

University of Mississippi

eGrove

Electronic Theses and Dissertations

Graduate School

2011

Use of Structure-And Ligand-Based Drug Design Tools for the Discovery of Small Molecule Inhibitors of Cysteine Proteases for the Treatment of Malaria and Sars Infection

Falgun Shah

Follow this and additional works at: <https://egrove.olemiss.edu/etd>

 Part of the [Pharmacy and Pharmaceutical Sciences Commons](#)

Recommended Citation

Shah, Falgun, "Use of Structure-And Ligand-Based Drug Design Tools for the Discovery of Small Molecule Inhibitors of Cysteine Proteases for the Treatment of Malaria and Sars Infection" (2011). *Electronic Theses and Dissertations*. 260.

<https://egrove.olemiss.edu/etd/260>

This Dissertation is brought to you for free and open access by the Graduate School at eGrove. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of eGrove. For more information, please contact egrove@olemiss.edu.

USE OF STRUCTURE-AND LIGAND-BASED DRUG DESIGN TOOLS FOR THE
DISCOVERY OF SMALL MOLECULE INHIBITORS OF CYSTEINE PROTEASES FOR
THE TREATMENT OF MALARIA AND SARS INFECTION

A Dissertation

Presented for the

Doctor of Philosophy

In

Pharmaceutical Sciences

The University of Mississippi

Falgun Shah

November 2011

Copyright © 2011 by Falgun Shah
All rights reserved

ABSTRACT

A wide array of molecular modeling tools were utilized to design and develop inhibitors against cysteine protease of *P. Falciparum* Malaria and Severe Acute Respiratory Syndrome (SARS). A number of potent inhibitors were developed against cysteine protease and hemoglobinase of *P. falciparum*, referred as Falcipains (FPs), by the structure-based virtual screening of the focused libraries enriched in soft-electrophiles containing compounds. Twenty one diverse, non-peptidic, low micromolar hits were identified. A combined data mining and combinatorial library synthesis approach was performed to discover analogs of virtual screening hits and establish the structure-activity relationships (SAR). However, the resulting SAR of the identified hits was unusually steep in some cases and could not be explained by a traditional analysis of the interactions (electrostatics, van der Waals or H-bond). To gain insights, a statistical thermodynamic analysis of explicit solvent in the ligand binding domain of FP-2 and FP-3 was performed that explained some of the complex trends in the SAR. Furthermore, the moderate potency of a subset of FP-2 hits was elucidated using quantum mechanics calculations that showed reduced reactivity of the electrophilic center of these hits. In addition, solvent thermodynamics and reactivity analysis also helped to elucidate the complex trends in SAR of peptidomimetic inhibitors of FP-2 and FP-3 synthesized in our laboratory. Multi nanosecond explicit solvent molecular dynamics simulations were carried out using the docking poses of the known inhibitors in the binding site of SARS-3CL^{pro}, a cysteine protease important for replication of SARS virus, to study the overall stability of the binding site interactions as well as

identify important changes in the interaction profile that were not apparent from the docking study. Analysis of the simulation studies led to the identification of certain protein-ligand interaction patterns which would be useful in further structure based design efforts against cysteine protease (3CL^{pro}) of SARS.

DEDICATION

To my loving mom, dad and sister

LIST OF ABBREVIATIONS

3CIP^{pro} = 3C-like protease, C=Chymotrypsin

ADME = Absorption, Distribution, Metabolism, Elimination

Boc = t-butyloxycarbonyl

Cat = Cathepsins

Cbz = Carboxybenzyl

DMF = Dimethylformamide

EDCI = 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide

FCPI = Focused cysteine protease inhibitor library

Gscore = Glide score

HOBT =Hydroxybenzotriazole

HOMO = Highest occupied molecular orbital

HRMS = High resolution mass spectroscopy

HTS = High throughput screening

FMK = Fluoro methyl ketone

FP-2 = Falcipain-2

FP-3 = Falcipain-3

LBD = Ligand binding domain

LUMO = Lowest unoccupied molecular orbital

MD = Molecular dynamics

M^{pro} = Main protease

MLR = Multiple linear regression

MM-GBSA = Molecular mechanics Generalized Born surface area

MM-PBSA = Molecular mechanics Poission-Boltzmann surface area

NMR = Nuclear magnetic resonance

ORF = Open reading frames

OPLS = Optimized potential for liquid simulation

PDB = Protein data bank

P. falciparum = Plasmodium falciparum

PP = polyprotein

QSAR = Quantitative structure activity relationships

RMSD = Root mean square deviation

SAR = Structure-activity relationships

SARS = Severe acute respiratory syndrome

SBVS = Structure-based virtual screening

TEA = triethylamine

W2 = Chloroquine-resistant strain of *P. falciparum*

ACKNOWLEDGMENTS

I thank Dr. Mitchell A. Avery, my research mentor, for presenting me with multiple opportunities to work in a diverse medicinal chemistry projects in his group, for providing me a comprehensive training in the field of synthetic organic chemistry, medicinal chemistry and computational chemistry and for funding my research for five years. He let me pursue my own ideas in malaria project and helped me grow as a capable researcher. In addition to giving leadership for ongoing cross-disciplinary projects in his group, Dr. Avery also allowed me to pursue several research projects outside the university campus in the form of summer internships, giving me an opportunity to prove myself in the challenging research environments of both, pharmaceutical industry and renowned academic institutions.

I am thankful to Dr. Stephen J. Cutler for his constant motivation and appreciations most often via brief notes whenever I received internal or external awards, for allowing me to attend the scientific meetings and external training at times when our group had limited funding and for keeping me on track to conclude my dissertation. I would like to thank Dr. Babu Tekwani for his help with biological assays against cultured malaria parasites and his valuable insights in my research projects. I am thankful to Dr. Robert J. Doerksen for teaching fundamentals of molecular modeling as a part of my training as a computational chemist and his willingness to help in solving problems pertaining to computational chemistry. I also appreciate efforts from all committee members for scheduling my defense on a very short notice.

I thank Dr. John S. Williamson for providing up to date information with graduate

curriculum requirement, for his kind recommendation for the ACS Olemiss section Graduate Student Research Award and for his advices on the career selection. I thank Dr. Christopher R. McCurdy for revising my concepts of medicinal chemistry in a very elegant way during the first year of my Ph.D. and imprinting the importance of networking in individual's scientific career. I appreciate efforts from Dr. John M. Rimoldi for teaching me the fundamentals of bioorganic medicinal chemistry thoroughly, for his constructive comments during my original research proposal and most importantly, for notifying me five years back that I am accepted as a graduate student at the Department of Medicinal Chemistry at Olemiss.

I deeply appreciate efforts of my internship mentors, Dr. Alexander Tropsha at the University of North Carolina, Chapel Hill, Dr. Jack A. Bikker at Pfizer Groton site and Dr. Ruben Abagyan at the University of California, San Diego for providing me extensive training in diverse disciplines of computational chemistry.

Also, I highly appreciate efforts of Dr. Philip Rosenthal, for clearing my concepts of parasites biology, for his help in biological assays of falcipains and his critical evaluation and extensive edits for my first authored manuscripts. I also appreciate efforts of our collaborators Dr. Woody Sherman from Schrodinger Inc., and Dr. Sanjeev Krishna from St. George University of London, UK for sharing their expertise in the advancement of my research project on Malaria.

I am deeply indebted to Dr. Prasenjit Mukherjee, my immediate senior, a lab mate and a very close friend, for sharing his scientific acumen, for helping me understand principles of computational chemistry more closely and precisely, for his advise on a professional and

personal front. I thank Dr. Ronak Patel, an excellent post-doctoral fellow in Dr. Doerksen's research group, a very close friend and a great teacher, for his help in solving technical problems related to computational chemistry in my projects. I thank Dr. Yakamram Pedurri, Dr. Amar Chittiboyina, Dr. Yunshan Wu and Dr. Swapnil Kulkarni for helping me to polish my synthetic chemistry skills and giving me a sufficient confidence that I as a computational chemist am capable of making real molecules when required. I would also like to thank other members of Avery's research group for their continued support. Dr. Rohit Bhat, a close friend, my roommate and a constructive critic and appreciator of lot of my activities deserves a special vote of thanks.

Last but not least, I am grateful to my mother (Mrs. Jyotiben H. Shah), my father (Mr. Hemantbhai D. Shah) and my sister (Mrs. Foram N. Parmar) for their selfless love, continuous support and encouragement to overcome toughest hurdles in my life as I face it.

TABLE OF CONTENTS

CHAPTER	PAGE
1. TARGETING CYSTEINE PROTEASE OF <i>P. FALCIPARUM</i> MALARIA AND SARS FOR DRUG DISCOVERY.....	1
1.1. INTRODUCTION.....	2
1.2. MALARIAL CYSTEINE PROTEASE “FALCIPAIN”: GENESIS, FUNCTION, AND STRUCTURAL REQUIREMENTS.....	5
1.3. DRUG DESIGN APPROACHES AGAINST FPS	9
1.3.1. KNOWLEDGE-BASED DRUG DESIGN STRUCTURE-GUIDED DRUG DESIGN	9
1.3.2. STRUCTURE-GUIDED DRUG DESIGN.....	15
1.3.3. PEPTIDOMIMETICS DESIGN.....	16
1.3.4. STRUCTURE-BASED VIRTUAL SCREENING	18
1.3.5. NATURAL PRODUCT-BASED DRUG DISCOVERY.....	26
1.4. SARS CYSTEINE PROTEASE “3CL ^{PRO} OR M ^{PRO} ”: GENESIS, FUNCTION, AND STRUCTURAL REQUIREMENTS.....	29
1.5. DRUG DESIGN APPROACHES AGAINST SARS 3CL ^{PRO}	32
1.5.1. STRUCTURE-GUIDED DRUG DESIGN.....	32
1.5.2. KNOWLEDGE-BASED DRUG DESIGN.....	38
1.5.3. STRUCTURE-BASED VIRTUAL SCREENING.....	43
1.5.4. HIGHTHROUGHPUT SCREENING (HTS).....	49
1.5.4.1. EXPERIMENTAL HTS.....	50
1.5.4.2. VIRTUAL HTS.....	52
1.5.4.3. DYNAMIC LIGAND SCREENING (DLS).....	55

1.6. CONCLUSIONS.....	56
2. IDENTIFICATION OF FP-2 INHIBITORS USING STRUCTURE-BASED VIRTUAL SCREENING OF FOCUSED CYSTEINE PROTEASE INHIBITOR LIBRARY.....	58
2.1. INTRODUCTION.....	59
2.2. METHODS.....	61
2.2.1. COMPUTATIONAL TOOLS.....	61
2.2.2. BUILDING THE FOCUSED CYSTEINE PROTEASE INHIBITOR (FCPI) LIBRARY.....	61
2.2.3. PROTEIN PREPARATION.....	62
2.2.4. POSE VALIDATION, ENRICHMENT AND DOCKING-POSE BASED PHARMACOPHORE QUERY.....	63
2.2.5. STRUCTURE-BASED VIRTUAL SCREENING PROTOCOL.....	69
2.2.6. BIOLOGICAL EVALUATION.....	71
2.2.7. CALCULATION OF ELECTROPHILICITY OF SOFT-ELECTROPHILES IN IDENTIFIED HITS.....	73
2.3. RESULTS AND DISCUSSION.....	74
2.4. CONCLUSIONS.....	88
3. DESIGN, SYNTHESIS AND BIOLOGICAL EVALUATION OF BENZOTHAZOLE ANALOGS AS FALCIPAIN-2 INHIBITORS	91
3.1. INTRODUCTION	92
3.2. OPTIMIZATION STRATEGIES: CORE HOPPING, STRUCTURE-BASED DESIGN.....	93
3.3. SYNTHESIS SCHEMES.....	96
3.4. BIOLOGICAL EVALUATION.....	97
3.5. RESULTS AND DISCUSSION.....	98
3.6. CONCLUSIONS.....	105

4.	ANALOGS OF FP-2 HITS: VIRTUAL SCREENING, STRUCTURE-ACTIVITY RELATIONSHIPS, SOLVENT THERMODYNAMICS AND REACTIVITY ANALYSIS.....	106
4.1.	INTRODUCTION.....	107
4.2.	METHODS.....	109
4.2.1.	VIRTUAL SCREENING.....	109
4.2.2.	BIOLOGICAL SCREENING.....	111
4.2.3.	THERMODYNAMIC CHARACTERIZATION OF WATERS IN THE BINDING SITE.....	111
4.3.	RESULTS AND DISCUSSIONS.....	112
4.3.1.	SAR IN 1, 2, 3, 4-TETRAZOLE SERIES.....	112
4.3.2.	SAR IN 1, 2, 4-TRIAZOLE SERIES.....	114
4.3.3.	SAR IN QUINAZOLINE SERIES.....	119
4.4.	CONCLUSIONS.....	142
5.	UNDERSTANDING BINDING AFFINITY OF ALPHA-KETO SUBSTITUTED PEPTIDOMIMETICS FOR THE INHIBITION OF <i>P. FALCIPARUM</i>	145
5.1.	INTRODUCTION.....	146
5.2.	METHODS.....	149
5.2.1.	WATERMAP CALCULATIONS.....	149
5.2.2.	CALCULATION OF LUMO DENSITY USING ATOMIC FUKUI INDICES.....	150
5.3.	RESULT AND DISCUSSIONS.....	150
5.4.	CONCLUSIONS.....	163
6.	UNDERSTANDING THE BINDING AFFINITY OF INHIBITORS OF SARS-3CL ^{PRO} BY EXPLICIT MOLECULAR DYNAMICS SIMULATIONS.....	165
6.1.	INTRODUCTION.....	166

6.2. METHODS.....	170
6.3. RESULT AND DISCUSSIONS.....	172
6.4. GUIDELINES FOR FUTURE STRUCTURE-BASED DRUG DESIGN OF SARS- 3CL ^{PRO} INHIBITORS.....	184
6.5. CONCLUSIONS.....	185
APPENDICES.....	186
BIBLIOGRAPHY.....	193
VITAE.....	214

LIST OF TABLES

TABLE	PAGE
1.3.4. Non-peptidic inhibitors identified against FP-2 by virtual screening approach.....	25
1.5.3. Non-peptidic inhibitors identified against SARS-CoV 3CL ^{pro} by virtual screening approach.....	48
2.2.4.1. Reported IC ₅₀ of previously identified FP-2 hits.....	68
2.3.1. Biological evaluation of compounds selected hits from SBVS against FP-2 and cultured parasites.....	74
2.3.2. The pair-wise Tanimoto similarity matrix generated using Canvas, version 1.2 for 21 identified hits. The maximum similarity was 0.5 between compound pairs 4 and 21 which share same core.....	88
2.3.3. The IC ₅₀ values of FP-2 hits against selected human cathepsin peptidases.....	89
3.5.1. Biological activity of benzothiazole analogs.....	98
3.5.2. The IC ₅₀ values of FP-2 hits against selected human cathepsin (Cat) peptidases, K, L1 and B.....	104
4.3.1. Biological evaluation of analogs of virtual screening hit 1 with 1, 2, 3, 4-tetrazole core.....	114
4.3.2. Biological evaluation of analogs of virtual screening hit 2 with 1, 2, 4-triazole core...	117
4.3.3. Biological evaluation of analogs of virtual screening hit 3 with quinazoline core.....	122
4.3.4. Evaluation of selected FP-2 hits against the mammalian cysteine proteases, cathepsins K, L1, and B.....	122
4.3.5. Predicted occupancy and thermodynamic properties of selected WaterMap waters in FP-2.....	128

4.3.6.	Predicted occupancy and thermodynamic properties of selected WaterMap waters in FP-3.....	128
4.3.7.	Biological evaluation of analogs of aryl and alkyl nitriles against FP-2 and FP-3.....	140
5.1.	Activity of compounds 15-50 against falcipain-2 (FP-2), falcipain -3 (FP-3) and cultured malaria parasites. All values shown are IC ₅₀ s in μM.....	148
5.3.1.	Predicted occupancy and thermodynamic properties of selected WaterMap waters in FP-2.....	155
5.3.2.	Predicted occupancy and thermodynamic properties of selected WaterMap waters in FP-3.....	155
5.3.3.	Predicted LUMO density of the most electrophilic center (shown by asterisk *) in compound 42-45	163
6.1.	Results from the SARS-3CL ^{pro} enzyme inhibition assay.....	167

LIST OF FIGURES

FIGURE	PAGE
1.2. (a) Overall topology of FP-2; (b) Magnified view of the substrate-binding pocket of FP-2 with co-crystallized ligand E-64. Subsites S1, S1', S2 and S3 of the enzyme (black), the P1, P1', P2 and P3 side chains of E-64 (red) and residues of FP-2 binding sites (white) are shown.....	9
1.3.1. Representative peptidic inhibitors developed by knowledge-based drug design.....	14
1.3.2. Representative non-peptidic inhibitors developed by structure-guided drug design.....	16
1.3.3. Representative peptidomimetics developed against FP-2 and FP-3.....	18
1.3.4. Natural product-based inhibitors of FP-2.....	27
1.4. a) The overall topology of SARS 3CL ^{pro} . Individual domains, N-finger and catalytic residues, Cys 145 and His 41 are shown; b) magnified view of the substrate-binding site in surface representation. P1-P5 and P1' groups (red) of peptidic inhibitor (gray), subsites S1', S1, S2 and S4 (black), and residues forming the substrate-binding pockets (white) are labeled.....	31
1.5.1. Inhibitors of SARS-M ^{pro} identified by structure-guided drug design.....	37
1.5.2. Inhibitors of SARS-CoV M ^{pro} obtain by knowledge-based drug desig approach.....	42
1.5.4.1. Non-peptidic inhibitors of SARS-CoV M ^{pro} obtained by an experimental HTS.....	52
1.5.4.2. Proposed inhibitors of SARS-CoV M ^{pro} by virtual HTS and DLS approach.....	54
2.1. Representative FP-2 inhibitors (a) covalent irreversible inhibitors; (b) covalent reversible inhibitors and, (c) non-peptidic small molecule inhibitors.....	60
2.2.2. Soft-electrophiles used in the substructure search to build FCPI library.....	62
2.2.4.1. Superimposition of FP-2 (shown in cyan) and Cruzain (shown in orange) using DaliLite server.....	64
2.2.4.2. (a) Magnified view of the substrate-binding pocket of FP-2 (PDB code: 2GHU). Subsites	

