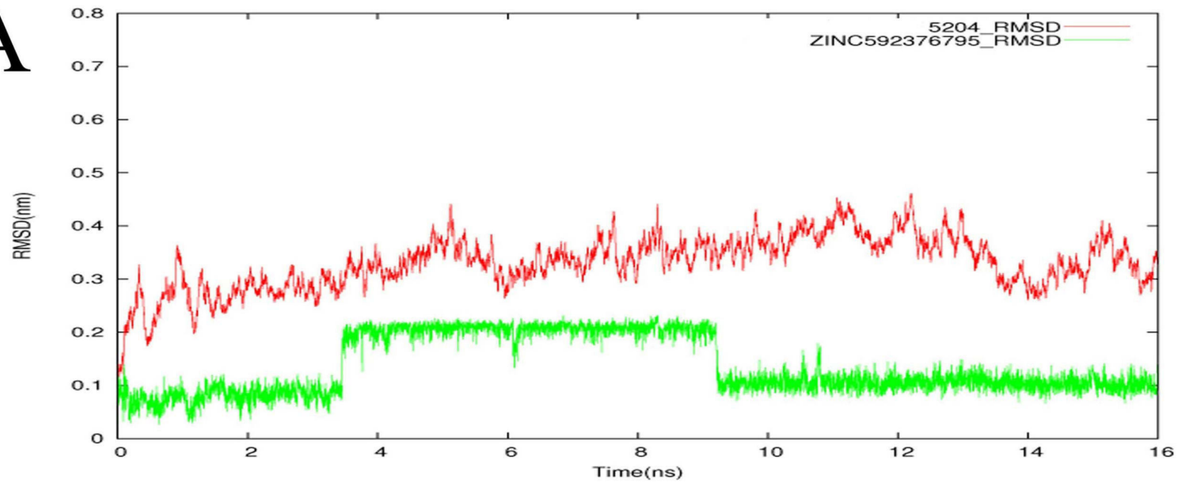
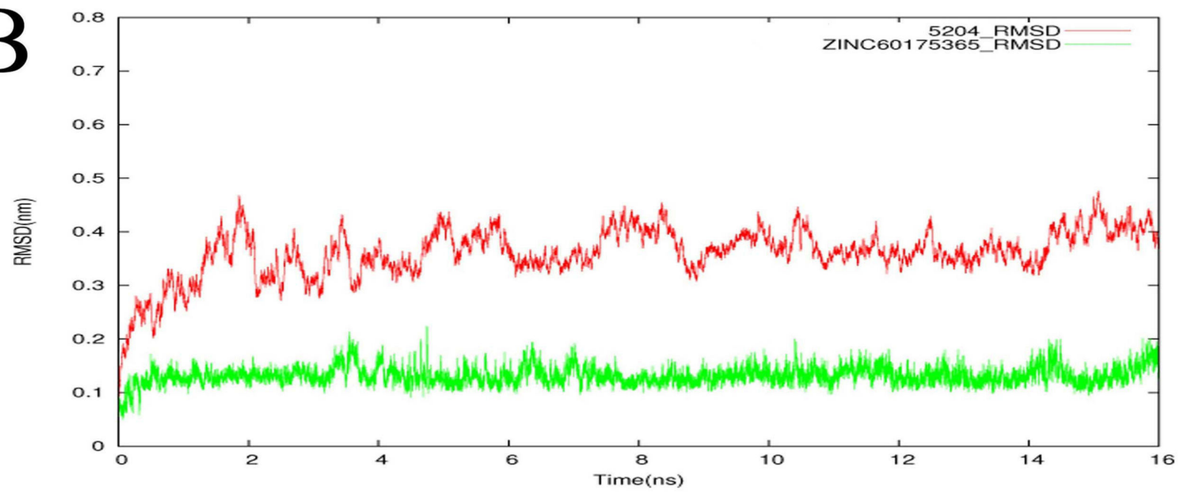

5204-PBP2B
R6-PBP2B
RMSF(nm)
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
100
200
300
400
500
600
700
Residues

Figure. 5

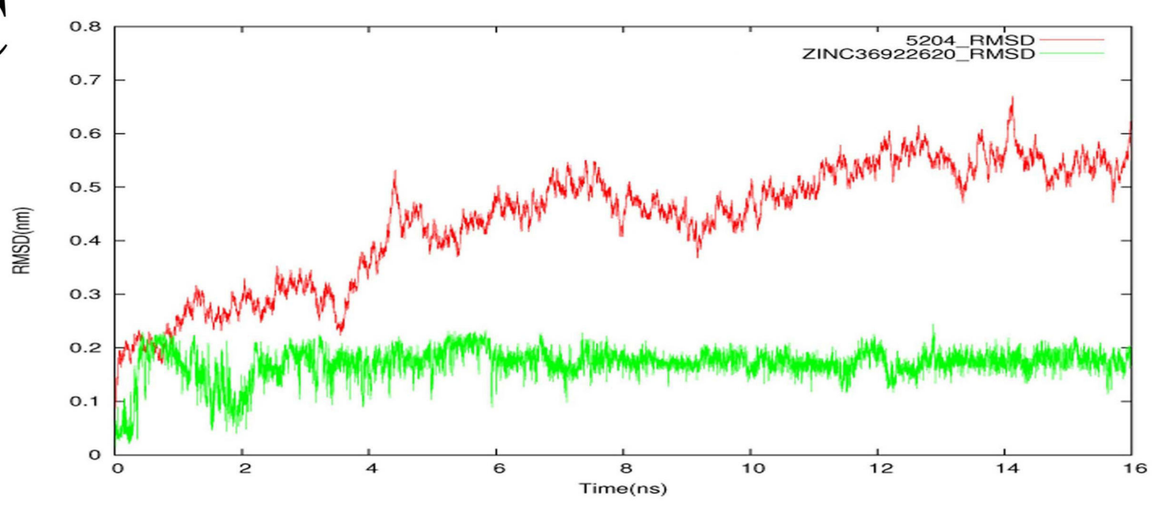
A



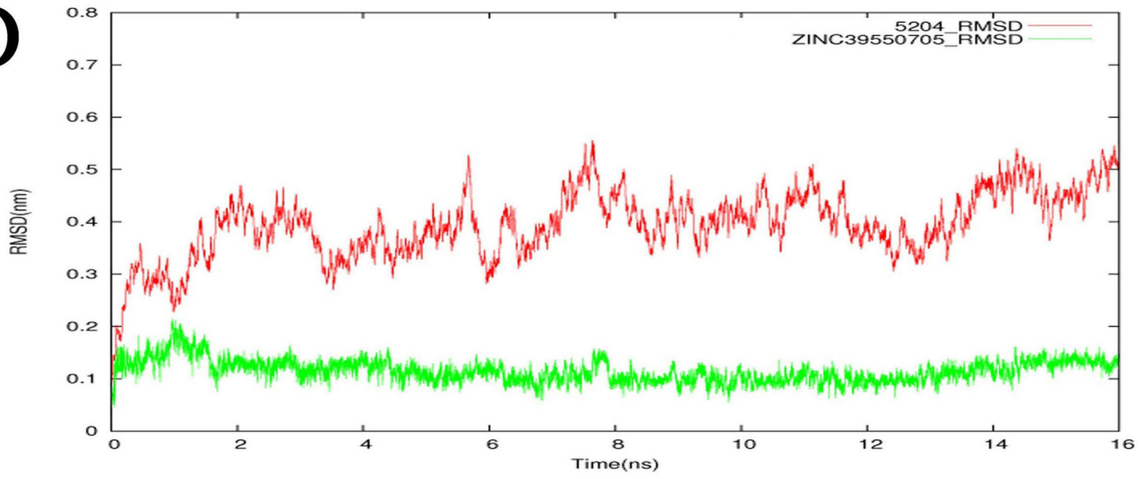
B



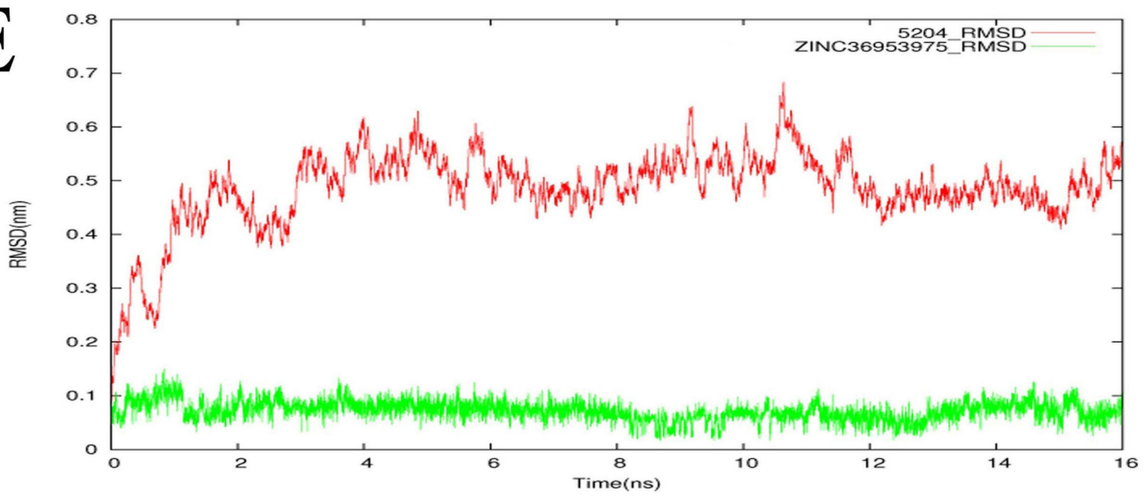
C



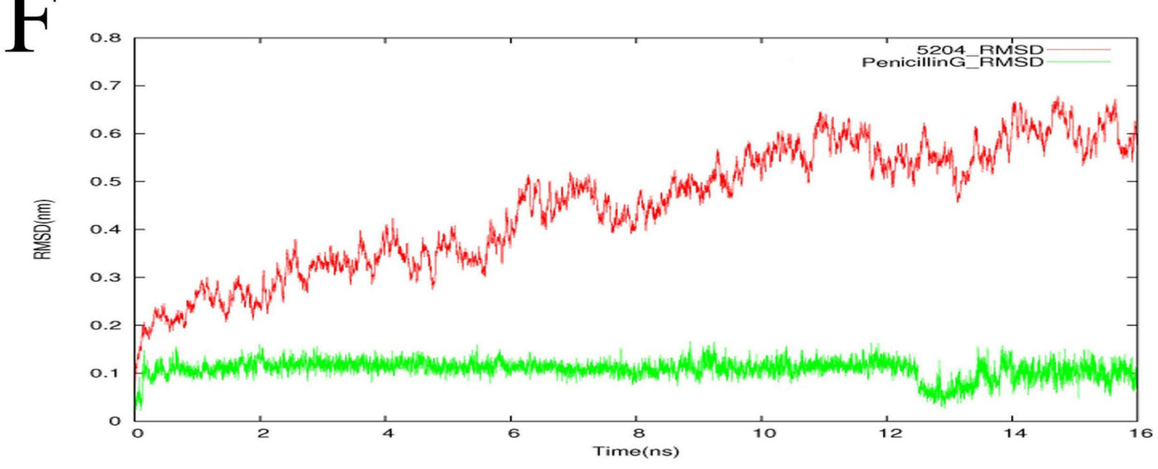
D



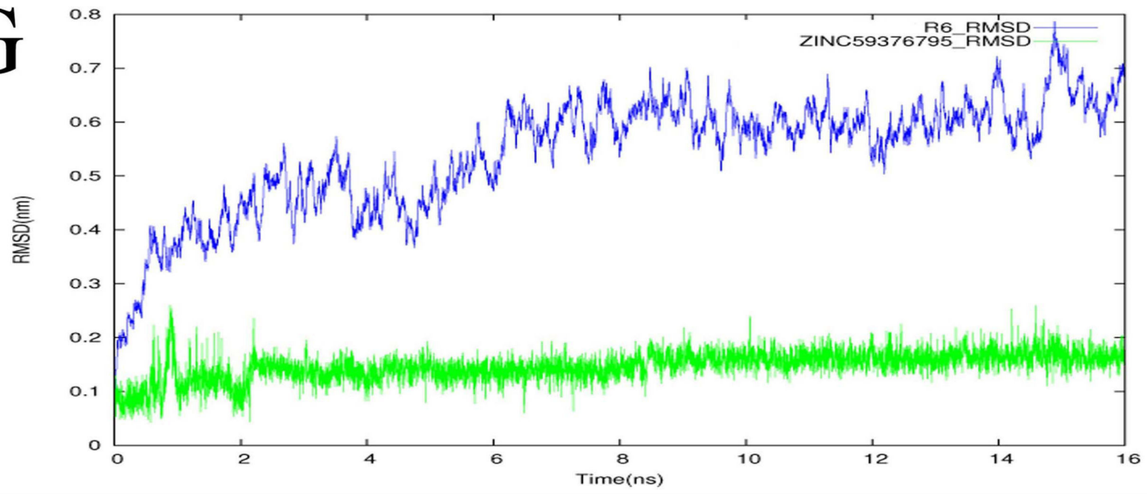
E



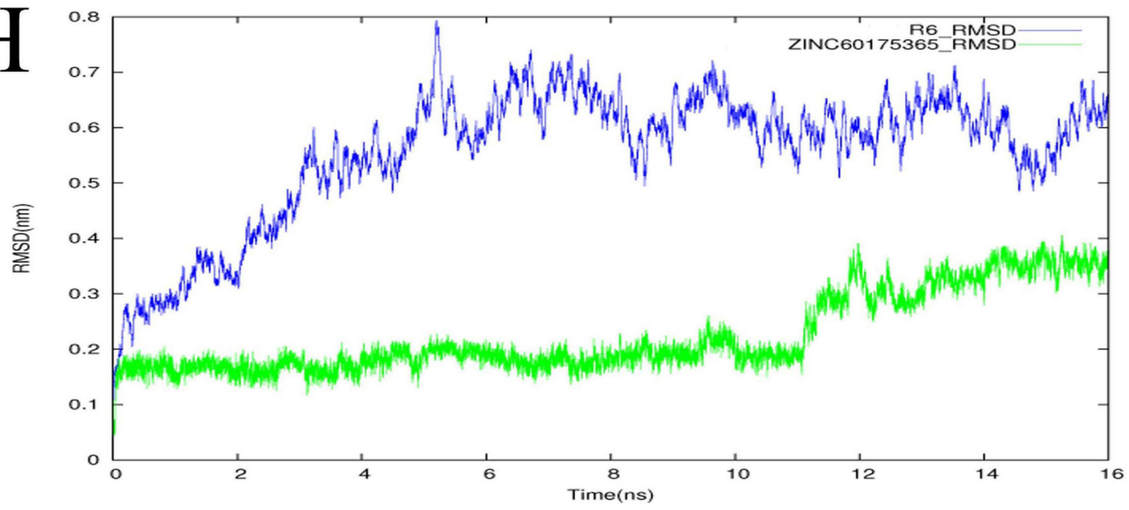
F



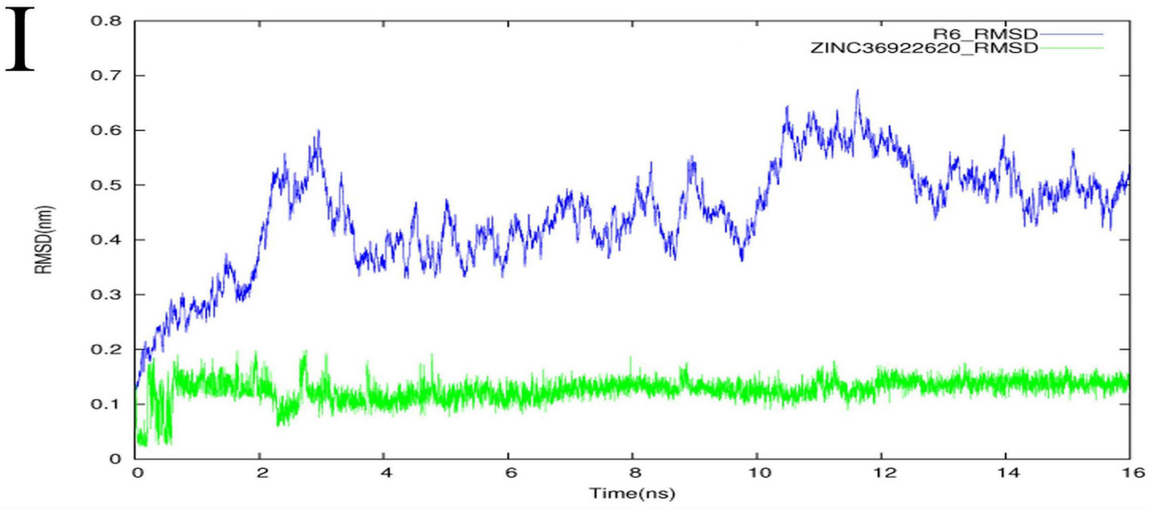
G



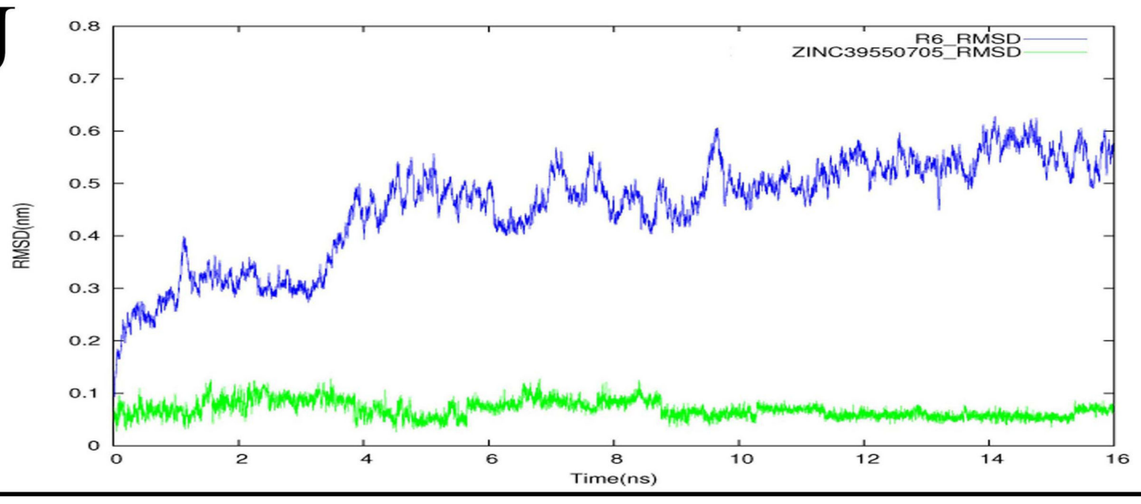
H



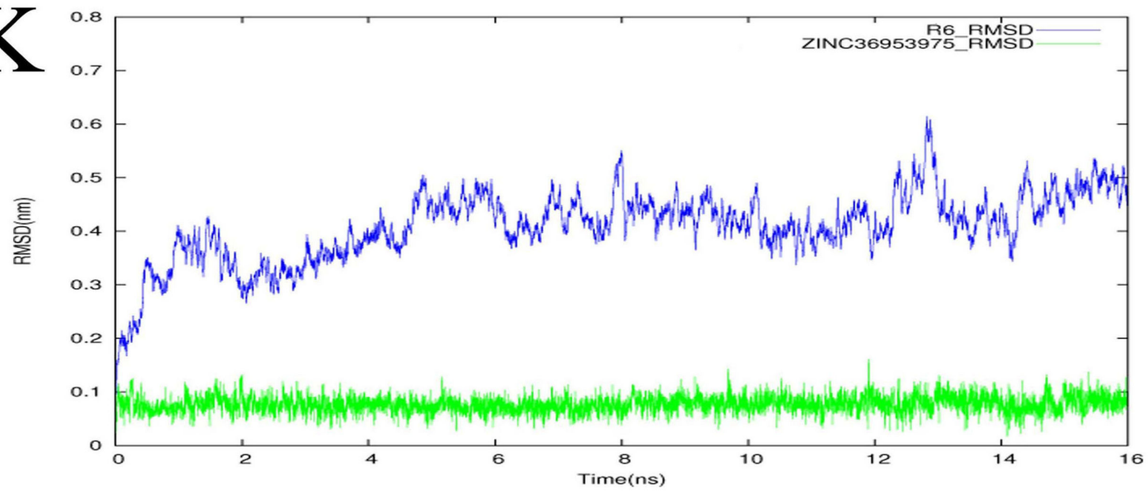
I



J



K



L

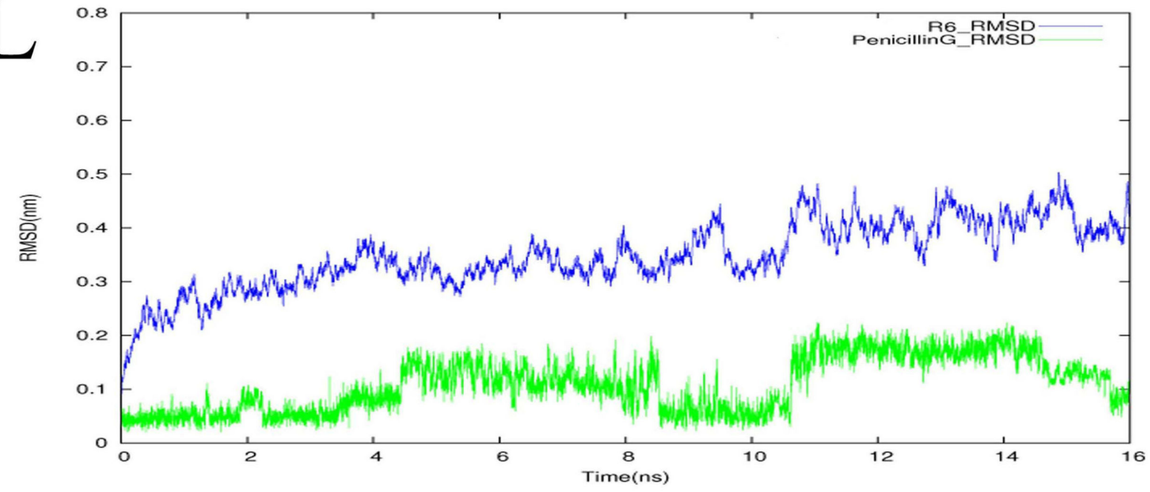
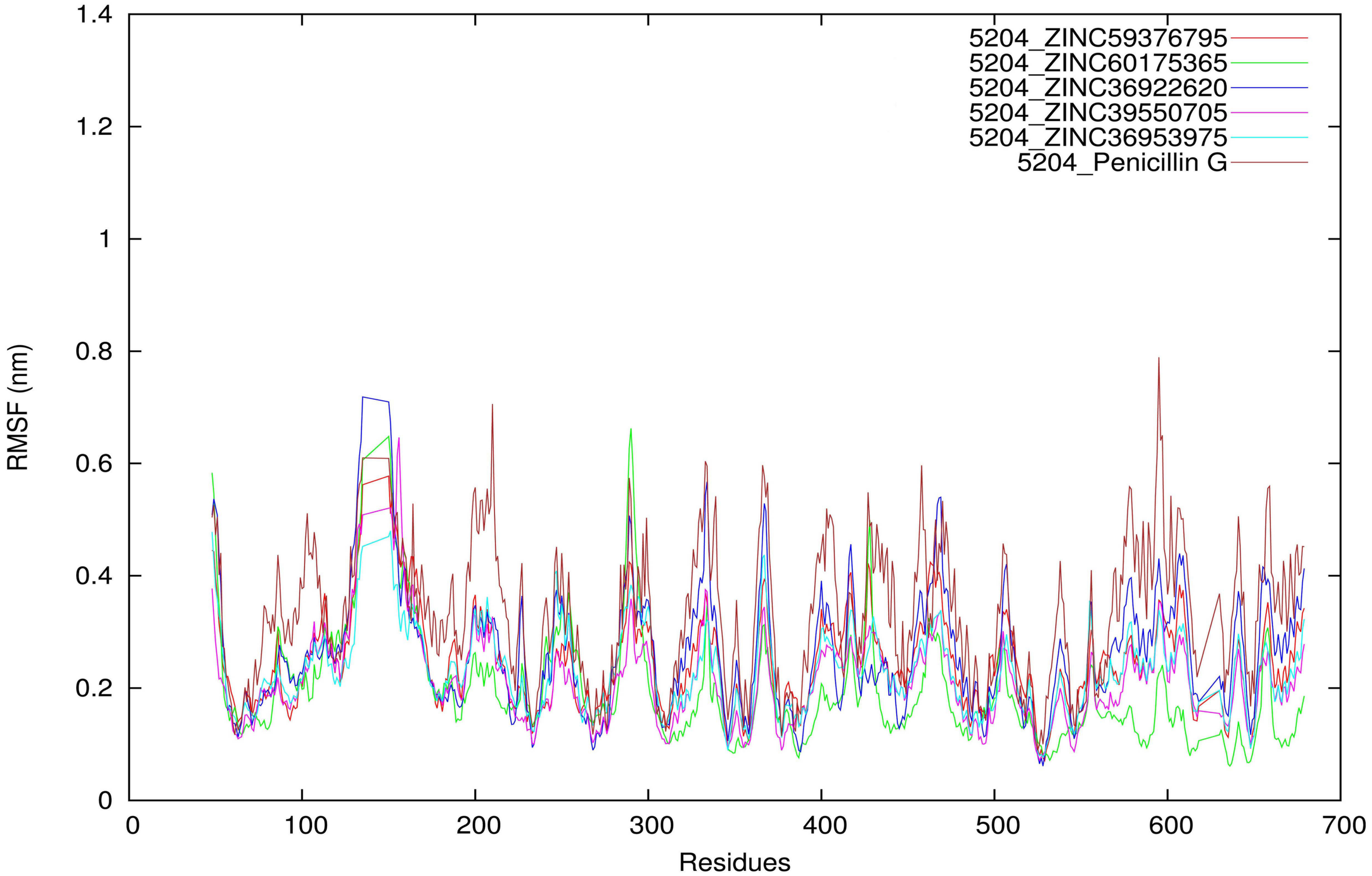


Figure. 6

A


5204_ZINC59376795
5204_ZINC60175365
5204_ZINC36922620
5204_ZINC39550705
5204_ZINC36953975
5204_Penicillin G
RMSF (nm)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
0
100
200
300
400
500
600
700
Residues

B

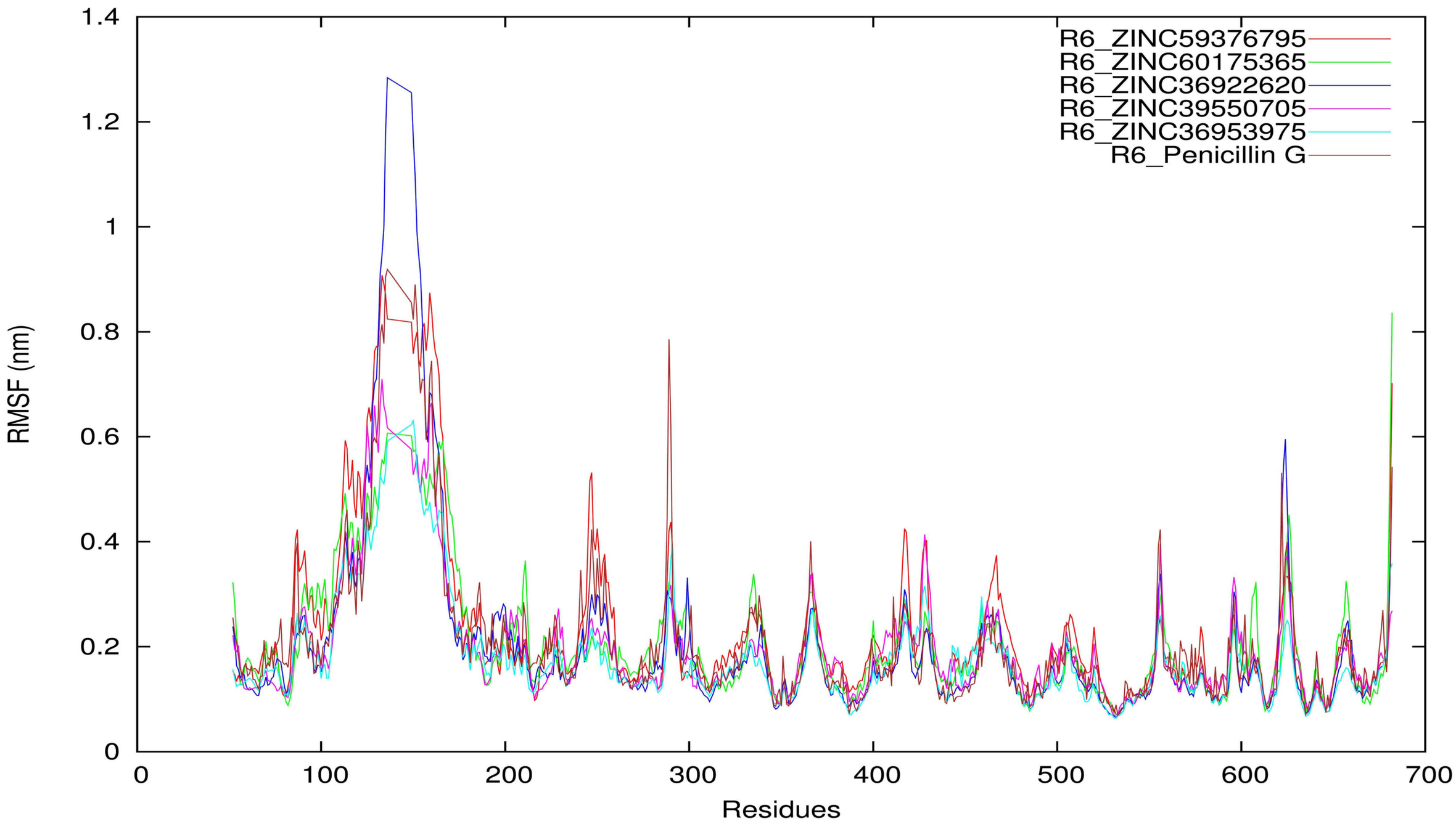
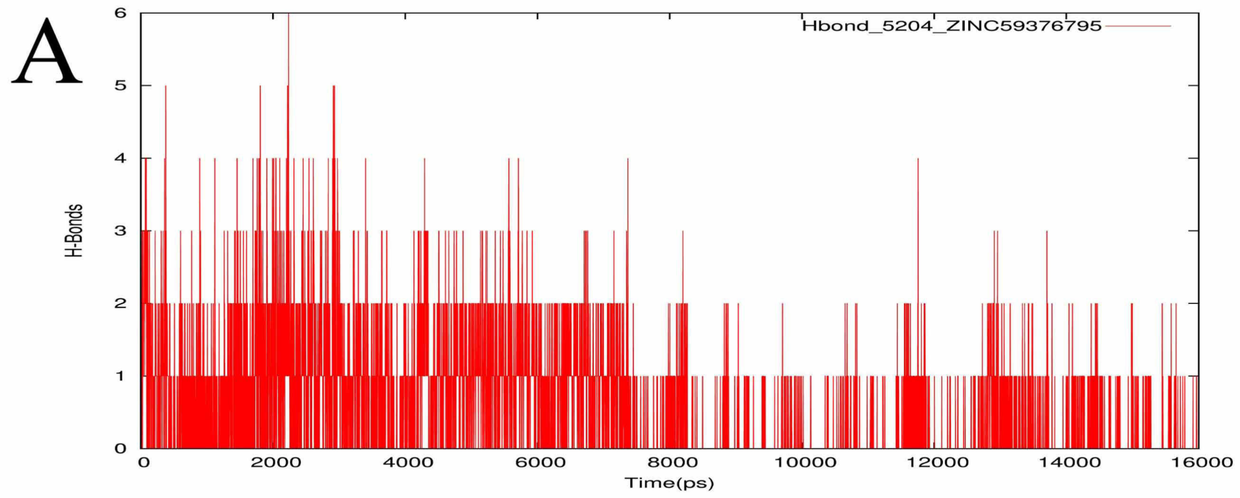
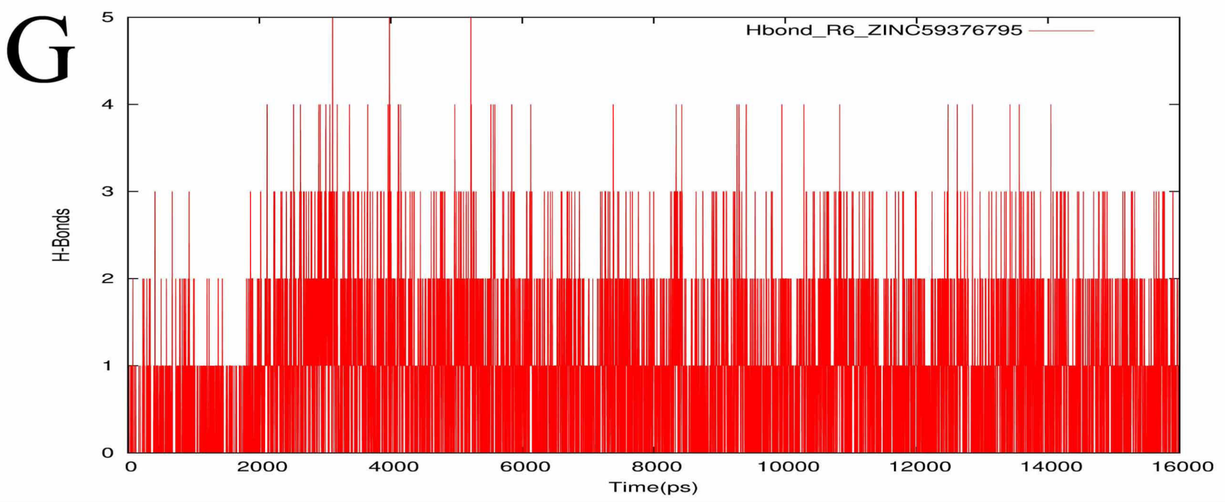
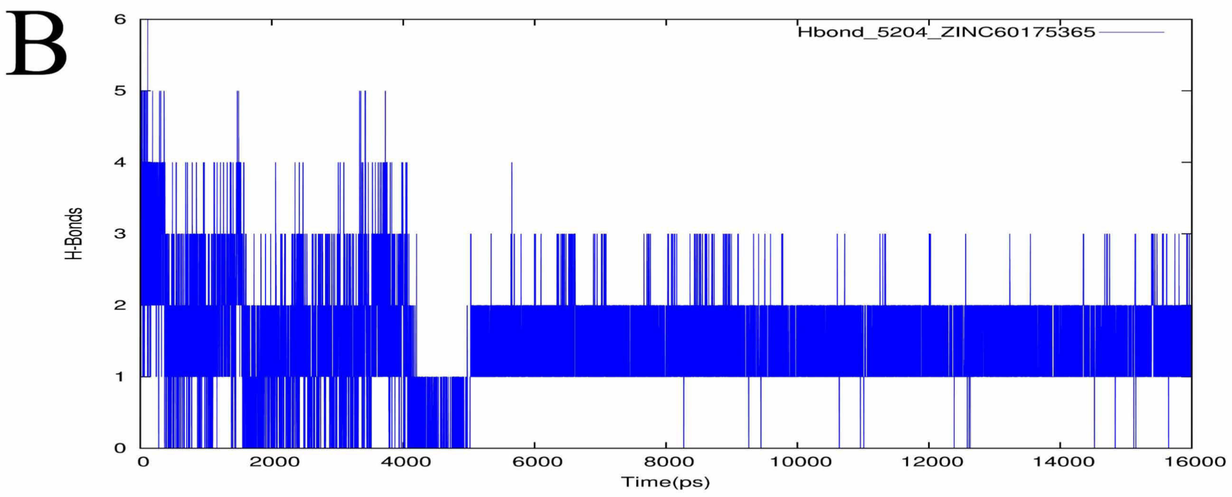
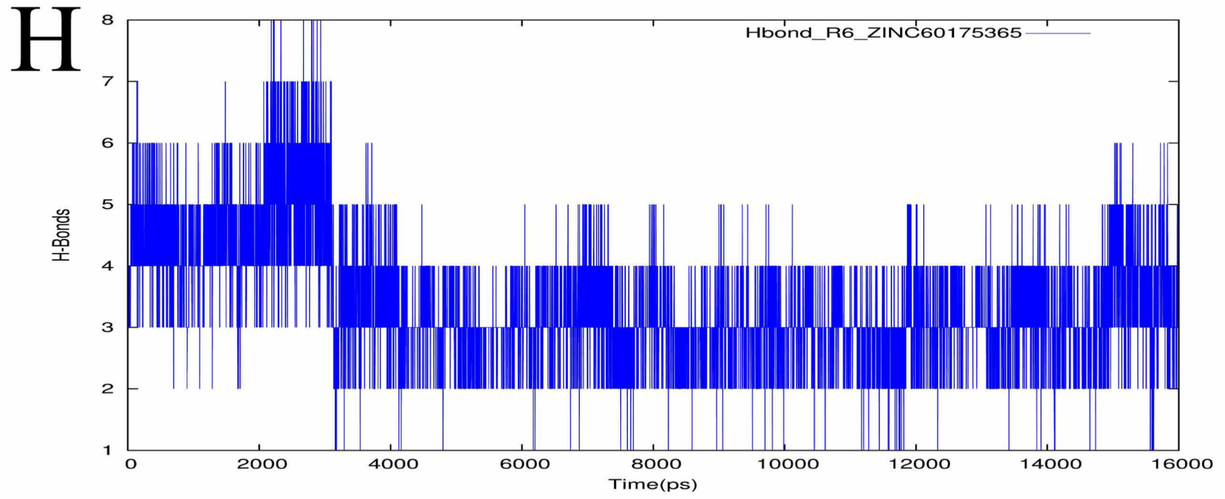