S1, S1', S2 and S3 of the enzyme are shown; (b) Pose validation study: superimposition of co-crystallized pose (green) and the top-ranked pose (cyan) obtained through docking; (c) Enrichment study: a bar graph showing percentage of actives recovered at 5% (blue), 10% (red) and 30% (green) of the ranked database with Gscore and Emodel scoring functions; (d) Docking pose-based pharmacophore query: HBD= hydrogen bond donor; HBA=hydrogen bond acceptor; HY=hydrophobic. P1, P1', P2 and P3 side chains of VS and interacting residues Gln36, Cys42, Gly83, Tyr78, Leu84, Leu172 and Ala175 of FP-2 binding site are shown.....	66
2.2.4.3. Chemical structures of previously reported FP-2 hits (IC ₅₀ ≤30μ M) used in enrichment study.....	67
2.2.5. A flow chart depicting the virtual screening protocol utilized in the study. W2 stands for chloroquine resistance strain of <i>P.falciparum</i>	71
2.3.1. Structures of the FP-2 inhibitors identified through SBVS of FCPI library Substructure containing electrophiles of interest are highlighted in blue.....	76
2.3.2. Multiple sequence alignment of FP-2 with homologous cysteine proteases performed using Clustal 2.0.12. Sequence information is from SWISS-PROT (accession numbers: Q9NAW2, Q9NB39, P43235, P07711 and P07858). The conserved residues among all papain-family cysteine proteases are shown in green. The amino acid residues of the S2 pocket differing between FPs and mammalian cysteine proteases are highlighted in red. Catalytic residues are highlighted in “*”.....	78
2.3.3. An evaluation of regional Fukui functions to predict the electrophilicity of atomic site of reference compounds (R1-R3) and selected FP-2 inhibitors is shown here. The most electrophilic center in the compounds is shown by the arrows. The corresponding value of the Fukui index, F _{NN} LUMO, for the reactive atoms (scaled from zero to one) is shown in blue.....	81
2.3.4. Predicted binding mode of virtual screening hits (a) 1 (b) 2 (c) 6 in the FP-2 binding site.....	83
2.3.5. Comparison of similarity of hits identified in the current study with previously identified hits (structure shown in 2.2.4.3) using dbcmp utility of Sybyl 8.1. The median Tanimoto coefficient is about 0.33 (where the graph crosses the 50% threshold), and all of compounds in the current dataset have a Tanimoto of 0.5 or less with their nearest neighbour in the reference database suggesting identified hits are structurally novel....	85

3.1.	Structures and biological activity of 1 against the cysteine proteases: FP-2 and FP-3, the chloroquine-resistant (W2) strain of <i>P. falciparum</i> , and the mammalian cysteine proteases cathepsins (Cat) K, L, and B are shown.....	93
3.2.1.	Docking pose of representative compounds from benzimidazole (3) and thioacetamide (7) series.....	94
3.2.2.	(top) docking pose of compounds 1 in the FP-2 binding site; (bottom) core hopping of compound 1 and 2 with tetrazole and benzothiazole cores (highlighted in blue), respectively are shown. The tetrazole core of 1 was modified to triazole core to take an advantage of click chemistry. The benzothiazole core of 2 was modified to commercially available benzimidazole and thioacetamide core containing compounds (3-10) with anticipated gain in potency.....	95
3.5.1.	Docking pose of 42 (a), 31 (c) in FP-2 and 42 in FP-3 (b).....	101
3.5.2.	The proposed mechanisms of action for the α -thioketone soft-electrophiles containing compounds.....	104
4.1.	FP-2 inhibitors identified by structure-based virtual screening of the Focused Cysteine Protease Inhibitor library($IC_{50}<10\ \mu M$). W2 stands for the chloroquine resistant strain of <i>P. falciparum</i> parasites. Cathepsin (Cat) K, L and B are human cysteine proteases of the papain-family. α -heteroacetamide soft-electrophile is highlighted in blue.....	109
4.2.	Virtual screening workflow to find hits related to compounds 1-3. Substructure queries for 1-3 are shown in blue.....	110
4.3.1.	Chemical structures of compounds with 1, 2, 3, 4-tetrazole core. Original subgraph is highlighted in blue.....	113
4.3.2.	Chemical structures of compounds with 1, 2, 4-triazole core. Original subgraph is highlighted in blue.....	115
4.3.3.	Inactive compounds in 1, 2, 4-triazole series.....	116
4.3.4.	The proposed binding mode of 2 in a) FP-2; b) FP-3; c) Cat K; d) Cat L1 and e) Cat B.....	118
4.3.5.	Effect of R_3 substituents of 1, 2, 4-triazole series on selectivity to FPs and homologous mammalian cysteine proteases of the papain family: Cat K, Cat L1, and Cat B.....	119

4.3.6. Chemical structures of compounds from quinazoline core. Original subgraph is highlighted in blue.....	121
4.3.7. The WaterMap profile of (a) FP-2, (b) FP-3, (c) FP-2 in complex with E-64, and (d) vinyl sulfone in complex with FP-3. The thermodynamically interesting hydration sites important for the SAR are shown in spheres. Stable hydration sites ($\Delta G < 0 \text{ kcal mol}^{-1}$) are colored in green, whereas significantly unstable hydration sites ($\Delta G > 1 \text{ kcal mol}^{-1}$) are shown in purple. The water sites are labeled based on decreasing values of predicted ΔG . Key hydrogen bonding interactions of inhibitors (shown in yellow) with the residues of FP binding sites (shown in grey) are displayed as dotted lines.....	126
4.3.8. The predicted binding poses and thermodynamic hydration site profile of 4 in complex with (a) FP-2 and (b) FP-3; the predicted binding poses and thermodynamic profile of (c) 9 and (d) 14 in complex with FP-2. Water site labeling and color codes are the same as shown in 4.3.7.....	132
4.3.9. The predicted binding poses and thermodynamic hydration site profile of (a) 16 (b) 17; (c) 29 and (d) 36 in complex with FP-2. Water site labeling and color codes are the same as shown in 4.3.7.....	134
4.3.10. The predicted binding poses and thermodynamic profiles of (a) 54 (b) 57 (c) 67 and (d) 68 in complex with FP-2. Water site labeling and color codes are the same as shown in 4.3.7.....	137
4.3.11. Potent inhibitors of FP-2, FP-3, and W2 from the 2-pyrimidinecarbonitrile series reported by Obella et al.....	139
4.3.12. Aryl- or aliphatic nitrile-containing compounds discovered by substructure search in the Chembridge database. Alkyl nitrile electrophiles are highlighted in blue.....	140
4.3.13. The predicted binding poses and thermodynamic hydration site profiles of (a) CN-1 (b) CN-2; (c) CN-3, and (d) 30 in complex with FP-2. Water site labeling and color codes are identical to those shown in 4.3.7.....	141
5.1.1. Overlap of cruzain inhibitor 1 with design template 2.....	146
5.1.2. Chemical structures of synthesized peptidomimetics 15-50.....	147

5.3.1.	The WaterMap profile of (a) FP-2 and (b) FP-3 LBD in complex with 42 is shown. The thermodynamically interesting hydration sites important for the SAR are shown in spheres. Stable hydration sites ($\Delta G < 0 \text{ kcal mol}^{-1}$) are colored in green, whereas significantly unstable hydration sites ($\Delta G > 1 \text{ kcal mol}^{-1}$) are shown in purple. The water sites are labeled based on decreasing values of predicted ΔG	153
5.3.2.	The WaterMap profile of 42 in (a) falcipain -2 and (b) falcipain -3 is shown. The thermodynamically interesting hydration sites important for SAR are shown in spheres. The unstable hydration sites ($\Delta G > 1 \text{ kcal mol}^{-1}$) are shown in purple. The water sites are labeled based on decreasing value of predicted ΔG . Key hydrogen bonding interactions of 42 (shown in yellow) with the residues of falcipain binding (shown in cyan) sites are displayed as dotted blue lines.....	157
5.3.3.	The predicted binding poses and interesting hydration sites for (a) 15 (b) 23 (c) 24 and (d) 49 in FP-2 are shown. Hydration sites labeling and color codes are the same as shown in 5.3.2.....	161
6.1.1.	SARS-3CL ^{pro} inhibitors identified by combined structure/ligand-based drug design approach.....	167
6.1.2.	Interaction profile and surface representation of the poses of the identified hits (a)(b) PJ207, (c)(d) PJ224, (e)(f) PJ225 and (g)(h) PJ316.....	169
6.3.1.	Potential energy plot for the production phase of dynamic simulation of ligand/SARS-3CL ^{pro} complex.....	172
6.3.2.	(a) Docking pose of PJ207 showing the H-bond (red) and hydrophobic (green) monitors used during the simulation; (b) Heavy atom RMSD of the side chain residues of enzyme's binding site (BS), PJ207 and side chain of four residues of BS that showed maximum fluctuations during the simulation; (c) H-bond (H-A) distances and (d) D-H-A angle of His163[-NH]-[O]PJ207 and Glu166[-NH]-[O]PJ207; (e) Hydrophobic [CH-CH] interactions of PJ207 with Thr25, Leu27, Met49 and M165 and a cation(pi)-pi interaction with His172.....	175
6.3.3.	(a) Docking pose of PJ224 showing the H-bond (red) and hydrophobic (green) monitors used during the simulation; (b) Heavy atom RMSD of the side chain residues of enzyme's binding site (BS), PJ224 and side chain of four residues of BS that showed maximum fluctuations during the simulation; (c) H-bond (H-A) distances and (d) D-H-A angle of Asn142[O]-PJ224[NH], His163[-NH]-[O]PJ224 and Gly-1[NH]-PJ224[O]; (e)	

	Hydrophobic [CH-CH] interactions of PJ224 with Thr25, Met49, and Met165 and a cation(pi)-pi interaction with His172.....	177
6.3.4.	(a) Docking pose of PJ225 showing the H-bond (red) and hydrophobic (green) monitors used during the simulation; (b) Heavy atom RMSD of the side chain residues of enzyme's binding site (BS), PJ225 and side chain of four residues of BS that showed maximum fluctuations during the simulation; (c) H-bond (H-A) distances and (d) D-H-A angle of Asn142[NH]-PJ225[O], Ser144[OH]-PJ225[O] and His163[NH]-[O]PJ225; (e) Hydrophobic [CH-CH] interactions of PJ225 with Thr25, Met49, Phe140, Leu141 and Met165 and a cation(pi)-pi interaction with His172.....	180
6.3.5.	(a) Docking pose of PJ316 showing the H-bond (red) and hydrophobic (green) monitors used during the simulation; (b) Heavy atom RMSD of the side chain residues of enzyme's binding site (BS), PJ316 and side chain of four residues of BS that showed maximum fluctuations during the simulation; (c) H-bond (H-A) distances and (d) D-H-A angle of Phe140 [O]-[NH]PJ316, G143[-NH]-[CO]PJ316, Ser144 [NH]-[O]PJ316 and H163[-NH]-[O]PJ316; (e) Hydrophobic [CH-CH] interactions of PJ316 with Thr25, Met49, Leu141 and Met165 and a cation(pi)-pi interaction with His172.....	182

LIST OF SCHEMES

SCHEMES	PAGE
3.1.....	97
3.2.....	97

Chapter 1

Targeting Cysteine Protease of *P. Falciparum* Malaria and SARS for Drug Discovery

Content of this chapter is published in: Shah, F., Mukherjee, P., Desai, P., Avery M. A., Computational approaches for the discovery of cysteine protease inhibitors against Malaria and SARS. *Current Computer Aided Drug Designing*, 2010, 6, 1.

1.1. Introduction

Proteases are an important class of enzymes involved in the hydrolysis of peptide bonds in protein. Proteases can be subdivided into five major categories based on their mechanism of peptide hydrolysis: Aspartic, Serine, Cysteine, Threonine and Metalloprotease ². They play a vital role in a variety of cellular and physiological events in lower order, viruses extending up to higher organisms including humans, making them attractive targets for further considerations ³. Among these proteases, cysteine proteases, with their unique mechanism and diverse array of functions, have become important targets for drug design.

Cysteine proteases are classified in three structurally distinct groups, namely papain-like (e.g., clan CA, the cathepsins, falcipain), ICE-like (e.g., clan CD, the caspases) and picornain-like (e.g. clan PA (C), SARS 3CL^{pro}) ⁴. Different cysteine proteases are associated with a variety of clinical conditions in humans, for example, Caspase-I (psoriasis, inflammation), Cathepsin S (autoimmune or inflammatory disorders) and Cathepsin K (osteoporosis) ⁵. They are also present in the life cycle of protozoa, for example, falcipain from *P. falciparum* (malaria) and cruzain from *T. cruzi* (chaga's disease), as well as in viruses such as HRV 3CL^{pro} from the rhinovirus (common cold) ⁶ and SARS 3CL^{pro}, an etiological agent responsible for the Severe Acute Respiratory Syndrome (SARS) infection ⁷.

The proteolytic mechanisms of cysteine proteases involve a nucleophilic attack of the thiolate anion from the catalytic cysteine on the carbonyl carbon of the labile peptide bond. This leads to the formation of an acyl-enzyme (tetrahedral) intermediate, stabilized by the hydrogen-bonding to the residues of the oxyanion hole. Subsequent hydrolysis of the tetrahedral intermediate releases the free enzyme ³. Although, selectivity among the members of cysteine protease from the same family is of great concern, some inhibitors have now entered into various

phases of pre-clinical or clinical development ⁵, which has attracted the attention of researchers for further drug design against this target.

The present chapter highlights various rational approaches adopted for the discovery of peptidic and non-peptidic inhibitors against parasitic cysteine proteases: falcipains, a papain-like protozoal cysteine protease, and SARS 3CL^{PrO}, a chymotrypsin-like viral cysteine protease. Although these two proteases are quite different from each other in structure and specific biological functions they exhibit in protozoa or virus, similarity exist in drug design approaches against both proteases. The drug design approaches discussed below include knowledge-based drug design, structure-guided drug design, virtual and high-throughput screening, peptidomimetic design and natural product-based drug discovery that provide an absolute understanding of drug discovery process against both the targets. Particular emphasis has been given to the computational approaches adopted for the design of inhibitors or understanding the mechanism of inhibitor binding. The drug design strategies discussed in the next sections are introduced here briefly in the context of cysteine proteases.

Knowledge-based drug design initiates drug discovery efforts based on the similarity of a cysteine protease under consideration with other members of the same or different protease family. This approach merges information obtained from the knowledge of the substrate requirement of a given protease with the known inhibitors of the homologous proteases while design of inhibitors against the protease under consideration. Knowledge-based drug design is particularly important when only a limited option exists for the treatment against a particular disease. For example, during the outbreak of SARS, in the absence of efficacious drugs, analogues of rupintrivir (a rhinovirus protease inhibitor belonging to the C3 family and

approved for the clinical trials in humans for treatment of cold) served as a starting point for the drug design against SARS-CoV M^{pro} (member of the C30 family of cysteine proteases) ⁸.

Structure-guided drug design considers the design of the inhibitors based on the X-ray crystal structure ⁹⁻¹¹ or homology model of the given protease ^{12, 13} taking into consideration the nature of the specificity subsites and interactions with the key residues of the active site. Rupintrivir developed by Pfizer Inc. is the best example of the application of structure-guided drug design in lead discovery ¹⁴. Peptidomimetics design against cysteine protease usually involves an address element that mimics the natural peptide sequence and sometimes the secondary structures of protein such as β -turn ¹⁵. The addressing element is important for subsite recognition and binding to the target enzyme. A warhead group or electrophilic deficient center is present in addition to the addressing element, which replaces the scissile amide bond of the substrate and is capable of reacting in reversible or irreversible fashion with the active site thiol ¹⁶. Improved pharmacokinetics properties of the peptidomimetics due to their stability against hydrolytic enzymes as well as reduced toxicity to the human host suggest the preference of peptidomimetics design over the natural peptide ¹⁷.

Highthroughput screening (HTS) provides an opportunity to simultaneously screen a large number of compounds originated from the diverse chemotypes with the help of a sensitive and robust assay method¹⁸. Structure-based virtual screening involves in silico evaluation of a database of molecules against an experimentally determined structure or a homology model of protein target using a docking program. The docking poses thus obtained are ranked based on a scoring function to identify putative binders. Thus, only a small set of compounds are selected and evaluated in a biological assay as compared to the high throughput screening of an entire compound library ¹⁹.

The following sections describe comprehensive drug discovery efforts conducted against FPs and SARS-3CL^{Pro} based on rationale drug design approaches discussed above. The drug development strategies described below could also be applied for the design of inhibitors of cysteine protease present in existing and future emerging pathogens, humans or design of inhibitors for all cysteine proteases in general.

1.2. Malarial cysteine protease “falcipain”: genesis, function and structural requirements

Malaria is a global health issue, causing close to a million deaths annually according to recent death statistics from the World Health Organization (WHO) ²⁰. It is caused by four major species of the *Plasmodium* (*P*) genus in humans: *P. falciparum*, *P. vivax*, *P. ovale*, and *P. malariae*. The most virulent strain of malaria-causing pathogen is *P. falciparum*, responsible for the highest morbidity, mortality and spread of the disease ²¹. The regions where malaria is endemic include tropical and subtropical regions of Africa, Asia and South America.²² Falciparum malaria is a particular threat for Africans, especially children under the age of five. Approximately 90% of malaria cases are reported from developing countries in Asia, South America, and Sub-Saharan Africa, and so there is a particular need for inexpensive, easy to use, and well-tolerated malaria treatments²³. A number of drugs are currently available to treat malaria²⁴, however, treatment is complicated by the toxicity, high cost, and diminishing efficacy of different agents. The first known treatment of malaria occurred in 1631 with the bark of the cinchona tree, from which quinine was later purified²⁵. As quinine is not well tolerated, new therapeutics were developed, including chloroquine and related aminoquinolines, antifolates, and mefloquine. However, strains of *P. falciparum* resistant to all these therapeutics have been observed.²⁶⁻²⁸ Artemisinin, a natural product from *Artemisia annua*, and its derivatives have been identified as rapidly active

and highly efficacious antimalarials²⁸. In recent years the standard of care for the treatment of uncomplicated falciparum malaria has moved to artemisinin-based combination therapy, including an artemisinin component (artesunate, artemether, or dihydroartemisinin) combined with a longer acting agent²⁹. However, *P. falciparum* strains have been found showing decreased responsiveness to artemisinin therapy, in particular near the Thailand-Cambodia border²⁶. This finding validates a continued need for effective alternative therapeutics targeting key pathways in the life-cycle of malaria parasites.

Several proteases play an essential role in the life cycle of the *P. falciparum*, which includes aspartic protease “plasmepsins”, cysteine protease “falcipains” and the metalloprotease “falcilysin”³⁰⁻³². Among these proteases, falcipains have gained increasing attention due to their critical roles in the survival of protozoa. The *P. falciparum* genome contains 33 open reading frames (ORFs) expressing four different cysteine proteases of the papain family (clan CA), collectively known as falcipains³³⁻³⁶. These include falcipain-1 (FP-1), falcipain-2 (FP-2), falcipain-2’ (FP-2’) and falcipain-3 (FP-3)³⁷.

The biological roles of falcipains are reviewed elsewhere in detail³⁸. In brief, their suggested roles involve erythrocytic invasion, hydrolysis of hemoglobin and erythrocytic rupture. Among FPs, FP-2 and FP-3, encoded by nearby genes, are localized to the *P. falciparum* food vacuole, the site of hydrolysis of large quantities of hemoglobin by erythrocytic parasites. FP-2 and FP-3 have been shown to be hemoglobinases³⁸, and they appear to be responsible for the hydrolysis of the host erythrocyte hemoglobin to provide amino acids required for protein synthesis in protozoa^{34, 36, 39}. Knockout of FP-2 led to parasites with diminished hemoglobin hydrolysis³⁹. Knockout of FP-3 was not possible, strongly suggesting that this protease is

essential for erythrocytic parasites³⁷. Functions of the other falcipains are less well understood. FP-1 activity was associated with erythrocyte invasion by *P. falciparum* merozoites⁴⁰, but knockout of the protease did not alter erythrocytic parasites⁴¹, and its primary function is more likely in mosquito-stage parasites⁴². FP-2' is a near-identical copy of FP-2, but knockout of FP-2' had no apparent phenotype, indicating that the protease is not essential³⁷. In particular, inhibition of FP-2 and -3 by covalent irreversible cysteine protease inhibitors blocked hemoglobin hydrolysis and formed characteristic swollen, dark-stained food vacuoles due to accumulation of unhydrolysed globin^{43, 44}. These morphological abnormalities are typical characteristics of cysteine protease inhibition not elicited by inhibitors of other food vacuole proteases such as aspartic protease⁴⁵. In any event, inhibition of FP-2 and FP-3, blocks parasite development, enabling these enzymes as viable targets for the development of antimalarial chemotherapy.

During the early phases of the drug discovery against these plasmodial hemoglobinase in the absence of the X-ray crystal structure, several homology models were reported^{13, 46-48}. Multiple crystal structures of FP-2 and FP-3 became available later on⁴⁹⁻⁵² which not only revealed the overall topology of FP-2 and FP-3 but also provided an opportunity to rationally explore the structural determinant of falcipains⁴⁹. The crystal structure of FP-2 (3BPF, protein data bank (PDB) code) shown in Figure 1.2.1a) has been discussed briefly here. FP-2 has a distorted ellipsoidal shape, a typical attribute of the papain-like cysteine proteases. In addition to a classic papain-like fold, FP-2 has two distinct domains, left domain (L) and right domain (R) (Figure 1.2.1a). The L domain is made up of α -helices whereas the R domain is mainly composed of antiparallel β -sheets⁴⁹. Both FP-2 and FP-3 have two unique structural features designated as N-terminal extension and hemoglobin binding insert that distinguish them from all

other structurally characterized papain family enzymes ⁵⁰. N-terminal extension is involved in the folding of enzyme into its mature form whereas hemoglobin binding insert captures the host erythrocyte hemoglobin ⁴⁹. The substrate-binding site is located at the junction of these two domains bearing a catalytic triad formed by Cys-42, His-174 and Asn-204. Figure 1.2.1b displays a magnified view of the substrate-binding site of FP-2 in complex with an irreversible epoxysuccinate (E-64) inhibitor. Subsites S1, S1', S2, S3 and the P1, P1', P2 and P3 side chain of E-64 are shown in Figure 1.2.1b.

Both FP-2 and FP-3 share a high sequence identity of 66.7% ⁵⁰. The homology models and the X-ray crystal structures of FP-2 and FP-3 revealed major differences in the S2 pocket of the active site with more flexible Glu 243 in FP-3 as compared to Asp 234 in FP-2 as well as bulkier Tyr 93 in FP-3 with corresponding Leu 84 in FP-2 ^{13, 50}. The presence of the bulkier residues narrow down the S2 binding pocket of FP3 as compared to FP2 and might be responsible for the kinetic difference between these closely related enzymes ⁵⁰. These two proteases contribute almost equally to the hemoglobin degradation with the former expressing at the earlier stages in the life cycle of protozoa ³⁶. The similarity in the functions exhibited by FP-2, FP-2' and FP-3 point towards the combined drug discovery efforts against these subtypes ^{35, 53}. Substrate-cleavage assays suggest marked preference for hydrophobic residue, particularly Leu at P2 for FP-2 ³⁴ and FP-3 ³⁶. These studies gave important insights for the designing of inhibitors against both the enzymes.

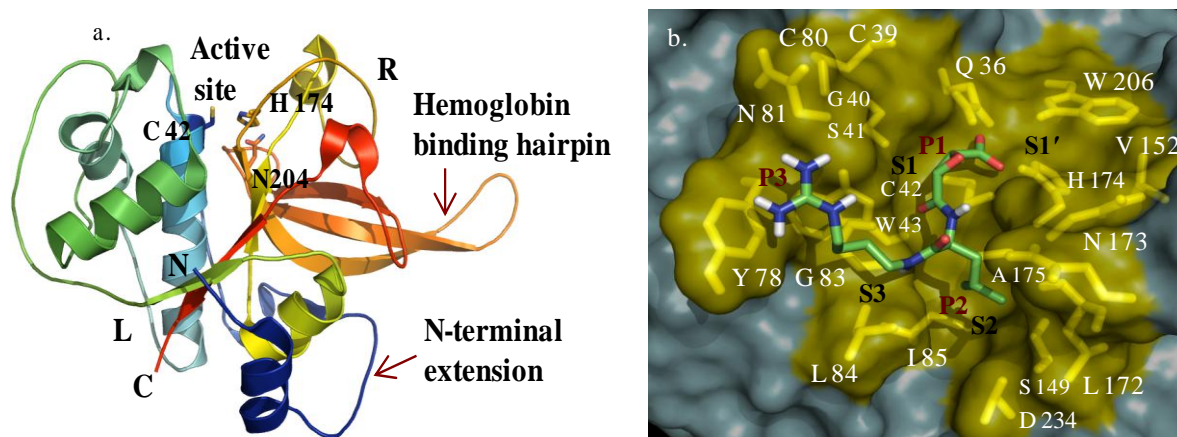


Fig. 1.2. (a) Overall topology of FP-2; (b) Magnified view of the substrate-binding pocket of FP-2 with co-crystallized ligand E-64. Subsites S1, S1', S2 and S3 of the enzyme (black), the P1, P1', P2 and P3 side chains of E-64 (red) and residues of FP-2 binding sites (white) are shown.

1.3. Drug Design Approaches Against FPs

1.3.1. Knowledge-Based Drug Design

The drug design efforts were initiated against FPs with the previous knowledge of inhibitors of other cysteine proteases such as papain, cathepsins, calpain, and cruzain having reversible electrophilic isostere such as peptidic aldehyde⁵⁴, α -keto carbonyl^{55, 56}, nitriles⁵⁷ as well as irreversible warhead groups such as halo methyl ketones⁵⁸, epoxysuccinyl⁵⁹ and vinylsulphones⁶⁰ etc.

Early efforts for the validity of FP-2 (formerly known as trophozoite cysteine proteinase or TCP) by Rosenthal and colleagues as a possible chemotherapeutic target led to the evaluation of a panel of peptidyl fluoromethyl ketones (FMKs) against TCP. Benzylloxycarbonyl (Z)-Phe-Arg-CH₂F (**1**, Figure 1.3.1), was found to inhibit TCP proteolytic activity in the sub-nanomolar concentration ($IC_{50} = 0.36$ nM) and parasite killing ($IC_{50} = 64$ nM) in the nanomolar range⁴³. In addition, Morpholine urea (Mu)-Phe-Homophe(HPhe)-CH₂F (**2**), one of the FMK inhibitors, cured 80% of *P. vinckei*-infected mice under test conditions having biochemically similar

cysteine protease as *P. falciparum*⁶¹. However, severe toxicity associated with the use of FMKs in animal models restricted their further development⁶².

Epoxysuccinyl peptides derived from E-64 (**3**, Figure 1.3.1) containing *trans*-epoxysuccinic acid as electrophilic element are selective inhibitors of cysteine proteases from the papain superfamily and involve the irreversible activation of the enzyme by alkylation of the active site cysteine⁶³. Several groups have reported the efficacy of E-64 and its derivatives as an antiplasmodial agent with activities in the nanomolar range against falcipains as well as cultured parasites^{34, 43, 64, 65}. However, their lack of selectivity against different members of the papain family limits their further development as drugs.

Various mechanism-based vinylsulphone inhibitors were reported against closely related cysteine proteases during the late 90's⁶⁰. Rosenthal *et al.* performed the synthesis of a previously identified potent cysteine protease inhibitor **4** as well as closely related compounds with different side chains of P1 and P2 amino acids and evaluated the antimalarial potential of vinyl sulphones⁶². Among the synthesized compounds, compound **4** effectively inhibited the *P. falciparum* (IC₅₀= 3 nM) cysteine protease as well as blocked hemoglobin hydrolysis and development of parasites at nanomolar concentration. Further, orally bioavailable derivatives of vinyl sulphones were reported by replacing the morpholine urea group with the N-methyl piperazine urea⁶⁶. Also, the phenyl vinyl sulphone was replaced with a naphthalene vinyl sulphone. Compound **5** (IC₅₀ of 5 nM for *P. falciparum* cysteine protease) was active *in-vivo* against murine malaria (cured 40% of infected mice after oral doses of 50mg/kg twice a day)⁶⁶. Shenai *et al.* synthesized 39 new compounds having vinyl sulphones, vinyl sulfonamide and vinyl sulfonate esters as warhead groups to establish the structure activity relationships⁶⁷. Inhibitors with Leu at

the P2 position and homophenylalanine (HPhe) at the P1 position displayed potent inhibitory activity against both the enzymes as well as cultured parasites, which point towards importance of these residues for substrate specificity against FP-2. Vinyl sulphonate esters were the most active (compound **6**, FP-2 IC₅₀ = 0.9 nM, W2 IC₅₀ = 9.7 nM), followed by vinyl sulfonamides (compound **7**, FP-2 IC₅₀ = 2.3, W2 IC₅₀ = 4.4 nM), and finally phenyl vinyl sulfones (compound **8**, FP-2 IC₅₀ = 6.9 nM, W2 IC₅₀ = 3.9 nM) among the synthesized series of compounds⁶⁷ (Figure 1.3.1). (Note: W2 and D6 stand for chlorquine resistant and chloroquine sensitive strains of the *P. falciparum*, respectively).

Further, conformationally restricted analogues of sulfones were designed with vinyl sulfone attached to the dipeptide sequence in order to improve their potency and metabolic stability. Compound **9** with trans (*E*) geometry of the double bond inhibited the enzyme as well as cultured parasite in micromolar range (FP-2 IC₅₀ = 13.7 μM, W2 IC₅₀ = 7.8 μM)⁶⁸. The substantial reduction in the activity of these compounds compared to corresponding acyclic analogs (for example **4**) prompted the authors to predict the binding mode of **9** in the active site of the homologous enzyme, cruzain. The predicted conformation of compound **9** disfavored the Michael addition of cysteine to the β-carbon of the vinyl-sultam due to interference of the ring substitution on the α-carbon of the Michael acceptor and, therefore, could be responsible for a lower activity⁶⁸ (Figure 1.3.1).

The most concerning drawback of vinyl sulphone inhibitors includes irreversible alkylation of the active site cysteine residue by conjugate addition. To encounter these issues, closely related peptides bearing reversible warhead groups such as aldehyde and α-ketoamide were developed⁶⁹. Representative compounds from the aldehyde (compound **10** native FP-2 IC₅₀

= 1 nM, W2 IC₅₀ = 1.1 nM) and ketoamide series (compound **11** native FP-2 IC₅₀ = 1 nM, W2 IC₅₀ = 2.9 nM) are shown in Figure 1.3.1. Both of these compounds were active against multiple strains of *P.falciparum* ⁶⁹.

Phenolic Mannich bases were previously reported to possess anticancer as well as antimalarial properties ⁷⁰. The bioactivities of these compounds were thought to involve *in situ* generation of α , β -unsaturated ketones possessing a higher affinity for cysteine thiols. Also, thiosemicarbazone scaffold-bearing compounds were reported to have a broad spectrum of chemotherapeutic activity including anti-malarial activity ⁷⁰. These compounds are Schiff bases, having iron-chelator properties as well as acting as electrophilic warheads to attract cysteine thiol. Chipeleme *et al.* appended the phenolic mannich base group to various (thio) semicarbozaone entities and linked them with aminoquinolines (known to inhibit hemazoin formation). Compound **12** from this series exhibited potent inhibition of FP-2 (IC₅₀ = 0.63 μ M) and W2 strain of *P. falciparum* (IC₅₀ = 0.27 μ M, Figure 1.3.1) in the lower micromolar range ⁷⁰.

Aziridines compounds are selective and irreversible inhibitors of cysteine protease. They are aza analogues of epoxides ⁷¹. These compounds were previously evaluated against the cysteine protease of SARS, Leishmania, and *Trypanosoma brucei*, as well as cathepsin K, and they inhibited the protease by covalent adduct formation with cysteine thiol of the active site. Schulz *et al.* evaluated three class of aziridines, aziridine-2-carboxylic acid derivatives, *N*-acylated aziridine-2, 3-dicarboxylic acid derivatives and aziridines-2-carboxylate containing a lysine residue, against FP-2 and FP-3 ⁷². Two of the most active compounds, compound **12** from the aziridines 2-carboxylate series (FP2 IC₅₀ = 2.5 μ M, FP3 IC₅₀ = 3.8 μ M, W2 IC₅₀ = 9.7 μ M) and compound **13** from the *N*-acylated aziridine-2, 3-dicarboxylic acid series (FP2 IC₅₀ = 0.079

μM , FP-3 $\text{IC}_{50} = 0.25 \mu\text{M}$, W2 $\text{IC}_{50} = 0.43 \mu\text{M}$) are shown in Figure 1.3.1 ⁷².

Recently, a multiple ligand approach was reported against FP-2, which involved the combination of artemisinin, a sesquiterpene lactone, and a potent antimalarial compound with dipeptidyl vinyl sulphones via a linker to combat multi-drug resistance in *P. falciparum* malaria ⁷³. The compound **15** exhibited potent inhibitory activity of FP-2 (FP2 $\text{IC}_{50} = 0.21 \text{ nM}$). Also, it had a superior activity profile against multiple strains of malaria as compared to artemisinin and chloroquine ⁷³.

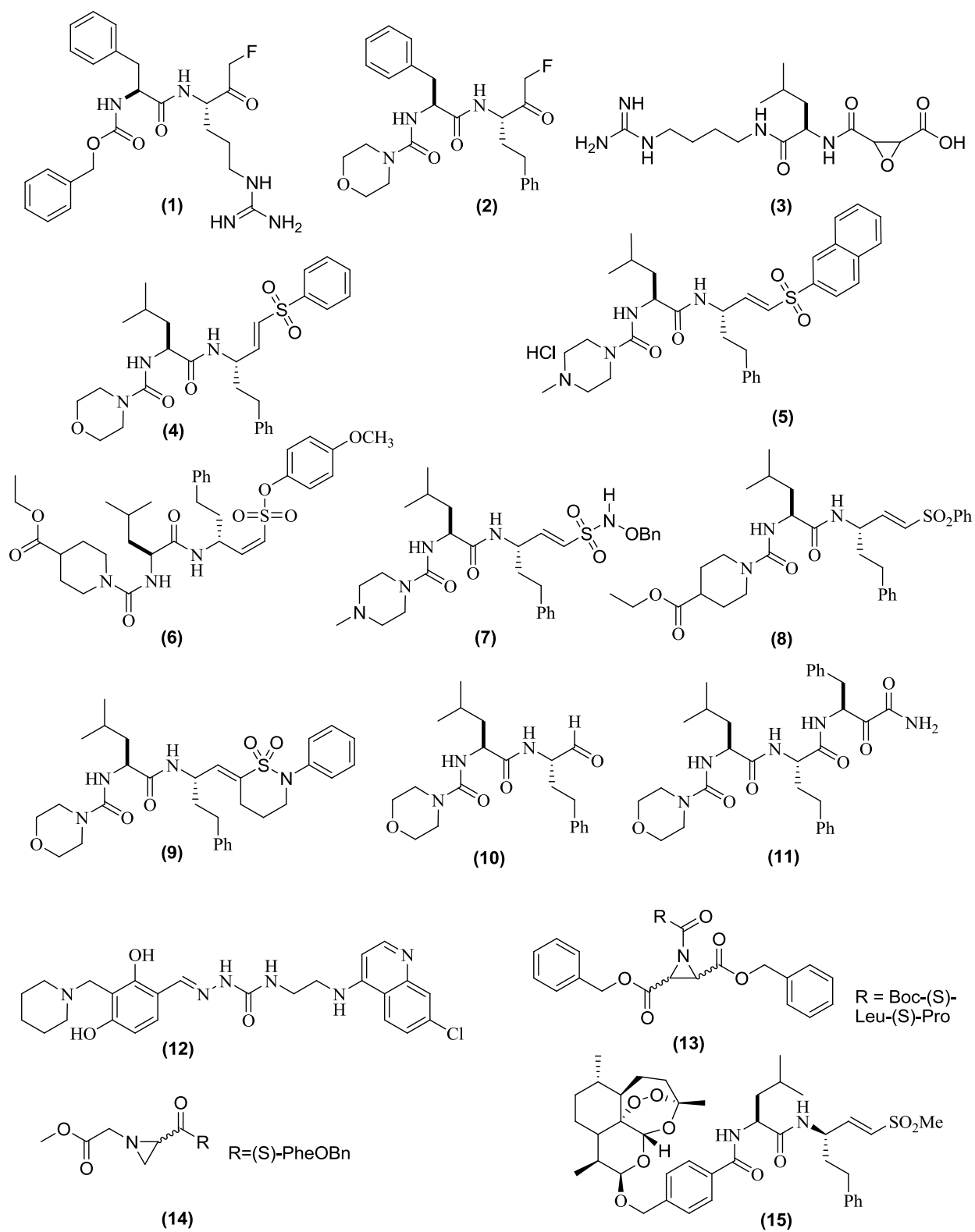


Figure 1.3.1. Representative peptidic inhibitors developed by knowledge-based drug design.