R6_ZINC59376795
R6_ZINC60175365
R6_ZINC36922620
R6_ZINC39550705
R6_ZINC36953975
R6_Penicillin G
RMSF (nm)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
0
100
200
300
400
500
600
700
Residues

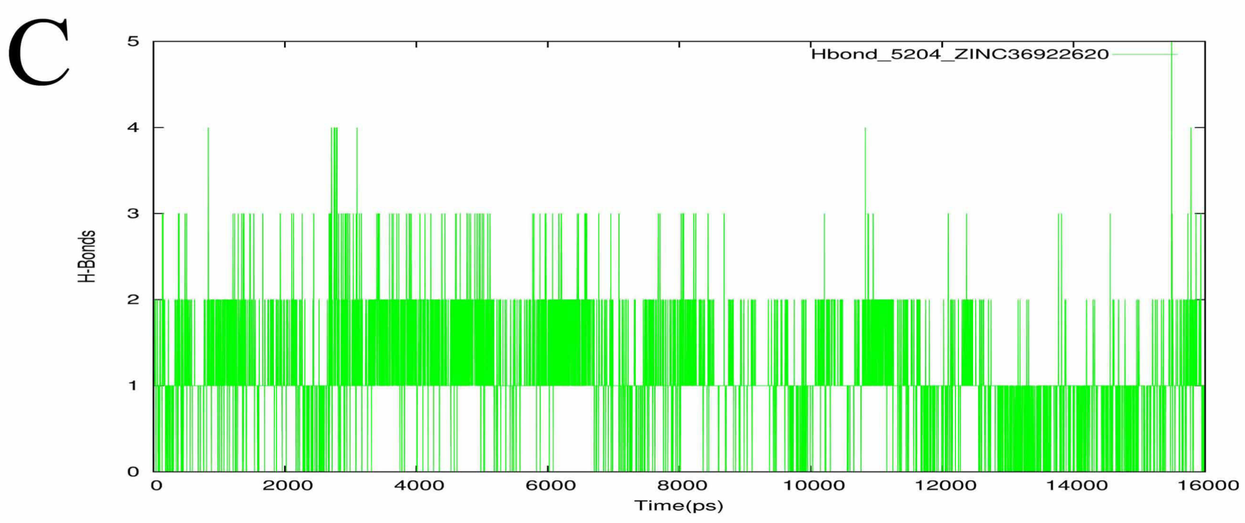
Figure. 7

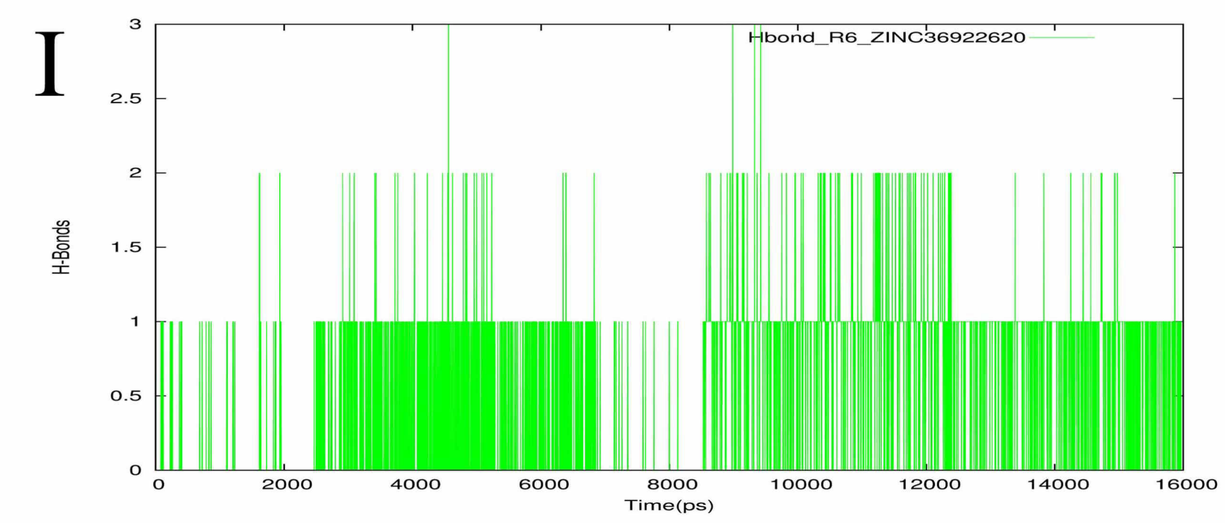

A
Hbond_5204_ZINC59376795
H-Bonds
Time(ps)

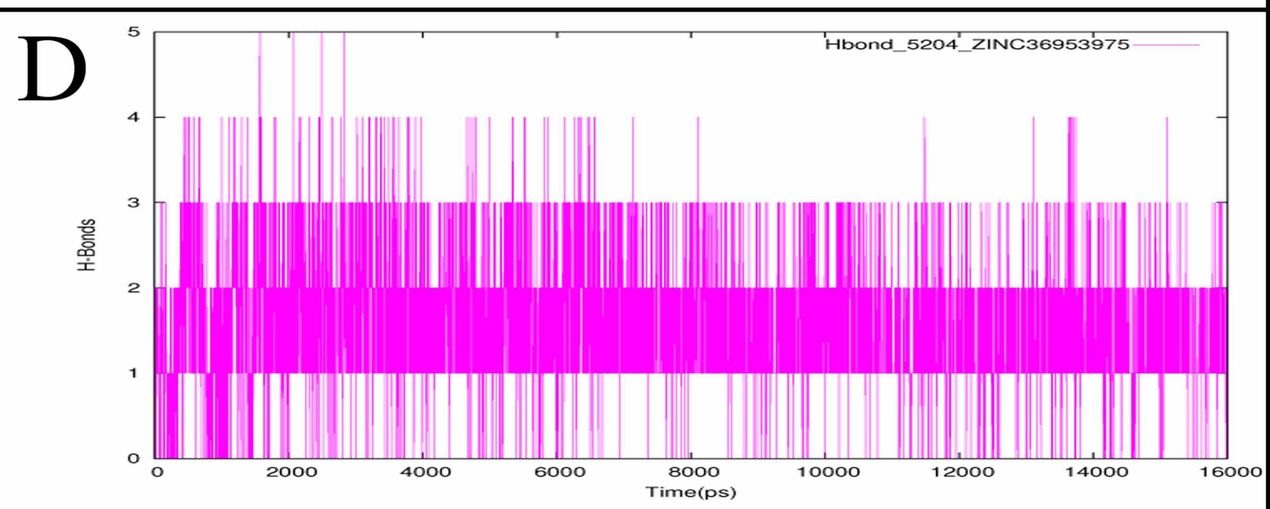

G
Hbond_R6_ZINC59376795
H-Bonds
Time(ps)

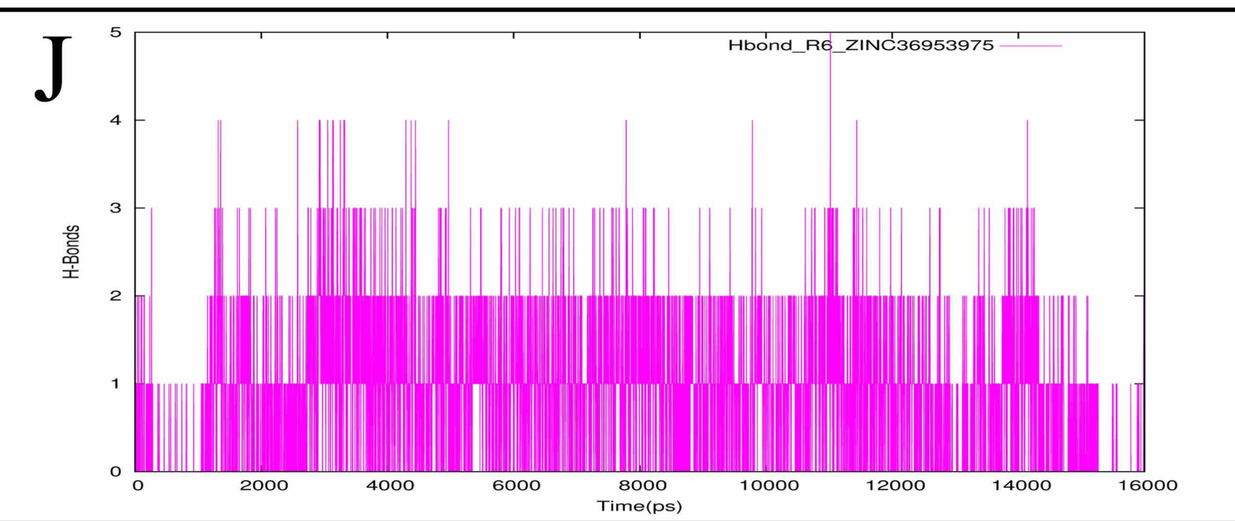

B
Hbond_5204_ZINC60175365
H-Bonds
Time(ps)

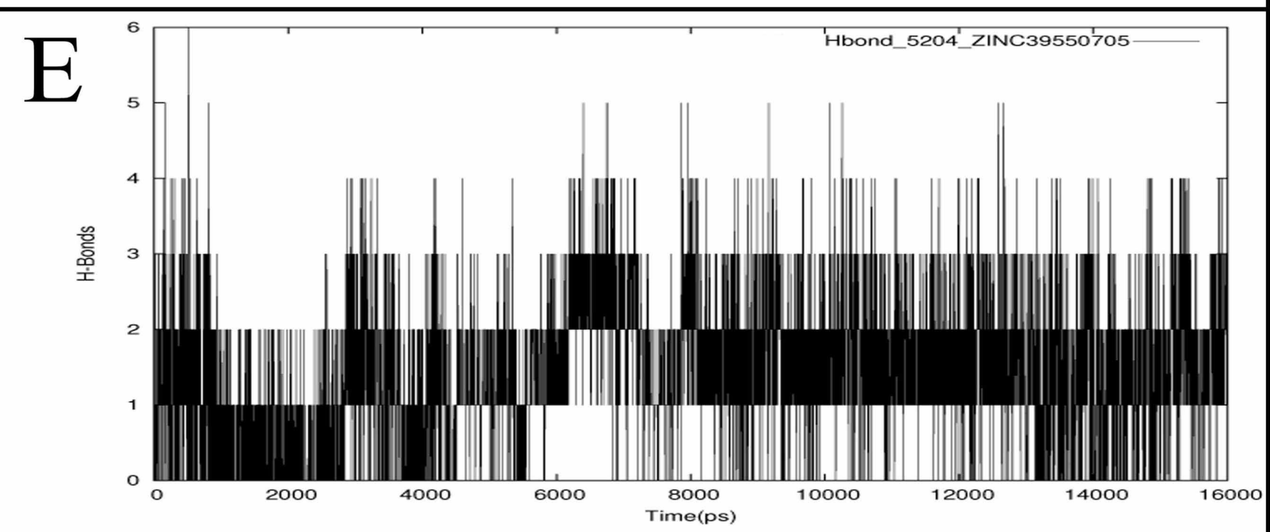

H
Hbond_R6_ZINC60175365
H-Bonds
Time(ps)

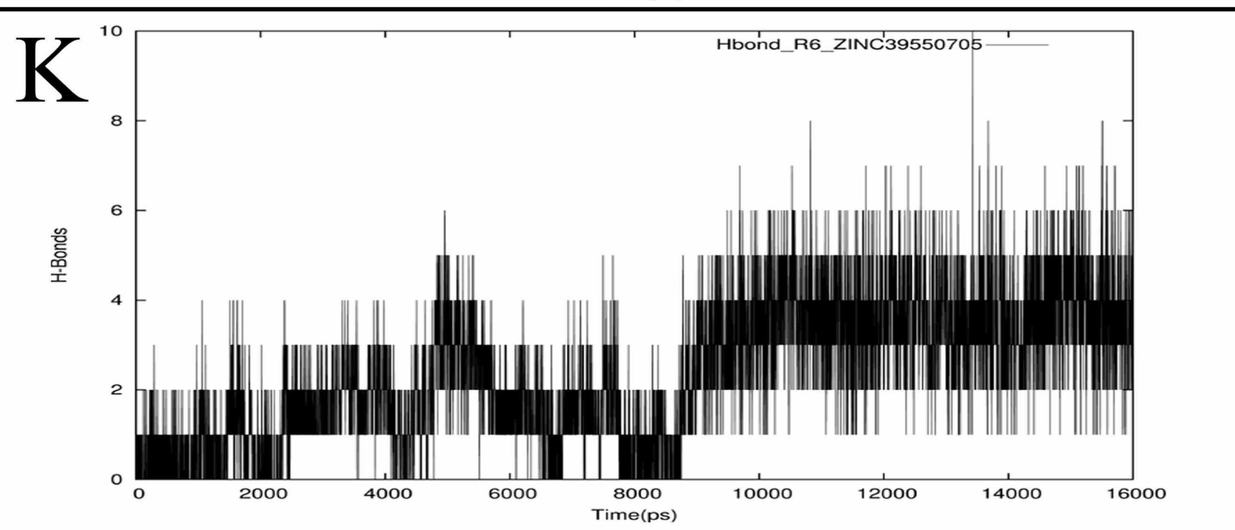

C
Hbond_5204_ZINC36922620
H-Bonds
Time(ps)

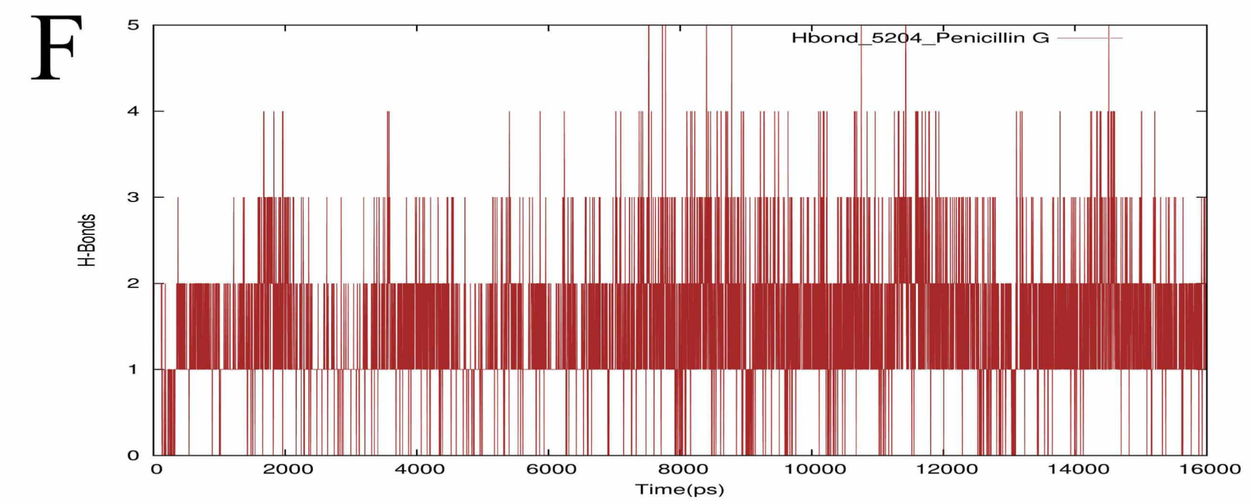

I
Hbond_R6_ZINC36922620
H-Bonds
Time(ps)


D
Hbond_5204_ZINC36953975
H-Bonds
Time(ps)


J
Hbond_R6_ZINC36953975
H-Bonds
Time(ps)


E
Hbond_5204_ZINC39550705
H-Bonds
Time(ps)


K
Hbond_R6_ZINC39550705
H-Bonds
Time(ps)


F
Hbond_5204_Penicillin G
H-Bonds
Time(ps)

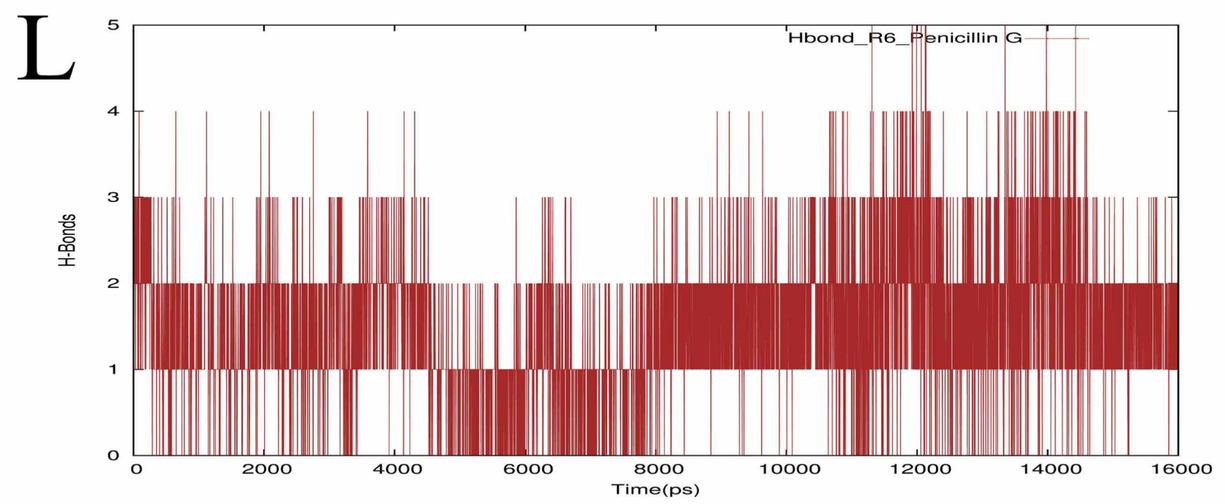
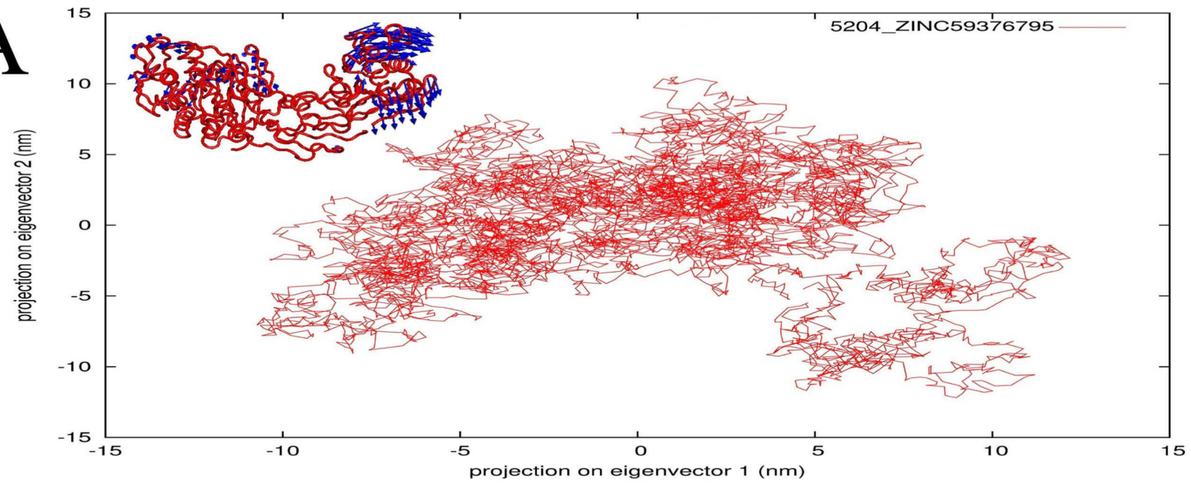
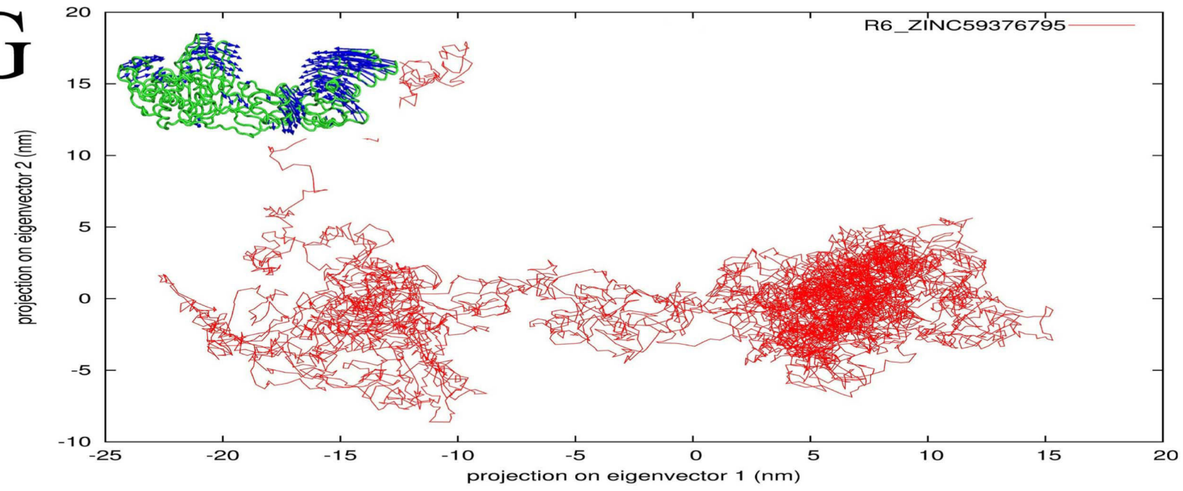

L
Hbond_R6_Penicillin G
H-Bonds
Time(ps)

Figure. 8

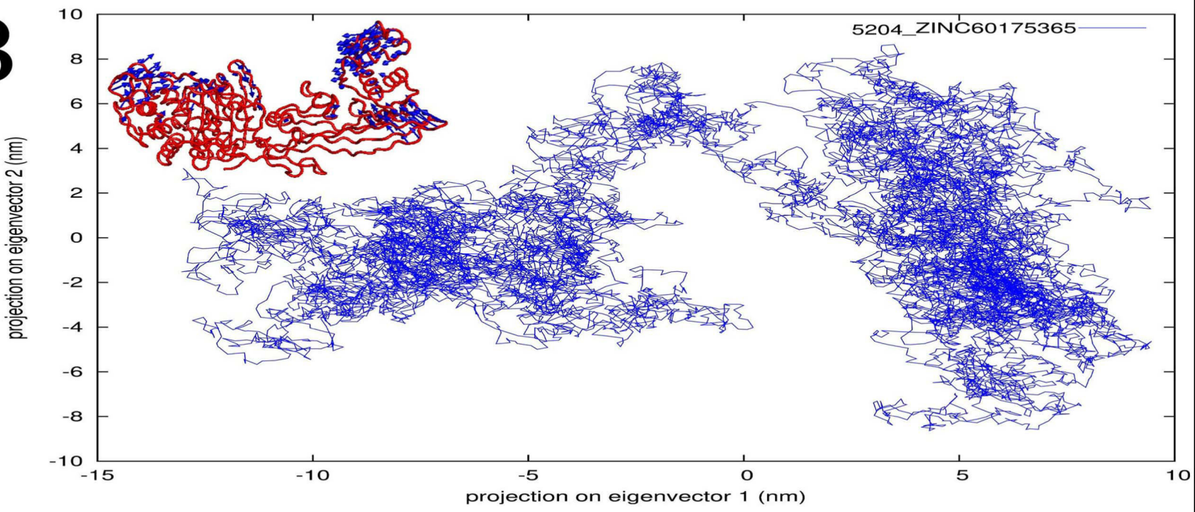
A



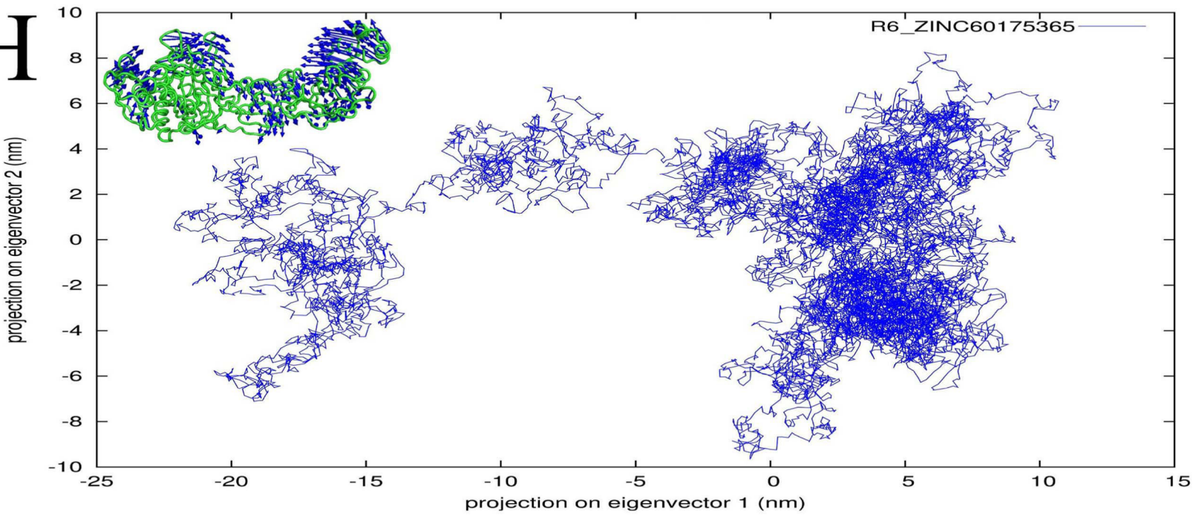
G



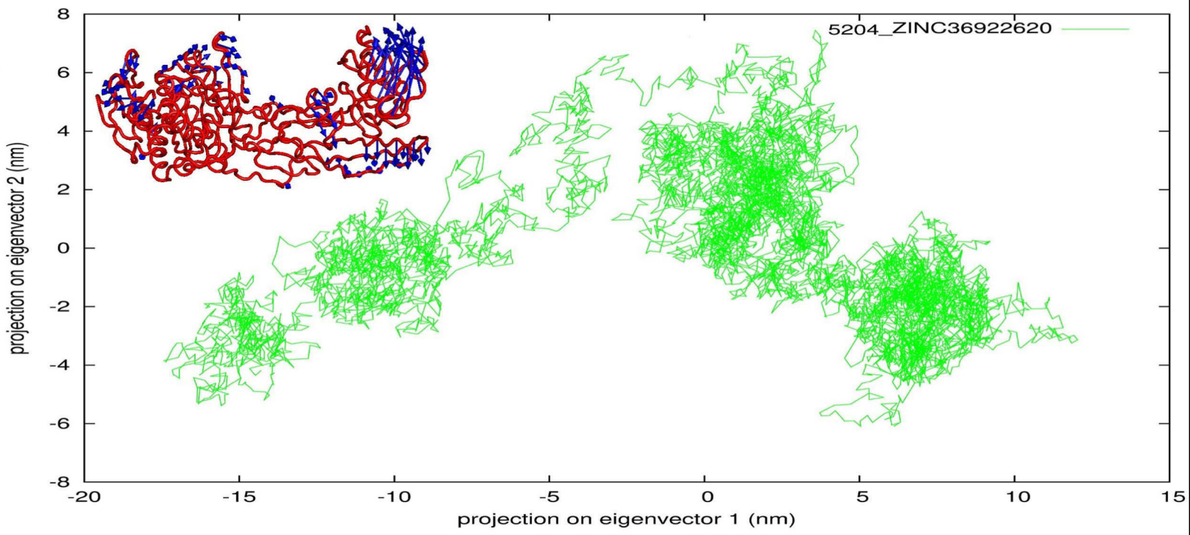
B



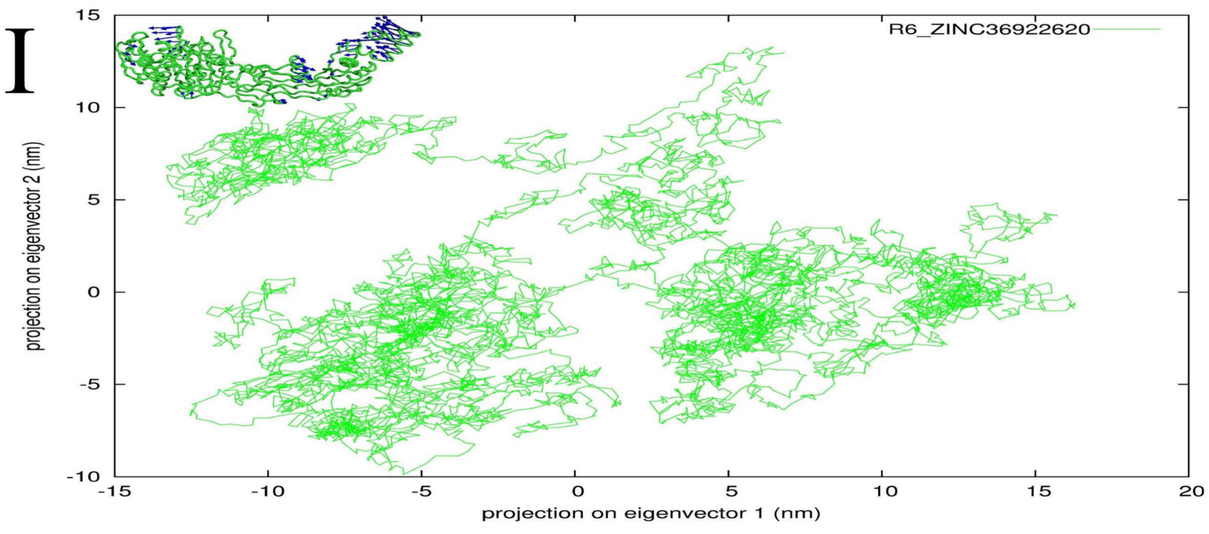
H



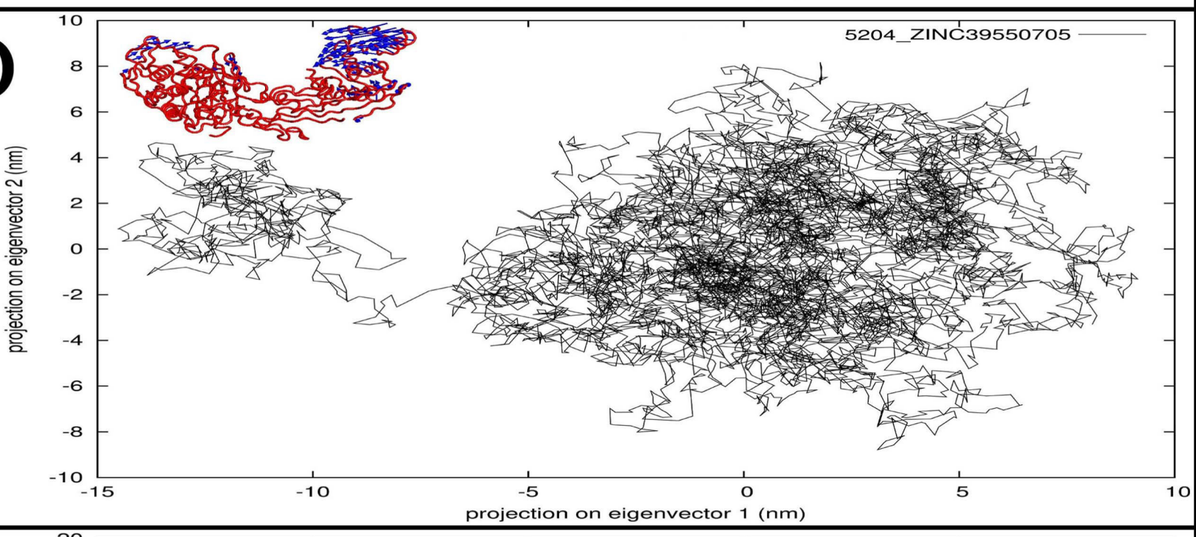
C



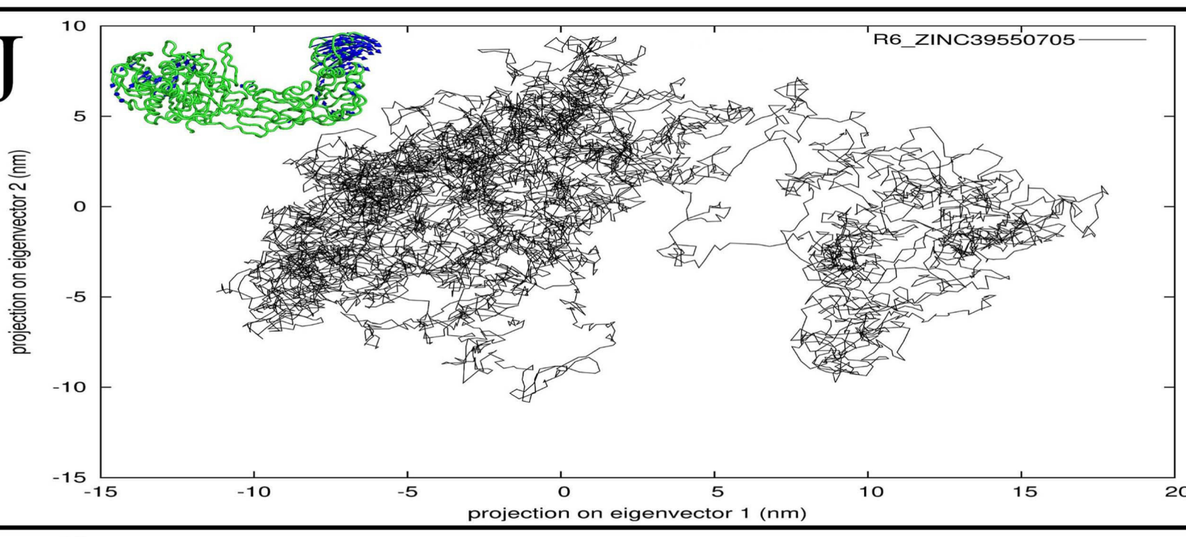
I



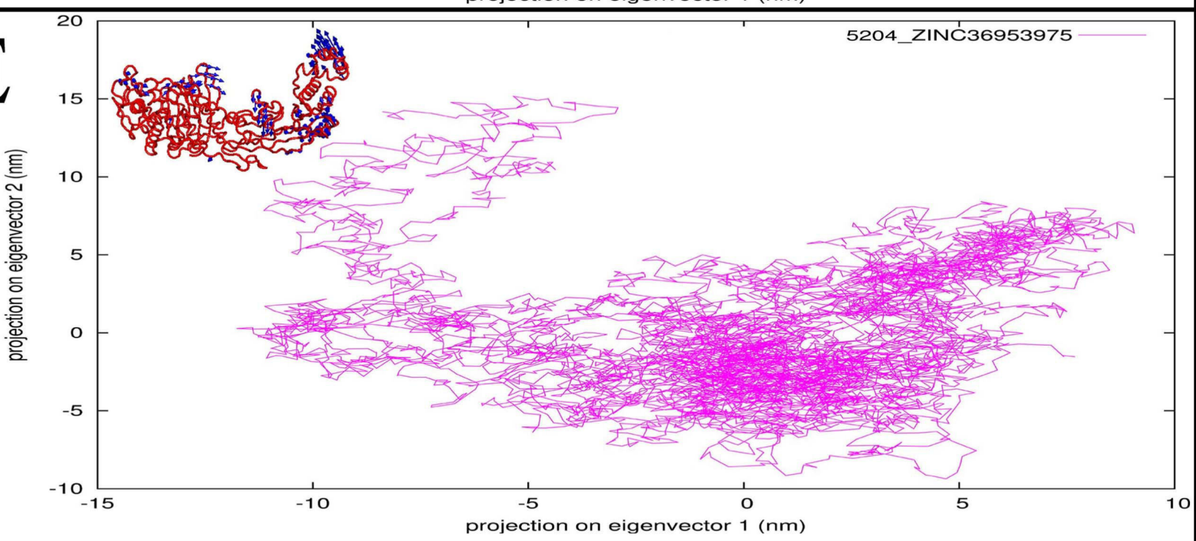
D



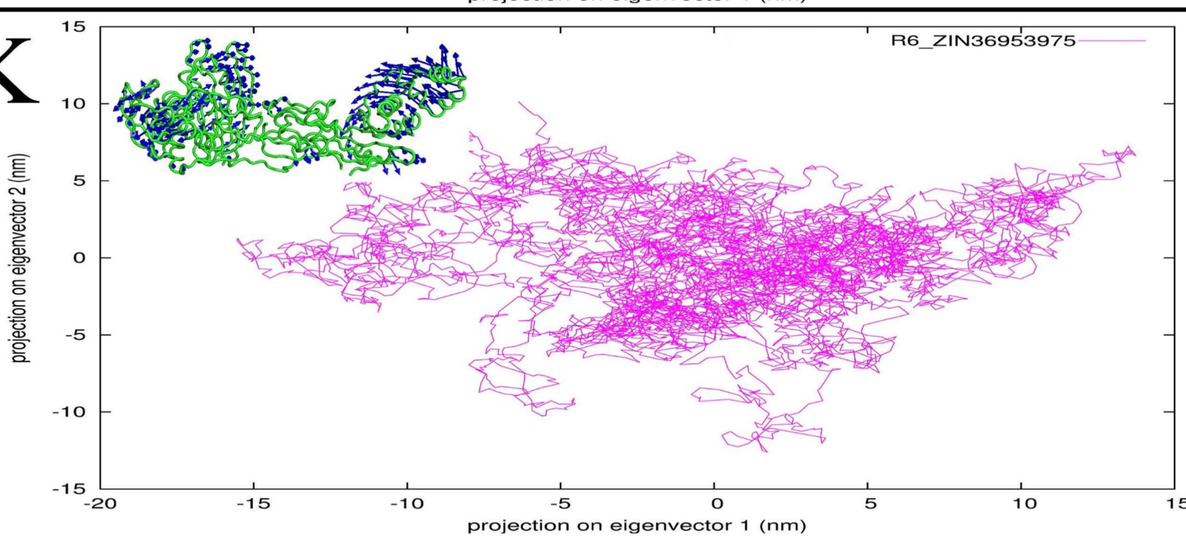
J



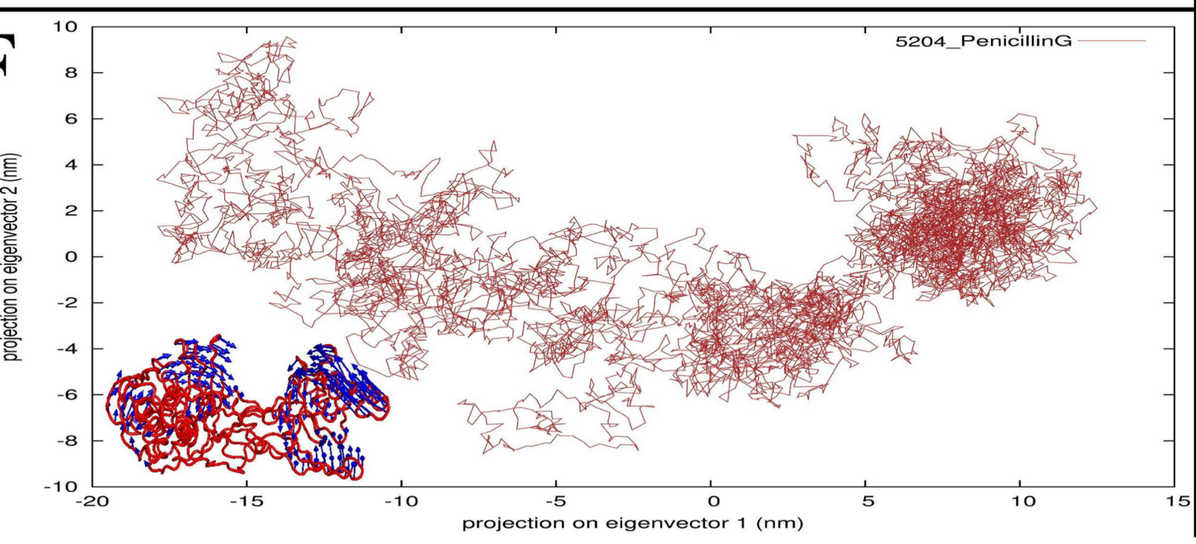
E



K



F



L

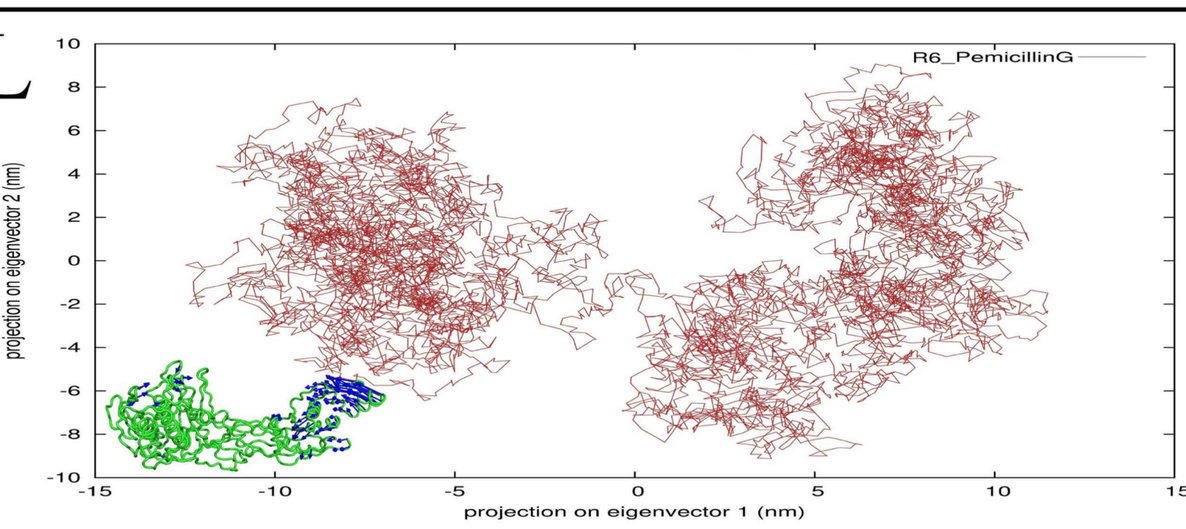
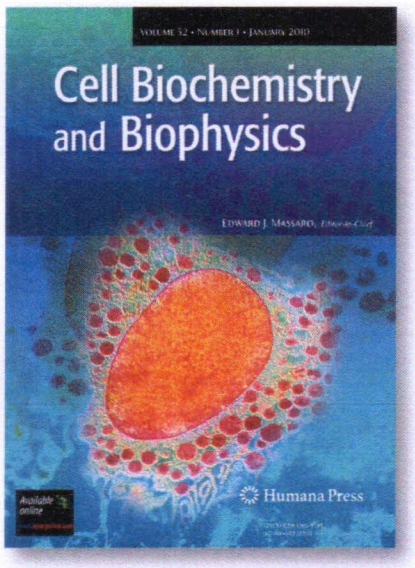


Figure. 9



Cell Biochemistry and Biophysics

Editor-in-Chief: L. Berliner

- Presents original information on the properties, structure, function, behavior, and interactions of cells and their molecular/macromolecular constituents
- Publishes letters, editorials, one-pagers on negative results, book reviews, and meeting/workshop announcements
- Coverage includes genetic and biomolecular engineering; computer-based analysis of tissues, cells, cell networks, organelles, and molecular/macromolecular assemblies
- 96% of authors who answered a survey reported that they would definitely publish or probably publish in the journal again

Cell Biochemistry and Biophysics fosters progress in comprehending the nature of the biochemical and biophysical mechanisms underlying the control of cellular physiological homeostasis and the consequences of its perturbation. Reports span the disciplines of modern biochemistry and chemistry, biophysics and cell physiology, physics and engineering, molecular and structural biology and the medical sciences. Emphasis is placed on the relationship between molecular structure and the nature of the specific property/function under investigation. CBB publishes articles describing the quantitative utilization of and innovative developments in: genetic and biomolecular engineering; computer analysis of tissues, cells, cell networks, organelles, and molecular/macromolecular assemblies; research in and development of photometric, spectroscopic, microscopic, mechanical, and electrical methodologies/techniques in analytical cytology and cytometry, as well as innovation in instrument design.

Impact Factor: 1.627 (2015), Journal Citation Reports®, Thomson Reuters

On the homepage of **Cell Biochemistry and Biophysics** at **springer.com** you can

- Sign up for our Table of Contents Alerts
- Get to know the complete Editorial Board
- Find submission information

4 issues/year

Electronic access

- springerlink.com

Subscription information

- springer.com/librarians

Cell Biochemistry and Biophysics

Analysis of species selectivity of human, mouse and rat Cytochrome P450 1A and 2B subfamily enzymes using molecular modeling, docking and dynamics simulations

--Manuscript Draft--

Manuscript Number:	CBBI-D-16-00083R1
Full Title:	Analysis of species selectivity of human, mouse and rat Cytochrome P450 1A and 2B subfamily enzymes using molecular modeling, docking and dynamics simulations
Article Type:	Original Article
Keywords:	Cytochrome P450; Species selectivity; Homology modeling; Glide docking; Schrödinger software suite; Molecular dynamics simulations
Corresponding Author:	Parthasarathy Subbiah, Ph.D., Bharathidasan University Tiruchirappalli, Tamil Nadu INDIA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Bharathidasan University
Corresponding Author's Secondary Institution:	
First Author:	Karthikeyan Shanmugam Bagavathy, M.Sc.,
First Author Secondary Information:	
Order of Authors:	Karthikeyan Shanmugam Bagavathy, M.Sc.,
	Suvaithenamudhan Suvaiyarasan, M.Sc., M.Phil.,
	Akbarsha Abdulkader Mohammad, Ph.D.,
	Parthasarathy Subbiah, Ph.D.,
Order of Authors Secondary Information:	
Funding Information:	
Abstract:	Cytochrome P450 (CYP) 1A and 2B subfamily enzymes are the important drug metabolizing enzymes that are highly conserved across several species. Though these enzymes are highly conserved at the sequence level, their structural and macromolecular features govern species and substrate-selectivity. Therefore, species selectivity of CYP1A and CYP2B subfamily proteins across human, mouse and rat was analyzed using molecular modeling, docking and dynamics simulations. The chiral molecules quinine and quinidine were used as ligands. The three-dimensional structures for 17 proteins of CYP1A and CYP2B subfamilies of mouse and rat were predicted through homology modeling, using the available structures of CYP1A and CYP2B proteins of human as template. Molecular docking and dynamics simulations of CYP1A subfamily proteins with quinine and quinidine reveal the existence of species-selectivity across human, mouse and rat. In case of CYP2B subfamily proteins, the absence of the role of chirality between quinine and quinidine was observed when these ligands form complexes with CYP2B subfamily proteins of human, mouse and rat. Our findings reveal the role of active site amino acid residues of CYP1A and CYP2B subfamily proteins and provide insights about species-selectivity of these enzymes across human, mouse and rat.
Response to Reviewers:	Response to the reviewer comments Reply to the reviewer’s comments on our manuscript (ID: CBBI-D-16-00083) entitled “Analysis of species selectivity of human, mouse and rat Cytochrome P450 1A and 2B subfamily enzymes using molecular modeling, docking and dynamics simulations” Response to the Comments of Reviewer 1 We thank the reviewer for the valuable comments. We have carefully gone through all

major and minor comments, and incorporated the necessary changes in the revised manuscript. The comments of the reviewer were very much helpful in improving our manuscript.

1. In this work authors studied the species selectivity of CYP1A and CYP2B subfamily proteins among human, mouse and rat using molecular modeling, docking and dynamics simulations studies using the chiral molecules, quinine and quinidine as ligands. Homology modelling is used to predict the threedimensional structures for 17 proteins of CYP1A and CYP2B subfamilies of mouse and rat using the available structures of CYP1A and CYP2B proteins of human as template. In their conclusions, authors found that CYP1A subfamily proteins exhibit species differences across human, mouse and rat, not the CYP2B subfamily proteins. In general the study is well executed and the manuscript is well written. Methods section is detailed and clear and the citations in all sections are adequate.
The manuscript can be considered for publication provided that the following points are addressed: The results of this sort of work are critically dependent on the validity/reliability of the initial computationally generated structures (ligand protein complexes) since the experimental Xray structures of most of the studied systems are unavailable. This necessitates careful validation of the modeled structures (homology models, active site prediction as well as docked structures) before processing with any subsequent MD or analyses. This part is largely lacking in the current form of the manuscript. Therefore, authors need to thoroughly validate their homology models, active site prediction methods and docking results, otherwise the outcome of this study is highly questionable. For validation of homology models, several bioinformatics tools are available for validation of homology models, the least of which is Ramachandran plots but others tools are reported in literature. For docking results, crossvalidation and RMSD calculations for reported Xray structures are the simplest approaches being used for validation of docking results. Before docking calculations, validation of active site prediction is something tricky, however, most critical, therefore, authors need to validate the active site residues using different reported tools in literature, however, I would suggest at least cross validation using two different approaches would be acceptable.

Reply:
As suggested by the reviewer, all three validations (homology models, active site prediction and docked structures) have been performed and included in the manuscript. i) Validation of the predicted 17 three-dimensional protein structures using Ramachandran plot was performed. The results shows that <0.5% of amino acids are in the outlier region and >95% of amino acids are in the favored region. The results are presented as Supplementary Figures S1-S3 and Supplementary Table T4.(Page No.11, Line Nos.250-253).ii) The sitemap module of Schrödinger was used to predict the active sites, and the results were further validated by predicting the active site residues using sitehound server as well as by comparing the predicted active sites with the literature and a thorough consensus prediction was done (Page No.7, Line Nos.158-160). The results are presented in Supplementary Tables T5-T7. iii) Validation of docking results was done by calculating RMSD values between the reported x-ray structures (PDB Id: 418V, 2HI4 and 3QOA) and docked complexes. The results of calculated RMSD values are included in Tables 2 and 4 and explained in ‘Results and discussion’ section (Page No.13, Line Nos. 296-297; Page No.13, Line Nos. 307-308 and Page No. 17, Line Nos. 408-410).

2. Minor issues: Abstract: This sentence require revising: "The results reveal that there are species differences in CYP1A subfamily proteins when complex with quinine and quinidine"

Reply: We have modified the above sentence as “Molecular docking and dynamics simulations of CYP1A subfamily proteins with quinine and quinidine reveal the existence of species selectivity across human, mouse and rat” in the abstract section as suggested by the reviewer.(Page No. 2, Line Nos. 40-42).

With these modifications, we hope that the revised manuscript may be acceptable for publication.

Reply to the reviewer's comments on our manuscript (ID: CBBI-D-16-00083) entitled "Analysis of species selectivity of human, mouse and rat Cytochrome P450 1A and 2B subfamily enzymes using molecular modeling, docking and dynamics simulations"

Response to the Comments of Reviewer 2
We thank the reviewer for the valuable comments on our manuscript. We have carefully gone through all the major and minor comments of the reviewer and incorporated the necessary changes in the revised manuscript. The comments of the reviewer are very much helpful to improve our manuscript.

1.The manuscript (MS) by Karthikeyan et al. presents a well-designed computational study of an interesting problem regarding ligand binding features of cytochromes P450 from human, mouse and rat. The MS reports a competent work with a clearly written text and with significant results well supported by a considerable amount of data in tables and figures. However, I have some reservations that the authors should address before recommending publication of their MS. The major one concerns the "Conclusion" section. In its current form, it includes many methodological points which have been already presented in the MS and that might be avoided here. On the other side, statements such as …"computational approaches reveals the influence of structural and molecular features in studying species selectivity of CYP enzymes" (p. 18) are pointless because of their obviousness. This section must focus on major conclusions and hence, it should be considerably shortened.

Reply:
As suggested by the reviewer, the sentence "computational approaches reveals the influence of structural and molecular features in studying species selectivity of CYP enzymes" in 'Conclusion' section is removed and further, the methodological points are critically shortened and the conclusion section is thoroughly modified (Page No.20 Line Nos.460-495).

Minor comments:
2. Except for readers familiar with Glide docking, the Glide XP score is not very informative. Values of this score are included in Tables 2-5 but when it is presented (page 8), a brief explanation on its meaning and magnitude should be given.

Reply:
As suggested by the reviewer, a description about glide XP docking is included in 'Materials and methods' section in page no. 8 and line no's. 177-179. The results of glide XP docking, i.e glide XP score is explained and included in 'Results and discussion' section in page no. 11, line nos'. 264-266; Page no. 12, 272-274 and page no. 16, line nos'. 382-386.

3. Same RMSD scales should be used for all the plots in a same panel. Thus, (0.0-0.4)in Fig. 4 and (0.0-0.35)in both Fig. 8 and Suppl. Fig. S2 should be more convenient (the authors have done so in Fig. 6).

Reply:
As suggested by the reviewer, the RMSD scales (Y axis) are presented uniformly for all the plots in the same panel figures such as figures 4, 8 and Supplementary Figure S8.

4.Avoid "0.5" divisions in Yaxislabels in Figs. 5, 7, and 9 as they make no sense for the number ofhydrogen bonds over the trajectories.

Reply:
As suggested by the reviewer, the Y axis (Number of hydrogen bonds) divisions are changed in figures 5, 7 and 9.

5.Given the precision with which binding energies are computed, reporting values in kcal/mol with threeseems excessive (one decimal is enough!)

Reply:
As suggested by the reviewer, the values of binding energies are rounded to one decimal in Tables 1-4.

6. Remove the sentence starting at line 12 in page 17 ("The g_hbond utility …") as it is irrelevant.

Reply:
The sentence, as suggested by the reviewer, is removed.

7. Correct the algorithm name (it must be steepest "descent") in page 9, lines 17 and 53.

Reply:
As suggested by the reviewer,the algorithm name is corrected as steepest descent in page no. 9 and 10, line nos. 203 and 223, respectively.

8. Correct the algorithm name (it must be "Parrinello"Rahman)in page 9, line 19.

Reply:
The algorithm name is changed to Parrinello-Rahman barostat algorithm, as suggested by the reviewer, in page no.9, line no.210.

9. In page 9, line 58 it is said that …"the total system was equilibrated at constant temperature andpressure using NVT and NPT ensemble…": remove "NVT" if the calculations were in fact made at constant pressure.

Reply:
In MD simulations, the total system was equilibrated involving two phases: one with constant number of particles, volume and temperature employed using NVT at 100ps, and another using constant pressure using NPT ensemble at 100 ps. The sentence is now reframed to improve clarity (Page no. 10, line no. 225-227).
With these modifications and improvements, we hope that the revised manuscript may be acceptable for publication.

Reply to the reviewer's comments on our manuscript (ID: CBBI-D-16-00083) entitled "Analysis of species selectivity of human, mouse and rat Cytochrome P450 1A and 2B subfamily enzymes using molecular modeling, docking and dynamics simulations"

Response to the Comments of Reviewer 3
1. I feel there is a lack of experimental evidence referenced in this paper. The quinidine ligands are particularly interesting because they have been shown to inhibit CYP 2D6, but not induce a low spin state. Rather their binding causes a high spin state change which suggests the axial water molecule bound to the heme iron is displaced. This should give a good hint to the binding mode (in 2D6) and I would like to see similar evidence referenced for the CYPS in the paper CYP1A and 2B. Similar to this there is insufficient detail about the heme parameterization used in the molecular dynamics. What is the charge model used for the heme? Is it a low spin or high spin state (looks from the pictures to be high)? Does the proposed binding mode displace the axial water? If so why is it not a substrate. These are the questions I would have. Also it is not clear to me why the focus is on the 1A and 2B families.

Reply:
In our molecular docking results of quinine and quinidine, we observed a close proximity of ligand towards heme when binding to CYP1A and CYP2B subfamily proteins. We have used GROMACS96 43a1 force field in GROMACS 4.5.5 package for parameterization of heme. The net charge of heme is -2. The topology of the heme added to the protein as well as protein-ligand complex topology files and energy minimization were performed (Page no. 9, line no. 206-208). On analysis of the docking poses and trajectories of MD simulations, we observed that quinine and quinidine binds closely to the active site residues in the proximity of heme, displace water and higher transition state is induced as suggested by Jovanovic et al.(2005) and Capoferri et al.(2016) and the proteins were well stabilized by forming stable hydrogen bonds. This suggests that these ligands may act as substrates. The objective of our study has been to analyze species selectivity of CYP1A and CYP2B subfamily enzymes using the chiral molecules quinine and quinidine using in silico approaches. Therefore, separate experimental studies are required to confirm whether these

ligands act as substrates of CYP1A and CYP2B subfamily proteins. Our choice of the ligands quinine and quinidine; CYP1A and CYP2B subfamily proteins is justified below.

In a recent study published by our group entitled “Network analysis and cross species comparison of protein–protein interaction networks of human, mouse and rat cytochrome P450 proteins that degrade xenobiotics”(DOI: 10.1039/C6MB00210B), we analyzed protein–protein interaction networks of human, mouse and rat CYP enzymes and reported important hub proteins of human, mouse and rat CYPs. CYP1A and CYP2B subfamily proteins are the important key proteins which act as hub proteins in the network. Also, CYP1A and CYP2B subfamily proteins are important drug metabolizing CYP enzymes, where CYP1A2 enzyme is involved in the metabolism of nearly 4% of drugs available in the market (Zuber et al. 2002) and CYP2B6 enzyme is involved in the metabolism of nearly 25% of drugs on the market (Xie et al. 2001) (Page No. 3; Line Nos.56-59).Though quinidine and quinine are reported as inhibitors of CYP2D6 protein, studies carried out by Hutzler et al. (2003), Venhorst et al. (2003), Chauret et al. (1997) and Edmund et al. (2013) used both quinine and quinidine in addressing species selectivity of several CYP enzymes as they are chiral molecules. Also, a study by Ching et al. (2001) reported the differential inhibition of human CYP1A1 and CYP1A2 by quinine and quinidine and Walsky et al. (2006) reported the use of quinine and quinidine for inhibition of CYP2B6 in vitro. In this study, we have used combination of molecular modeling, docking and dynamics simulations approaches to address species selectivity of CYP1A and CYP2B subfamily enzymes using quinine and quinidine(Page No. 4, Line Nos.81-83). Recently, a study by Kesharwaniet al. (2016) reported substrate specificity of CYP1A1, CYP1A2 proteins using molecular docking and dynamics simulations and concluded that specificity of a particular substrate depends upon the type of the active site residues, and several other macromolecular features though they are highly conserved at the sequence level (Page No. 3 and Line Nos.70-73).

With these modifications and improvements, we hope that the revised manuscript may be acceptable for publication.