1.3.2. Structure-Guided Drug Design

Sabnis *et al.* developed homology models of FP-2 and FP-3 and validated these models with the help of various structures/geometry tools as well as by the docking of substrate/known vinyl sulphone inhibitors^{13, 46}. The validated model was used to design novel non-peptidic analogs having clinically validated isoquinolines core. Two compounds were designed based on the intuitions from the docking studies. *In-vitro* evaluation of these compounds against FP-2 showed activity in the micromolar range (most active compound **16** with FP-2 IC₅₀ of 8 μM)¹³. Several isoquinolines and dihydroisoquinolines analogs were designed to further optimize the activity of compound **16**, considering the interaction of designed analogs with Asp234 of the S2 pocket as essential¹². The synthesized compounds vary in the IC₅₀ range from 3 to 10 μM (representative compound **17**, FP2 IC₅₀ = 3μM, Figure 1.3.2). Docking predictions of the compounds from these series were in an excellent agreement with the *in-vitro* activity, having correlation coefficient of 0.94.

Goud *et al.* addressed the *de-novo* design of novel non-peptidic inhibitors bearing the thiazoline scaffold⁷⁴. The thiazoline core was chosen because of the favored key interactions and the spatial arrangement of the designed trisubstituted thiazoles in the active site of FP-2. Also, the thiazole core was able to gain a planar conformation and could mimic the amino acid portion of the peptide bond in the molecular modeling studies. In addition to this, docking studies predicted a similar binding mode and interaction energies of these designed thiazole compounds to the known vinyl sulphone inhibitors. This led to the dedicated synthesis and biological evaluation of a series of trisubstituted thiazole analogs against FP-2 and FP-3⁷⁴. Compound **18** was a dual inhibitor of both enzymes (FP-2 IC₅₀=6.6 μM, FP-3 IC₅₀=29.4 μM) as well as

showing activity against cultured malarial (W2 EC_{50} = 13.8) and leishmanial parasites (EC_{50} = 4.5 μ M) (Figure 1.3.2).

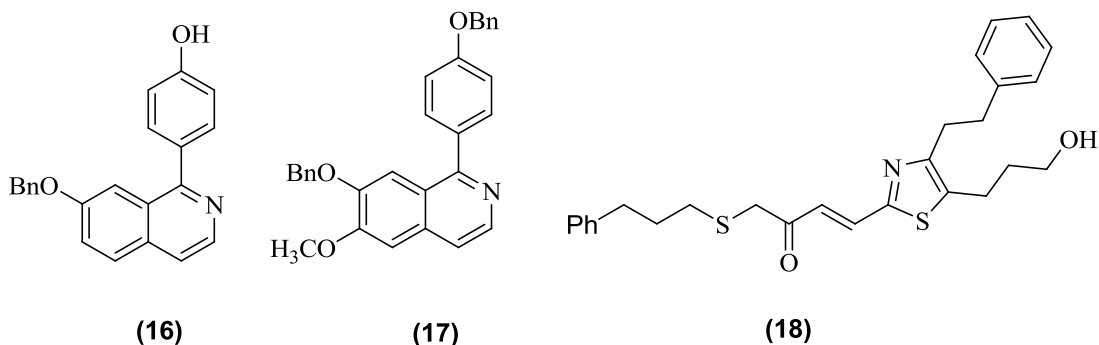


Fig.1.3.2. Representative non-peptidic inhibitors developed by structure-guided drug design.

1.3.3. Peptidomimetics Design

Verissimo *et al.* undertook modeling studies to compare the low energy conformations of their proposed peptidomimetics with vinyl sulphone inhibitor **4** by substituting leucine at P2 of vinyl sulphone with pyridone⁷⁵. Their designed molecules adopted a similar low energy conformation to that of a vinyl sulphone inhibitor in 3D space. Synthesis of the designed peptidomimetics with pyridone as a P2 recognition element and different electrophilic isostere at P1' positions were performed. Compounds **19** and **20** with aldehyde and vinyl sulphone warhead groups showed an IC_{50} value of 10.9 μ M and 19.0 μ M, respectively, against FP-2⁷⁵ (Figure 1.3.3). Although molecular modeling studies showed an excellent overlay of the peptidic counterparts of these compounds, the lower activity of the designed compounds suggested that the rigid pyridone moiety chosen was not suitable for the backbone modifications.

Nicola *et al.* designed aldehyde-based peptidomimetics with benzodiazepines (BDZ) scaffold as a mimetic of the fragment D-Ser-Gly of peptide and the aspartic aldehyde building block at C-terminal that was known to form a reversible covalent bond with the enzyme active

site ⁷⁶. They also considered the excellent bioavailability, tolerability and β -turn mimetic properties of BDZ scaffold in their peptidomimetics design. Effort towards the synthesis of the designed peptidomimetics discovered compound **21** with IC_{50} of 10.29 μ M and 19.31 μ M against FP-2 and FP-2', respectively. Also, these compounds were selective to malarial protease as evident by their inactivity against recombinant human caspases 1-9 for up to 50 μ M concentration ⁷⁶.

Further, the P1 and P1-P1' positions of the most active compound **21** were modified to explore the structural requirements for this class of peptidomimetics. The aspartic aldehyde building block at the P1 position of **21** was replaced with an electrophilic vinyl sulphone moiety spanning the P1-P1' position (for covalent trapping of cysteine thiol of the active site) with different aromatic substitution (to explore the size and characteristics of the lipophilic P1' pocket) ¹⁵. A second modification involved HPhe at the P1 position (a residue known to boost the potency of FP-2 inhibitors) and, finally, the placement of glycine (Gly) at the same position (to evaluate the process of ligand recognition as well as its impact on inhibitory kinetics by the enzyme) ¹⁵. As expected, compounds with the HPhe at P1 position showed higher second order rate constant as compare to glycine at the same position ¹⁵. The representative compound (**22**, Figure 1.3.3) from this series showed the highest inhibition of FP-2 coupled with good inhibition potency against the cultured parasites (W2 IC_{50} = 9.1 μ M). Attempts were made to replace the vinyl sulphone moiety with vinyl phosphates, but only resulted in compounds exhibiting poor activity and selectivity profile ⁷⁷. Recently, another strategy presented by the same group focused on the evaluation of the strength and efficacy of different α , β -unsaturated electrophilic withdrawing groups such as vinyl-ketones, amides, esters and nitriles against cysteine thiol ⁷⁷.

Among the synthesized compounds, vinyl ester **23** displayed potent inhibition of FP-2 ($K_i = 17$ nM) as well as significant inhibition of the *P. falciparum* FCBR strain ($IC_{50} = 12$ μ M). However, this compound was also selective towards human cysteine protease, cathepsin B and L.

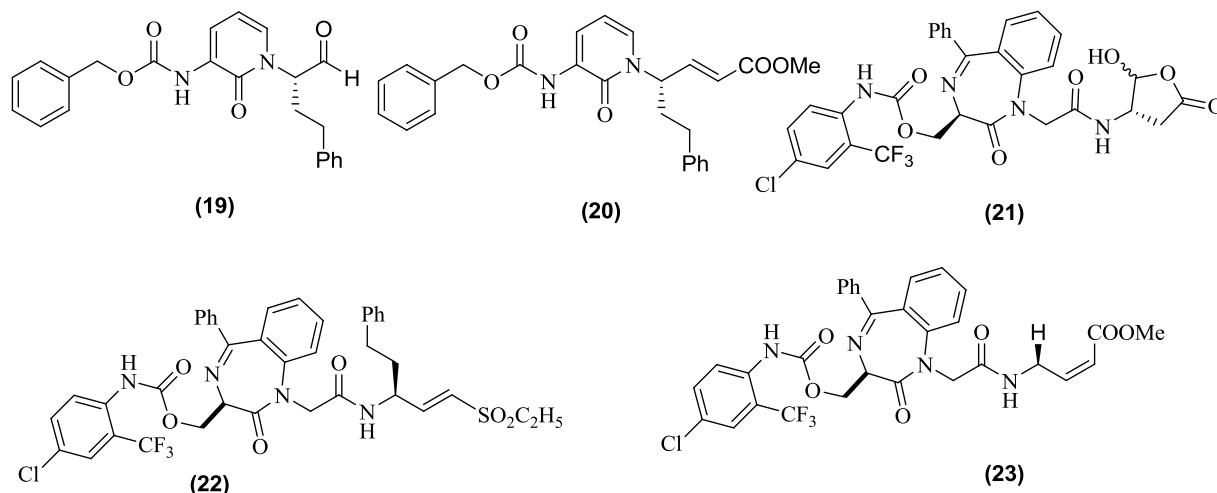


Fig.1.3.3. Representative peptidomimetics developed against FP-2 and FP-3.

1.3.4. Structure-Based Virtual Screening

This section will review the discovery of the non-peptidic inhibitors developed against FP-2 and FP-3 by a structure-based virtual screening approach. Table 1.3.4 summarizes the commercial databases, methods, softwares and scoring functions used for virtual screening, hit ratio along with chemical structures of hits obtained.

Ring *et al.* initiated the structure-based virtual screening efforts against cercarial elastase (a serine protease responsible for schistosomiasis) and TCP using a developed homology model⁷⁸. An earlier version of DOCK was used to screen the fine chemical directory of ~50,000 compounds. A total of 8800 molecules were retained after the initial phase of docking, considering the shape complementary and force field-based scoring scheme. In the final stage of visual inspection, compounds exhibiting good fitness as well as potential H-bonding and electrostatic interactions were retained with reduced priority on the scoring scheme. A total of 31

compounds were biologically evaluated against TCP. All four inhibitors identified against TCP and cercarial elastases were from the shape-based and force field-based scoring scheme, respectively. The shape-based scoring functions worked well for the hydrophobic site of TCP whereas force-field based scoring performed well for the charged amino acid cavity (polar) of cercarial elastase. Compound **24**, obtained with the shape-based scoring scheme, was found to have an IC₅₀ value of 6 μM against TCP (Table 1.3.4, row 1). The docking pose of **24** showed one naphthol group occupying the S2 pocket whereas another naphthol group was found to interact with the Trp173 of the S1'pocket ⁷⁸. Further, structural modifications of compound **24** were carried out to reduce the conformational flexibility, enhance water solubility and electrostatic interactions, improve overall chemical/metabolic stabilities and to analyze the size and electronic characters of the subsite specificity pocket ⁷⁹. The structural modifications resulted in a synthesis of a series of novel chalcone derivatives with good potency against the W2 and D6 strains of *P. falciparum*. (Representative compound **25**, W2 IC₅₀ = 0.24 μM, D6 IC₅₀ = 0.19 μM). Molecular modeling tools as well as experimental assay revealed the inhibitory action of this compound by the inhibition of TCP ⁷⁹ (**Table 1.3.4**, row 1).

Joachimiak *et al.* developed a homology model of FP-2 using the X-ray crystal structure of cathepsin K zymogen (PDB code: 1BY8) as a template for virtual screening of compounds from the Available Chemical Library (ACD) using DOCK 4.0 ⁴⁷. 5000 compounds were retained from the energy and shape-scoring schemes implemented in DOCK after initial docking. The final selection of the set of compounds was based on the knowledge of the specificity sites of FP-2 along with the consideration of drug-like properties of small molecules (such as hydrophobicity, molecular weight and absence of reactive chemical functionalities with the tendencies to form covalent bond). A set of selected compounds were evaluated in a

fluorescence-based assay using a recombinant FP-2 enzyme. Eight inhibitors **26-33** were identified (IC₅₀ of 1 to 7 μM. Table 1.3.4, row 2) in the enzyme assay, three of which **28, 29, 33** were also active against the chloroquine resistant (W2) strain of *P.falciparum* (IC₅₀ range < 26 μM)⁴⁷.

FP-2, FP-3, cruzain and cysteine protease from *Leishmania donovani* (*L. donovani*) share more than 40% homology in their mature domain and about 90% in the binding sites³³. To identify a broad spectrum inhibitor against these protozoa, Desai *et al.*, screened the CHEMBRIDGE database against homology models of FP-2 and FP-3 developed in their laboratory^{13, 46, 80, 81}. Protein preparation for screening was carried out at pH 5 to mimic the acidic environment of the food vacuole of *P. falciparum*, where the enzyme is located. Prior to the docking studies, the database was processed to eliminate metal and counter ions, neutralize charged molecules, eliminate compounds with poor ADME properties and identify compatibility with the Lipinski rule. The docking protocol was pre-validated by predicting the experimental pose of a vinyl sulphone inhibitor in the available crystal structure of cruzain (1F2A, PDB code). The filtered databases were subjected to docking using GOLD software in three different settings: 7-8 times speed up mode, 2 times speed up mode and standard mode. As docking of the large databases can be computationally expensive, multispeed docking protocol was adopted to discard the compounds quickly in the initial stages, retaining only those which were complementary to the binding site topology. In the consequent stages (with the standard docking mode), more exhaustive docking runs were performed for predicting the binding modes and interactions of the remaining compounds with the active site. Considering the high sequence similarity in the active sites of these two enzymes with major differences in the size of S2 pocket

(narrower in FP-3 as compared to FP-2), it was thought that the compounds binding to FP-3 would most likely fit into FP-2. Hence, the initial two rounds of docking were only conducted in FP-3 to reduce the computational time. The top 10% of compounds at the end of the second run were docked in both homology models using more rigorous standard mode of GOLD. Ten vinyl sulphone inhibitors were seeded as a positive control for periodic measure of the docking procedure efficiency at each stage. Finally, the top common hits for the two enzymes were visually inspected based on the geometry of the binding pose, proximity of the electrophilic center of the ligand to the catalytic cysteine and complementarity between ligand and protein surfaces in terms of spatial occupancy and hydrophobic/hydrophilic contacts. After visual inspection, eighty four compounds were biologically evaluated, twenty four of which showed inhibition against one or more cysteine protease. Twelve compounds were found as dual inhibitors of FP-2 and FP-3 (for example: **34-41**, IC₅₀ range: 1 to 63.4 μM for FP-2, 4.9 to 62.2 μM for FP-3, Table 1.3.4, raw 3), with compound **34** having the highest dual activity. Compound **40** showed an excellent profile against the chloroquine-sensitive (D6) and chloroquine resistant (W2) strains of *P. falciparum*, (W2 IC₅₀ = 6.6 μM, D6 IC₅₀ = 6.9 μM), falcipains (FP-2 IC₅₀ = 4.6 μM, FP-3 IC₅₀ = 5.8 μM), *L. donovani* cysteine protease and *L. donovani* promastigotes with IC₅₀ values of 112.9 and 8.8 μM, respectively. Compound **40**, along with other compounds (for example, **34** and **35**), suggested the possibility of developing broad spectrum antiprotozoal drugs⁸⁰ (see Table 1.3.4, raw 3).