Reply to the reviewer’s comments on our manuscript (ID: CBBI-D-16-00083) entitled “Analysis of species selectivity of human, mouse and rat Cytochrome P450 1A and 2B subfamily enzymes using molecular modeling, docking and dynamics simulations”

Responseto the Comments of Reviewer 4

1. Table T1: Providing a sequence alignment would help, especially including the active sites discussed in this manuscript.

Reply:

As suggested by the reviewer, sequence alignment of human, mouse and rat CYP1A and CYP2B subfamily proteins was performed and results are given as Supplementary Figures S4, S5 and S6. The active site residues that are discussed in the manuscript is highlighted in the figures and explained in ‘Results and discussion’ section (Page no. 19, line no. 439-447).

2. The last paragraph of Introduction is too long. Much of its content should be in the Methods or Results sections.

Reply:

The last paragraph of the introduction section is shortened as suggested by the reviewer (Page no.5, line nos. 103-117).

3. I suggest reorganizing the results: instead of grouping the results according to docking methods, I think Docking, IFD, and MD of the same protein/ligand should be grouped together.

Reply:

As suggested by the reviewer, the results are grouped according to CYP subfamily in the ‘Results and discussion’ section (Page nos.11-19, line nos. 254-437).

With these modifications and improvements, we hope that the revised manuscript may be acceptable for publication.

Reply to the editor's comments on our manuscript (ID: CBBI-D-16-00083) entitled "Analysis of species selectivity of human, mouse and rat Cytochrome P450 1A and 2B subfamily enzymes using molecular modeling, docking and dynamics simulations"

Response to the Comments of the editor
1. Abstract needs to be reframed

Reply:
As suggested by the editor, the abstract of the MS is completely reframed (Page no.2, line nos. 32-46).

2. Author has used two different patterns for reference style in introduction part

Reply:
We have modified the reference style to uniformity throughout the manuscript according to the journal format.

3. Author has covered part of the experimental and R & D in the Introduction, infact there is repetition of the text. Authors could have given good significant background of the work.

Reply:
We have modified/shortened the 'introduction' section, especially the last paragraph which appeared like a methodological and with repeated texts (Page no.5, line nos. 103-117).

4. To make the manuscript strong and impactful, authors could have covered the rationale to select and use quinine and quinidine as ligands???

Reply:
CYP1A subfamily proteins are highly conserved at the sequence level, but not at the active site regions. Recently, a study conducted by Kesharwaniet al. (2016) reported that substrate specificity of CYP1A1 and CYP1A2proteins using molecular docking and dynamics simulations. They concluded that the role of active site residues and other macromolecular features influences on substrate selectivity (Page No. 3, Line Nos. 70-73). Quinine and quinidine are the chiral molecules which exhibit differences only in their 3D structure. Though they are well known inhibitors of CYP2D6 protein, they have beenextensively used to study species selectivity of various CYP enzymes by Hutzler et al. (2003), Venhorst et al. (2003), Chauret et al. (1997) and Edmund et al. (2013) as explained in Page No. 4, line nos. 79-102. Ching et al. (2001) reported the differential inhibition of human CYP1A1 and CYP1A2 by quinine and quinidine and Walsky et al. (2006) reported the use of quinine and quinidine for inhibition of CYP2B6 in vitro(Page No. 4, Line Nos. 81-83).Thus,the chiral molecules quinine and quinidine were used as ligands in addressing species selectivity of CYP1A and CYP2B subfamily enzymes using the combination of molecular modeling, docking and dynamics simulations.

5. Author mentioned the version of Prime (Schrodinger) in the introduction part, and missed to mention in experimental section, and it was an equally quite important to mention in experimental methods.

Reply:
We have used Prime module of version 3.2 and this is included in 'Materials and methods' section as suggested by the editor (Page no.6, line no. 133).

6. Standard formatting as per journal guidelines is missing. Use of the English language especially in results and discussion is not in commensuration with the standard of the journal. There are small grammatical mistakes throughout the MS

Reply:
We have thoroughly revised and formatted the manuscript. The language and grammatical mistakes were also corrected in the revised manuscript.

	With these modifications and improvements, we believe that the manuscript has been improved to the best. We request that the revised manuscript may be accepted for publication.

Response to the reviewer comments

Reply to the reviewer's comments on our manuscript (ID: CBBI-D-16-00083) entitled "Analysis of species selectivity of human, mouse and rat Cytochrome P450 1A and 2B subfamily enzymes using molecular modeling, docking and dynamics simulations"

Response to the Comments of Reviewer 1

We thank the reviewer for the valuable comments. We have carefully gone through all major and minor comments, and incorporated the necessary changes in the revised manuscript. The comments of the reviewer were very much helpful in improving our manuscript.

1. In this work authors studied the species selectivity of CYP1A and CYP2B subfamily proteins among human, mouse and rat using molecular modeling, docking and dynamics simulations studies using the chiral molecules, quinine and quinidine as ligands. Homology modelling is used to predict the threedimensional structures for 17 proteins of CYP1A and CYP2B subfamilies of mouse and rat using the available structures of CYP1A and CYP2B proteins of human as template. In their conclusions, authors found that CYP1A subfamily proteins exhibit species differences across human, mouse and rat, not the CYP2B subfamily proteins. In general the study is well executed and the manuscript is well written. Methods section is detailed and clear and the citations in all sections are adequate.

The manuscript can be considered for publication provided that the following points are addressed: The results of this sort of work are critically dependent on the validity/reliability of the initial computationally generated structures (ligand protein complexes) since the experimental Xray structures of most of the studied systems are unavailable. This necessitates careful validation of the modeled structures (homology models, active site prediction as well as docked structures) before processing with any subsequent MD or analyses. This part is largely lacking in the current form of the manuscript. Therefore, authors need to thoroughly validate their homology models, active site prediction methods and docking results, otherwise the outcome of this study is highly questionable. For validation of homology models, several bioinformatics tools are available for validation of homology models, the least of which is Ramachandran plots but others tools are reported in literature. For docking results, crossvalidation and RMSD calculations for reported Xray structures are the simplest approaches being used for validation of docking results. Before docking calculations, validation of active site prediction is something tricky, however, most critical, therefore, authors need to validate the active site residues using different reported tools in literature, however, I would suggest at least cross validation using two different approaches would be acceptable.

Reply:

As suggested by the reviewer, all three validations (homology models, active site prediction and docked structures) have been performed and included in the manuscript. i) Validation of the predicted 17 three-dimensional protein structures using Ramachandran plot was performed. The results shows that <0.5% of amino acids are in the outlier region and >95% of amino acids are in the favored region. The results are presented as Supplementary Figures S1-S3 and Supplementary Table T4.(Page No.11, Line Nos.250-253).ii) The sitemap module of Schrödinger was used to predict the active sites, and the results were further validated by predicting the active site residues using sitehound server as well as by comparing the predicted active sites with the literature and a thorough consensus prediction was done (Page No.7, Line Nos.158-160). The results are presented in Supplementary Tables T5-T7. iii) Validation of docking results was done by calculating RMSD values between the reported x-ray structures (PDB Id: 418V, 2HI4 and 3QOA) and docked complexes. The results of calculated RMSD values are included in Tables 2 and 4 and explained in 'Results and discussion' section (Page No.13, Line Nos. 296-297; Page No.13, Line Nos. 307-308 and Page No. 17, Line Nos. 408-410).

2. Minor issues: Abstract: This sentence require revising: "The results reveal that there are species differences in CYP1A subfamily proteins when complex with quinine and quinidine"

Reply: We have modified the above sentence as "Molecular docking and dynamics simulations of CYP1A subfamily proteins with quinine and quinidine reveal the existence of species selectivity across human, mouse and rat" in the abstract section as suggested by the reviewer.(Page No. 2, Line Nos. 40-42).

With these modifications, we hope that the revised manuscript may be acceptable for publication.

Reply to the reviewer's comments on our manuscript (ID: CBBI-D-16-00083) entitled "Analysis of species selectivity of human, mouse and rat Cytochrome P450 1A and 2B subfamily enzymes using molecular modeling, docking and dynamics simulations"

Response to the Comments of Reviewer 2

We thank the reviewer for the valuable comments on our manuscript. We have carefully gone through all the major and minor comments of the reviewer and incorporated the necessary changes in the revised manuscript. The comments of the reviewer are very much helpful to improve our manuscript.

1. *The manuscript (MS) by Karthikeyan et al. presents a well-designed computational study of an interesting problem regarding ligand binding features of cytochromes P450 from human, mouse and rat. The MS reports a competent work with a clearly written text and with significant results well supported by a considerable amount of data in tables and figures. However, I have some reservations that the authors should address before recommending publication of their MS. The major one concerns the "Conclusion" section. In its current form, it includes many methodological points which have been already presented in the MS and that might be avoided here. On the other side, statements such as …"computational approaches reveals the influence of structural and molecular features in studying species selectivity of CYP enzymes" (p. 18) are pointless because of their obviousness. This section must focus on major conclusions and hence, it should be considerably shortened.*

Reply:

As suggested by the reviewer, the sentence "computational approaches reveals the influence of structural and molecular features in studying species selectivity of CYP enzymes" in 'Conclusion' section is removed and further, the methodological points are critically shortened and the conclusion section is thoroughly modified (Page No.20 Line Nos.460-495).

Minor comments:

2. *Except for readers familiar with Glide docking, the Glide XP score is not very informative. Values ofthis score are included in Tables 2-5 but when it is presented (page 8), a brief explanation on itsmeaning and magnitude should be given.*

Reply:

As suggested by the reviewer, a description about glide XP docking is included in 'Materials and methods' section in page no. 8 and line no's. 177-179. The results of glide XP docking, i.e glide XP score is explained and included in 'Results and discussion' section in page no. 11, line nos'. 264-266; Page no. 12, 272-274 and page no. 16, line nos'. 382-386.

3. *Same RMSD scales should be used for all the plots in a same panel. Thus, (0.0-0.4)in Fig. 4 and (0.0-0.35)in both Fig. 8 and Suppl. Fig. S2 should be more convenient (the authors have done so in Fig. 6).*

Reply:

As suggested by the reviewer, the RMSD scales (Y axis) are presented uniformly for all the plots in the same panel figures such as figures 4, 8 and Supplementary Figure S8.

4. *Avoid "0.5" divisions in Yaxislabels in Figs. 5, 7, and 9 as they make no sense for the number ofhydrogen bonds over the trajectories.*

Reply:

As suggested by the reviewer, the Y axis (Number of hydrogen bonds) divisions are changed in figures 5, 7 and 9.

5. *Given the precision with which binding energies are computed, reporting values in kcal/mol with threeseems excessive (one decimal is enough!)*

Reply:

As suggested by the reviewer, the values of binding energies are rounded to one decimal in Tables 1-4.

6. *Remove the sentence starting at line 12 in page 17 ("The g_hbond utility ...") as it is irrelevant.*

Reply:

The sentence, as suggested by the reviewer, is removed.

7. *Correct the algorithm name (it must be steepest "descent") in page 9, lines 17 and 53.*

Reply:

As suggested by the reviewer,the algorithm name is corrected as steepest descent in page no. 9 and 10, line nos. 203 and 223, respectively.

8. *Correct the algorithm name (it must be "Parrinello"Rahman)in page 9, line 19.*

Reply:

The algorithm name is changed to Parrinello-Rahman barostat algorithm, as suggested by the reviewer, in page no.9, line no.210.

9. *In page 9, line 58 it is said that …"the total system was equilibrated at constant temperature andpressure using NVT and NPT ensemble…": remove "NVT" if the calculations were in fact made at constant pressure.*

Reply:

In MD simulations, the total system was equilibrated involving two phases: one with constant number of particles, volume and temperature employed using NVT at 100ps, and another using constant pressure using NPT ensemble at 100 ps. The sentence is now reframed to improve clarity (Page no. 10, line no. 225-227).

With these modifications and improvements, we hope that the revised manuscript may be acceptable for publication.

Reply to the reviewer's comments on our manuscript (ID: CBBI-D-16-00083) entitled "Analysis of species selectivity of human, mouse and rat Cytochrome P450 1A and 2B subfamily enzymes using molecular modeling, docking and dynamics simulations"

Responseto the Comments of Reviewer 3

1. *I feel there is a lack of experimental evidence referenced in this paper. The quinidine ligands are particularly interesting because they have been shown to inhibit CYP 2D6, but not induce a low spin state. Rather their binding causes a high spin state change which suggests the axial water molecule bound to the heme iron is displaced. This should give a good hint to the binding mode (in 2D6) and I would like to see similar evidence referenced for the CYPS in the paper CYP1A and 2B. Similar to this there is insufficient detail about the heme parameterization used in the molecular dynamics. What is the charge model used for the heme? Is it a low spin or high spin state (looks from the pictures to be high)? Does the proposed binding mode displace the axial water? If so why is it not a substrate. These are the questions I would have. Also it is not clear to me why the focus is on the 1A and 2B families.*

Reply:

In our molecular docking results of quinine and quinidine, we observed a close proximity of ligand towards heme when binding to CYP1A and CYP2B subfamily proteins. We have used GROMACS96 43a1 force field in GROMACS 4.5.5 package for parameterization of heme. The net charge of heme is -2. The topology of the heme added to the protein as well as protein-ligand complex topology files and energy minimization were performed (Page no. 9, line no. 206-208). On analysis of the docking poses and trajectories of MD simulations, we observed that quinine and quinidine binds closely to the active site residues in the proximity of heme, displace water and higher transition state is induced as suggested by Jovanovic et al.(2005) and Capoferri et al.(2016) and the proteins were well stabilized by forming stable hydrogen bonds. This suggests that these ligands may act as substrates. The objective of our study has been to analyze species selectivity of CYP1A and CYP2B subfamily enzymes using the chiral molecules quinine and quinidine using *in silico* approaches. Therefore, separate experimental studies are required to confirm whether these ligands act as substrates of CYP1A and CYP2B subfamily proteins. Our choice of the ligands quinine and quinidine; CYP1A and CYP2B subfamily proteins is justified below.

In a recent study published by our group entitled "Network analysis and cross species comparison of protein–protein interaction networks of human, mouse and rat cytochrome P450 proteins that degrade xenobiotics"(DOI: 10.1039/C6MB00210B), we analyzed protein–protein interaction networks of human, mouse and rat CYP enzymes and reported important

hub proteins of human, mouse and rat CYPs. CYP1A and CYP2B subfamily proteins are the important key proteins which act as hub proteins in the network. Also, CYP1A and CYP2B subfamily proteins are important drug metabolizing CYP enzymes, where CYP1A2 enzyme is involved in the metabolism of nearly 4% of drugs available in the market (Zuber et al. 2002) and CYP2B6 enzyme is involved in the metabolism of nearly 25% of drugs on the market (Xie et al. 2001) (Page No. 3; Line Nos.56-59).Though quinidine and quinine are reported as inhibitors of CYP2D6 protein, studies carried out by Hutzler et al. (2003), Venhorst et al. (2003), Chauret et al. (1997) and Edmund et al. (2013) used both quinine and quinidine in addressing species selectivity of several CYP enzymes as they are chiral molecules. Also, a study by Ching et al. (2001) reported the differential inhibition of human CYP1A1 and CYP1A2 by quinine and quinidine and Walsky et al. (2006) reported the use of quinine and quinidine for inhibition of CYP2B6 *in vitro*. In this study, we have used combination of molecular modeling, docking and dynamics simulations approaches to address species selectivity of CYP1A and CYP2B subfamily enzymes using quinine and quinidine(Page No. 4, Line Nos.81-83). Recently, a study by Kesharwaniet al. (2016) reported substrate specificity of CYP1A1, CYP1A2 proteins using molecular docking and dynamics simulations and concluded that specificity of a particular substrate depends upon the type of the active site residues, and several other macromolecular features though they are highly conserved at the sequence level (Page No. 3 and Line Nos.70-73).

With these modifications and improvements, we hope that the revised manuscript may be acceptable for publication.

Reply to the reviewer's comments on our manuscript (ID: CBBI-D-16-00083) entitled "Analysis of species selectivity of human, mouse and rat Cytochrome P450 1A and 2B subfamily enzymes using molecular modeling, docking and dynamics simulations"

Response to the Comments of Reviewer 4

1. *Table T1: Providing a sequence alignment would help, especially including the active sites discussed in this manuscript.*

Reply:

As suggested by the reviewer, sequence alignment of human, mouse and rat CYP1A and CYP2B subfamily proteins was performed and results are given as Supplementary Figures S4, S5 and S6. The active site residues that are discussed in the manuscript is highlighted in the figures and explained in 'Results and discussion' section (Page no. 19, line no. 439-447).

2. *The last paragraph of Introduction is too long. Much of its content should be in the Methods or Results sections.*

Reply:

The last paragraph of the introduction section is shortened as suggested by the reviewer (Page no.5, line nos. 103-117).

3. *I suggest reorganizing the results: instead of grouping the results according to docking methods, I think Docking, IFD, and MD of the same protein/ligand should be grouped together.*

Reply:

As suggested by the reviewer, the results are grouped according to CYP subfamily in the 'Results and discussion' section (Page nos.11-19, line nos. 254-437).

With these modifications and improvements, we hope that the revised manuscript may be acceptable for publication.

Reply to the editor's comments on our manuscript (ID: CBBI-D-16-00083) entitled "Analysis of species selectivity of human, mouse and rat Cytochrome P450 1A and 2B subfamily enzymes using molecular modeling, docking and dynamics simulations"

Responseto the Comments of the editor

1. *Abstract needs to be reframed*

   Reply:

   As suggested by the editor, the abstract of the MS is completely reframed (Page no.2, line nos. 32-46).

2. *Author has used two different patterns for reference style in introduction part*

Reply:

We have modified the reference style to uniformity throughout the manuscript according to the journal format.

3. *Author has covered part of the experimental and R & D in the Introduction, infact there is repetition of the text. Authors could have given good significant background of the work.*

Reply:

We have modified/shortened the 'introduction' section, especially the last paragraph which appeared like a methodological and with repeated texts (Page no.5, line nos. 103-117).

4. *To make the manuscript strong and impactful, authors could have covered the rationale to select and use quinine and quinidine as ligands???*

Reply:

CYP1A subfamily proteins are highly conserved at the sequence level, but not at the active site regions. Recently, a study conducted by Kesharwaniet al. (2016) reported that substrate specificity of CYP1A1 and CYP1A2proteins using molecular docking and dynamics simulations. They concluded that the role of active site residues and other macromolecular features influences on substrate selectivity (Page No. 3, Line Nos. 70-73). Quinine and quinidine are the chiral molecules which exhibit differences only in their 3D structure. Though they are well known inhibitors of CYP2D6 protein, they have beenextensively used to study species selectivity of various CYP enzymes by Hutzler et al. (2003), Venhorst et al. (2003), Chauret et al. (1997) and Edmund et al. (2013) as explained in Page No. 4, line nos. 79-102. Ching et al. (2001) reported the differential inhibition of human

CYP1A1 and CYP1A2 by quinine and quinidine and Walsky et al. (2006) reported the use of quinine and quinidine for inhibition of CYP2B6 *in vitro*(Page No. 4, Line Nos. 81-83).Thus,the chiral molecules quinine and quinidine were used as ligands in addressing species selectivity of CYP1A and CYP2B subfamily enzymes using the combination of molecular modeling, docking and dynamics simulations.

5. *Author mentioned the version of Prime (Schrodinger) in the introduction part, and missed to mention in experimental section, and it was an equally quite important to mention in experimental methods.*

Reply:

We have used Prime module of version 3.2 and this is included in 'Materials and methods' section as suggested by the editor (Page no.6, line no. 133).

6. *Standard formatting as per journal guidelines is missing. Use of the English language especially in results and discussion is not in commensuration with the standard of the journal. There are small grammatical mistakes throughout the MS*

Reply:

We have thoroughly revised and formatted the manuscript. The language and grammatical mistakes were also corrected in the revised manuscript.

With these modifications and improvements, we believe that the manuscript has been improved to the best. We request that the revised manuscript may be accepted for publication.

Analysis of species selectivity of human, mouse and rat Cytochrome P450 1A and 2B subfamily enzymes using molecular modeling, docking and dynamics simulations

Bagavathy Shanmugam Karthikeyan[a,b], Suvaiyarasan Suvaithenamudhan[a], Mohammad Abdulkader Akbarsha[b] and Subbiah Parthasarathy[a*]

*Corresponding author

[a]Department of Bioinformatics, School of Life Sciences, Bharathidasan University, Tiruchirappalli – 620 024, Tamil Nadu, India
Email: bdupartha@gmail.com, partha@bdu.ac.in
Phone: +91 431 2407071
Fax: +91 431 2407045
Mobile: + 91 9443533095

[b]Mahatma Gandhi-Doerenkamp Center (MGDC) for Alternatives to Use of Animals in Life Science Education, Bharathidasan University, Tiruchirappalli – 620 024, Tamil Nadu, India
Email: mgdcaua@yahoo.in, akbarbdu@gmail.com
Phone: +91 431 2407117
Mobile: +91 9790995854

Analysis of species selectivity of human, mouse and rat Cytochrome P450 1A and 2B subfamily enzymes using molecular modeling, docking and dynamics simulations

Abstract

Cytochrome P450 (CYP) 1A and 2B subfamily enzymes are the important drug metabolizing enzymes that are highly conserved across several species. Though these enzymes are highly conserved at the sequence level, their structural and macromolecular features govern species and substrate-selectivity. Therefore, species selectivity of CYP1A and CYP2B subfamily proteins across human, mouse and rat was analyzed using molecular modeling, docking and dynamics simulations. The chiral molecules quinine and quinidine were used as ligands. The three-dimensional structures for 17 proteins of CYP1A and CYP2B subfamilies of mouse and rat were predicted through homology modeling, using the available structures of CYP1A and CYP2B proteins of human as template. Molecular docking and dynamics simulations of CYP1A subfamily proteins with quinine and quinidine reveal the existence of species-selectivity across human, mouse and rat. In case of CYP2B subfamily proteins, the absence of the role of chirality between quinine and quinidine was observed when these ligands form complexes with CYP2B subfamily proteins of human, mouse and rat. Our findings reveal the role of active site amino acid residues of CYP1A and CYP2B subfamily proteins and provide insights about species-selectivity of these enzymes across human, mouse and rat.

Keywords: Cytochrome P450, Species selectivity, Homology modeling, Glide docking, Schrödinger software suite, Molecular dynamics simulations.

Introduction

Cytochrome P450 (CYP) are the major enzymes responsible for numerous oxidative reactions that play prominent roles in metabolism of xenobiotics and drugs. These enzymes are derived from a single ancestral gene about 1.36 million years ago [1]. CYP1A and CYP2B subfamily enzymes are important among CYP enzymes, where CYP1A2 enzyme is involved in the metabolism of nearly 4% of drugs available in the market [2] and CYP2B6 enzyme is involved in the metabolism of nearly 25% of drugs in the market [3]. In spite of all members of the CYP protein family possessing conserved regions of amino acid sequences, there are evidences for species differences and substrate-specificities which may arise due to small differences in their amino acid sequences [4]. The structure-based computational techniques, such as molecular modeling and dynamics simulation studies can help effectively in analyzing the role of structural features in species selectivity. Skopalik et al. and Hritz et al. carried out molecular dynamics simulations of CYP3A4, CYP2A6 and CYP2D6 proteins and explained their substrate preferences [5,6]. These authors found that residues on the active site and their topology critically regulate these enzymes. Rosales-Hernández et al. studied the metabolism of CYP1A1 and CYP2B1 proteins using different aryl derivatives and it was found that the results of *in silico* approaches such as docking and dynamics simulations were in good agreement with experimental results [7]. Recently, Kesharwani et al. reported substrate-specificity of CYP1A1, CYP1A2 and CYP1B1 proteins using molecular docking and dynamics simulations and concluded that the role of active site residues and its type is important in studying substrate selectivity [8]. Mukherjee et al. developed a server to predict the sites of metabolism (SOM) of CYPs (CYP1A2, CYP2C9, CYP2C19 and CYP3A4). The methodology comprises of docking, binding energy calculations, analyses of docked complexes and molecular orbitals of ligand molecules. Their methodology thus achieved success rate of 87% in identifying the

experimentally known SOM of the xenobiotics [9]. Thus structure-based computational approaches are gaining momentum in the recent times in predicting the metabolism of drugs.

From the previous studies carried out by Hutzler et al., Venhorst et al., Chauret et al. and Edmund et al., it was observed that the chiral molecules quinine and quinidine could be used as ligands in studying species selectivity of different CYP proteins [10-13]. Ching et al. reported the differential inhibition of human CYP1A1 and CYP1A2 by quinine and quinidine and Walsky et al. reported the use of quinine and quinidine for inhibition of CYP2B6 *in vitro*[14,15]. Apart from use in studying species-selectivity of CYP enzymes, quinine and quinidine are known for their use as anti-malarial drugs[16].Other than CYP proteins, these compounds interact with human ether-a-go-go-related gene (HERG), 5-HT(3), GABA(A) receptors, GST M1 and GST P1 proteins [17-20]. Hutzler et al. used quinine and quinidine as ligands and demonstrated that these structural analogues are involved in inhibition of CYP2D6 activity [10]. Chauret et al. used quinine and quinidine, in addition to several other compounds *in vitro*, to study the metabolic activity of CYP enzymes across different species [12]. Edmund et al. studied species-selectivity of human and rat CYP2D subfamily proteins using these chiral molecules through molecular interaction approaches [13]. Venhorst et al. calculated experimental IC_{50} values for quinine and quinidine in rat and human CYP2D subfamily enzymes and demonstrated ligand binding specificities [11].Edmund et al. demonstrated inhibition of human CYP2D6 protein by quinine and quinidine by calculating experimental and predicted K_i (inhibitory constant) values [13]. These authors found that rat and human CYP2D subfamily proteins overall share high sequence identity, but at the active site regions they share very less sequence identity. From the study carried out by Venhorst et al. it was found that quinidine interacts with Phe120, Asp301 and Ser304 of human-CYP2D6 [11], whereas Edmund et al. found that quinidine interacts only with Met304 (replaces Ser304) of rat-CYP2D2 [13].Quinine interacts with Asp301 of human-

CYP2D6, and with rat-CYP2D2 at Asp216, Thr217 and Met304 and, thus, showed its species selective nature[11].

In this paper, analysis of species selectivity of CYP1A and CYP2B subfamily proteins across human, mouse and rat using molecular modeling, docking and dynamics simulations was carried out and it is first of its kind to the best of our knowledge. Three-dimensional structures of CYP1A and CYP2B subfamily proteins of mouse and rat are not available, until now. Hence, homology modeling approach was adopted to predict the three dimensional structures of seventeen CYP1A and CYP2B subfamily proteins of mouse and rat. The chiral molecules quinine and quinidine were used as ligands to dock with CYP1A and CYP2B subfamily proteins of human, mouse and rat using Glide and Induced Fit Docking (IFD) modules of Schrödinger software. The docked complexes were further subjected to molecular dynamics simulations using GROMACS package version 4.5.5 [21] to reveal the stability and dynamic behavior of CYP1A and CYP2B subfamily proteins of human, mouse and rat with quinine and quinidine. The outcome of this study gives insights about species-selectivity of CYP1A and CYP2B subfamily proteins across human, mouse and rat, infers on the role of structural features, the influence of active site residues in forming hydrogen bonds and further helps in better understanding the drug metabolism and cross species extrapolations.

Materials and methods

In the present study, CYP1A and CYP2B subfamily enzymes of human (*Homo sapiens*), mouse (*Mus musculus*) and rat (*Rattus norvegicus*) were used. In particular, CYP2B subfamily proteins CYP2B6 and CYP2B7 of human; CYP2B1, CYP2B2, CYP2B3, CYP2B12, CYP2B13, CYP2B15, and CYP2B21 of rat; and CYP2B9, CYP2B10, CYP2B13, CYP2B19, and CYP2B23 of mouse were used. CYP1A subfamily proteins, CYP1A1 and CYP1A2 of human, mouse and rat were considered and tabulated (Supplementary Table T1). Experimentally

determined three-dimensional structures are available only for three proteins, CYP1A1, CYP1A2 and CYP2B6 of human and their structures with PDB IDs: 418V, 2HI4 and 3QOA, respectively, were obtained from Protein Data Bank PDB (http://www.rcsb.org/pdb) [22]. For the remaining 17 CYP1A and CYP2B subfamily proteins, homology modeling approach was adopted to predict the three-dimensional structures. Supplementary Table T1 furnishes the updated list of CYP proteins published by Karthikeyan et al.[23].

Homology modeling using Prime, Schrödinger

Full length protein sequences of the above mentioned CYPs of human, mouse and rat were retrieved from UniProt database [24]. Prime module (version 3.2) of Schrödinger, (Schrödinger, LLC, New York, NY, 2013) was used to predict three-dimensional structures of i) human-CYP2B7, ii) mouse CYP2B9, CYP2B10, CYP2B13, CYP2B19 and CYP2B23, iii) rat CYP2B1, CYP2B2, CYP2B3, CYP2B12, CYP2B13, CYP2B15 and CYP2B21, and iv) mouse and rat CYP1A1, CYP1A2 proteins. Totally, 17 three-dimensional protein structures were generated. Template structures were selected based on the percentage identity with mouse and rat proteins. The resolution of the experimental structure was also considered in the template selection. Human CYP2B6 enzyme with PDB ID 3QOA [25] was used as template to generate 13 models of CYP2B subfamily proteins of mouse and rat. Similarly, the two human CYP1A1 and CYP1A2 with PDB Ids 418V and 2HI4,respectively,were used as templates for predicting three-dimensional structures of CYP1A1 and CYP1A2 of mouse and rat proteins(Supplementary Tables T2 and T3).

Protein preparation and optimization for docking

Before docking, the modeled proteins were prepared and optimized using protein preparation wizard of Schrödinger (Schrödinger, LLC, New York, NY, 2013). All the predicted 17 three dimensional protein structures were validated by using the Ramachandran plot available

within protein preparation wizard, where the residues that lie in allowed regions and outliers were analyzed (Supplementary Table T4). Hydrogen was added and water molecules and ligands bound to the complex were removed. Finally, proteins were minimized using OPLS-2005 force field.

Active site studies

Sitemap module available within Schrödinger (Schrödinger, LLC, New York, NY, 2013) was used to generate a series of protein-ligand binding sites for the template structures of human as well as modeled protein structures. Top-ranked potential active sites of the receptor were selected based on the best site score, size and volume of the active site (Supplementary Tables T5-T7). For cross validation of the predicted active site residues, Sitehound server was used to predict the active site residues [26], as well as active sites defined in the literature were also considered to make consensus prediction [25,27,28].

In addition to predicting active sites, the conserved residues among the CYP1A and CYP2B subfamily proteins across human, mouse and rat were investigated. ClustalX version 2.1 was used for multiple sequence alignment of protein sequences.

Ligand preparation and optimization

The structures of the epimers quinine (CID: 8549) and quinidine (CID: 441074) considered for the study, as retrieved from NCBI-PubChem compound database (http://pubchem.ncbi.nlm.nih.gov/), is shown in Figure 1. The ligands were prepared and optimized using ligand preparation workflow (Schrödinger, LLC, New York, NY, 2013). After neutralizing charged groups, ionization and tautomeric states were generated using Epik module. One low energy conformer per ligand was generated and used further.

Molecular docking

Glide docking (Schrödinger, LLC, New York, NY, 2013) was carried out after constructing the grid by selecting centroid of active site (Supplementary Tables T5-T7) residues at the radius of 3Å that resulted from the sitemap. The default parameters were selected by keeping ligand flexible with docking calculations set as XP extra-precision. XP docking performs more extensive sampling than HTVS and SP docking and eliminates false positives obtained from SP docking (Schrödinger, LLC, New York, NY, 2013). Formation of hydrogen bond between the ligand and the residues of the active site and its length along with Glide XP score were recorded. Binding energy calculation was effected using EMBrAcE minimization procedure (Schrödinger, LLC, New York, NY, 2013). The method followed here is distance-dependent electrostatic treatment with OPLS 2005 force field set to 1000 iterations at the maximum.