Desai *et al.* adopted similar protocol as described before using the ACD in their second attempt of virtual screening against parasitic cysteine proteases⁸¹. Additionally, a toxicity filter was added in the initial processing of the database to remove potentially toxic functionalities

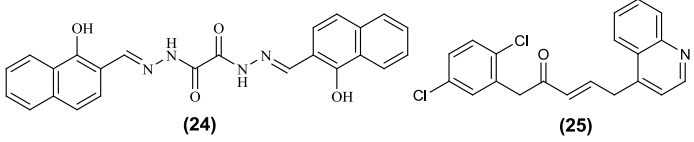
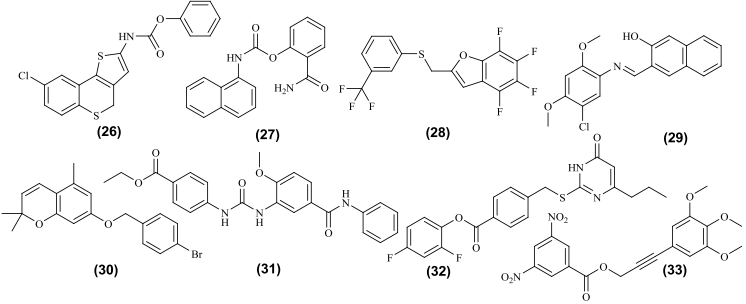
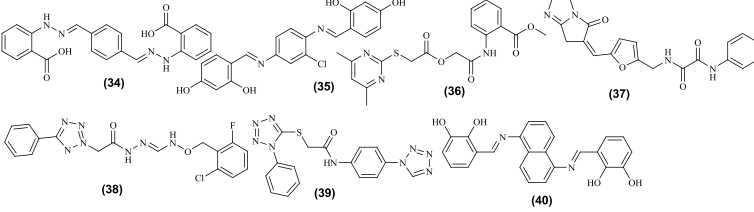
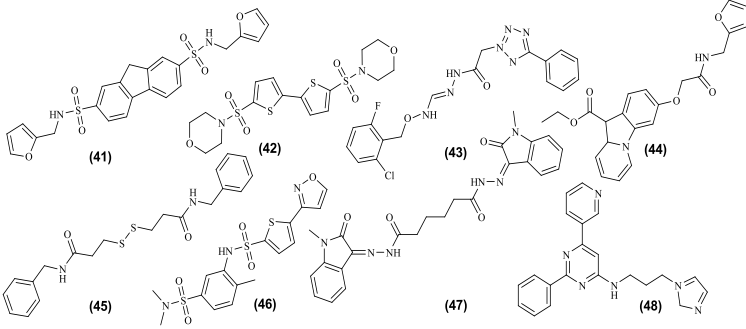
such as N-oxides, aldehydes, nitrogen and sulfur mustards, chloramines and isocyanides. The study led to the identification of 22 non-peptidic small molecule inhibitors of the parasitic cysteine protease, **18** were active against falcipains (representative examples **41-47** in Table 1.3.4, row 3, IC₅₀ range: 1.4-54.3 μM), out of which eight compounds were common inhibitors of FP-2 and FP-3 (the most active compound **41**, FP-2 IC₅₀= 1.4 μM, FP-3 IC₅₀= 11.4 μM). Four compounds inhibited *L. donovani* cysteine protease in micromolar concentration, and six compounds exhibited *in-vitro* inhibition of both *P. falciparum* and *L. donovani* promastigotes in the micromolar range (the most active compound **48**, W2 IC₅₀ = 0.44 μM, *L. donovani* promastigotes IC₅₀= 7 μM). Further, to identify common features of the diverse hits active against FP-2, a pharmacophore model was proposed using a Hip-Hop module of Catalyst. A four-feature hypothesis, which correlated well with the ligand-receptor interactions observed in the docking studies, was selected. The pharmacophoric features include two hydrogen bond acceptors and two hydrophobic regions. Mapping of the chemical features of previously identified FP-2 hits (from the chembridge database) on the developed model showed good fitting for more than 60% of the FP-2 inhibitors identified from the previous study. However, the developed model was unable to distinguish compounds with dual inhibitory (FP-2 and FP-3) activity from the exclusive inhibitors of FP-2⁸¹ (Table 1.3.4, row 4).

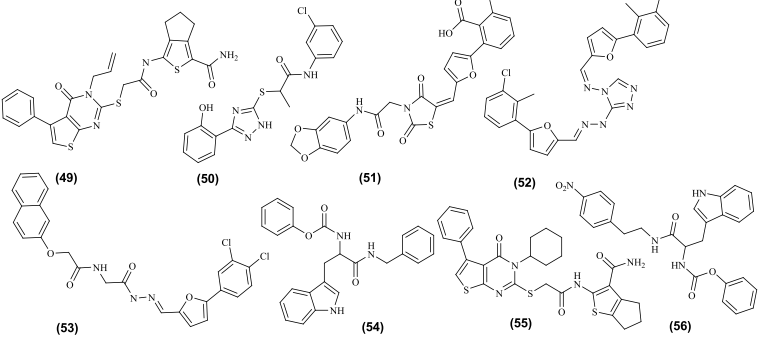
Recently, a successful application of virtual screening was reported by Li *et al.* using two parallel docking routes with Glide and GAsDock docking programs to screen the SPECS database against the crystal structure of FP-2 (2GHU, PDB code)⁸². The rationale behind the screening of the database with two different docking programs was to highlight and explore the different aspects of the ligand binding. All the compounds were initially subjected to drug-like

filters⁸³. The top thousand compounds were selected from both routes after ranking the compounds with the Glide score (GScore) in case of Glide standard precision (SP) docking and with the energy score in GAsDock. Redocking of the compounds obtained after the first step in Glide docking was carried out with the more accurate Glide extra precision (XP) mode. The top 200 compounds selected based on the GScore were visually inspected for complementarity between ligands and the hydrophobic S2 pockets and H-bond networks of ligands within the active site. 53 compounds retrieved from the Glide docking run were evaluated against FP-2. Similarly, after initial evaluation of the compounds with energy score of GAsDock, compounds were reranked based on the consensus scoring function (CScore) of five. The top 154 compounds thus selected were visually inspected using the criteria mentioned above. A total of 56 compounds obtained from the GAsDock route were purchased from SPECS. 28 non-peptide molecules were identified from the pool of 81 compounds tested as non-peptidic inhibitors of FP-2 with IC₅₀ ranging from 2.7 to 52.4 μM (see representative compounds **49-54** in Table 1.3.4, row 5). The GAsDock outperformed the Glide in terms of hit ratios (38.8% for GAsDock and 31% for Glide) in this study. The authors concluded that GOLD and DOCK scores performed well for apolar binding sites such as of FP-2 whereas FlexX and Chemscore-based GScore work well for polar binding sites. Their conclusion was based on the previous reports by Desai *et al.*⁸⁰,⁸¹ in conjunction with the analysis of binding site and the results obtained in their screening⁸². Similarity analysis of the 28 active hits, calculated using the Tanimoto similarity index (average value= 0.25), showed structural diversity in identified hits. The shape analyses of each pair of the docked poses calculated using Ultrafast Shape Recognition (USR, which consider the shape of the molecules based on the relative position of its atoms) showed an average similarity score of

0.73. This finding raised the possibility of discovering novel chemical scaffold against FP-2 with similar steric shape and binding site requirements from chemically diverse datasets ⁸². Further, optimization and library synthesis of the most potent inhibitor **49** (FP-2 IC₅₀= 2.7 μM) obtained from the virtual screening resulted in compound **55** having an IC₅₀ value of 1.46 μM ⁸⁴. Due to the ease of chemical modifications of compound **54**, it was also considered for the library design and synthesis. The most active from this series, **56**, had an IC₅₀ of 10.0 μM against FP-2 ⁸⁵ (Table 1.3.4, raw 5).

Table 1.3.4. Non-peptidic inhibitors identified against FP-2 by virtual screening approach

Database	Software/ Scoring function	^a Hit ratio (%)	Structure of the representative compounds
Fine chemical library (55,313) <small>78, 79</small>	DOCK 3.0/ shape and force field based scoring method	12.9	 <p>(24) (25)</p>
Available chemical library (195419) <small>47</small>	Dock 4.0/ Energy- scoring and shape- scoring scheme	15.9	 <p>(26) (27) (28) (29) (30) (31) (32) (33)</p>
Chembridg e database,20 01 (241000) <small>80</small>	GOLD/ Gold score	28.5	 <p>(34) (35) (36) (37) (38) (39) (40)</p>
Available chemical library (~355000) <small>86</small>	GOLD/ GOLD Score	22	 <p>(41) (42) (43) (44) (45) (46) (47) (48)</p>

<p>SPECS (~287000) 82, 84, 85</p>	<p>GLIDE SP/XP/ GAsDock, Energy score, Cscore</p>	<p>34.5</p>	
---	---	-------------	--

^a hit ratio is calculated by dividing the number of active hits retrieved against FP-2 by the number of compounds evaluated in the biological assay. The number in bracket in database column shows number of compounds available in the database at the time of virtual screening.

1.3.5. Natural Product-Based Drug Discovery

E-64, (N-(L-3-*Trans*-carboxyoxiran-2-carobnyl)-L-leucyl)-amido(4-guanido)butane), isolated from the culture of *Asperigillus japonicas*, was an irreversible broad-spectrum inhibitor of papain-like cysteine proteases (clan CA, family C1) (compound **3**),⁶³. Recently, the X-ray crystal structure of FP-2 was published in a complex with E-64 (shown in Figure 1.2) that showed a covalent, irreversible hemithioketal formation between thiolate of Cys 42 in the FP-2 active site and the epoxy carbon of E-64⁵⁰.

Several chalcone derivatives were developed after the discovery of the antimalarial activity of lipochalcone A, a natural product obtained from Chinese liquorice root⁸⁷. These chalcone derivatives exhibited potent antimalarial activity^{79, 88, 89}. Molecular modeling studies as well as experimental evaluation of the chalcones indicated that they exert antimalarial activity via multiple mechanisms, one of which involved the inhibition of FP-2. Representative compound (**57**) from the chalcone series is shown in figure 1.3.5 (compound **57**, FP2 IC₅₀ = 1.8 μM)⁹⁰.

Isatin represents a natural product from the plants of the *Isatis* genus, which was a biologically validated starting point for the design and synthesis of chemical libraries against cysteine and serine proteases^{91, 92}. Several efforts were pursued to synthesize and evaluate the Isatin derivatives against FP-2 as well as cultured parasites. Resultant compounds vary in range from higher to lower micromolar activity^{93, 94}. Representative members of this class are shown in Figure 1.3.5 (compound **58**, FP2 IC₅₀ = 4.4 μM, compound **59**, FP2 IC₅₀ = 0.375μM). Both compounds also inhibited cysteine proteases of trypanosomes, rhodesin and cruzain.

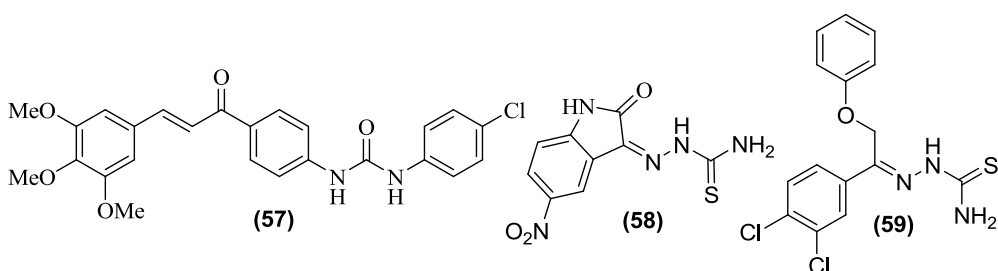


Figure 1.3.4. Natural product-based inhibitors of FP-2.

In brief, *P. falciparum* cysteine protease ‘falcipains’, can be a promising target for the development of novel malarial chemotherapy. Several peptidic inhibitors/peptidomimetics mainly from vinyl sulphone class as well as structurally diverse non-peptidic inhibitors displayed an excellent activity against falcipains and cultured parasites. Recently, published crystal structures of FP-2-E-64 (epoxysuccinate inhibitor) and FP-3-leupeptin (peptidic aldehyde inhibitor) provide more insights for structure guided drug designing. As noted before, the P2-Leu and P1-HPhe boost the inhibitory potencies towards FP-2. In addition, the polar residues (Asp 234 and Ser 149 in FP-2, and Glu 243 and Ser 158 in FP-3) buried in the hydrophobic S2 pocket should be considered while designing selective inhibitors against these enzymes. Also, difference in the size of S2 pockets of FP-2 and FP-3, can be addressed for designing selective inhibitors of

these enzymes, although, combined drug discovery efforts targeting different subtypes of falcipains might be beneficial in reducing level of cultured parasites. Selection of suitable scoring functions plays major role in structure-based virtual screening approach. Published virtual screening studies against falcipains pinpointed suitability of GOLD and DOCK scoring functions for apolar binding sites of FP-2 over GScore. These findings are particularly important in future structure-based virtual screening studies against FP-2. Moreover, the sequence similarity of falcipains with several homologous parasitic cysteine proteases can be used to develop broad spectrum inhibitors against these targets as previously reported by our group^{80,81}.

The cysteine proteases discussed here are of the utmost importance for drug design considerations. For instance, several efforts are currently underway to develop effective antimalarials for encountering the parasite resistance to all front line drugs, including the most potent artemisinin derivatives⁹⁵. In such scenarios, falcipains manifests an emerging drug target due to its unique function in hemoglobin degradation and survival of the protozoa. Moreover, effective peptidic and non-peptidic inhibitors of falcipains (summarized above) with good inhibitory potency project this enzyme as a promising target for the development of antimalarial chemotherapy.

In the next sections, various drug design approaches adopted against highly pathogenic SARS virus are discussed.

1.4. SARS cysteine protease “3CL^{pro} or M^{pro}”: genesis, function and structural requirements

SARS is a life threatening upper respiratory tract illness which reached an epidemic status in 2003. Within the short span of a year, this disease proliferated from the Guangdong

province of China, where it originated, to 32 countries of Asia, North America and Europe. The SARS epidemic infected nearly 8000 people and resulted in approximately 800 casualties from November 2002 to August 2004⁹⁶⁻⁹⁹. A novel form of the coronavirus (CoV), SARS-CoV, was identified as an etiological agent for the disease^{7, 100-102}.

The SARS-CoV genome has been reported to contain fourteen functional ORFs, which provide accessories for encoding the proteins for virus replication and transcription. Two large ORFs connected by a ribosomal frame shift constitute the replicase gene which transcribes two overlapping polyprotein (pp), pp1a (~450 kDa) and pp1b (~750 kDa), necessary for viral propagation¹⁰³. These polyproteins are cleaved by virus proteases to generate the functional proteins required for replication of the virus. The SARS-CoV proteases responsible for this function involve papain-like proteases (PL^{pro}) and a chymotrypsin-like protease (3C-like protease or 3CL^{pro}) due to their distant relationship with 3C protease of picornaviruses. The 3CL^{pro} is also called the main protease (M^{pro}) because of its dominant role in the processing of viral polyproteins and control of replicase complex activity¹⁰³⁻¹⁰⁶. Due to the pivotal role of M^{pro} in viral replication and infection process, the enzyme demonstrates itself as an attractive target for the design of antivirals⁸.

During the outbreak of SARS, several homology models of SARS M^{pro} were developed by different groups shortly after its genome sequencing, based on its sequence similarity with other known CoV-like proteases such as human CoV (HCoV) or porcine transmissible gastroenteritis (TGEV)^{8, 103, 107}. Within a short time of the outbreak, multiple crystal structures of SARS M^{pro} were published^{104, 108-117}. The published structures revealed remarkable degree of conservation in the substrate-binding sites during SARS mutation and provide an opportunity for

the design and screening of inhibitors for anti-SARS activity.

The structure of the monomeric unit of SARS M^{pro} is shown in Figure 1.4a. SARS M^{pro} is a functional homodimer in which each subunit is composed of three structural domains: I (residues 8-101), II (residues 102-184), and III (residues 201-301). Domains I and II show a β -barrel fold resembling trypsin-like serine protease whereas domain III is α -helical. The substrate-binding site is located in a cleft between domains I and II. The enzyme is atypical cysteine protease containing a catalytic dyad (His-41-Cys145) instead of a triad, which is required for proteolytic activity^{8, 116, 118}. The N-terminal is comprised of seven residues (N1-N7) known as “N-finger”. A 16-residue loop (residues 195-200) forms a bridge between domains II and III (see fig.1 (a)). While domains I and II are important for catalytic activity, domain III plays an essential role in the dimerization of M^{pro} and the maintenance of quaternary structure^{116, 119, 120}. N-finger is more important for enzyme activity than dimerization as confirmed by site-directed mutagenesis or deletion studies^{118, 121-123}. Figure 1.4b shows the magnified view of the active site of SARS-CoV M^{pro} with co-crystallized peptidic ligand (2AMD, PDB code)¹¹⁵.

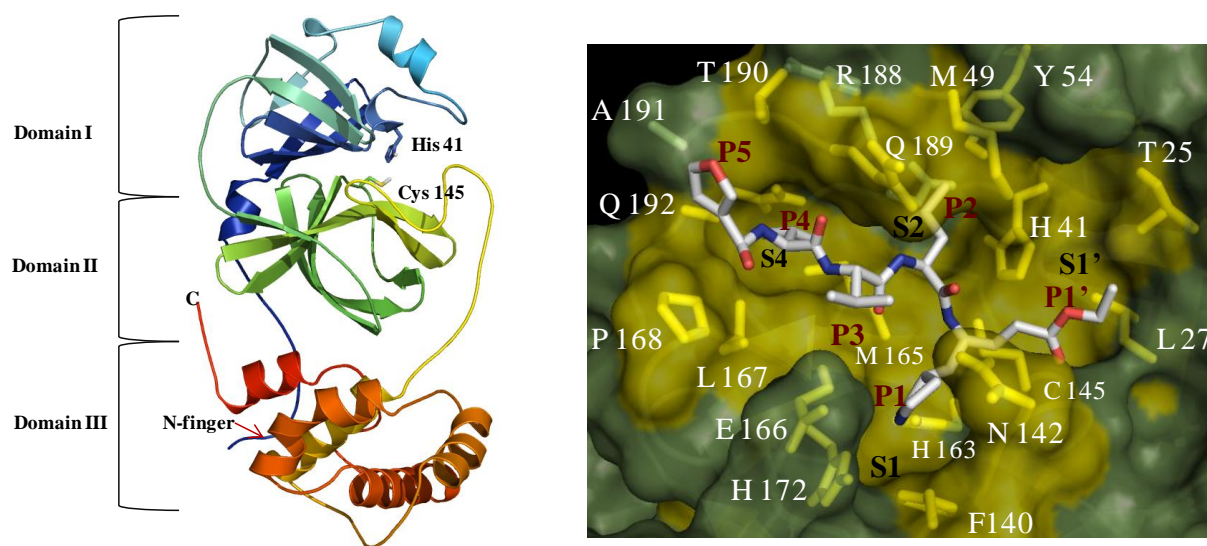


Fig.1.4 a) The overall topology of SARS 3CL^{pro}. Individual domains, N-finger and catalytic

residues, Cys 145 and His 41 are shown; b) magnified view of the substrate-binding site in surface representation. P1-P5 and P1' groups (red) of peptidic inhibitor (gray), subsites S1', S1, S2 and S4 (black), and residues forming the substrate-binding pockets (white) are labeled.

The substrate binding pockets S4, S2 and S1 are highly conserved among the CoV¹²⁴. Successful cloning and expression of recombinant SARS-CoV M^{pro} has unveiled important insights on the substrate-specificity data for this enzyme. Along with the absolute specificity of Glu at P1, recent data elucidated an equal preference for the P1-His containing substrate^{125, 126}. The S2 subsite of SARS-CoV can accommodate the side chains of Leu, Phe, Val, Met and Thr residues as opposed to traditional preference of only Leu/Ile at P2 for other CoV^{125, 126}. The side chain of P3 residue is mainly solvent-exposed in all CoV M^{pro} and is capable of accommodating side chains of a broad range of functionalities¹¹⁵. Small residues such as Ser, Thr, Val, Pro and Ser, Ala, Gly can be tolerated at P4 and P1', residues respectively¹²⁴. The knowledge of substrate specificities has not only increased the understanding of polyprotein processing mechanism by the enzyme but also provided important clues for drug design¹²⁵.

1.5. DRUG DESIGN APPROACHES AGAINST SARS-COV M^{PRO}

During the epidemic of SARS, combinations of drugs were given to individuals suffering from this life threatening disease and these multi-drug treatments resulted in several post-therapy side effects^{127, 128}. As more information became available for the key components of the virus life-cycle, target-oriented drug design efforts were pursued to prepare against the possibility of the next epidemic of this disease. Early phases of drug discovery against SARS M^{pro} were centered on the structural modification of the known drugs with established safety profiles in humans in order to avoid painstaking efforts of finding and characterizing new drug entities in

the scenario of a global outbreak of SARS. The drug design efforts pursued against SARS 3CL^{pro} are summarized below.

1.5.1. Structure-guided drug designing

Anand *et al.* published a homology model of SARS M^{pro} using the X-ray crystal structures of HCoV 229E M^{pro} and TGEV M^{pro} (40 and 44% sequence identity respectively) as a template to stimulate drug discovery efforts against this promising target⁸. Also, they unveiled remarkable degrees of conservation among the substrate-binding sites of CoV as supported by the cleavage of the TGEV M^{pro} substrate by the recombinant SARS-CoV M^{pro}. During the outbreak of SARS, rupintrivir (formerly designated as AG7088), a potent inhibitor of human rhinovirus (HRV) 3C protease, was in the early stages of drug development by Pfizer Inc. for the treatment of the common cold¹²⁹⁻¹³¹. Anand *et al.* addressed that Rupintrivir could partially fit into the putative binding pocket of SARS-CoV 3CL^{pro} based on the identical orientation of the hexapeptidyl chloromethyl ketone (CMK) inhibitor **60** in TGEV M^{pro} and Rupintrivir **61** in HRV2 3C^{pro} and suggested a possibility to use rupintrivir as a model to develop inhibitors against SARS-CoV 3CL^{pro}⁸ (Figure 1.5.1).

Ghosh *et al.* carried out the modification of the P2 side chain of rupintrivir with P2-benzyl **62** and prenyl **63** considering the P2-Phe substrate-specificity of SARS M^{pro}¹³². Also, they developed another series by replacing the ketoethylene of compound (**67**) with hydroxyethylene **64**. Compound **63**, having a P2-prenyl group, was more potent than the corresponding analogue **62** with a P2-phenylmethyl group, which further indicated the possibility of accommodating other small side chain residue at the P2 position. The inhibitor with hydroxyethylene isostere **64** showed little inhibition of the SARS M^{pro} enzyme. To gain molecular level insights, an X-ray structure of SARS M^{pro} co-crystallized with **63** was resolved.

The crystal structure revealed the covalent binding mode for the inhibitor as expected. Also, the crystal structure offered insights about poor inhibitory activity of **64** was due to the loss of hydrogen bond interaction between the backbone amine nitrogen of Glu166 and the carbonyl group of the inhibitor ¹³² (Figure 1.5.1).

Shie *et al.* designed novel analogs of rupintrivir and pursued combinatorial synthesis of a series of ketomethylene isosteres (representative compounds **65**, **66** as well as tripeptidomimetic α , β unsaturated esters (representative compounds **67**, **68** having lactam-glutamate or Phe at P1 ¹³³. Interestingly, compounds with L-phenylalanine at P1 were found to have enhanced activity compared to corresponding lactamglutamate (Compound **66**, $M^{pro} IC_{50} = 13 \mu M$, Compound **68**, $M^{pro} IC_{50} = 11 \mu M$). Although, rupintrivir **61** was found to be inactive at the concentration $>100 \mu M$ against SARS M^{pro} ¹³³ as well as in a cell-based assay ¹³⁴, the higher activity of its analogue **66** motivated the authors to predict the binding mode of these compounds by molecular docking studies using previously reported crystal structures ¹⁰⁴. Docking studies depicted that P1 and P2-Phe of **66** fit well into the S2 and S3 pockets respectively, placing the conjugated ester moiety in the S1 pocket to interact with His163 and Gly 143. Also, **66** was stabilized by multiple H-bond interactions with the active site residues. However, the conjugated ester was still too far ($> 4.5 \text{ \AA}$) from Cys145 for a covalent bond formation. In contrast, the docking of rupintrivir **61** showed an improper placement of the terminal ester group in the S1' pocket as a result of poor fitting of the P1-lactam moiety in the S1 pocket. In addition, P2-fluorophenyl and P4- isoxazolyl moiety of **61** exhibited steric clashes with the residues of S2 and S4 ¹³³. This might be responsible for the inactivity of **61** as well as the lower activity of corresponding lactamglutamate analogs (Figure 1.5.1).

Based on the higher activity of Phe-Phe dipeptide inhibitors **66** and **68**, several Phe-Phe dipeptide inhibitors were designed with two Michael acceptors at N and C-terminals mimicking pseudo *C*-2 symmetry. Dedicated synthesis of compounds from this series identified **69** (JMF1521) with an IC_{50} value of 1 μ M and a K_i value of 0.52 μ M¹³³. Further, synthesis of a series of peptide anilide with L-phenylalanine as P1 residue attached with a 2-chloro-4-nitroaniline moiety (considering the activity of niclosamide against replication of SARS-CoV¹³⁵) was pursued¹³⁶. Among the synthesized compounds, anilide **70** functioned as a potent inhibitor of enzymes having K_i and IC_{50} values of 0.06 and 0.03 μ M, respectively. SAR of various analogs of **70** confirmed that chloro, nitro and dimethylamino groups in 2-chloro-4-nitroaniline of **70** were indispensable for the activity. The key interactions of **70** as revealed by the docking studies involved interactions of a nitro group with the NH of Ala 46, 3 Å distance of chlorine from γ -S of Cys145 and ϵ -N2 of His41, as well as hydrophobic interaction of a (dimethylamino) phenyl moiety in the S2 pocket. Also, authors indicated the possibility of modifying P1 phenyl residue of **70** in the S1 pocket to enhance interactions with Phe140, His163, and Glu166 to further enhance activity¹³⁶ (Figure 1.5.1).

In search of the effective drug candidates, Chou *et al.* performed docking studies using previously developed homology model of SARS M^{pro}⁸ by modifying the bulkier *p*-fluorobenzyl side chain of **61** with *p*-fluorophenyl (KZ7088, **71**) and the octapeptide AVLQSGFR to understand the binding interactions of SARS-CoV M^{pro} with its ligand¹³⁷. The octapeptide was chosen by considering its size to occupy protease-specificity sites, specificity of SARS-CoV for Gln↓(Ser, Ala, Gly) and for the rational designing of new peptidic analogs based on the “distorted key approach.” The distorted key approach involved modifications of the substrate-

cleavage site by keeping other residues of the peptide (substrate) intact for tight binding. Such a substrate could behave as a distorted key which binds effectively in the enzyme active site but blocks the function of the enzyme. Also, this approach could help to identify the competitive inhibitors against this enzyme¹³⁸. Docking studies of **71** and AVLQSGFR revealed the presence of multiple hydrogen bonds between the SARS proteinase and respective ligands¹³⁷. However, the experimental binding affinity of **71** was not reported. Gen *et al.* confirmed the efficacy of octapeptide for the inhibition of the replication of SARS-CoV ($EC_{50} = 0.031 \mu\text{M}$) with no detectable toxicity on vero cells under experimental conditions. Sirios *et. al.* encoded the receptor interactions of **71** in pharmacophoric features and conducted a pharmacophoric search over 3.6 millions of compounds to identify only 0.03% of potential compounds for further experimental analysis¹²⁷ (Figure 1.5.1).

Du *et al.* studied catalytic interactions between the experimentally cleavable octapeptide AVLQSGFR and SARS M^{pro} using molecular mechanics (MM) and quantum mechanics (QM) calculation with the help of a previously developed homology model of SARS-CoV M^{pro}^{8, 139}. The residues of catalytic dyad His 41 and Cys 145 were found to attract electron density of peptide bond between Gln and Ser, which accelerated the positive charge on C(CO) of Gln and the negative charge on N (NH) of Ser. The replacement of the carbonyl of Gln to CH₂ or CF₂ distorted the π -bond system of the peptide bond, making it uncleavable by the enzyme, situation analogous to the “distorted key” approach discussed above¹³⁹. The modified octapeptide showed tight binding with the enzyme and represents a good starting point for the design of peptide based anti-SARS drugs (Figure 1.5.1).

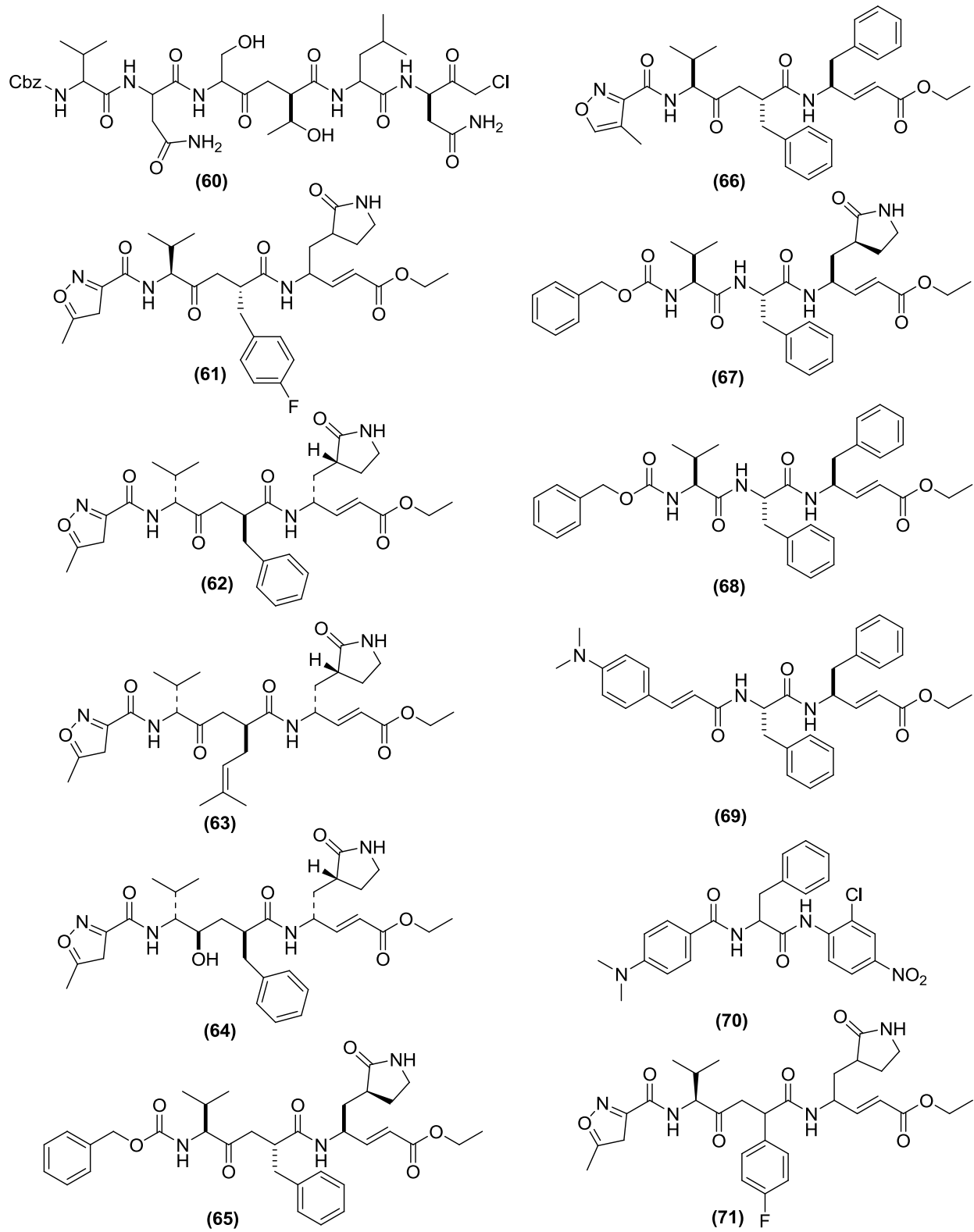


Figure 1.5.1. Inhibitors of SARS-M^{pro} identified by structure-guided drug design.

1.5.2. Knowledge-Based Drug Design

Anand *et al.* pinpointed the use of the rhinovirus protease inhibitor, rupintrivir, as a starting point for the design of inhibitors against SARS 3CL^{pro}. Further development of inhibitors based on the structural modifications of rupintrivir has already been discussed in previous sections.

Jain *et al.* performed the synthesis and biological evaluation of a series of keto-glutamine analogs with a phthalhydrazido group at the α -position against SARS M^{pro} ¹⁴⁰. Their knowledge-based approach was based on a previously identified lower micromolar reversible inhibitor **72** of the hepatitis A virus (HAV) 3C proteinase bearing a phthalhydrazido moiety ¹⁴¹ and preference of P1-glutamine by SARS M^{pro}. Initial efforts for the synthesis of *N, N*-dimethylglutamine analogs with leucine, threonine, and valine at the P2, P3, and P4 positions respectively, resulted in compounds having higher micromolar activity (representative compound **73**, M^{pro} IC₅₀ = 28 μ M). Further, optimization and structure-activity relationships (SAR) identified cyclic glutamine analogs, among which compound **74** exhibited an IC₅₀ of 0.6 μ M against SARS M^{pro} ¹⁴². Docking studies of **74** indicated that the active site of the enzyme had enough room to accommodate the bulkier phthalhydrazide group with minimal rearrangement of protein (Figure 1.5.2).

Bacha *et al.* identified a cluster of serine residues (Ser139, Ser144 and Ser147) close to catalytic residues of the active site ¹⁴³. Sequence alignment of this region among the proteases from other CoV revealed the presence of a conserved serine cluster among CoV. This observation motivated the authors toward designing inhibitors targeting serine cluster of the active site. Based on the knowledge of the reactivity of boronic acid compounds to the hydroxyl group of serine, inhibitory potencies of several boronic acid compounds were evaluated against

SARs-CoV M^{pro}. Compound **75** found to have an inhibition constant (K_i) of 40 nM. Further studies revealed a reversible binding mode of **75** in an energetically favorable fashion, defining it as an attractive scaffold for further optimization ¹⁴³ (Figure 1.5.2).

Meanwhile, Zhang *et al.* extended their knowledge from the studies of active caspase (another cysteine protease) inhibitor **76** toward the discovery of potential SARS-CoV M^{pro} inhibitors ¹⁴⁴. A series of dipeptides analogs were synthesized with *N*, *N*-dimethyl glutamyl FMK at the P1 site and different amino acids (leucine/valine/phenylalanine) at the P2 site from the knowledge of a peptide-substrate recognition sequence by the enzyme. *N*, *N*-dimethyl-glutamine analogs were specifically chosen to prevent possible cyclization of the free glutamine side chain with FMK ^{145, 146}. Compound **77** was found to have an EC₅₀ value of 2.5 μM in the CPE inhibition assay against 6109 strains of SARS-CoV with a selectivity index (SI) of >40, presumably through the inhibition of SARS CoV M^{pro} ¹⁴⁴ (Figure 1.5.2).

The nucleophilic attack of cysteine thiol to trifluoromethyl ketone (TFMK) leads to a tetrahedral adduct formation which mimics the substrate-enzyme intermediate ^{147, 148} and compounds bearing TFMK have a potential to be a serine/cysteine protease inhibitor. Syndes *et al.* designed and synthesized derivatives of Gln and Glu possessing a TFMK moiety considering the utility of TFMK as a warhead and P1-Glu preference of the enzyme ¹⁴⁹. Compounds **78** and **79** displayed K_i values of 134.5 μM and 116.1 μM, respectively. Further, structure-based optimization of **79** led to synthesis of mono-, di- and tripeptide analogs with modification in the halogen content of warhead as well as in the P1 side chain ¹⁰⁸. Five inhibitors were identified with activity in the nanomolar range. One of the inhibitor **80** showed unexpected reversible inhibition of SARS 3CL^{pro} by forming a thioether complex with Cys145 of the active site, supported by the co-crystallized structure of **80** with SARS M^{pro} ¹⁰⁸. Regnier *et al.* noticed the

moderate activity of compounds might be due to possible cyclization of the free acid/amine side chains of **78** and **79** at P1 with TFMK as reported previously with FMK derivatives¹⁴⁴, which resulted in the poor binding of the cyclized compounds in the active site¹⁵⁰. Several modifications were carried out in P1, P2, P3 side chains as well as warhead groups. Compound **81** containing P1 pyrrolidone and thiazolyl units as a warhead showed an IC₅₀ of 2.2 μM against SARS M^{pro}¹⁵⁰. Recently, Shao *et al.* synthesized a series of TFMK by variation in the amino acid at P1-P4 chains as well as incorporation of alkyl chains at P2-P4¹⁵¹. Compound **82** with an identical arrangement of P1-P4 amino acids to the substrate inhibited SARS-CoV 3CL^{pro} in a time dependent manner with an IC₅₀ of 0.8 μM and a K_i of 0.29 μM after a four hour incubation. The computational model of **82** pointed that the formation of a covalent bond between the cysteine thiol and inhibitor **82**, as well as other key interactions of inhibitor with the active site residues, were favored only with the [S, S, S, S] isomer of **82** with R configuration of carbonyl carbon adjacent to the trifluoromethyl group¹⁵¹ (Figure 1.5.2).