CYP1A (CYP1A1 human, mouse and rat; CYP1A2 human, mouse and rat) and CYP2B (human-CYP2B6, human-CYP2B7, mouse-CYP2B9 and rat-CYP2B1) subfamily proteins were selected for IFD. The active site residues at around 3Å at the centroid of receptors were selected for the IFD with Glide redocking calculations (Schrödinger, LLC, New York, NY, 2013) set as XP extra precision with a maximum of 10 poses generated [29]. The best low energy conformer was selected among ten poses for binding energy calculation. The formation of hydrogen bond and its length were also considered in selecting the best pose. EMBrAcE minimization procedure (Schrödinger, LLC, New York, NY, 2013) was used for binding energy calculation and the method followed here is the same as mentioned vide supra. To validate the results of induced fit docking, the Root Mean Square Deviation (RMSD) between the reported x-ray structures [25,27,28] and the quinine-protein and quinidine-protein complexes was calculated.

Molecular dynamics (MD) simulations of CYP1A and CYP2B subfamilies

Molecular dynamics simulations: Protein in water

The MD simulations were carried out for the proteins of CYP1A [CYP1A1 (human, mouse and rat)], [CYP1A2 (human, mouse and rat)] and CYP2B subfamilies (Rat-CYP2B1, human-CYP2B6, human-CYP2B7 and mouse-CYP2B9). GROMACS 4.5.5 package installed in Biolinux platform was used for the complete study [21]. Energy minimization of the proteins was performed using steepest descent algorithm using GROMACS96 43a1 force field [30]. An aqueous environment was provided to the protein with a cubic box set at a size of 1.0 nm from the edge and at least 2.0 nm between any two periodic images of a protein. The system was neutralized by adding appropriate counter-ions. For parameterization of heme, GROMACS96 43a1 force field [30] was used and topology of heme was included to the topology of protein for MD simulations. Equilibration was carried out in two phases. NVT ensemble with constant number of particles, volume and temperature was employed at 100 ps and NPT ensemble with a constant pressure was employed at 100 ps using the Parrinello-Rahman barostat algorithm [31]. Final production MD was performed at 10 ns after equilibration of the system at constant temperature and pressure. The trajectories were recorded for further analyses.

Molecular dynamics simulations: Protein-ligand complexes

The MD simulation studies were carried out for the protein-ligand (quinine and quinidine) complexes of CYP1A [CYP1A1 (human, mouse and rat), CYP1A2 (human, mouse and rat)] and CYP2B subfamilies (Rat-CYP2B1, human-CYP2B6, human-CYP2B7, mouse-CYP2B9). GROMACS 4.5.5 package with GROMACS96 43a1 force field was used for the dynamics simulations [21,30]. The topology parameters of the proteins were built by GROMACS program. The topology parameter of ligands (quinine and quinidine) was developed using PRODRG server [32]. The protein was solvated with explicit water molecules in a defined

simple point charge (SPC) as a water model in a cubic cell. The system was neutralized by adding proper counter-ions. Energy minimization of the system was performed with the help of 50000 steps using steepest descent method and at a tolerance of 1000 kJ mol^{-1} nm^{-1}. Restraining of ligands and treatment of temperature coupling groups were made before equilibration process. The total system was equilibrated involving two phases, one with constant number of particles, volume and temperature employed using NVT at 100 ps and another with constant pressure using NPT ensemble at 100 ps. The final MD was performed at 10 ns after equilibration. The trajectories of protein-ligand complexes were recorded for further analyses.

Trajectory analysis

The trajectory files that resulted from MD simulations were analyzed using GROMACS utilities such as g_rms (RMSD analysis) and g_hbond (Hydrogen bond analysis). GNU plot version 4.4 was used to plot the graphs [33].

Hardware and software

The docking was carried out on a personal computer with 64 bit processor and 2 GB DDR RAM using Centos-Linux operating system installed with Maestro, Schrödinger (version 9.6) and the molecular dynamics work was carried out in GROMACS package version 4.5.5 on a 2GB DDR RAM using Biolinux platform.

Results and discussion

Modeling and validation of predicted models

Supplementary Table 1 provides information about the updated list of drug-metabolizing CYP enzymes published by Karthikeyan et al. [23]. From Supplementary Table T2, it was inferred that human-CYP1A1 and human-CYP1A2 proteins show above 70% sequence identity with mouse and rat CYP1A1 and CYP1A2 proteins. Human-CYP2B7 protein shows 92% sequence identity with human-CYP2B6 protein (Supplementary Table T3). The three-

dimensional structures of mouse and rat-CYP2B subfamily proteins were predicted by choosing human-CYP2B6 as template structure with PDB-ID: 3QOA. In addition, Supplementary Tables T2 and T3 list percentage of sequence identity of all the 17 CYP protein sequences with the chosen template structures, along with their sequence length and PDB ID. Totally, 17 three-dimensional protein structures of interest were generated.

The predicted structures validated using Ramachandran plot show that more than 95% of the amino acid residues in the predicted proteins were in the favoured region and less than 0.5% of the amino acid residues were in the outlier region (Supplementary Figs. S1-S3 and Supplementary Table T4).

CYP1A subfamily proteins:

Glide docking of ligands with human, mouse and rat CYP enzymes CYP1A1 and CYP1A2

The results of Glide docking are tabulated in Table 1, which reveal the overall selective nature of ligands in binding with residues. Though the CYP1A subfamily (CYP1A1 and CYP1A2) proteins have more conserved residues on the active site [34] than other subfamily proteins, differences in ligand binding with these proteins were observed in our investigation (Table 1).

When looked at the subfamily level, species difference, selective binding, and conserved binding were observed (Table 1). Gly316 and Gly320 are the residues conserved among CYP1A1 of human, mouse and rat with hydrogen bonds of length ranging from 1.70Å to 2.10Å (Table 1). From Table 1, it is observed that quinine interacts with rat-CYP1A1 with the highest glide XP score -9.99, whereas quinidine interacts with human-CYP1A1 with the highest glide XP score -9.66.

The ligands quinine and quinidine bind with human-CYP1A2 protein at Asp313 and Thr124, respectively (Supplementary Figs. S7 (a) and S7 (d)). Quinine and quinidine do not

show any hydrogen bond formation with mouse-CYP1A2 protein (Supplementary Figs. S7 (b) and S7 (e); Table 1). In the case of rat-CYP1A2 protein, quinine forms hydrogen bond at Cys456 but quinidine does not show any hydrogen bond formation (Supplementary Figs. S7 (c) and S7 (f); Table 1). From Table 1, it is observed that among CYP1A2 protein, quinine has the highest glide XP score of -8.74 with mouse-CYP1A2 and quinidine interacts with human-CYP1A2 with glide XP score of -8.86. The difference in binding of ligands between CYP1A1 and CYP1A2 is observed which may be because of the difference in the active site cavity and topology as suggested by Porubsky et al. [35].

Thus, overall, selective binding of ligands are observed. At the subfamily level conserved nature of the ligand binding is observed for CYP1A1 protein across species. Species difference and selective binding of ligand are observed for CYP1A2 proteins. IFD was carried out to further rationalize this selective nature of ligands.

Induced fit docking and binding energy calculation of ligands with human, mouse and rat CYP enzymes CYP1A1 and CYP1A2

Induced fit docking was carried out to further rationalize the species selectivity and ligand selectivity. As observed in glide docking results, overall selective binding of ligands was observed (Table 1). But, at the subfamily level, it was observed that ligands exhibit more selective binding and less conserved binding.

CYP1A1 of human, mouse and rat

The ligands quinine and quinidine exhibit substrate selective binding on human-CYP1A1 protein. For example, quinine forms hydrogen bond with Gly316 residue but quinidine binds at Ser122 and Ala317 of human-CYP1A1 protein (Fig. 2(a), 2(b) and Table 2). Quinine and quinidine form hydrogen bond with mouse-CYP1A1 protein at Gly320 (Table 2). With rat-CYP1A1 protein, quinine and quinidine form hydrogen bonds at Thr501 and Thr126,

respectively (Table 2). Selective binding of ligands among CYP1A1 of human, mouse and rat was observed. Quinine and quinidine show effective binding with human-CYP1A1 protein with binding energy values -56.6 Kcal/mol and -56.3 Kcal/mol, respectively, compared to mouse and rat CYP1A1 proteins (Table 2). As observed from Table 2, the RMSD between the reported x-ray structure and CYP1A1-ligand (quinine and quinidine) complex is less than 0.1 Å.

CYP1A2 of human, mouse and rat

From Table 2, it is observed that the ligands exhibit species-selective binding as well as substrate-selective binding with human, mouse and rat CYP1A2 protein. Quinine binds with human-CYP1A2 at Gly316 but quinidine binds with the same protein at Thr321 and Thr498. In mouse-CYP1A2, quinine forms hydrogen bonds at Thr123 and Gly314, but quinidine forms hydrogen bond at Thr117 (Table 2). Quinine binds with rat-CYP1A2 at Arg107, Arg454 and Cys456, but quinidine binds at Leu448 and Cys456 (Table 2). From the binding energy results, it is noted that quinine binds effectively with mouse-CYP1A2 with binding energy -138.5 Kcal/mol, and quinidine binds effectively with human-CYP1A2 with binding energy -67.9Kcal/mol (Table 2). As observed from Table 2, the RMSD between the reported x-ray structure and CYP1A2-ligand (quinine and quinidine) complex is less than 0.2 Å.

MD simulations of CYP1A1 protein-ligand complex

MD simulations for a period of 10 ns were seeded in GROMACS 4.5.5 to compare the structural behavior and flexibility of CYP1A and CYP2B subfamily proteins among human, mouse and rat with ligands (quinine and quinidine). The RMSD and hydrogen bond analysis were performed using the g_rms, and g_hbond GROMACS utilities, respectively, and trajectories were plotted. The RMSD of protein backbone was calculated for both the native and complex forms to reveal the stability of the proteins with these ligands and also to compare it among different species. The ligands quinine and quinidine are represented as a line graph as a

ligand-positional RMSD throughout the analysis. Intermolecular hydrogen bonding plays a major role in the stable binding of the ligand with the protein. Time-dependant hydrogen bonding was monitored to understand the nature of binding of ligands with the protein and also to compare it across the species. The average RMSD and hydrogen bonds formed from the simulations are presented in Table 5.

Backbone RMSD analysis was performed using g_rms GROMACS utilities. The trajectories were analyzed for a period of 10 ns. There was no significant difference among protein backbones of CYP1A1 of human, mouse and rat till 2.5 ns. A drift occurred after 2.5 ns at 0.2 nm (Supplementary Fig. S8 (a)). The ligand-positional RMSD of quinine and quinidine were generated and analyzed to ensure the stability of ligands with proteins and to facilitate the comparison among different species (Fig. 4). From Table 5 and Figure 4, it could be inferred that notable difference exists among human, mouse and rat RMSD from the native protein. Proteins with quinine complex showed a difference in average RMSD of approximately 0.03 to 0.05 nm from its native protein structure. Proteins with quinidine complex showed small difference of about 0.01 to 0.02 nm from its native protein structure (Table 5). From Figure 4, it is inferred that binding of both the ligands with CYP1A1 protein influences the stability of the protein backbone except when quinidine complexes with mouse-CYP1A1 protein (Fig. 4(d)).

The time-dependant hydrogen bonding pattern throughout 10 ns was obtained for human, mouse and rat proteins with the ligands quinine and quinidine. From Table 5 and from Figure 5 it is clear that the average number of hydrogen bonds formed by quinidine with human, mouse and rat protein is more compared to quinine complex with human, mouse and rat-CYP1A1 protein. Quinidine formed an average of three hydrogen bonds with human, mouse and rat-CYP1A1 whereas quinine formed only an average of one hydrogen bond (Table 5).

From the RMSD and hydrogen bond analysis, it was observed that quinidine binds efficiently with CYP1A1 protein compared to quinine. Human-CYP1A1 shows less adaptability to these ligands. There exists less significant difference in the structural behaviour between quinine and quinidine when binding with rat-CYP1A1 protein and more significant difference in structural behaviour between quinine and quinidine when binding with human and mouse-CYP1A1 protein.

MD simulations of CYP1A2 protein-ligand complex

There was no significant difference among the stability of the protein backbones of CYP1A2 of human, mouse and rat till 6 ns and a drift occurred after 6 ns at 0.25 nm (Supplementary Fig. S8(b)). From Table 5 it is inferred that the difference between RMSD of native protein and protein complexes of human, mouse and rat is ~0.01 to 0.02 nm. It is observed from Figure 6 that the stability of protein backbone of human-CYP1A2 changes in accordance with ligand binding especially after 6 ns (Fig. 6(a), (b)). No noticeable changes were observed for stability of protein backbone of mouse-CYP1A2 with ligands as observed from Figure 6(c), (d). But significant differences existed in RMSD of rat-CYP1A2 protein backbone in native and ligand-bound states (Fig. 6(e), (f)).

From Table 5 and Figure 7 it is inferred that quinine forms more average number of hydrogen bonds (3) with human-CYP1A2 compared to mouse and rat. Quinidine forms more average number of hydrogen bonds (3) with mouse-CYP1A2 compared to human and rat. As observed from the trajectories quinine forms stable hydrogen bonds with rat-CYP1A2 only after 7 ns (Fig. 7(e)) and quinidine forms stable hydrogen bonds with mouse-CYP1A2 only after 4 ns (Fig. 7(d)).

From the above analysis it is inferred that quinine and quinidine bind efficiently with human-CYP1A2, and significant difference exists among human, mouse and rat CYP1A2 when binding with quinine and quinidine.

CYP2B subfamily proteins:

Glide docking of ligands with human, mouse and rat CYP2B enzymes

The ligand does not form any hydrogen bond with human CYP2B6 and CYP2B7 proteins except one at Thr-302, but it forms hydrogen bonds with other mouse and rat CYP2B proteins as given in Table 3.

The ligand quinine binds with mouse-CYP2B9 at Pro428, Leu362 and with mouse-CYP2B19 at His398, respectively. Quinidine forms hydrogen bonds only with mouse-CYP2B9 at Pro428 and these ligands do not bind with other mouse-CYP2B proteins (Table 3).

Quinine forms hydrogen bonds with rat-CYP2B1 at Leu362, with rat-CYP2B3 at Pro428, with rat-CYP2B15 at Thr299 and with rat-CYP2B21 at Pro428 and Leu362. Quinine did not show any appreciable hydrogen bond formation with the rest of the rat-CYP2B proteins. Quinidine forms hydrogen bonds with rat-CYP2B1, rat-CYP2B3 and rat-CYP2B15 at Pro428, Ser363 and Thr299, respectively. With the rest of the proteins, quinidine does not show any appreciable hydrogen bond formation. As reported by Zhang & Yang, similar substrate-specificity was not observed even between rat-CYP2B1 and rat-CYP2B2 [36], i.e., with rat-CYP2B1, quinine and quinidine form hydrogen bond at Leu362 and Pro428, respectively, but the ligands do not form any hydrogen bond with rat-CYP2B2 protein.

From Table 3, it is observed that among CYP2B proteins, the highest glide XP score of -9.43 was found for mouse-CYP2B23 when docked with quinine, and the highest XP glide score of -9.87 was found for rat-CYP2B13 when docked with quinidine. The least glide XP score of -

2.84 and -2.55 was observed for mouse-CYP2B19 when binding with quinine and quinidine respectively (Table 3).

There is no relationship between chirality and bond formation. For example, when quinine binds with mouse-CYP2B9 protein, quinidine forms hydrogen bond but with difference in binding with the active site residues (Table 3). Thus, the role of chirality in ligand binding was not observed for CYP2B subfamily proteins.

Induced fit docking and binding energy calculations of ligands with CYP2B enzymes of human, mouse and rat

From the glide docking, species selectivity for CYP2B enzymes was analyzed. IFD was carried out further in order to figure out the species selectivity of CYP2B subfamily proteins. It is to be noted that both the ligands, quinine and quinidine, do not bind to human-CYP2B6 (Fig. 2(c), 3(a) and Table 4) but form hydrogen bonds with human-CYP2B7 (Fig. 2(d), 3(b) and Table 4). It should be noted that the ligands do not exhibit substrate selective binding among CYP2B enzymes.

As shown in Table 4, while quinine binds with human-CYP2B7 at Ser210 and Phe297, quinidine also binds with the same protein at Ser210, Phe297 and Thr302 with a very small difference in their hydrogen bond lengths. Thus, there is no selectivity in the binding of quinine and quinidine with human-CYP2B6 and CYP2B7 proteins, and the trend is similar for mouse-CYP2B9 and rat-CYP2B1 proteins.

Quinine binds with mouse-CYP2B9 at Pro428 and Val363, while quinidine binds with the same protein at Pro428 and Gln357 (Table 4). With rat-CYP2B1 protein, quinine binds at Pro428 and Gln357 whereas quinidine binds at Arg98, Pro428 and Cys436 (Fig. 3(c), 3(d)). From the binding energy as well as hydrogen bond formation it is observed that quinidine binds effectively with mouse-CYP2B9 compared to the other proteins (Table 4).As observed from

Table 4, the RMSD between the reported x-ray structure and CYP2B subfamily protein-ligand (quinine and quinidine) complex is less than 0.15 Å.

The chiral nature of ligands does not have any role in ligand binding with CYP2B subfamily of enzymes.

MD simulations of CYP2B subfamily proteins-ligand complex

The backbone RMSD of proteins before ligand binding shows the stability of the modeled proteins. At around 1.8 ns there is a drift in the stability of the protein backbone (Supplementary Fig. S8(c)). Quinidine impacts more structural changes than quinine in the protein backbone among human, mouse and rat proteins as observed from the ligand positional trajectories (Fig. 8). Overall, there is no significant difference observed in backbone RMSD of proteins in native and complex forms (Table 5; Fig. 8) except for mouse-CYP2B9 (Fig. 8(e), (f)).

The hydrogen bonding pattern of ligands with protein was recorded throughout the simulation period. Significant difference exists between quinine and quinidine in forming hydrogen bonds only with human-CYP2B6 protein, but with other proteins, like human-CYP2B7, mouse-CYP2B9 and rat-CYP2B1, there was no much difference (Table 5; Fig. 9). The average number of hydrogen bonds formed by quinine with human-CYP2B6 over the period is 2. But with the same protein quinidine forms stable hydrogen bond only after 5 ns (Fig. 9(b)) and the average number of bond formed is 1. From the trajectory analysis it is observed that quinine forms the highest average number of hydrogen bonds with human-CYP2B7 (3) (Table 5; Fig. 9(c)) and quinidine forms the highest average number of hydrogen bonds with mouse-CYP2B9 (3) (Table 5; Fig. 9(f)). Approximately two hydrogen bonds are formed by quinine and quinidine with human-CYP2B7, mouse-CYP2B9 and rat-CYP2B1 (Table 5; Fig. 9).

From RMSD and hydrogen bond trajectory analyses, it is inferred that the ligands quinine and quinidine exhibit no substrate-selective binding (except for human-CYP2B6) and there is no

critical role of chiral nature of ligands is reflected here. This result is in complete agreement with the IFD results.A close approach in binding of ligands in the proximity of heme is observed for CYP1A and CYP2B subfamily proteins suggesting high spin state transition which provides insights about the role of heme in determining the binding efficiency of ligands quinine and quinidine [37, 38].

Binding site analysis

The results of multiple sequence alignment of CYP1A and CYP2B subfamilies across human, mouse and rat shows the conserved amino acid residues. From Supplementary Fig. S4 it is observed that Thr126, Gly320, Ala321, Thr325, Val386, Leu500 and Thr501 are the residues conserved across CYP1A1 protein of human, mouse and rat. Arg108, Gly316, Ile384, Arg454, Cys456, Ile457 and Gly458 are the amino acid residues that are conserved in CYP1A2 protein across human, mouse and rat (Supplementary Fig. S5). Arg98, Ile114, Phe115, Trp121, Arg125, Phe298, Thr303, Pro428, Ser430, Arg434 and Cys437 are the conserved amino acid residues as inferred from multiple sequence alignment of CYP2B subfamily proteins of human, mouse and rat (Supplementary Fig. S6).

Also, the binding sites were analyzed after the ligands bind with the receptor and form hydrogen bonds. It is observed that the following active site residues in CYP1A1 and CYP1A2 proteins (Supplementary Tables T5 and T6) are involved in binding with the ligands, residues Arg107, Ser122, Thr123, Thr124, Thr126, Asp313, Gly314, Gly316, Ala317, Gly320, Thr321, Leu448, Arg454, Cys456, Thr498 and Thr501. When these residues are absent, the ligands quinine and quinidine do not bind to the receptor.

Ser210, Thr299, Phe297, Thr302, Gln357, Leu362, Val363, Ser363, His398, Pro428 and Cys436 are the key residues involved in hydrogen bond formation with CYP2B proteins. Since these residues are absent in the active site of human-CYP2B6, mouse-CYP2B10, mouse-

CYP2B13, mouse-CYP2B23, rat-CYP2B2, rat-CYP2B12, and rat-CYP2B13, quinine and quinidine do not form hydrogen bonds (Supplementary Table T7).

Conclusion

In the present study, we investigated species selectivity of CYP1A and CYP2B subfamily proteins among human, mouse and rat using molecular modelling, docking and dynamics approaches. Our analysis revealed that species differences and selectivity exist among human, mouse and rat CYP1A subfamily proteins when binding with quinine and quinidine. In CYP2B subfamily proteins, species differences and a role for chirality between ligands are not observed.

Three-dimensional structures of CYP1A and 2B subfamily proteins of human, mouse and rat are not available until now and hence, homology modeling approach was used to predict the three dimensional structures of 17 CYP1A and CYP2B subfamily proteins of human, mouse and rat. Glide and induced fit docking of CYP1A and CYP2B subfamily proteins with quinine and quinidine reveals that the binding of ligands is more conserved for CYP1A1 protein. The amino acid residues Ser122, Thr126, Gly316, Ala317, Gly320 and Thr501 are the ones involved in binding of CYP1A1 proteins with ligands. Species difference and role of chirality in ligand selective binding are observed for CYP1A2 protein. For CYP2B subfamily proteins, the ligand quinine binds only with CYP2B9 and CYP2B19 of mouse, and CYP2B1, CYP2B3, CYP2B15 and CYP2B21 of rat. The ligand quinidine binds only with CYP2B7 of human, CYP2B9 of mouse and CYP2B1, CYP2B3 and CYP2B15 of rat. Here, chirality does not seem to play any role in the binding of ligands with CYP2B subfamily proteins.

Molecular dynamics analysis carried using GROMACS revealed the structural behavior of CYP1A and CYP2B subfamily proteins with quinine and quinidine which facilitated cross species comparisons. The results revealed that quinidine binds more efficiently with CYP1A1 protein than quinine. From the RMSD and hydrogen bond analyses, it is observed that the

ligands quinine and quinidine bind efficiently with rat-CYP1A1 protein than the human and mouse CYP1A1 proteins. Quinine and quinidine bind more efficiently with human-CYP1A2 protein than with mouse and rat CYP1A2 proteins. The analysis of stability and intermolecular interactions revealed that there is a significant difference in the structural behaviour among human, mouse and rat CYP1A2 proteins when binding with quinine and quinidine. The stability analysis showed the structural and dynamic behaviour of CYP2B subfamily proteins (human-CYP2B6, human-CYP2B7, mouse-CYP2B9 and rat-CYP2B1) with quinine and quinidine. From RMSD and time-dependant hydrogen bond analyses, it is inferred that quinine and quinidine exhibit no substrate-selective binding (except for human-CYP2B6). There is no significant difference in the structural behaviour among human, mouse and rat CYP2B proteins when quinine and quinidine bind with these proteins, it shows less species- and substrate-selective behaviour when compared to CYP1A subfamily proteins. In conclusion, this study provides insights about the role of active site residues and information about species difference and substrate selectivity of CYP1A and CYP2B subfamily enzymes and reiterates caution in respect of extrapolating data from animal models to humans.

References

1 Lin, JH., & Lu, AY. (1997). Role of pharmacokinetics and metabolism in drug discovery and development. *Pharmacological Reviews*, *49*, 403-449.

2 Zuber, R., Anzenbacherová, E., & Anzenbacher, P. (2002). Cytochromes P450 and experimental models of drug metabolism. *Journal of Cellular and Molecular Medicine*, *6*(2), 189-198.

3 Xie, W., & Evans, RM. (2001). Orphan nuclear receptors: the exotics of xenobiotics. *Journal of Biological Chemistry*, *276*, 37739-37742.

4 Martignoni, M., Groothuis, GM., & de, Kanter, R. (2006). Species differences between mouse, rat, dog, monkey and human CYP-mediated drug metabolism, inhibition and induction. *Expert Opinion on Drug Metabolism and Toxicology*, *2*, 875-894.

5 Skopalik, J., Anzenbacher, P., & Otyepka, M. (2008). Flexibility of human cytochromes P450: molecular dynamics reveals differences between CYPs 3A4, 2C9, and 2A6, which correlate with their substrate preferences. *Journal of Physical Chemistry B*, *112*, 8165-8173.

6 Hritz, J., de, Ruiter, A., & Oostenbrink, C. (2008). Impact of plasticity and flexibility on docking results for cytochrome P450 2D6: a combined approach of molecular dynamics and ligand docking. *Journal of Medicinal Chemistry*, *51*, 7469-7477.

7 Rosales- Hernández, MC., Mendieta-Wejebem, JE., Trujillo-Ferrara, JG., &Correa-Basurto, J. (2010). Homology modeling and molecular dynamics of CYP1A1 and CYP281 to explore the metabolism of aryl derivatives by docking and experimental assays. *European Journal of Medicinal Chemistry*, *45*, 4845-4855.

8 Kesharwani, SS., Nandekar, PP., Pragyan, P., Rathod, V., & Sangamwar, AT. (2016). Characterization of differences in substrate specificity among CYP1A1, CYP1A2 and CYP1B1: an integrated approach employing molecular docking and molecular dynamics simulations. *Journal of Molecular Recognition*. *29*(8), 370-390

9 Mukherjee, G., Lal, Gupta, P., & Jayaram, B. (2015). Predicting the binding modes and sites of metabolism of xenobiotics. *Molecular Biosystems*,*11*(7), 1914-1924.

10 Hutzler, JM., Walker,GS., & Wienkers, LC. (2003). Inhibition of cytochrome P450 2D6: structure-activity studies using a series of quinidine and quinine analogues. *Chemical Researchin Toxicology*, *16*(4), 450-459.

11 Venhorst,J., Ter, Laak, AM., Commandeur, JN., Funae, Y., Hiroi, T., & Vermeulen, NP. (2003). Homology modeling of rat and human cytochrome P450 2D (CYP2D) isoforms and

computational rationalization of experimental ligand-binding specificities. *Journal of Medicinal Chemistry*, *2*, 46 (1), 74-86

12 Chauret, N., Gauthier, A., Martin, J., & Nicoll-Griffith, DA. (1997). In vitro comparison of cytochrome P450-mediated metabolic activities in human, dog, cat, and horse. *Drug Metabolism and Disposition*, *25*(10), 1130-1136.

13 Edmund, GH., Lewis, DF., & Howlin, BJ. (2013). Modelling species selectivity in rat and human cytochrome P450 2D enzymes. *PLoS One*, *8*, e63335.

14 Ching, MS., Blake, CL., Malek, NA., Angus, PW., & Ghabrial, H. (2001). Differential inhibition of human CYP1A1 and CYP1A2 by quinidine and quinine. *Xenobiotica*. *31*(11), 757-767.

15 Walsky, RL., Astuccio, AV., & Obach, RS. (2006). Evaluation of 227 drugs for in vitro inhibition of cytochrome P450 2B6. *The Journal of Clinical Pharmacology*, *46*(12), 1426-1438

16 Trampuz, A., Jereb, M., Muzlovic, I., & Prabhu, RM. (2003). Clinical review: Severe malaria. *Critical Care*. *7*(4), 315-323

17 Thompson, AJ., Lochner, M., & Lummis, SC. (2007). The antimalarial drugs quinine, chloroquine and mefloquine are antagonists at 5-HT3 receptors. *British Journal of Pharmacology*. *151*(5), 666-677

18 Thompson, AJ., & Lummis, SC. (2008). Antimalarial drugs inhibit human 5-HT(3) and GABA(A) but not GABA(C) receptors. *British Journal of Pharmacology*. *153*(8),1686-1696.

19 Sănchez-Chapula, JA., Ferrer, T., Navarro-Polanco, RA,, & Sanguinetti, MC. (2003). Voltage-dependent profile of human ether-a-go-go-related gene channel block is influenced by a single residue in the S6 transmembrane domain. *Molecular Pharmacology*, *63*(5), 1051-1058

20 Mukanganyama, S., Widersten, M., Naik, YS., Mannervik, B., & Hasler, JA. (2002). Inhibition of glutathione S-transferases by antimalarial drugs possible implications for circumventing anticancer drug resistance. *International Journal of Cancer*, *10*, 97(5),700-705

21 Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, MR., Smith, JC., Kasson, PM., Van, der, Spoel, D., &Hess, B., & Lindahl, E. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, *29*, 845-854.

22 Berman, H., Henrick, K., Nakamura, H., & Markley, JL. (2007). The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Research*, *35*, D301-303.

23 Karthikeyan, BS., Akbarsha, MA., & Parthasarathy, S. (2016). Network analysis and cross species comparison of protein-protein interaction networks of human, mouse and rat cytochrome P450 proteins that degrade xenobiotics. *Molecular Biosystems*, *21*;12(7), 2119-34

24 Magrane, M., & Consortium, U. (2011). UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)*, *29*, bar009.

25 Shah,MB., Pascual, J., Zhang, Q., Stout, CD., & Halpert, JR. (2011). Stuctures of cytochrome P450 2B6 bound to 4-benzylpyridine and 4-(4-nitrobenzyl) pyridine: insight into inhibitor binding and rearrangement of active site side chains. *Molecular Pharmacology*, *80*, 1047-1055.

26 Hernandez, M., Ghersi, D., & Sanchez, R. (2009). SITEHOUND-web: a server for ligand binding site identification in protein structures. *Nucleic Acids Research*. *37* (Web Server issue):W413-6

27 Sansen, S., Yano, JK., Reynald, RL., Schoch, GA., Griffin, KJ., Stout, CD., & Johnson, EF. (2007). Adaptations for the oxidation of polycyclic aromatic hydrocarbons exhibited by the structure of human P450 1A2. *Journal of Biological Chemistry*, *282*, 14348-14355.

28 Walsh, AA., Szklarz, GD., & Scott, EE. (2013). Human cytochrome P450 1A1 structure and utility in understanding drug and xenobiotic metabolism. *Journal of Biological Chemistry*, *288*, 12932-12943.

29 Sherman, W., Day, T., Jacobson, MP., Friesner, RA., & Farid, R. (2006). Novel procedure for modeling ligand/ receptor induced fit effects. *Journal of Medicinal Chemistry*, *49*, 534-553.

30 Gunsteren, WFV., Billeter, SR., Eising, AA., Hunenberger, PH., Kruger, PK., Mark, AE., Scott, WRP., & Tironi, IG. (1996). Biomolecular stimulation: The {GROMOS96} manual and user guide. *Verlag der Fachvereine, Zurich*, 1–1024.

31 Parrinello, M., & Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, *52*, 7182–7190.

32 Schuttelkopf, AW., van, & Aalten, DM. (2004). PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica D Biological Crystallography*, *60*, 1355-1363.

33 Williams, T., & Kelley, C. (2011). Gnuplot. An Interactive Plotting Program available at http://www.gnuplot.info/

34 Wang, A., Savas, U., Stout, CD., & Johnson, EF. (2011). Structural characterization of the complex between alpha-naphthoflavone and human cytochrome P450 1B1. *Journal of Biological Chemistry*, *286*, 5736-5743.

35 Porubsky, PR., Meneely, KM., & Scott, EE. (2008). Structures of human cytochrome P-450 2E1. Insights into the binding of inhibitors and both small molecular weight and fatty acid substrates. *Journal of Biological Chemistry*, *283*, 33698-33707.

36 Zhang, YY., & Yang, L. (2009). Interactions between human cytochrome P450 enzymes and steroids: physiological and pharmacological implications. *Expert Opinion on Drug Metabolism and Toxicology*, *5*, 621-629.