Several Aza-peptide epoxides (APEs) were reported as inhibitors of clan CD proteases, for example, leguminases¹⁵² and caspases¹⁵³. The architecture of APE involves an epoxide moiety attached to the carbonyl group of the P1 residue, nitrogen at the C α position of P1 and target-specific side chain of P1 residue. Synthesis of APE possessing an azaglutamine (AGln) as the P1 residue was performed to evaluate their efficacy against SARS-CoV M^{pro}¹⁵⁴. Micromolar quantities of the [S, S] diastereomer **83** showed strong inhibition for the cleavage of a peptidic substrate whereas the corresponding R, R diastereomer was inactive against M^{pro}. Later, the crystal structures of the [S, S] diastereomer with M^{pro} revealed the formation of a covalent bond between the S ^{γ} atom of catalytic Cys145 and the epoxide C3 atom of the aza-peptide component of the inhibitor, suggesting the irreversible mode of action of this class of cysteine protease

inhibitors. Models of different diastereomers of APEs revealed the importance of *S* configurations of the epoxide C3 atom of APE to avoid clashes with the active site residues and other components of azapeptide. Also, the *S* configuration of epoxide provided better geometry for a nucleophilic attack by active-site cysteine thiol¹⁵⁴ (Figure 1.5.2).

Peptidyl aldehydes are effective against multiple proteases including serine protease^{155, 156}, aspartic protease^{157, 158}, and cysteine protease^{159, 160}. Moreover, they possess reversible electrophilic isostere and are capable of mimicking the endogenous substrate by retaining an H-bonding framework of backbone and addressing different subsites of enzyme with their side chains. Considering these advantages of the designing inhibitor with peptidic aldehyde scaffold, Al-gharabli *et al.* employed an efficient method for the synthesis of their designed peptidic library against SARS 3CL^{pro}¹⁶¹. They considered two approaches for the design of their peptidic aldehyde library. In the first approach, they carried out a sequence homology search using the extracted binding pockets of SARS 3CL^{pro}. The information gained from the bound ligand fragments in homologous proteins were used for the modeling of the putative side chains for the corresponding sub-pockets of enzyme. In their second approach, natural and non-natural amino acids from the ACD and Sigma-Aldrich library were docked in the S1, S2 and S4 pocket of SARS 3CL^{pro} using the FlexX docking program. Amino acids which scored well for the respective pockets were selected for the designing. The peptide backbone was retained as in the complex structure of SARS-CoV 3CL^{pro} with a CMK inhibitor (1UK4, PDB code). Evaluation of the designed library identified **84** and **85** with IC₅₀ of 7.5 μM as a reversible inhibitor of SARS M^{pro}¹⁶¹. The presence of the hydrophilic residue at the P2 position prompted the authors to predict non-canonical binding modes of compounds **84** and **85** as observed in the active

monomer A of SARS 3CL^{pro} (1UK4, PDB code) (Figure 1.5.2).

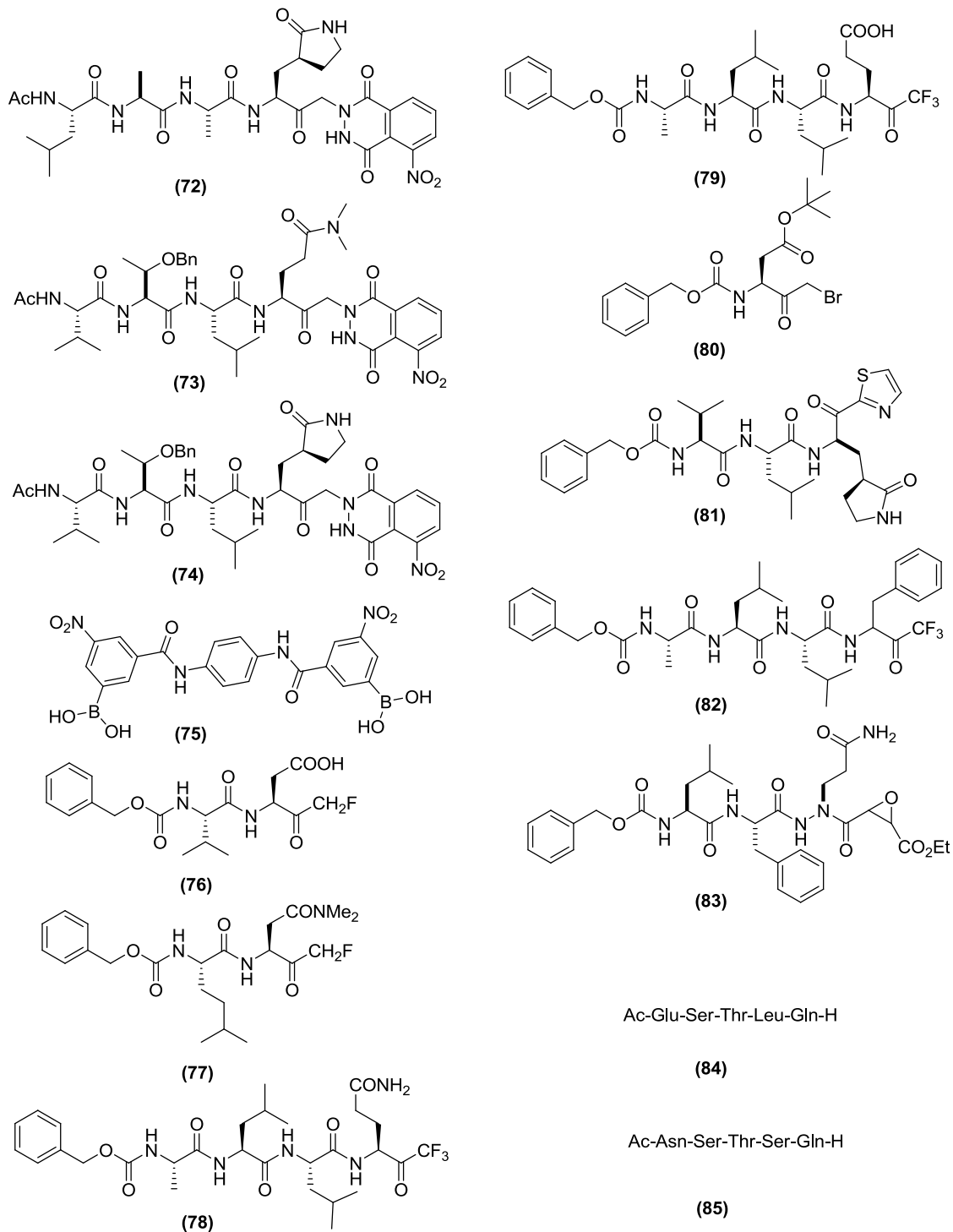


Figure 1.5.2. Inhibitors of SARS-CoV M^{pro} obtain by knowledge-based drug design approach.

1.5.3. Structure-Based Virtual Screening

This section will review the discovery of the non-peptidic inhibitors developed against SARS-CoV M^{pro} by a structure-based virtual screening approach. Table 1.5.3 summarizes the commercial databases, methods, softwares and scoring functions used for virtual screening, hit ratio along with the chemical structures of hits obtained.

In 2005, Liu *et al.* screened chemical databases (listed in Table 1.5.3, raw 1) against the “flexible homology model” developed by combined molecular dynamics (MD) simulation and multicanonical sampling protocol¹⁶². The top 10% of docked compounds were retained based on their dock energy-score and contact-score. Subsequent filtration steps involved removal of compounds with $pK_i > 4.0$ (to remove score bias caused by a single empirical function) followed by selection of the top 2000 compounds by score and Xcscore scoring functions¹⁶². Final pools of compounds were subjected to a pharmacophore filter (developed using the POCKET module from LigBuilder) and evaluated using drug-like criteria for similarity to the original drug framework as well as to known antiviral compounds. After visual inspections, 48 compounds were subjected to an enzyme assay, out of which three compounds **86-88** including calmidazolium **86**, a well known antagonist of calmodulin, were found to inhibit the enzyme with k_i ranging from 61 to 178 μM ¹⁶² (Table 1.5.3, raw 1).

Chen *et al.* carried out the docking studies of compounds from the CMC database (comprised of drugs which have been used or are currently used against human) against previously developed homology model of SARS M^{pro} to expedite the screening process against the SARS epidemic^{107, 163}. The potential candidates were ranked using a consensus scoring function along with the scoring function implemented in AutoDock3.0. Cinanserin **89**, a well

characterized serotonin antagonist found to inhibit the catalytic activity of SARS-CoV M^{pro} with IC₅₀ of 5 μM,) also interfered with SARS-CoV replication in cell culture (IC₅₀ = 31 μM) (Table 1.5.3, row 2).

Chen *et al.* screened the SPECS database using dock 4.0. After the first phase of docking, the top 2000 compounds were selected based on the highest dock score and subjected to consensus scoring¹⁶⁴ and drug-like property calculations⁸³. Finally, a set of 256 compounds were evaluated in an integrated assay platform comprised of a Surface Plasmon Resonance (SPR) assay (for measuring binding affinity of the candidate compound with the SARS-CoV M^{pro}) and fluorescence resonance energy transfer (FRET) technologies-based assays (to measure the inhibition of proteolytic activity of SARS-CoV M^{pro}). Eight novel inhibitors of SARS-CoV M^{pro} were discovered (representative compounds **90-96**), resulting in IC₅₀ values of 6.86 to 80.46 μM¹⁸ (Table 1.5.3, row 3).

Lu *et al.* performed a virtual screening of the Maybridge database using a published X-ray structure of SARS M^{pro} with a covalently bound CMK inhibitor (1UK4, PDB code)^{116, 165}. The molecules were first ranked by the fitness score component of GOLD score after docking to select the best pose followed by reranking of the compounds with the external H-bond energy term implemented in GOLD score for ranking the binding affinity of compounds. This was the first report, to the best of our knowledge, to use only H-bond energy term to rank and select compounds. The top 50 compounds were subjected to a SARS-CoV M^{pro} inhibition assay. Two active compounds **93** and **94** were identified from an enzyme inhibition assay. The core structures of these two hits, defined by the docking study, were used for further analogue searches. Twenty-one analogs derived from these two hits exhibited IC₅₀ values below 50 μM, with the most potent one, **95**, showing an IC₅₀ of 0.3 μM¹⁶⁵ (Table 1.5.3, row 4).

Dooley *et al.* presented a genome-to-drug-lead approach which uses terascale computing to model flexible regions of protein and thus encodes genetic information to identify drug leads¹⁶⁶. A homology model of SARS-CoV M^{pro} complexed with a substrate fragment (ATVRLQ^{p1}A^{p1'}) was developed by predicting the conformation of the flexible loop with a total of 200 simulations, and average structure (deposited in PDB, 2AJ5, PDB code) was used as a drug target in virtual screening for small-molecule inhibitors using the EUDOC docking program. A total of 3958 compounds were selected based on the total and Van der waals interaction energies. Twelve inhibitors were tested in a cell-based inhibition assay after removing compounds with poor solubility, poor cell permeability, multiple chiral centers and commercial unavailability. Compound **96** inhibited the SARS-CoV Toronto-2 strain with an EC₅₀ of 23 μM. Further, screening of a similar set of compounds against two published crystal structures of protease failed to identify compound **96**¹⁶⁶. This observation suggested that terascale computing could complement the crystallography, increase the hit rate in virtual screening and accelerate the process of drug discovery against emerging infectious diseases (Table 1.5.3, raw 5).

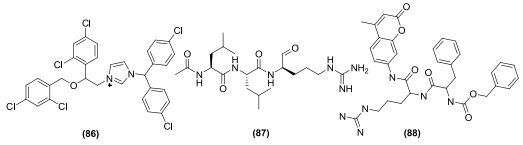
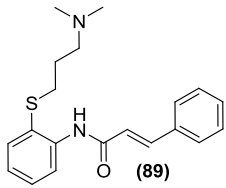
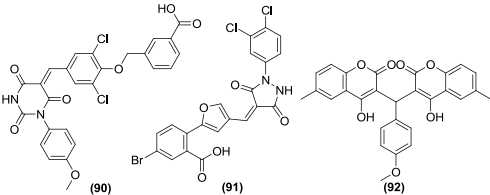
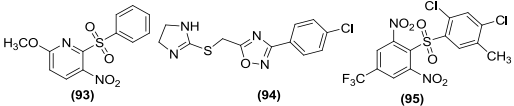
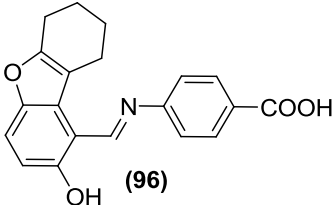
Tsai *et al.* carried out a structure-based virtual screening against Maybridge database¹⁶⁷. After the initial filtration of the database using drug-like filters, remaining compounds were docked into the co-crystallized structure of SARS-CoV M^{pro} with an octapeptide CMK inhibitor using DOCK4.0.2. 93 compounds were subjected for the inhibition assay, twenty one of these showed inhibition against SARS-CoV M^{pro} (IC₅₀ ≤ 30 μM), with three of them found having a common substructures. In order to search analogs of these compounds, 2D substructure searches were carried out in the Maybridge, ChemBridge, and SPECS databases, which further identified 25 compounds against SARS-CoV M^{pro} with an IC₅₀ of 3-1000 μM (most active compound **97**, IC₅₀ = 3 μM)¹⁶⁷ (Table 1.5.3, raw 6).

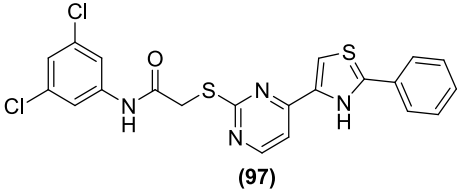
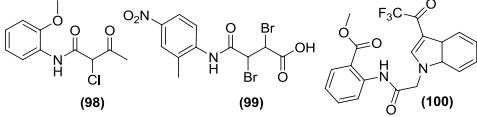
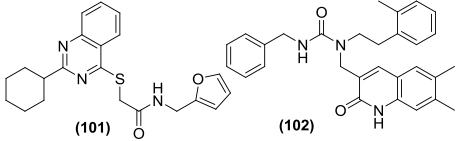
Bacha *et al.* generated a pharmacophore model using low energy conformation of the most active compounds from a previously developed peptide library (mostly peptidic halomethyl ketones) with the pharmacophore application module of MOE^{108, 149}. A threshold of 1000 μM was selected (active = <1000 μM , inactive = >1000 μM) based on the distribution of K_i data in their developed library of compounds for the generation of the qualitative pharmacophore. Exclusion spheres were added for improving the ability of pharmacophore model to discriminate between actives and inactives. A pharmacophore-based database search conducted in a commercial database implemented in MOE resulted in 40 potential compounds, 38 of which had K_i <1000 μM . The most active hits **98-100** from this screening are shown in a Table 1.5.3, raw 7 with $K_i = 4.5-27 \mu\text{M}$ ¹⁰⁸.

Mukherjee *et al.* conducted structure-based virtual screening of Asinex platinum collection against the crystal structures of SARS-CoV M^{pro} (PDB code: 1UK4, 2AMD) using the GOLD docking program¹⁹. The initial database was prefiltered using various drug-like and ADME (absorption, distribution, metabolism and excretion) filters. The filtered database was subjected to multi-stage cascade docking involving computationally less intensive and faster docking protocol (GOLD 7-8 times speed up mode) at the early stages of the screening with computationally complex protocol (GOLD 2-times speed up mode and GOLD standard mode) towards the end. After the cascade docking, top-ranked poses (based on GOLD score) were rescored using a cumulative score obtained by summing six scores (Fscore, Gscore, Pmfscore, Dscore, and Chemscore calculated using Cscore module of Sybyl 6.9), including the GOLD score. The diversity analysis of the top 500 hits obtained after rescoring retained 100 molecules that were visually inspected for their ability to obtain key substrate specificity sites, geometric quality of the binding pose, hydrophobic/ lipophilic mismatches and complementarity of the key

interactions. Meanwhile, publication of the co-crystallized ligand structure of SARS M^{pro} (2AMD, PDB code) as well as reports of several inhibitors prompted the authors to perform a scoring function-based enrichment study using the previously mentioned six scoring functions. Gscore followed by Dscore gave better enrichment of the true actives and were combined in the range-scaled format to generate a composite GDR score. In the extended phase of screening, pharmacophore filters based on the crystal pose of the peptidic inhibitor (2AMD, PDB code), docking pose-based descriptors as well as GDR scoring function were used to select the compounds for biological assay. Two hits were identified (**101, 102**) with IC₅₀ of 17 and 18 μM, respectively (Table 1.5.3, raw 8)¹⁹

Table 1.5.3. Non-peptidic inhibitors of SARS-CoV M^{pro} obtained by an experimental HTS

Database	Crystal structure/ Homology model	Method/ Software/ Scoring function used	^a Hit ratio (%)	Structure of the most active compounds
National chemical institute diversity set (~230 000), ACD 3D 2002.2 (~280 000), MDDR- 3D 2002.2 (~120 000)	flexible homology model, 1UK4, 1UK3	Docking / DOCK4.0/ Dock- energy-score, Contact-score, XCscore Pharmacophore and Drug like filter	7.5	
Comprehensive Medicinal Chemistry Database of Molecular Design Limited (MDL-CMC) (~8000)	Homology model, 1UK4	Docking/Dock4.01, Dock score, Cscore	0.013	
SPECS (81287)	1UK4	Docking/ DOCK 4.01/ Dock score, Cscore & scoring function of Autodock 3.0	3.12	
Maybridge (58 855)	1UK4	Docking, Similarity search/ GOLD2.1, fitness score and H-bond energy component of Gold score	4	
361413 small molecules	Homology model (2AJ5)	Docking/EUDOC	-	

Maybridge, ChemBridge, SPECS	1UK4	Docking, Similarity search/ DOCK 4.0.2/ Internal ligand- receptor binding energy term, electrostatic energy term	7.63 ^b	 (97)
MOE 2006.0804	-	Pharmacophore/ MOE	12.5	 (98) (99) (100)
Asinex platinum.2004	1UK4, 2AMD	GOLD 2.2 /composite GDR score	1.8	 (101) (102)

^ahit ratio is calculated by dividing the number of active hits retrieved against FP-2 by the number of compounds evaluated in the biological assay. ^bCompounds having IC₅₀ values below 30 μM were considered for the calculation of hit ratio.