37 Jovanovic, T., Farid, R., Friesner, RA., & McDermott, AE. (2005). Thermal equilibrium of high- and low-spin forms of cytochrome P450 BM-3: repositioning of the substrate? *Journal of the American Chemical Society*. *127*(39), 13548-52.

38 Capoferri, L., Leth, R., ter, Haar. E., Mohanty, AK., Grootenhuis, PD., Vottero, E., Commandeur, JN., Vermeulen, NP., Jørgensen, FS., Olsen, L., & Geerke, DP. (2016). Insights into regioselective metabolism of mefenamic acid by cytochrome P450 BM3 mutants through crystallography, docking, molecular dynamics, and free energy calculations. *Proteins*, *84*(3), 383-96.

Acknowledgments

MAA acknowledges the financial support from Doerenkamp-Zbinden Foundation, Switzerland, for establishing the MGDC. BSK acknowledges the research fellowship from MGDC. SP acknowledges the financial assistance in the form of research project funded by University Grants Commission (UGC), New Delhi (Grant No. F.41-965/2012(SR)). SS acknowledges the research fellowship from the above mentioned UGC-funded research project. We thank anonymous reviewers for improving the quality of the paper.

Correspondence address

Subbiah Parthasarathy

Department of Bioinformatics, School of Life Sciences, Bharathidasan University,

Tiruchirappalli – 620 024, Tamil Nadu, India

Email: bdupartha@gmail.com , partha@bdu.ac.in

Phone: +91 431 2407071

Fax: +91 431 2407045

Mobile: + 91 9443533095

List of Figures

Fig. 6 - Backbone RMSD profile of CYP1A2 protein in native and complex forms. Backbone RMSD of, (a) Human-CYP1A2 in native form and as complex with Quinine, (b) Human-CYP1A2 in native form and as complex with quinidine, (c) Mouse-CYP1A2 in native form and as complex with quinine, (d) Mouse-CYP1A2 in native form and as complex with quinidine, (e) Rat-CYP1A2 in native form and as complex with quinine, (f) Rat-CYP1A2 in native form and as complex with quinidine.

Fig. 7 - Total number of hydrogen bonds formed between ligands and CYP1A2protein. Total number of hydrogen bonds formed between, (a) Quinine and human-CYP1A2 protein, (b) Quinidine and human-CYP1A2 protein, (c) Quinine and mouse-CYP1A2 protein, (d) Quinidine and mouse-CYP1A2 protein, (e) Quinine and rat-CYP1A2 protein, (f) Quinidine and rat-CYP1A2 protein.

Fig. 8 - Backbone RMSD profile of CYP2B proteins in native and complex forms. Backbone RMSD of, (a) Human-CYP2B6 in native form and as complex withQuinine, (b) Human-CYP2B6 in native form and as complex with quinidine,(c) Human- CYP2B7 in native form and as complex with quinine, (d)Human-CYP2B7 in native form and as complex with quinidine, (e) Mouse-CYP2B9 in native form and as complex with quinine, (f) Mouse-CYP2B9 in native form and as complex with quinidine, (g) Rat-CYP2B1 in native form and as complex with quinine, (h) Rat-CYP2B1 in native form and as complex with quinidine.

Fig. 9 - Total number of hydrogen bonds formed between ligands and CYP2B proteins. Total number of hydrogen bonds formed between, (a) Quinine andhuman-CYP2B6 protein, (b) Quinidine and human-CYP2B6 protein, (c)Quinine and human-CYP2B7 protein, (d) Quinidine andhuman-CYP2B7protein, (e) Quinine and mouse-CYP2B9 protein, (f) Quinidine and mouse-CYP2B9 protein, (g) Quinine and rat-CYP2B1 protein, (h)

Quinidine and rat-CYP2B1 protein.

List of Tables

Supplementary Figures

Supplementary Fig. S1 –Ramachandran plots for CYP1A subfamily proteins, (a) Human-CYP1A1, (b) Mouse-CYP1A1, (c) Rat-CYP1A1, (d) Human-CYP1A2, (e) Mouse-CYP1A2, (f) Rat-CYP1A2.

Supplementary Fig. S2 - Ramachandran plots for CYP2B subfamily proteins, (a) Human-CYP2B6, (b) Human-CYP2B7, (c) Mouse-CYP2B9, (d) Mouse-CYP2B10, (e) Mouse-CYP2B13, (f) Mouse-CYP2B19, (g) Mouse-CYP2B23.

Supplementary Fig. S3 - Ramachandran plots for CYP2B subfamily proteins, (a) Rat-CYP2B1, (b) Rat-CYP2B2, (c) Rat-CYP2B3, (d) Rat-CYP2B12, (e) Rat-CYP2B13, (f) Rat-CYP2B15, (g) Rat-CYP2B21.

Supplementary Fig. S4 –Multiple sequence alignment of CYP1A1 protein of human, mouse and rat. The active site residues Thr126, Gly320, Ala321, Thr325, Val386, Leu500 and Thr501 conserved across species are highlighted in red boxes.

Supplementary Fig. S5 – Multiple sequence alignment of CYP1A2 protein of human, mouse and rat. The active site residues Arg108, Gly316, Ile384,Arg454, Cys456, Ile457 and Gly458 conserved across species are highlighted in red boxes.

Supplementary Fig. S6 – Multiple sequence alignment of CYP2B subfamily proteins ofhuman, Mouse and rat. The active site residues Arg98, Ile114, Phe115, Trp121, Arg125, Phe298, Thr303, Pro428, Ser430, Arg434and Cys437 Conserved across species are highlighted in red boxes.

Supplementary Fig. S7 - Results of Glide docking of (a) Quinine against human-CYP1A2. Hydrogen bond formed at Asp313 is shown, (b) Quinine against mouse-CYP1A2 (Ligand did not form any hydrogen bond), (c) Quinine against rat-CYP1A2.Hydrogenbond formed at Cys456 is shown, (d) Quinidine against human-CYP1A2. Hydrogen bond formed at Thr124 is shown, (e) Quinidine against mouse-CYP1A2. Ligand did not form any hydrogen bond, (f) Quinidine against rat-CYP1A2. Ligand did not form any hydrogen bond.

Supplementary Fig. S8 - Backbone RMSD profile of proteins in native form. Backbone RMSD of human, mouse and rat, (a) CYP1A1Protein (b) CYP1A2 protein (c) CYP2B protein.

Supplementary Tables

Supplementary Table T1 - The list of drug metabolizing CYP enzymes (from Karthikeyan et al. [23])

Supplementary Table T2 - The percentage of amino acid sequence identity of mouse and rat CYP1A1, CYP1A2 proteins against the template human-CYP1A1 (PDB ID: 418V) and human-CYP1A2 (PDB ID: 2HI4) proteins.

Supplementary Table T3 - The percentage of amino acid sequence identity of mouse and rat CYP2B subfamily proteins against the template human-CYP2B6 (PDB ID: 3QOA) protein.

Supplementary Table T4 –Validation of predicted structures using Ramachandran plots. The percentage of amino acid residues lies in favoured, allowed and outlier regions for (a) CYP1A subfamily proteins and (b) CYP2B subfamily proteins.

Supplementary Table T5 - Active site amino acid residues of CYP1A1 protein predicted using sitemap and sitehound server and compared with available literature.

Supplementary Table T6 - Active site amino acid residues of CYP1A2 protein predicted using sitemap and sitehound server and compared with available literature.

Supplementary Table T7 - Active site amino acid residues of CYP2B subfamily proteins predicted using sitemap and sitehound server and compared with available literature.

Figure 1

Click here to download Figure Fig1.jpg

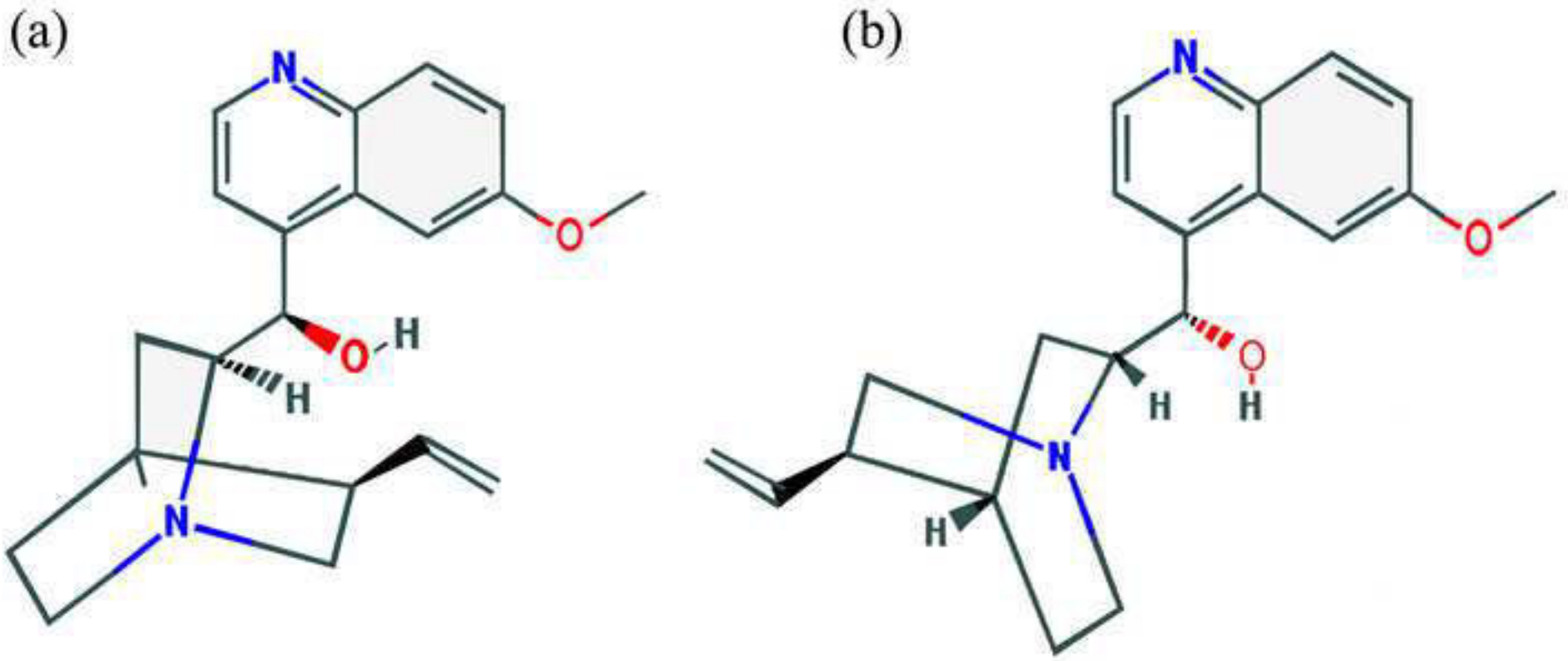


Fig. 1

Figure 2

Click here to download Figure Fig2.jpg

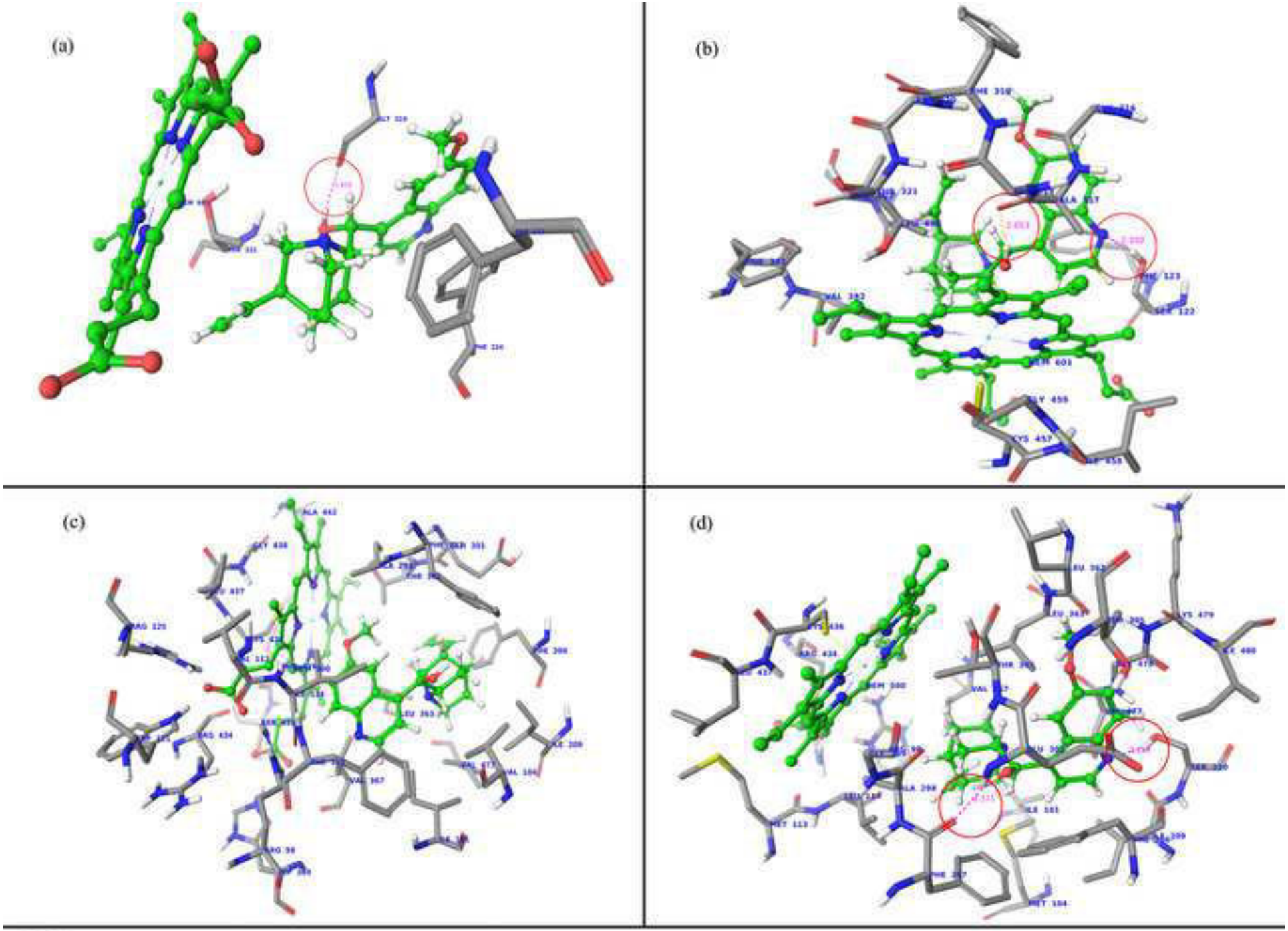


Fig. 2

Figure 3

Click here to download Figure Fig3.jpg

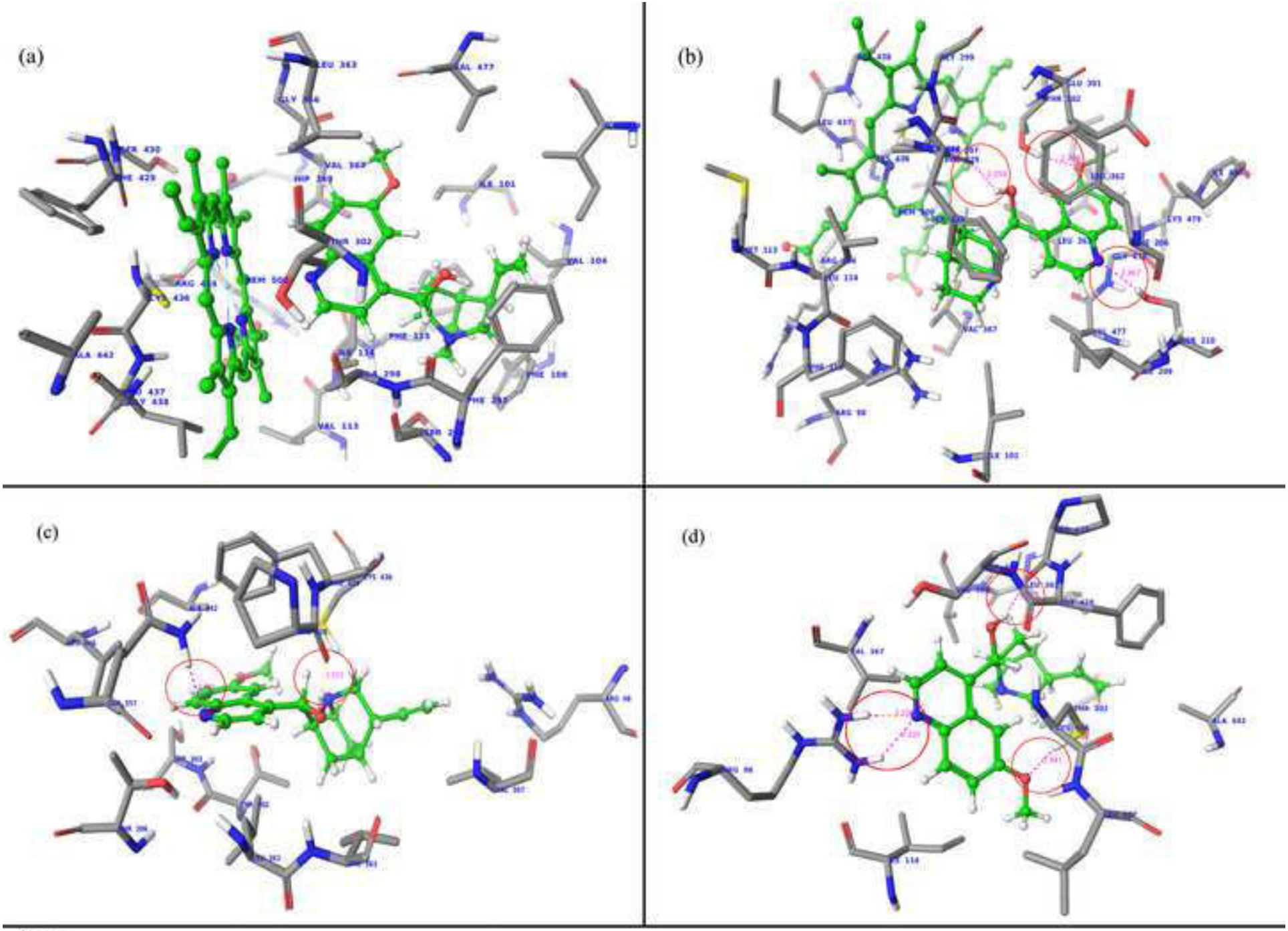


Fig. 3

Figure 4

Click here to download Figure fig4.jpg

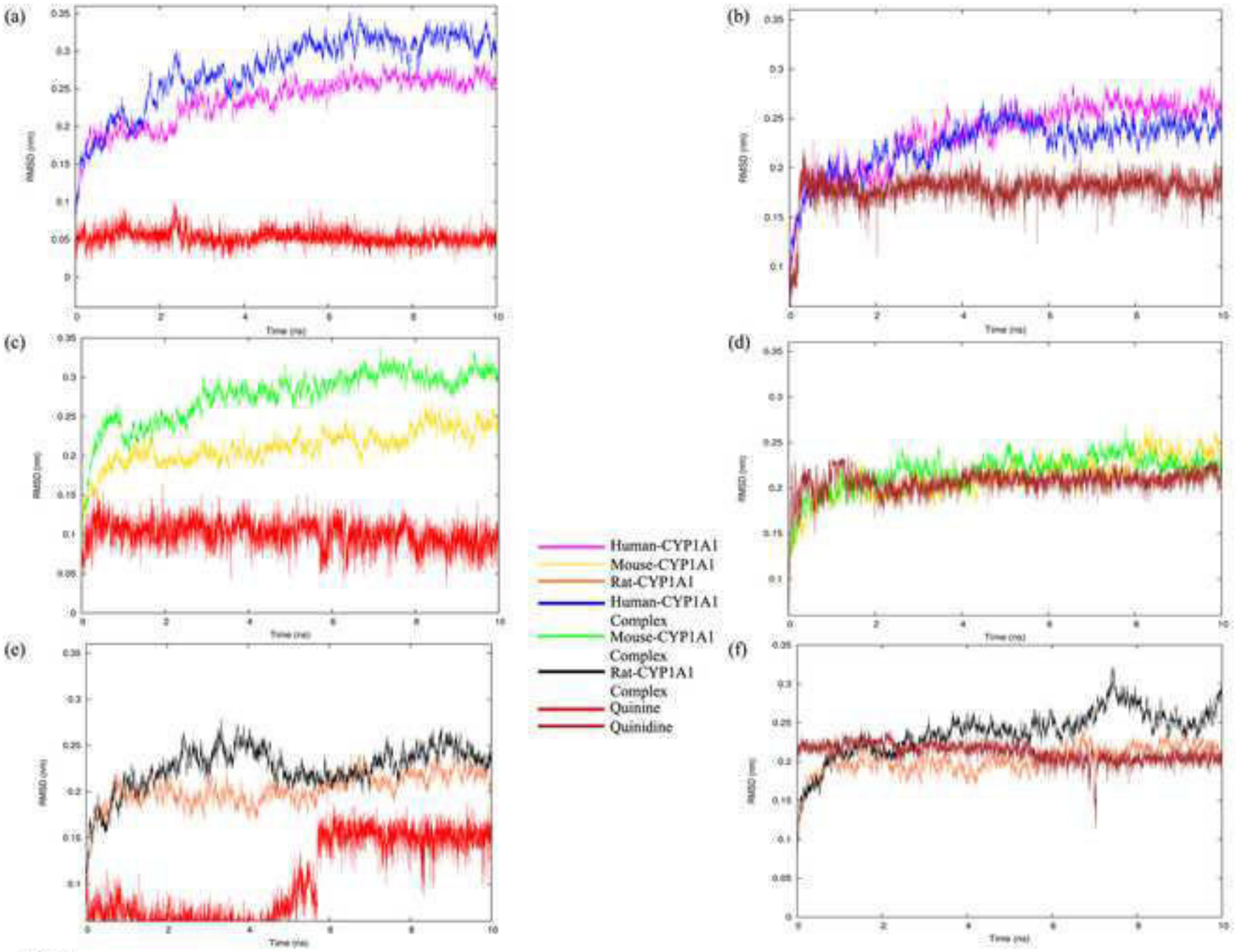


Fig. 4

Figure 5

Click here to download Figure Fig5.jpg

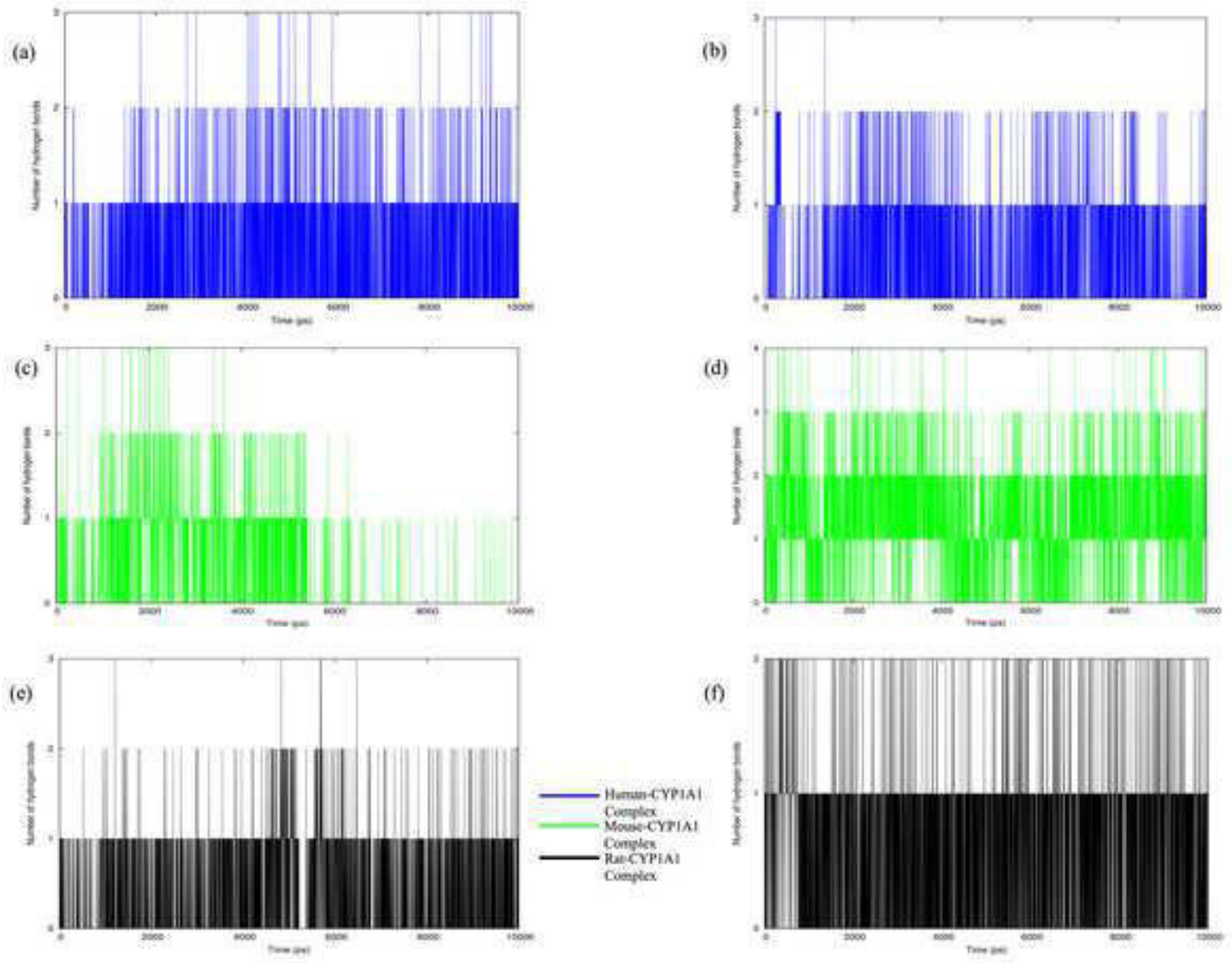


Fig. 5

Figure 6

Click here to download Figure Fig6.jpg

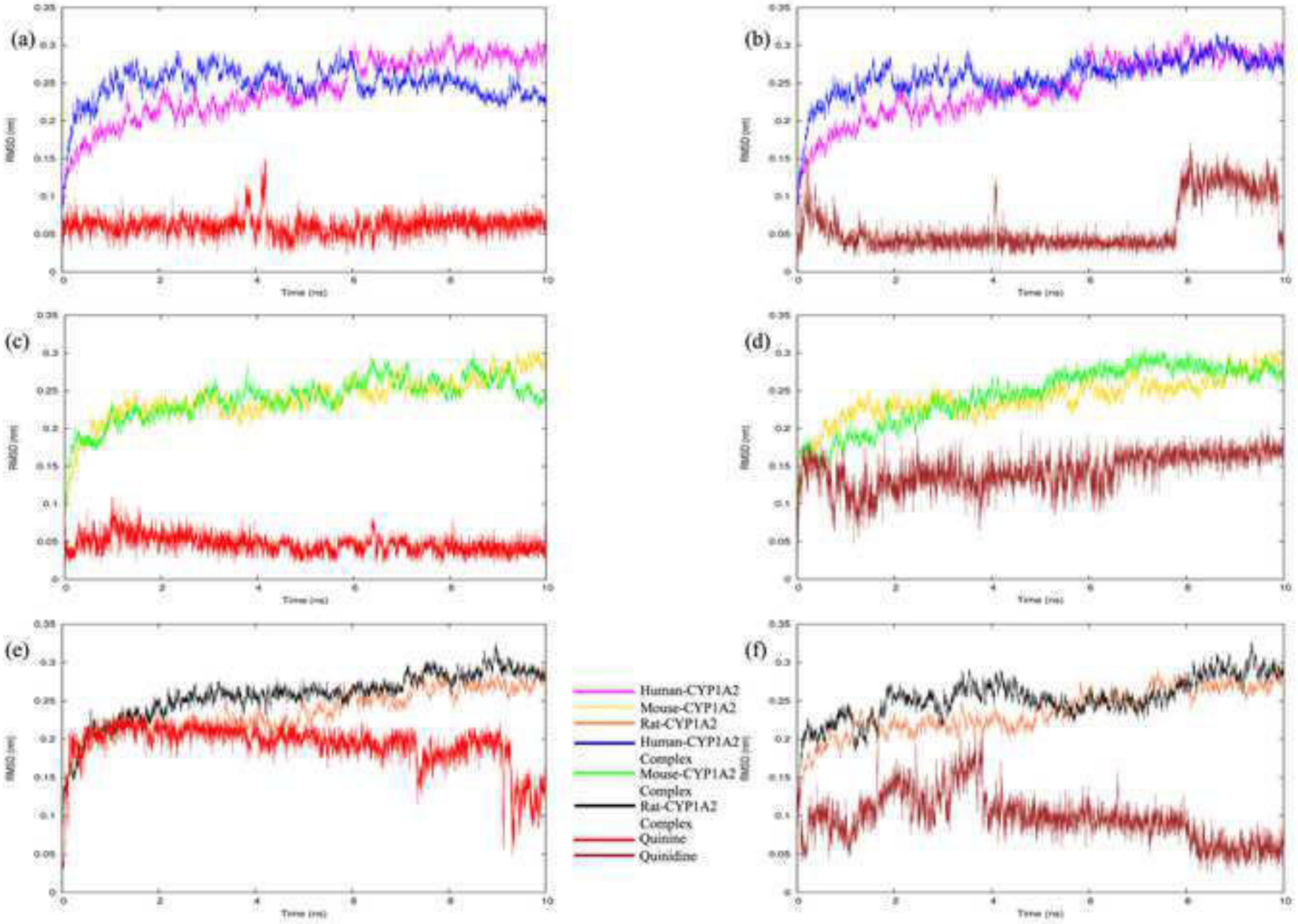


Fig. 6

Figure 7

Click here to download Figure Fig7.jpg

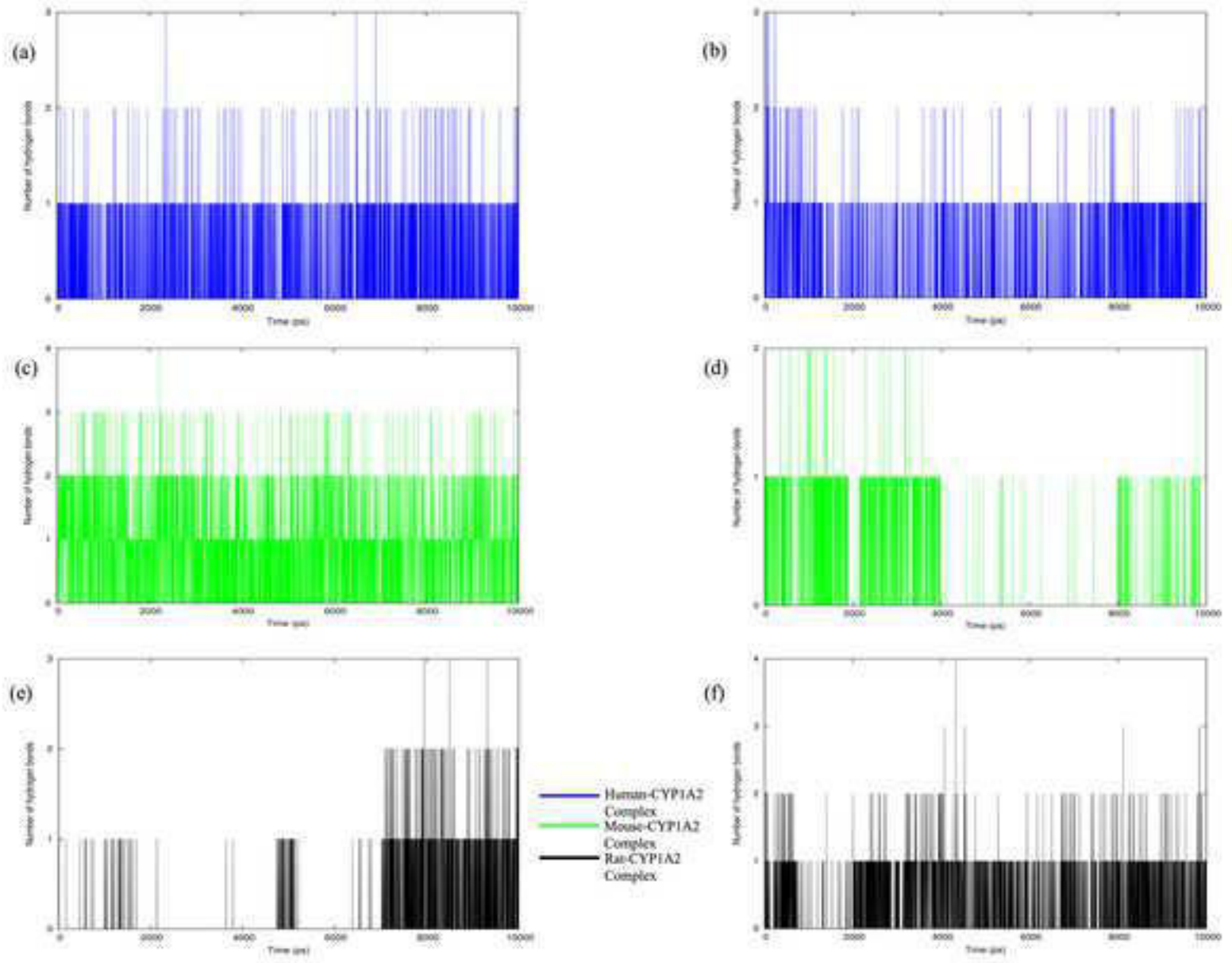


Fig. 7

Figure 8

Click here to download Figure Fig8.jpg

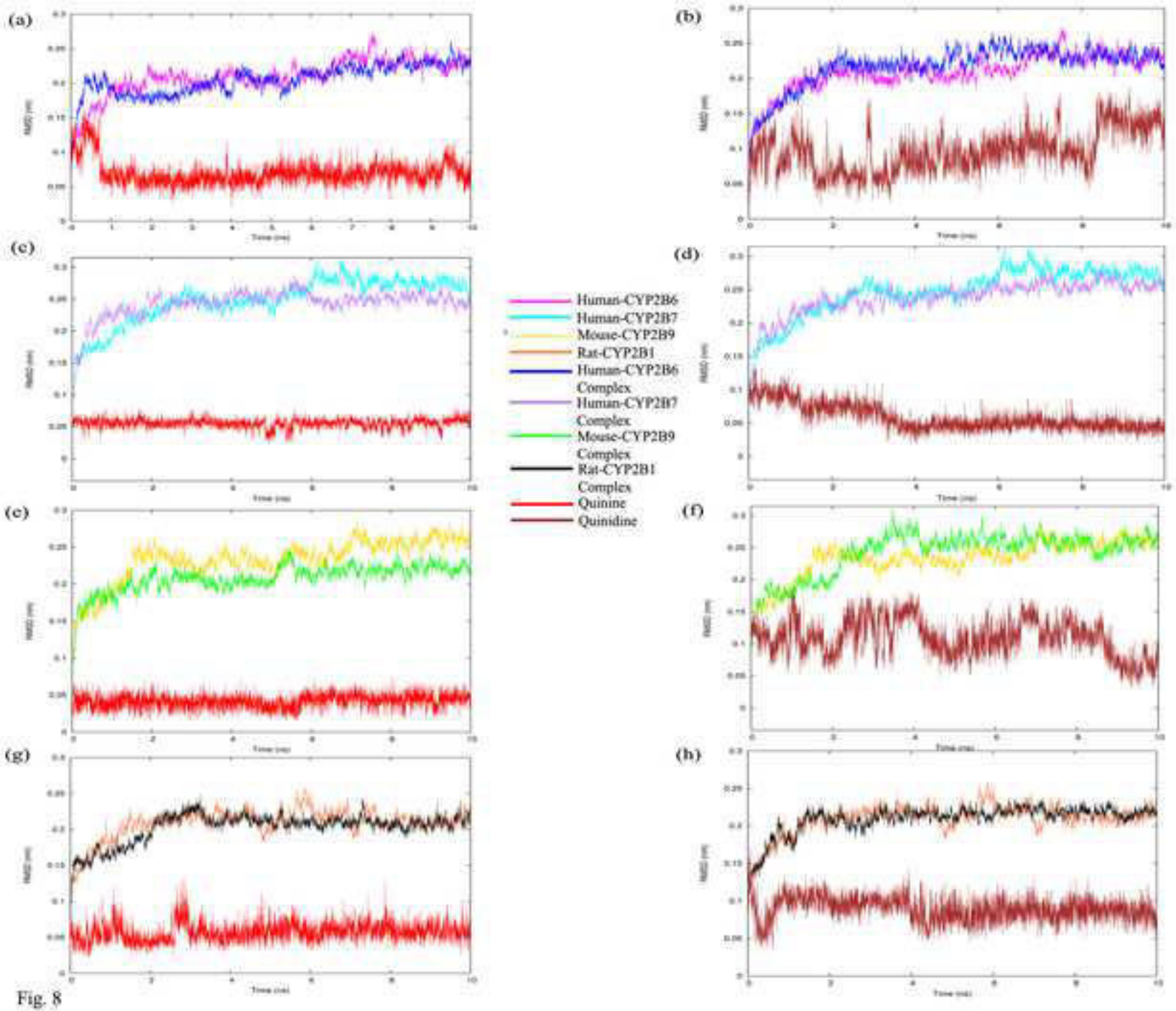


Fig. 8

Figure 9

Click here to download Figure Fig9.jpg

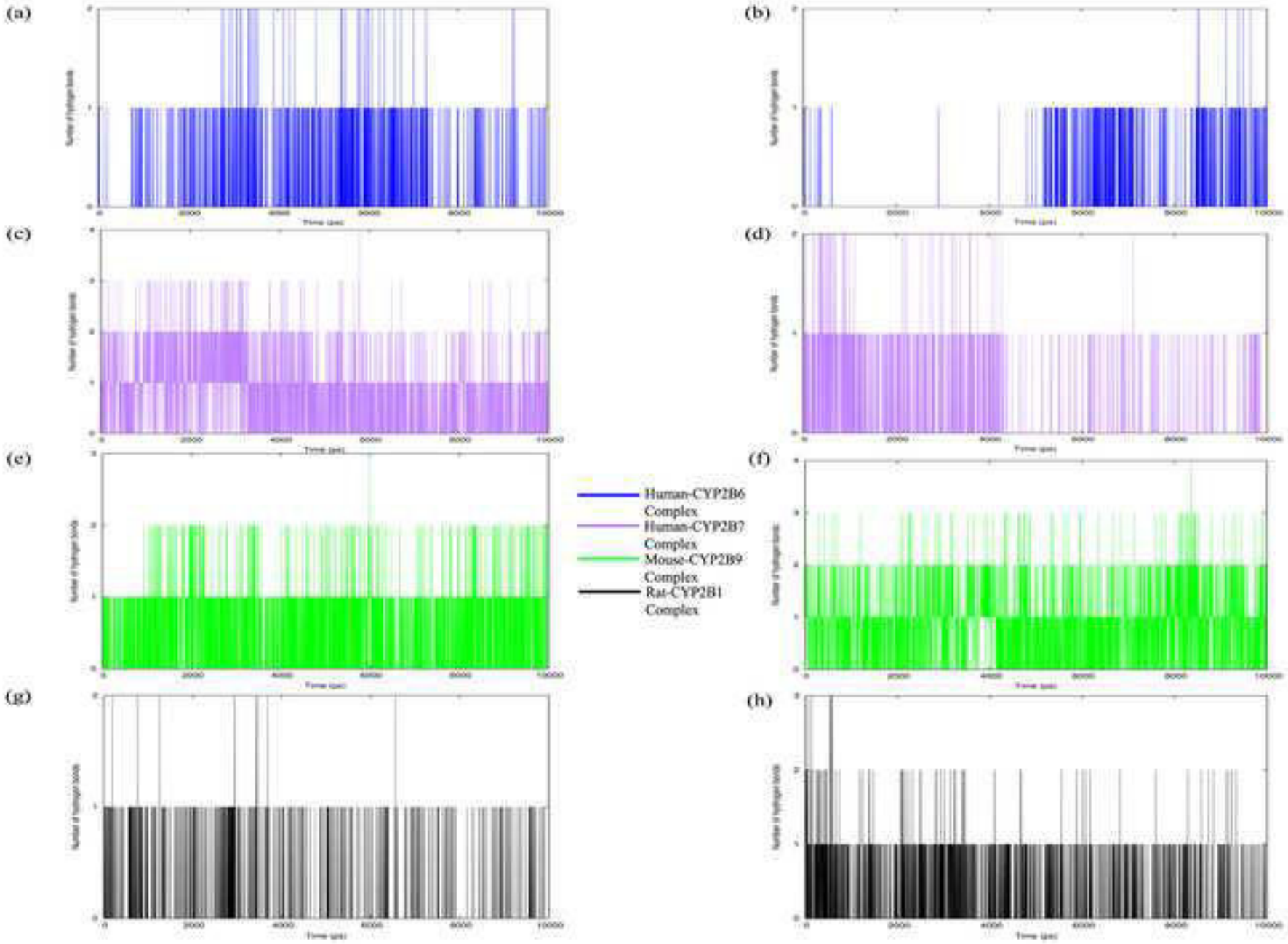


Fig. 9

Table 1.

Protein	Quinine				Quinidine			
	Glide XP score	Binding energy (Kcal/mol)	Residues	Hydrogen bond length (Å)	Glide XP score	Binding energy (Kcal/mol)	Residues	Hydrogen bond length (Å)
Human-CYP1A1	-9.48	-17.4	Gly316	1.95	-9.66	-52.4	Gly316	1.70
Mouse-CYP1A1	-9.34	-27.6	Gly320	1.86	-9.15	-34.6	Gly320	2.05
Rat-CYP1A1	-9.99	-16.3	Gly320	2.10	-9.10	-7.3	Gly320	1.94
Human-CYP1A2	-8.31	-3.3	Asp313	1.71	-8.86	-3.7	Thr124	1.66
Mouse-CYP1A2	-8.74	-21.3	NI	-	-8.51	-31.9	NI	-
Rat-CYP1A2	-6.22	-10.2	Cys456	2.23	-5.94	-26.7	NI	-

NI – No interaction

Table 2.

Protein	Quinine						Quinidine					
	Glide XP score	Binding energy (Kcal/ mol)	Residues	Hydrogen bond length (Å)	RMSD between reported x-ray structure and protein-quinine complex (Å)	Hydrogen bond (Atom number) Protein-ligand (Atom number)	Glide XP score	Binding energy (Kcal/ mol)	Residues	Hydrogen bond length (Å)	RMSD between reported x-ray structure and protein-quinidine complex (Å)	Hydrogen bond (Atom number) Protein-ligand (Atom number)
Human-CYP1A1	-9.90	-56.6	Gly316	1.93	0.08	(4401) O..H (7750)	-9.93	-56.3	Ser122 Ala317	2.10 2.05	0.09	(1351) O..H (7716) (4408) O..H (7750)
Mouse-CYP1A1	-10.46	-15.4	Gly320	1.68	0.09	(4420) O..H (7736)	-7.52	-27.5	Gly320	1.82	0.11	(4420) O..H (7655)
Rat-CYP1A1	-10.82	-49.1	Thr501	2.00	0.10	(7376) O..H (7762)	-10.23	-51.2	Thr126	2.18	0.10	(1341) H..N (7728)
Human-CYP1A2	-10.42	-15.5	Gly316	2.12	0.08	(4452) O..H (7814)	-10.51	-67.9	Thr321 Thr498	1.88 2.02	0.10	(4510) O..H (7814) (7779) N..H (7444)

Table 2 continued

Mouse-CYP1A2	-11.99	-138.5	Thr123 Gly314	1.81 1.87	0.16	(1424) H..O (7755) (4414) O..H (7791)	-9.90	-26.7	Thr117	2.00	0.17	(1339) H..O (7755)
Rat-CYP1A2	-7.82	-21.9	Arg107 Arg454 Cys456	1.83 1.93 2.34	0.16	(1174) H..O (7720) (6748) O..H (7757) (6803) H..N (7722)	-6.38	-57.6	Leu448 Cys456	2.14 2.10	0.17	(6654) O..H (7757) (6803) H..O (7721)

NI - No interaction

Table 3.

Protein	Quinine				Quinidine			
	Glide XP score	Binding energy (Kcal/mol)	Residues	Hydrogen bond length (Å)	Glide XP score	Binding energy (Kcal/mol)	Residues	Hydrogen bond length (Å)
Human-CYP2B6	-8.69	-66.2	NI	-	-8.31	-36.3	NI	-
Human-CYP2B7	-7.19	-40.9	NI	-	-6.12	-45.6	Thr302	2.36
Mouse-CYP2B9	-7.84	-58.6	Pro428 Leu362	1.87 1.93	-6.73	-17.3	Pro428	2.20
Mouse-CYP2B10	-9.33	-47.4	NI	-	-8.72	-35.5	NI	-
Mouse-CYP2B13	-9.08	-46.8	NI	-	-9.81	-62.3	NI	-
Mouse-CYP2B19	-2.84	-9.0	His398	2.44	-2.55	-12.1	NI	-
Mouse-CYP2B23	-9.43	-90.6	NI	-	-9.65	-66.4	NI	-
Rat-CYP2B1	-6.44	-67.3	Leu362	1.96	-6.93	-76.8	Pro428	1.82
Rat-CYP2B2	-8.63	-39.2	NI	-	-8.62	-62.7	NI	-
Rat-CYP2B3	-6.66	-61.8	Pro428	2.15	-6.64	-24.7	Ser363	2.00
Rat-CYP2B12	-6.62	-45.6	NI	-	-6.62	-22.3	NI	-
Rat-CYP2B13	-9.02	-69.7	NI	-	-9.87	-59.1	NI	-
Rat-CYP2B15	-7.58	-33.7	Thr299	2.31	-7.60	-30.9	Thr299	1.92
Rat-CYP2B21	-7.65	-28.4	Pro428 Leu362	1.79 2.04	-6.48	-62.9	NI	-

NI – No interaction

Table 4.

Protein	Quinine						Quinidine					
	Glide XP score	Binding energy (Kcal/mol)	Residues	Hydrogen bond length (Å)	RMSD between reported x-ray structure and protein-quinine complex (Å)	Hydrogen bond (Atom number) Protein-ligand (Atom number)	Glide XP score	Binding energy (Kcal/mol)	Residues	Hydrogen bond length (Å)	RMSD between reported x-ray structure and protein-quinidine complex (Å)	Hydrogen bond (Atom number) Protein-ligand (Atom number)
Human-CYP2B6	-9.92	-24.7	NI	-	0.07	-	-10.06	-12.8	NI	-	0.06	-
Human-CYP2B7	-11.68	-20.0	Ser210 Phe297	2.12 2.17	0.91	(2970) H..N (7558) (4374) O..H (7592)	-10.41	-32.4	Ser210 Phe297 Thr302	2.06 2.26 2.47	0.91	(2970) H..N (7558) (4374) O..H (7592) (4447) H..O (7556)
Mouse-CYP2B9	-8.08	-99.9	Pro428 Val363	2.13 2.11	0.11	(6458) O..H (7506) (5414) H..O (7470)	-9.57	-18.2	Pro428 Gln357	1.88 1.96	0.15	(6458) O..H (7506) (5320) H..N (7472)

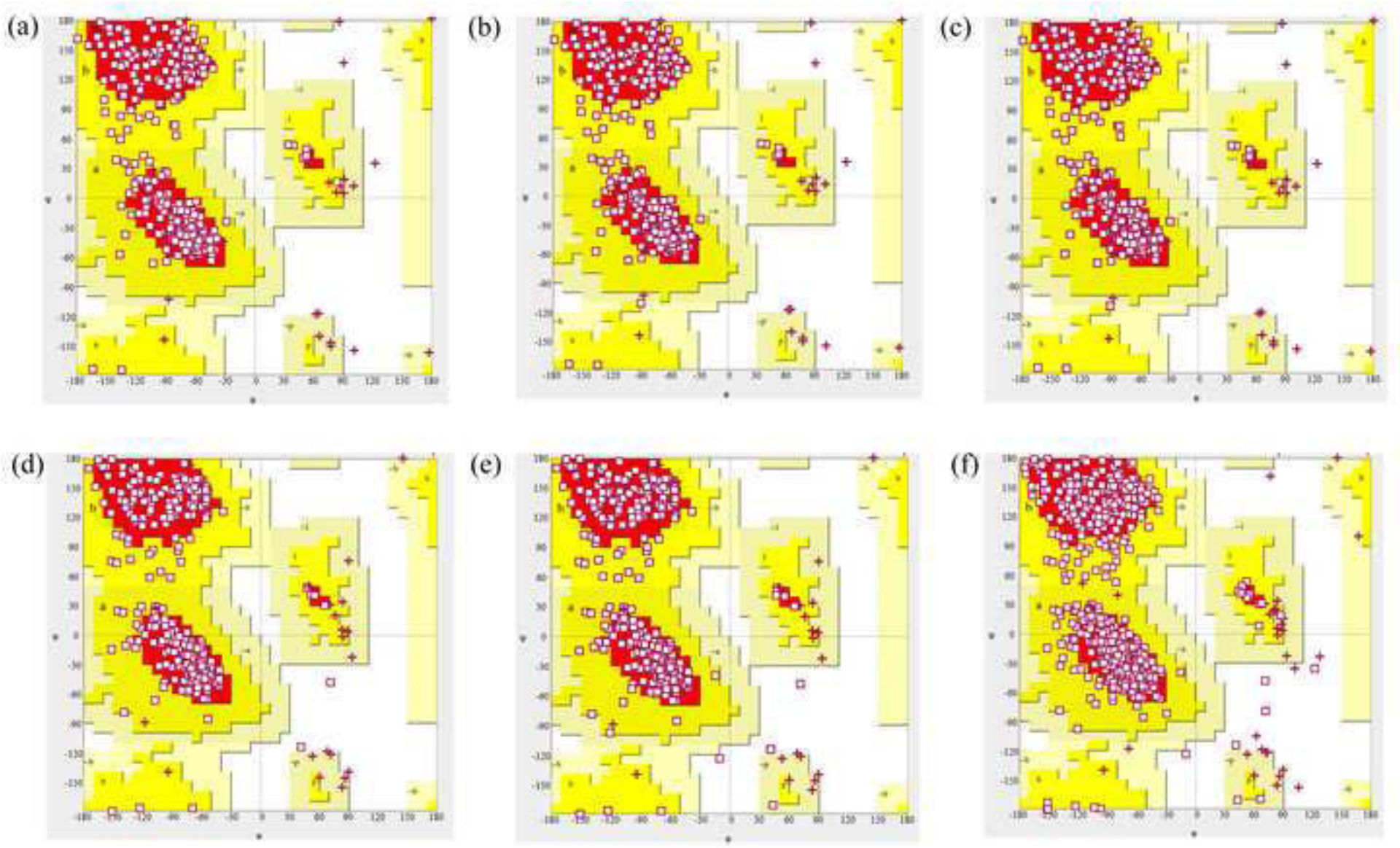
Table 4 continued.

Rat-CYP2B1	-7.79	-40.6	Pro428 Gln357	1.51 1.90	0.10	(6454) O..H (7495) (5337) H..N (7461)	-7.35	-59.2	Pro428 Cys436 Arg98 Arg98	1.91 1.94 2.24 2.32	0.12	(6454) O..H (7495) (6592) H..O (7459) (1092) H..N (7461) (1090) H..N (7461)

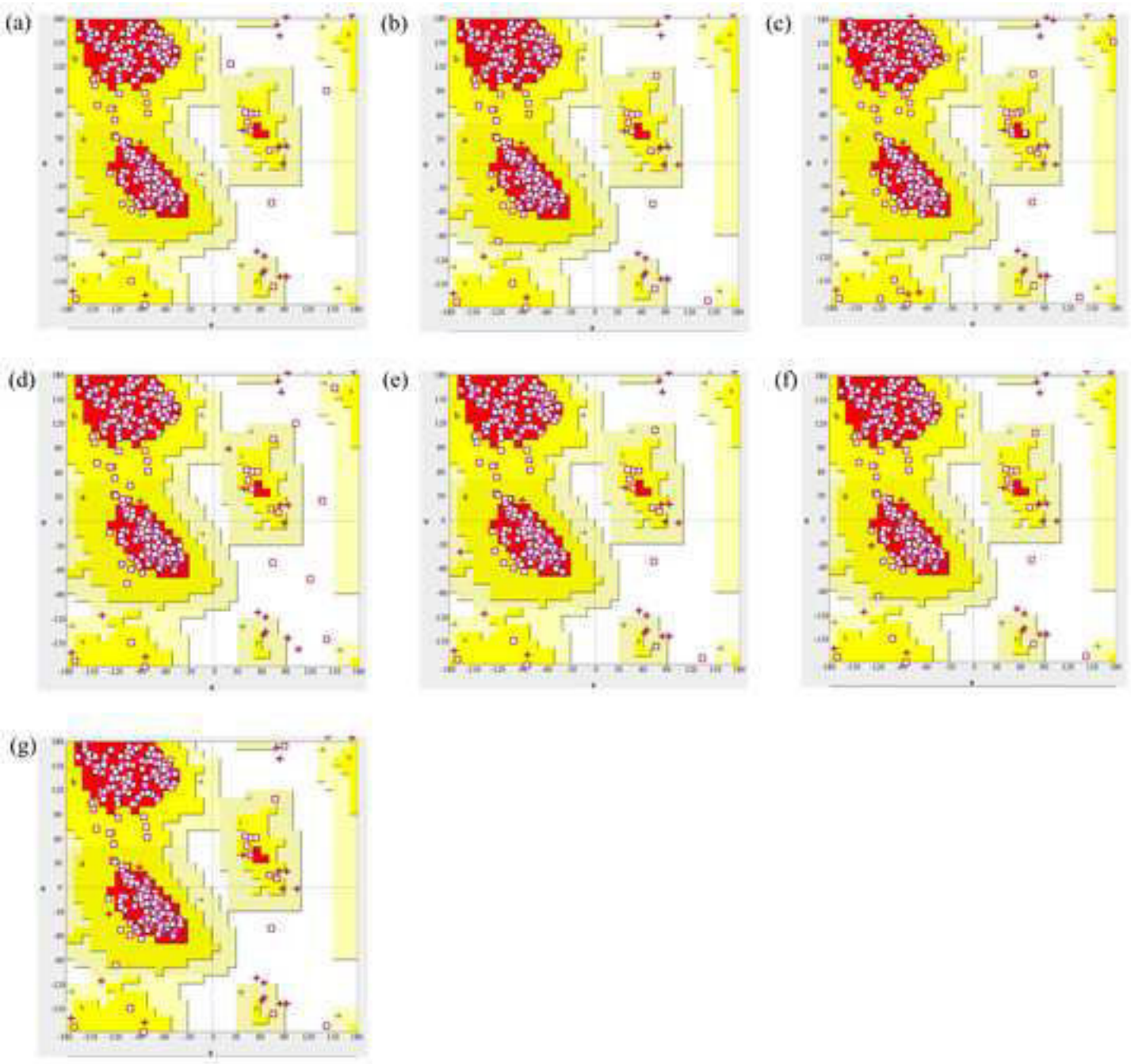
NI – No interaction

Table 5.

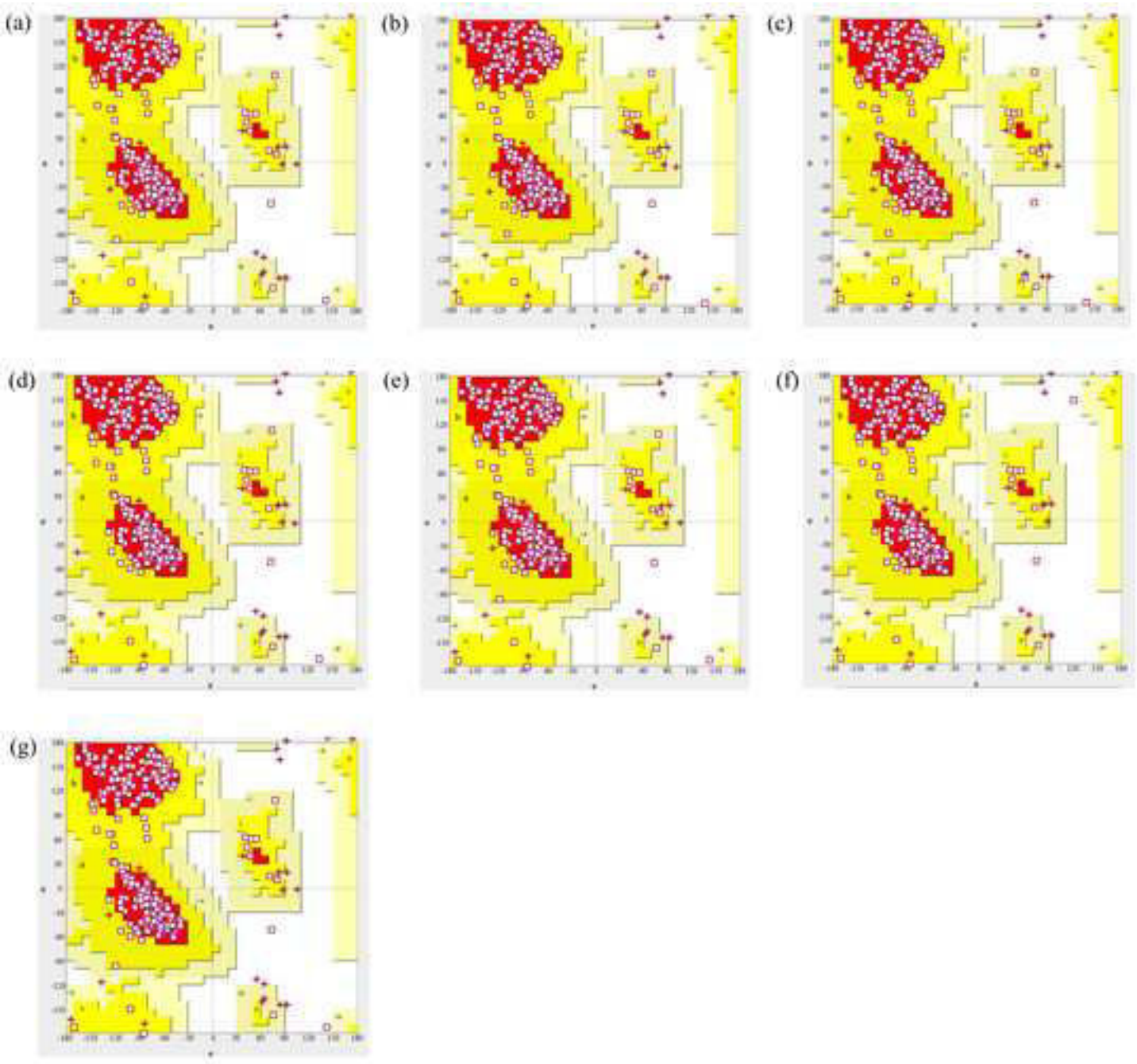
Protein		Average Root Mean Square Deviation (RMSD)			Average number of hydrogen bond interactions	
		Protein in native form (nm)	Complex with quinine (nm)	Complex with quinidine (nm)	Complex with quinine	Complex with quinidine
CYP1A1	Human-CYP1A1	0.23	0.28	0.22	0.48	2.95
	Mouse-CYP1A1	0.21	0.28	0.22	0.39	2.87
	Rat-CYP1A1	0.20	0.23	0.24	0.37	3.49
CYP1A2	Human-CYP1A2	0.24	0.25	0.26	3.40	2.23
	Mouse-CYP1A2	0.24	0.24	0.24	0.77	3.31
	Rat-CYP1A2	0.24	0.26	0.25	1.18	1.64
CYP2B Subfamily	Human-CYP2B6	0.21	0.20	0.22	2.43	0.31
	Human-CYP2B7	0.25	0.24	0.24	3.02	2.27
	Mouse-CYP2B9	0.23	0.21	0.24	2.02	2.92
	Rat-CYP2B1	0.21	0.20	0.21	1.45	2.13



Supplementary Fig. S1



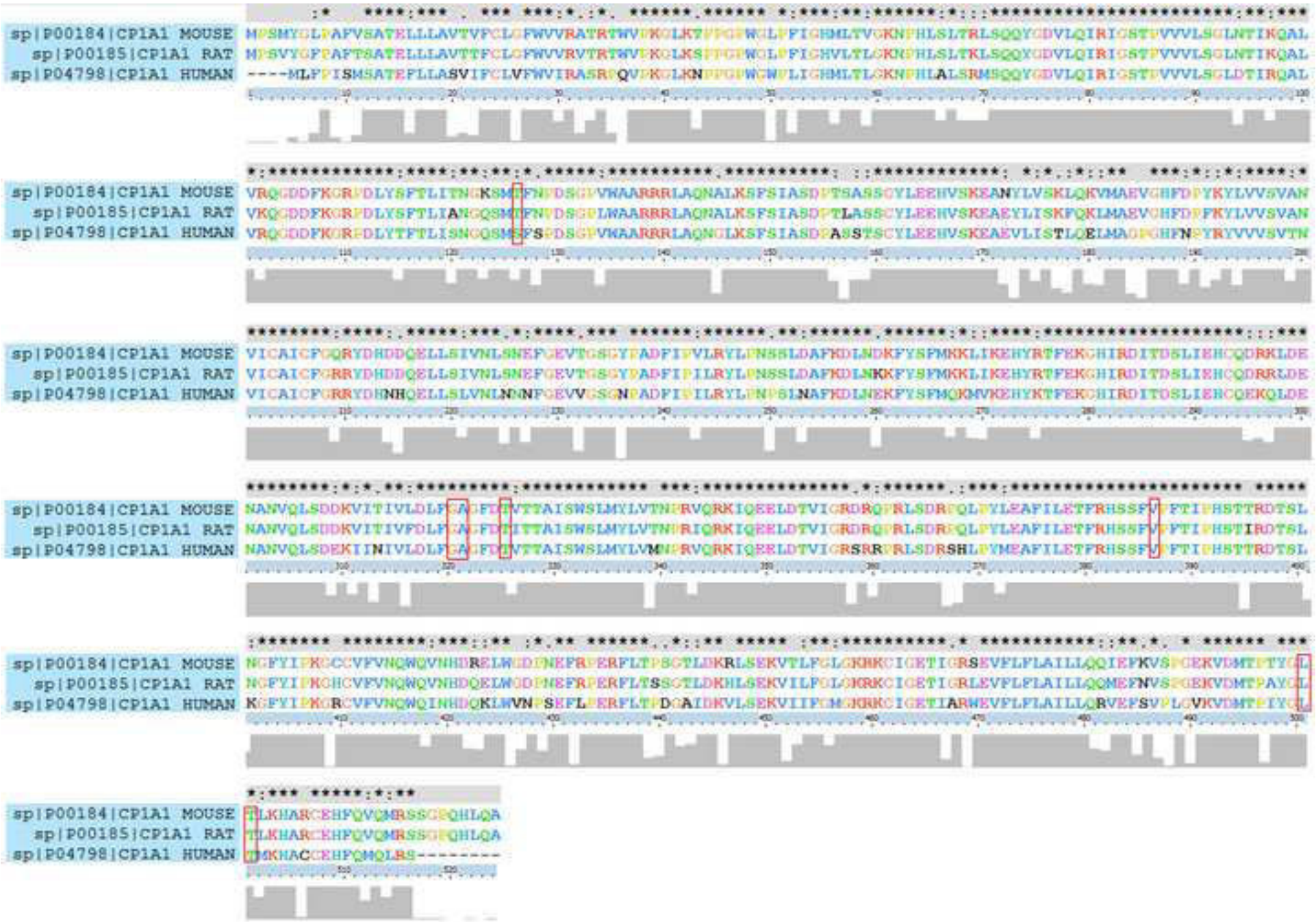
Supplementary Fig. S2



Supplementary Fig. S3

Supplementary Fig S4

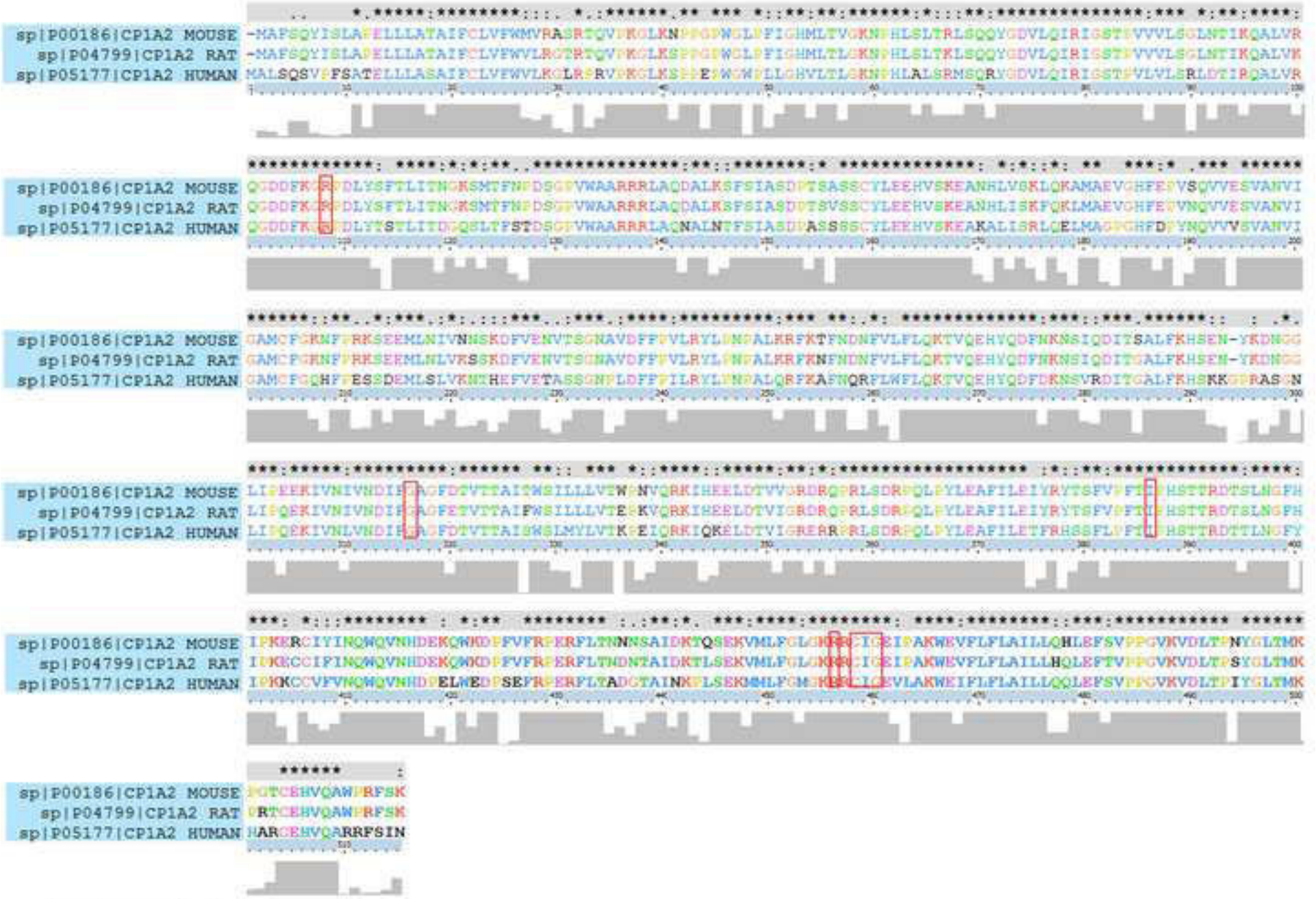
Click here to download Supplementary Material Supplementary FigS4.jpg



Supplementary Fig. S4

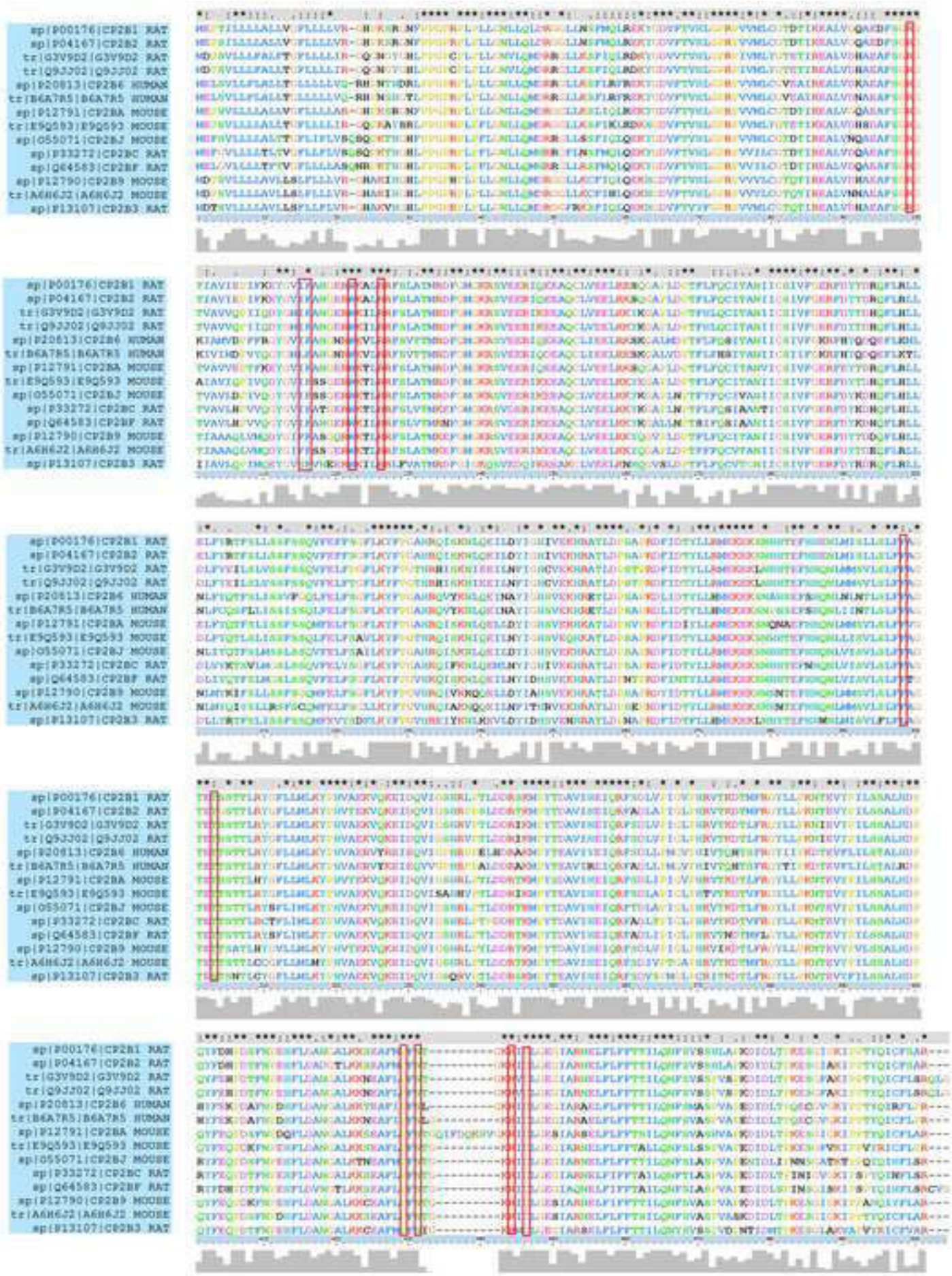
Supplementary Fig S5

Click here to download Supplementary Material Supplementary FigS5.jpg

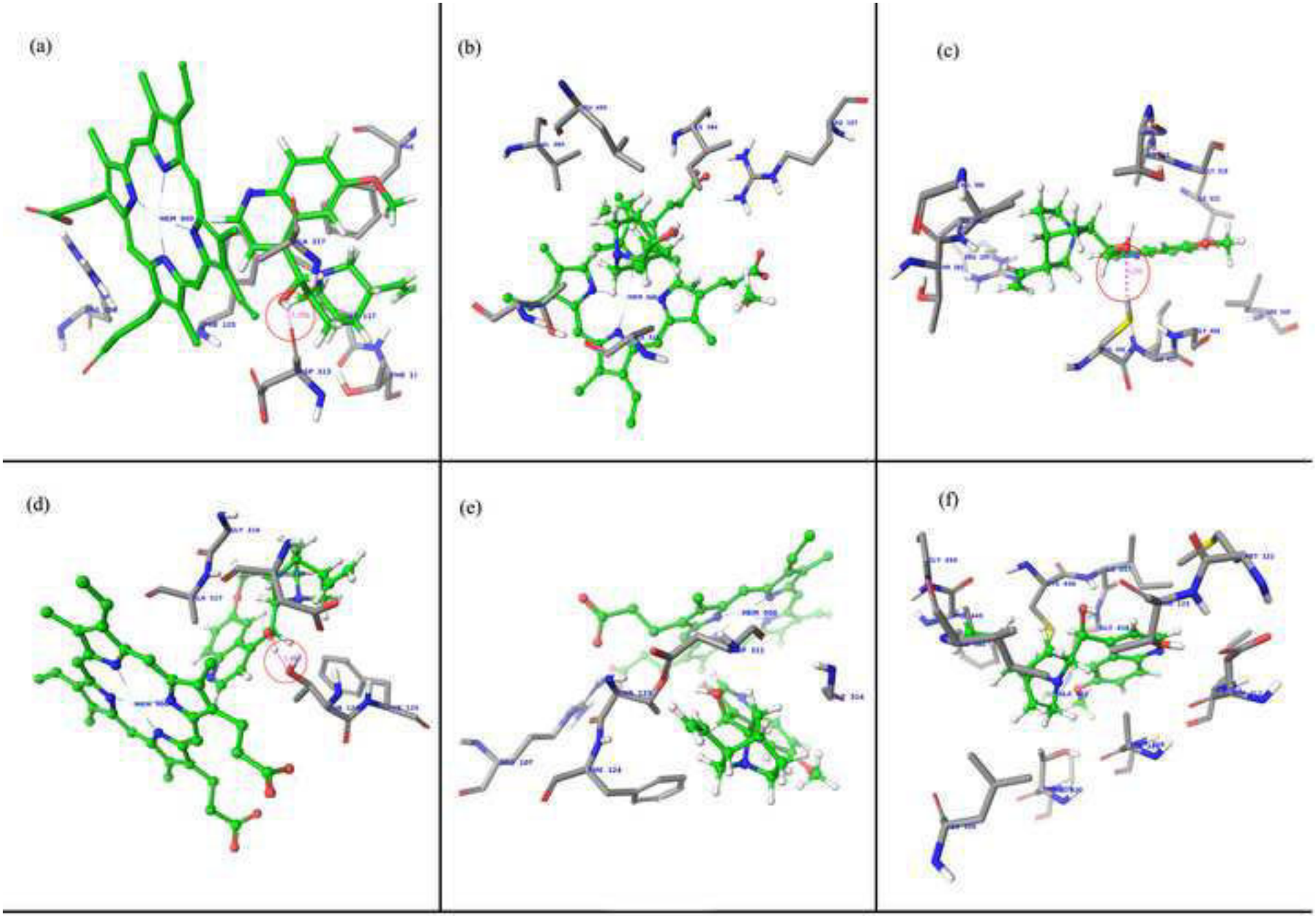


Supplementary Fig. S5

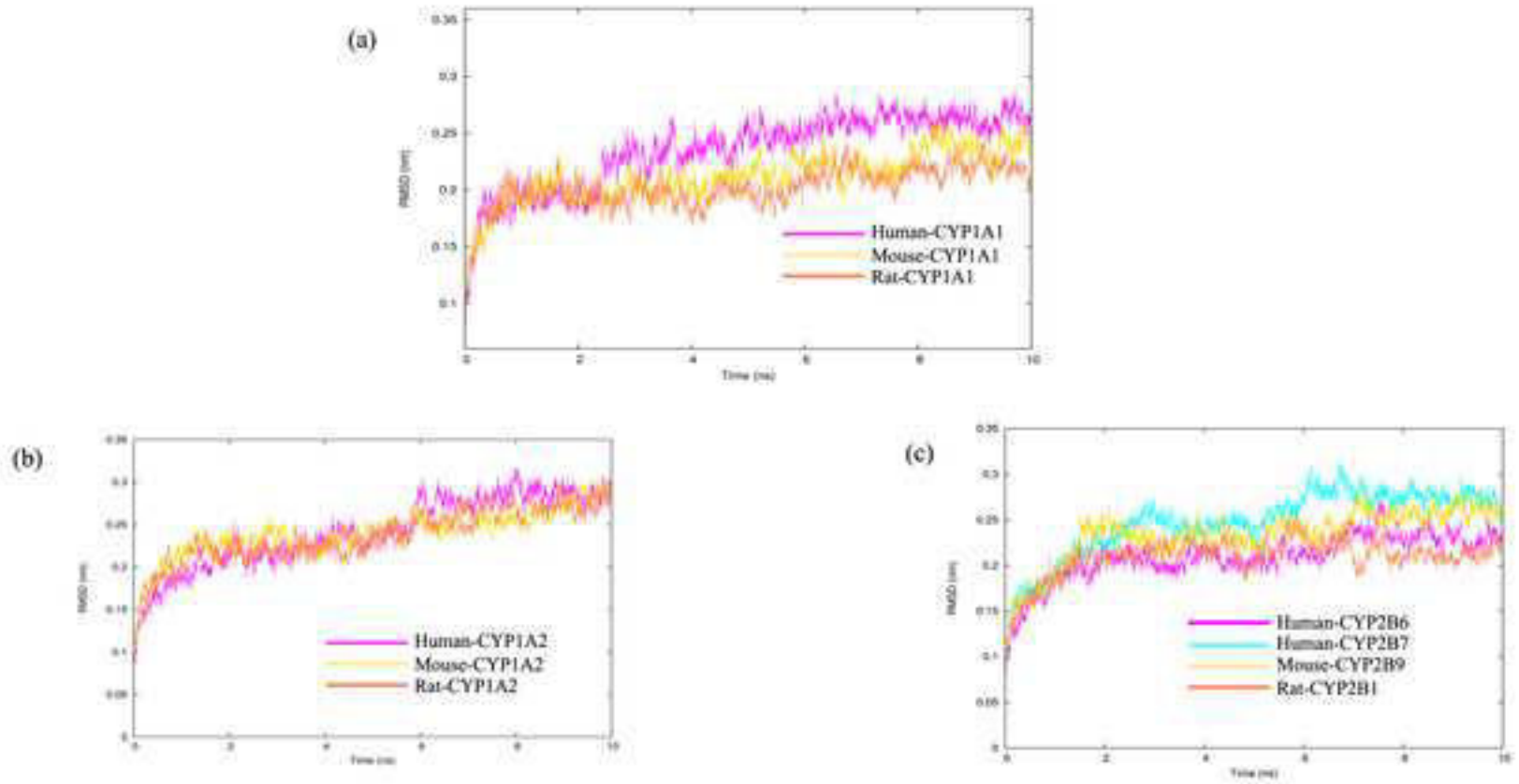
Click here to download Supplementary Material Supplementary FigS6.jpg



Supplementary Fig. S6



Supplementary Fig. S7



Supplementary Fig. S8

Supplementary Table T1

Family	Subfamily	Human	Mouse	Rat
CYP1	**1A** 1B	**1A1 1A2** 1B1	**1A1 1A2** 1B1	**1A1 1A2** 1B1
CYP2	2A	2A3 2A4 2A6 2A7 2A13	2A4 2A5 2A12 2A22	2A1 2A2 2A3
	2B	**2B6 2B7**	**2B9 2B10 2B13 2B19 2B23**	**2B1 2B2 2B3 2B12 2B13 2B15 2B21**
	2C	2C8 2C9 2C18 2C19	2C29 2C37 2C38 2C39 2C40 2C44 2C50 2C53 2C54 2C55 2C65 2C67 2C68 2C69 2C70	2C6 2C7 2C11 2C12 2C13 2C22 2C23 2C24 2C79 2C80
	2D	2D6 2D7 2D8	2D9 2D10 2D11 2D12 2D13 2D22 2D26 2D34 2D40	2D1 2D2 2D3 2D4 2D5 2D18
	2E	2E1	2E1	2E1
	2F	2F1	2F2	-
	2J	2J2	2J5 2J6 2J8 2J9 2J11 2J12 2J13	2J2/2J4 2J3/2J9 2J10 2J13 2J16
	2R	2R1	2R1	2R1
	2S	2S1	2S1	2S1
	2U	2U1	2U1	2U1
	2W	2W1	2W1	2W1
CYP3	3A	3A4 3A5 3A7 3A43	3A1 3A11 3A13 3A16 3A25 3A41 3A44 3A57 3A59	3A1/3A23 3A2 3A9 3A18 3A62 3A73

Supplementary Table T2

Protein family and subfamily	Human			Mouse			Rat		
	UniProt ID	Sequence length	Protein PDB ID and template for mouse and rat	UniProt ID	Sequence length	Sequence Identity (%)	UniProt ID	Sequence length	Sequence Identity (%)
CYP1A1	P04798	512	418V	P00184	524	81	P00185	524	81
CYP1A2	P05177	515	2HI4	P00186	513	72	P04799	513	75

Supplementary Table T3

Protein	UniProt ID	Sequence length	Sequence Identity (%)
Human-CYP2B6	P20813	491	Three dimensional structure available. PDB ID: 3QOA
Human-CYP2B7	BA6A7R5	491	92
Mouse-CYP2B9	P12790	491	72
Mouse-CYP2B10	P12791	500	75
Mouse-CYP2B13	A6H6J2	491	72
Mouse-CYP2B19	O55071	492	75
Mouse-CYP2B23	E9Q593	491	75
Rat-CYP2B1	P00176	491	76
Rat-CYP2B2	P04167	491	75
Rat-CYP2B3	P13107	491	67
Rat-CYP2B12	P33272	492	72
Rat-CYP2B13	G3V9D2	494	73
Rat-CYP2B15	Q64583	495	72
Rat-CYP2B21	Q9JJ02	494	73

Supplementary Table T4

(a) CYP1A subfamily proteins

Protein	Human			Mouse			Rat		
	No. of residues in favoured region (%)	No. of residues in allowed region (%)	No. of residues in outlier region (%)	No. of residues in favoured region (%)	No. of residues in allowed region (%)	No. of residues in outlier region (%)	No. of residues in favoured region (%)	No. of residues in allowed region (%)	No. of residues in outlier region (%)
CYP1A1	96.6	3.4	0.0	96.2	3.6	0.2	96.2	3.6	0.2
CYP1A2	95.8	4.0	0.2	95.0	4.2	0.8	95.0	4.2	0.8

Supplementary Table T4 contd.

(b) CYP2B subfamily proteins

Protein	No. of residues in favoured region (%)	No. of residues in allowed region (%)	No. of residues in outlier region (%)
Human-CYP2B6	98.0	2.0	0
Human-CYP2B7	97.2	2.4	0.4
Mouse-CYP2B9	96.8	2.4	0.9
Mouse-CYP2B10	95.8	2.5	1.7
Mouse-CYP2B13	96.5	2.6	0.9
Mouse-CYP2B19	97.0	2.4	0.6
Mouse-CYP2B23	96.5	2.6	0.9
Rat-CYP2B1	96.8	2.6	0.6
Rat-CYP2B2	96.8	2.6	0.6
Rat-CYP2B3	96.5	3.0	0.4
Rat-CYP2B12	97.2	1.9	0.9
Rat-CYP2B13	96.5	2.6	0.9
Rat-CYP2B15	97.4	2.2	0.4
Rat-CYP2B21	96.8	2.6	0.6

Supplementary Table T5.

CYP1A1					
Active site predicted using Sitemap for CYP1A1			Active site predicted using Sitehound for CYP1A1		
Human	Mouse	Rat	Human	Mouse	Rat
Arg106	Arg69	Arg110	Arg106	Arg110	Arg110
Ser122	Thr126*	Thr126*	Met121	Met125	Met125
Ile198	Phe127	Ile326	**Ser122**	Thr126*	Thr126*
Phe224	Phe228	Phe316	Trp131	Phe127	Trp135
Asp310	Leu258	Gly320*	Arg135	Trp135	Arg139
Leu314	Phe262	Ala321*	Leu314	Arg139	Leu318
Gly316	Leu318	Phe323	**Thr321**	Phe228	Ala321*
Ala317	Gly320*	Asp324	Ala325	Asp317	Thr325*
Asp320	Ala321*	Thr325*	Phe376	Leu318	Val386*
Thr321	Gly322	Val386*	Phe381	Gly320*	Thr389
Val382	Thr325*	Ile390	**Val382**	Ala321*	Ile390
Gly459	Val386*	Hip392	Thr385	Asp324	His392
Ala463	Leu500*	Leu500*	Ile386	Thr325*	Gln415
Leu496	Thr501*	Thr501*	His388	Val386*	Leu453
Thr497			Gln411	Thr389	Phe454
			Ile449	Ile390	Gly455
			Phe450	His392	Arg459
			Gly451	Gln415	Lys460
			Arg455	Leu453	Cys461
			Lys456	Phe454	Ile462
			Cys457	Gly455	Gly463
			Ile458	Leu456	
			Gly459	Arg459	
			Ala463	Lys460	
				Cys461	
				Ile462	
				Gly463	
				Leu500*	

Supplementary Table T6.

CYP1A2					
Active site predicted using Sitemap for CYP1A2			Active site predicted using Sitehound for CYP1A2		
Human	Mouse	Rat	Human	Mouse	Rat
Thr104	Arg107*	Arg107*	Arg108	Ile116	Arg107*
Arg108	Met122	Leu143	Leu123	Thr117	Met122
Ile117	Thr117	Ile199	**Thr124**	Thr123	Thr123
Thr118	Thr123	Ile312*	Trp133	Phe124	Trp132
Thr124	Asp311	Ala315	Arg137	Ile218	Arg136
Phe125	Ile312*	Gly316*	**Asp313**	Val219	Asp311
Asp313	Gly314	Thr319	**Ala317**	Ser222	Ala315
Ile314	Val320	Val380	**Leu382**	Phe225	Thr319
Gly316*	Ile384*	Thr383	Thr385	Val226	Ala323
Ala317	Arg455	Ile384*	Ile386	Phe255	Tyr374
Gly318	Arg454*	Leu448	His388	Asn256	Phe379
Thr321	Cys456*	Arg454*	Gln411	Phe259	Val380
Leu382	Ile457*	Cys456*	Leu450	Asn310	Thr383
Ile459	Gly458*	Ile457*	Phe451	Asp311	Ile384*
Gly460	Ala462	Gly458*	Gly452	Gly314	His386
Ala464		Leu495	Arg456	Ala315	Gln409
Thr498			Arg457	Phe317	Leu448
			Cys458	Asp318	Phe449
			Ile459	Thr319	Gly450
				Leu495	Arg454*
				Thr496	Arg455
					Cys456*
					Ile457
					Ala462

Supplementary Table T7.

(a) Human

Human			
Active site predicted using Sitemap for CYP2B6	Active site predicted using Sitehound for CYP2B6	Active site predicted using Sitemap for CYP2B7	Active site predicted using Sitehound for CYP2B7
Arg98*	Arg98*	Arg98*	Arg98*
Val104	Val113	Met113	Met113
Ile114*	Ile114*	Leu114*	Leu114*
Phe115*	Trp121	Phe115*	Arg125*
Trp121	Arg125*	Arg125*	Ala298
Arg125*	Ala298*	Phe206	Leu363
Ala298	Leu363	Ser210	Gly366
Thr302*	Gly366	Phe297	Val367
Val367	Val367*	Glu301	His369
Ser430*	His369	Thr302*	Leu390
Arg434*	Leu392	Thr303	Leu392
Leu437	Pro428	Ser430*	Pro428
Val477	Phe429	Lys433	Phe429
	Ser430*	Arg434*	Ser430*
	Arg434	Gly478	Arg434*
	Ile435	Ile480	Ile435
	Cys436		Cys436
	Leu437		Leu437
	Gly438		Gly438

Supplementary Table T7 continued

(b) Mouse

Mouse									
Active site predicted using Sitemap for CYP2B9	Active site Predicted using Sitehound for CYP2B9	Active site predicted using Sitemap for CYP2B10	Active site Predicted using Sitehound for CYP2B10	Active site predicted using Sitemap for CYP2B13	Active site Predicted using Sitehound for CYP2B13	Active site predicted using Sitemap for CYP2B19	Active site Predicted using Sitehound for CYP2B19	Active site predicted using Sitemap for CYP2B23	Active site Predicted using Sitehound for CYP2B23
Phe114*	Arg98	Arg98*	Leu88	Arg98*	Arg98*	Leu53	Arg99	Arg98*	Arg98*
Ala298	Ile113	Val101	Arg98*	Phe114*	Ile113	Asp312	Val114	Ile104	Val113
Thr302*	Phe114*	Ile114	Val113	Arg125*	Phe114*	Ile357	Ile115	Trp121*	Ile114
Thr303	Trp121	Phe115	Ile114	Tyr206	Arg125	Phe360	Arg126	Arg125*	Trp121*
Gln357	Arg125	Ala116	Trp121*	Ser294	Ser294	Ala364	Ala299	Leu362*	Arg125
Asp361	Val363	Trp121*	Arg125*	Ala298*	Leu295	Pro365	Gly367	Ala363	Ala298
Leu362*	Gly366	Arg125*	Val298	Thr302*	Ala298*	Ser394	Leu368	Ser430*	Leu362*
Val363	Leu367*	Ile363	Ile363	Val362	Val363	Ser395	His370	Arg434*	Ala363
Leu367*	His369	Gly366	Val367	Ala367	Gly366	His398	Leu393	Leu437*	Gly366
Leu396	Pro428	Val367	His369	Ser430*	Ala367	Pro406	Pro429	Leu367*	Leu367*
Pro428	Phe429	Arg443	Pro390	Arg434*	His369	Asp407	Phe430	Phe477	His369
Phe429	Ser430	Leu446	Leu392	Leu437*	Leu392	Ile472	Ser431		Leu392
Cys436	Arg434	Ile486	Phe429		Pro428	Pro473	Arg435		Pro428
Ala442	Ile435		Val440		Phe429	Ser476	Ile436		Phe429
	Cys436		Arg443		Ser430*	Lys480	Cys437		Ser430
	Leu437		Ile444		Arg434*		Leu438		Arg434
			Cys445		Cys436		Gly439		Ile435
			Leu446		Leu437*				Cys436
			Gly447		Gly438				Leu437
									Gly438

Supplementary Table T7 continued

c) Rat

Rat							
Active site predicted using Sitemap for CYP2B1	Active site Predicted using Sitehound for CYP2B1	Active site predicted using Sitemap for CYP2B2	Active site Predicted using Sitehound for CYP2B2	Active site predicted using Sitemap for CYP2B3	Active site Predicted using Sitehound for CYP2B3	Active site predicted using Sitemap for CYP2B12	Active site Predicted using Sitehound for CYP2B12
Arg98* Ile353 His354 Glu355 Ile356 Gln357* Arg358 Ser360 Asp361* Leu362* Ile365 Gly366* Leu392* Phe426 Met427 Pro428* Ser430* Cys436*	Arg98* Val113 Ile114 Arg125 Ala298 Gly366 Val367 His369 Pro428 Phe429 Ser430 Arg434 Ile435 Cys436* Leu437 Gly438	Arg98* Ile104 Phe115 Trp121* Arg125* Phe206 Ala298* Leu367* Ser430* Arg434 Leu437* Ile477	Arg98* Val113 Ile11 4 Arg125* Ala298 Gly366 Leu367* His369 Pro428 Phe429 Ser430* Arg434 Ile435 Cys436 Leu437* Gly438	Ser302 Thr306* Gln357* Ser363 Gly366* Leu367* Leu392* Pro428* Phe429* Ser430* Cys436* Ala442* Leu446	Leu88 Phe95 Ser96 Gly97 Arg98 Val113 Ser114 Arg125 Ala298 Gly366* Leu367 Cys369 Phe390 Leu392* Pro428* Phe429* Ser430* Ile431 Gly432 Lys433 Arg434 Met435	Arg99* Leu105 Ile115* Phe116 Met210 Ser295 Phe298* Ala299 Thr303* Leu363 Thr364 Leu368* Cys437*	Arg99* Val114 Ile115* Arg126 Ala299 Thr364 Gly367 Leu368 His370 Leu393 Pro429 Phe430 Ser431 Arg435 Ile436 Cys437* Leu438 Gly439

					Cys436* Leu437* Gly438		

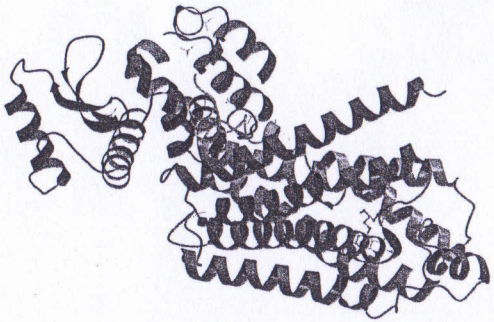
Supplementary Table T7 continued

c) Rat continued

Rat					
Active site predicted using Sitemap for CYP2B13	Active site Predicted using Sitehound for CYP2B13	Active site predicted using Sitemap for CYP2B15	Active site Predicted using Sitehound for CYP2B15	Active site predicted using Sitemap for CYP2B21	Active site Predicted using Sitehound for CYP2B21
Gly97 Arg98 Met113 Asn117 Arg120 Trp121* Leu124 Arg125* Leu367* Ser430* Cys436* Leu437*	Arg98 Met113 Arg125* Ala298 Val363 Gly366 Leu367* His369 Leu392 Pro428 Phe429 Ser430 Arg434 Val435 Cys436* Leu437* Gly438	Arg99* Val114 Ile115* Arg126 Phe207 Phe298* Thr299 Thr303* Ile364 Leu368* Arg425 Cys437* Leu438 Ile478	Arg99* Val114 Ile115* Trp122 Arg126 Thr299 Ile364 Leu368* His370 Phe430 Ser431 Arg435 Ile436 Cys437* Leu438 Gly439	Ala298* Thr302 Ser303 Thr306* Glu357 Asp361* Leu362* Val363 Leu367* Leu396 Pro428* Phe429* Lys436 Ala442*	Arg98 Met113 Ile114 Arg125 Ala298* Val363 Gly366 Leu367* His369 Leu392 Pro428* Phe429* Ser430 Arg434 Val435 Cys436 Leu437 Gly438

Annexure – 4

Certificate



This is to certify that

S. Suvaithenamudhan

Participated and presented a poster titled

On the New Inhibitors for the Resistant 5204 Strain of Penicillin Binding Protein 2B (PBP2B) of Streptococcus pneumoniae through Structure-Based Virtual Screening

in the conference on "Recent Advances in Computational Drug Design" held at J N Tata Auditorium, Indian Institute of Science Campus, Bangalore on 16th Sep -17th Sep 2013.

Organizers

Indian Institute of Science

SCHRÖDINGER®

Innovations in Computational Drug Design

RACDD : 503

CCMB

SciGenom Research Foundation

Institute of Bioinformatics

CERTIFICATE OF PARTICIPATION

This is to certify that

Mr./~~Ms.~~/~~Dr.~~ S. SUVAITHENAMUDHAN, BHARATHIDASAN UNIVERSITY, TRICHY.

Presented a poster at the **2015 NextGen Genomics, Biology, Bioinformatics and Technologies (NGBT) Conference** organized by SciGenom Research Foundation (SGRF), Centre for Cellular and Molecular Biology (CCMB) & Institute of Bioinformatics (IOB) at Hyderabad International Convention Centre (HICC), Hyderabad, India during 1st - 3rd October, 2015

George Thomas, Ph.D.
Director
SciGenom Research, India

Krishnaraj Rajalingam, Ph.D.
Professor
Johannes Gutenberg University, Germany