1.5.4. Highthroughput Screening (HTS)

HTS against SARS 3CL^{pro} implemented rapid and sensitive fluorescence resonance energy transfer (FRET) assay in most of the screening discussed below ¹⁶⁸. The anti-SARS activities of the potential inhibitor identified through HTS were evaluated in a plaque reduction assay (PRA) or cytopathic effect (CPE) inhibition assay. HTS against SARS-CoV 3CL^{pro} can be classified into three approaches: experimental HTS where the *in-vitro* binding affinities of the screened hits were measured; virtual HTS where the binding affinities of the screen hits were only predicted; and finally, a novel dynamic ligand screening (DLS) methodology which perform the screening in a highthroughput manner, and has several advantages over conventional HTS. These approaches are summarized in the following sections.

1.5.4.1. Experimental HTS

Kao *et al.* screened the Chembridge database of structurally diverse 50,240 compounds using a continuous FRET assay against recombinant SARS-CoV M^{pro}. Compound **103** inhibited SARS-CoV by an IC₅₀ value of 2.5 μM and exhibited an EC₅₀ value of 7 μM in cell-based antiviral assays¹⁶⁹. Blanchard and colleagues screened the small molecule library of 50,000 drug-like compounds using a quenched FRET assay. Through a series of experimental and virtual filters, five novel inhibitors of SARS M^{pro} were discovered (IC₅₀ = 0.5-7μM), the most potent compound **104** in their screening had an IC₅₀ of 0.5 μM¹⁷⁰ (Figure 1.5.4.1).

Wu *et al.* carried out the screening of >10,000 compounds from the diverse source based on ELISA, immunofluorescence, flow cytometry and Western blot analysis. As a result of their efforts, compound **105** was found to have a K_i value of 0.6 μM against SARS-CoV M^{pro}¹³⁴. The docking simulation of compound **105** revealed that the inhibitor is folded into a ring-like structure in the active site and was stabilized by multiple H-bonds and hydrophobic interactions with the active site residues. Compound **105** was previously reported as a transition state inhibitor of HIV protease and found to be a promising candidate for further development against SARS due to its established *in vivo* efficacy in cats for anti-AIDS effects¹³⁴ (Figure 1.5.4.1).

Goetz *et al.* conducted HTS of a library of compounds developed using known cysteine protease inhibitor scaffolds with the substrate-specificity data obtained by in-house screening of natural tetrapeptides against recombinant SARS-CoV 3CL^{pro}¹²⁶. WRR495 (**106**) with dipeptidyl epoxyketone scaffold inhibited enzyme and viral replication (>50%) at a concentration of 10 μM. Further optimization of **106** with the help of computational modeling as well as knowledge of substrate-specificity (valine at P3) identified WRR495 (**107**), which was 175 fold more potent than parent compound **106** with a k₃/k₁ of 0.5 M⁻¹S⁻¹ against SARS 3CL^{pro} and an EC₅₀ of 12 μM against SARS-CoV¹²⁶. The crystal structures of these compounds with SARS-CoV 3CL^{pro} were

developed which showed interesting binding mechanism. The cysteine mediated nucleophilic attack on epoxide C2 resulted in the formation of an alcohol group in the P1 position exhibiting direct as well as water-mediated H-bonds in the S1 pocket, but also shifted the alanine from P1 to the P2 position. This observation raised the possibility of developing non-covalent inhibitors using this scaffold. Also, the P2-Ala side chain was exposed to a solvent supporting the less stringent P2 specificity for SARS M^{pro}. The P3-Phe side chain seemed to form sulfur-arene interactions with Met 49 and Met 165 of the canonical S2 pocket. Higher activity of **107** might be due to the use of the preferred amino acid Val for the S3 pocket ¹²⁶ (Figure 1.5.4.1).

Recently, Kuo *et al.* evaluated a compound library of 6800 compounds to find out individual and common inhibitors of coronavirus and picornavirus M^{pro}. Five novel inhibitors of SARS M^{pro} were identified (IC₅₀ = 3.3 to 10.6 μM). Compound **108** inhibited all the proteases (SARS M^{pro}, 229E M^{pro}, CVB3 3CL^{pro}, EV71 3CL^{pro}, RV14 3CL^{pro}) with an IC₅₀ of < 10 μM ¹⁷¹. The synthesis of analogs of **108** identified **109** as a broad spectrum inhibitor against CoV proteases. Compound **109** exhibited an IC₅₀ of 2.5 M against SARS M^{pro} as well as inhibited other CoV proteases with an IC₅₀ of < 3 μM ¹⁷¹.

Other known compounds identified by HTS against SARS M^{pro} include metal ions containing compounds (phenyl nitrate, thimerosal) ¹⁷², hexachlorophene and its analogs ^{172, 173}, polyphenols from black tea ¹⁷⁴, diuretic ethacrynic acid ¹⁷⁵, isatin derivatives and plant-derived phenolic compound hesperetin ¹⁷⁶.

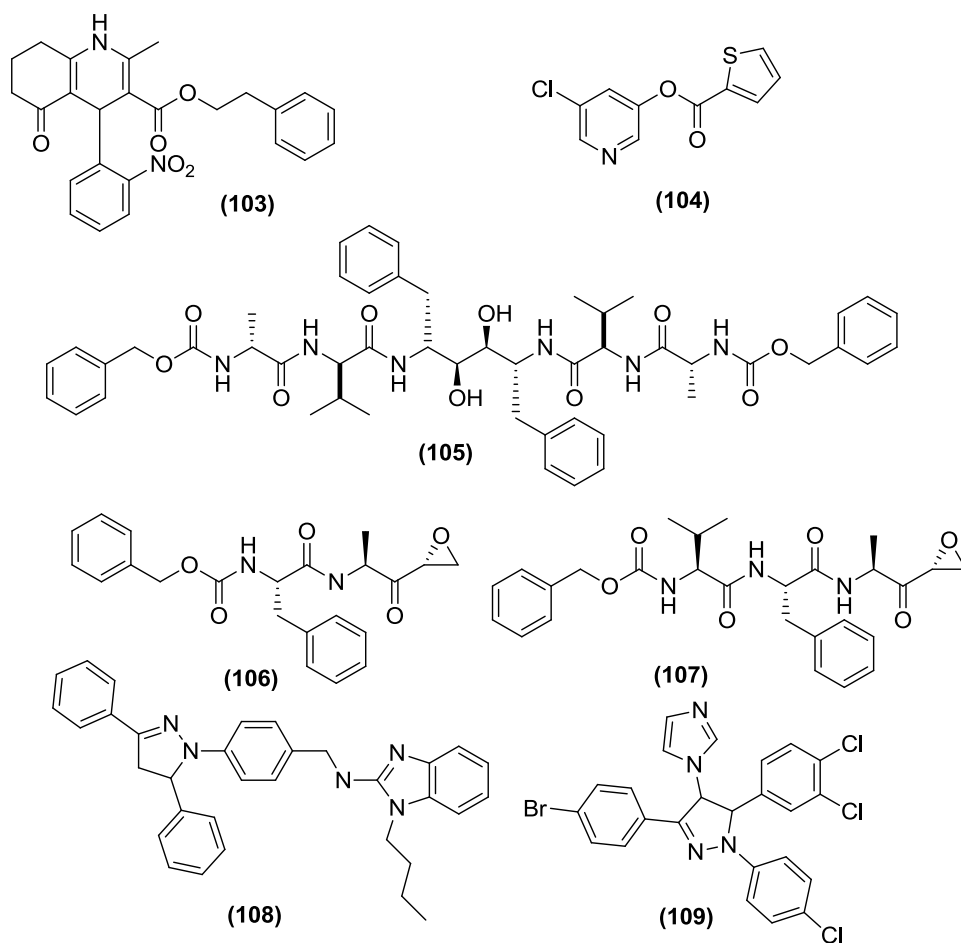


Figure 1.5.4.1. Non-peptidic inhibitors of SARS-CoV M^{pro} obtained by an experimental HTS

1.5.4.2. Virtual HTS

Jenwitheesuk *et al.*, carried out preliminary docking studies of 29 approved and experimental drugs against the homology model of the SARS-CoV proteinase as well as the experimental structure of the transmissible gastroenteritis virus (TGEV) proteinase to obtain an enzyme-inhibitor complex which was then refined by MD simulation protocol followed by final docking calculations to predict the binding affinity based on the lowest energy-solution¹⁷⁷. Docking calculations were performed using Auto Dock version 3.0.5, which employs the Lamarckian Genetic Algorithm (LGA) whereas MD simulations were carried out with the

NAMD (version 2.5b1) package in the TIP3 water shell using the X-PLOR force field. The inhibitory constant (K_i) was calculated according to Hess's law to define the binding affinity of the inhibitors. HIV-1 protease inhibitor, L-700417 **110** showed the highest binding affinity with an identical binding mode to the substrate and has proposed as a good starting point for the designing of SARS-CoV 3CL^{pro} inhibitors¹⁷⁷ (Figure 1.5.4.2).

Plewczynski *et al.* conducted structure-based *in silico* virtual HTS by extracting potential compounds from the protein-ligand complexes selected from the protein data bank (PDB) based on sequence similarity with SARS-CoV M^{pro}¹⁷⁸. Identified complexes mainly included the viral cysteine protease bearing the trypsin fold. Later, flexible docking was performed on the small molecules and peptides extracted from the protein-ligand complex in the crystal structure of SARS-CoV M^{pro} using eHITS (a flexible docking method for finding an accurate docking pose based on exhaustive positional and conformational searches in a timely manner). The best scoring compounds were then subjected to 2D similarity searches in Ligand.info Meta-Database (comprised of publically available database). A set of closely related structures bearing 2-methyl-2, 4-pentanediol **111** scaffolds were identified as lead candidate compounds against SARS-CoV M^{pro}¹⁷⁸ (Figure 1.5.4.2).

Zhang *et al.* performed the docking studies of existing HIV, psychotic and parasitic drugs (lopinavir, ritonavir, PNU, UC2, niclosamide and promazine) against the crystal structure of M^{pro} using the Hex docking program in a strategy of discovering new leads from old drugs. The estimated binding affinities of these compounds, especially lopinavir **112** and ritonavir **113** against SARS-CoV M^{pro}, indicated the potential of these drugs for designing new inhibitors of SARS-CoV M^{pro}¹⁷⁹. Meanwhile, Savarino *et al.* supported strong binding of ritonavir **113** with

SARS-CoV M^{pro} ($K_i \sim 10^{-5}$ M) stabilized by multiple H-bond network in their independent MD simulation¹⁸⁰. Zhang *et al.* developed a pharmacophore model based on the previously predicted pose of six drugs¹⁷⁹ and the crystal pose of the CMK inhibitor (1UK4, PDB code) in SARS-CoV M^{pro} using the POCKET module in the LIGBUILDER program¹⁸¹. The developed four-point pharmacophore model was used to search NCI database. A total of 30 drugs were selected after consideration of pharmacophoric criteria and preference for compounds with antiviral activity. Interestingly, six of the identified drugs already have known anti-SARS-CoV activity¹⁸¹. Meanwhile, Wu *et al.* also evaluated HIV protease inhibitors, currently in clinic as inhibitors of SARS-CoV M^{pro}, and discarded the possibility of ritonavir as a potential inhibitor of SARS-CoV M^{pro}. However, lopinavir **112** showed moderate activity against SARS M^{pro} with an IC₅₀ value of 50 μ M ($K_i = 14 \mu$ M)¹³⁴ (Figure 1.5.4.2).

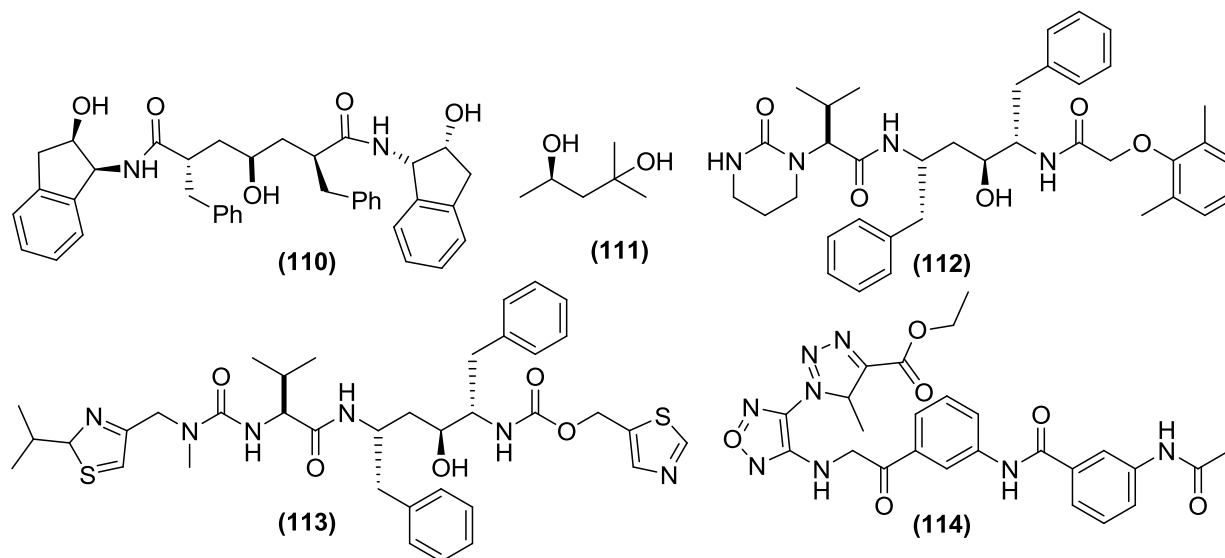


Figure 1.5.4.2. Proposed inhibitors of SARS-CoV M^{pro} by virtual HTS and DLS approach

1.5.4.3. Dynamic Ligand Screening (DLS)

Recently, Schmidh *et al.* reported the DLS approach for the discovery of non-peptidic inhibitors against SARS-CoV 3CL^{pro} ¹⁸². This approach involves site-directed identification of low-affinity inhibitory fragments against a particular target by *in situ* formation of inhibitory species followed by its detection with a fluorogenic reporter substrate of that target (readers are referred to original reference ¹⁸² for more detail). This method presents several benefits over conventional HTS or fragment-based approach using NMR or X-ray crystallography, for example, possible detection of low affinity fragments with inhibition constant in milimolar range, requirement of minimal amount of protein for assay, exploration of a wide range of the chemical space, minimum requirements equipments and transformation of a moderately active peptide to an entirely non-peptidic inhibitor with a micromolar range. A “reverted” DLS methodology adopted for screening against SARS 3CL^{pro} successfully identified a lower micromolar non-peptidic inhibitor **114** of the enzyme (Figure 1.5.4.2) with an IC₅₀ of 2 μM.

In short, although, SARS has been contained, the virus is not eradicated yet. In these circumstances, high potency of the developed inhibitors of SARS-CoV M^{pro} illustrates this enzyme suitable for the development of anti-SARS therapy. Also, the availability of multiple crystal structures of SARS-CoV M^{pro} in complex with peptidic/non-peptidic inhibitors from diverse classes presents unique opportunities to discover promising inhibitors of this enzyme by structure based drug designing. As mentioned before, P1-Glu/His triggers the potency of SARS inhibitors and should be focused while designing of peptidomimetics. Moreover, S1, S2 and S4 subsites which are highly conserved among Co-V can be targeted to develop broad spectrum inhibitors against Co-V related diseases. Also, structure analogs of Rupintrivir with established activity against 3C^{pro} of rhino and other Co-V protease can serve as a template for discovery of

common inhibitors of Co-V. In addition, with number of inhibitors in hand, efforts can be made to correlate the physical, chemical, biological properties of reported inhibitors with SARS-CoV M^{pro} activity in the form of 2D Quantitative Structure Activity Relationships (QSAR) model and can be used for virtual screening of chemical libraries to identify novel hits. Also, 3D QSAR modeling can be performed with the help of published crystal structures enabling structural alignment of known inhibitors.

1.6. Conclusions

This chapter highlights a variety of the rational approaches for drug discovery against two cysteine proteases, protozoal protease “falcipain” and viral protease “SARS 3CL^{pro}”. To the best of my knowledge, this chapter covers the peptidic and non-peptidic inhibitors from major classes of inhibitors reported against falcipain and SARS-CoV M^{pro} to date. The cysteine proteases discussed here are of the utmost importance for drug design considerations. For instance, several efforts are currently underway to develop effective antimalarials for encountering the parasite resistance to all front-line drugs, including the most potent artemisinin derivatives ⁹⁵. In such scenarios, falcipains manifests an emerging drug target due to its unique function in hemoglobin degradation and survival of the protozoa. Moreover, effective peptidic and non-peptidic inhibitors of falcipains (summarized above) with good inhibitory potency project this enzyme as a promising target for the development of antimalarial chemotherapy. Likewise, although SARS has been contained, possible reemergence of the disease with mutated forms of the virus still remains a potential threat to human health. From this view, structurally and biochemically well-characterized SARS-CoV M^{pro}, playing pivotal role in the virus replication, provide an opportunity for drug development in case of the next pandemic of SARS.

As discussed above, knowledge-based drug design approaches have been effective for the rapid discovery of inhibitors against these targets, for example, evaluation of inhibitors of other cysteine proteases such as papain, cathepsins, calpain or cruzain against FP-2, identified vinyl sulphone class of inhibitors, the most potent inhibitors reported till date against this enzyme. Structure-guided drug design helped to optimize the potential or known compounds against the target protease with the help of three-dimensional structure and knowledge of the substrate specificity. For instance, structure modifications of rupintrivir (known rhinovirus protease inhibitor with no efficacy against SARS) identified potent inhibitors of SARS 3CL^{pro} with a nanomolar to lower micromolar activity range. In addition, structure-based virtual screening and HTS helped to discover non-peptidic inhibitors with diverse scaffolds against both enzymes. Taken together, although the cysteine protease discussed here are quite different from each other, the approaches presented here provide general guidelines for drug design against the cysteine protease of existing and possible emerging pathogens as well as those present in humans. In addition, the rational approaches described for developing inhibitors against SARS set successful examples to initiate drug design efforts against other viral pandemics such as the swine flu (H1N1) virus.

Chapter 2

Identification of Falcipain-2 Inhibitors Using Structure-Based Virtual Screening of Focused Cysteine Protease Inhibitor Library

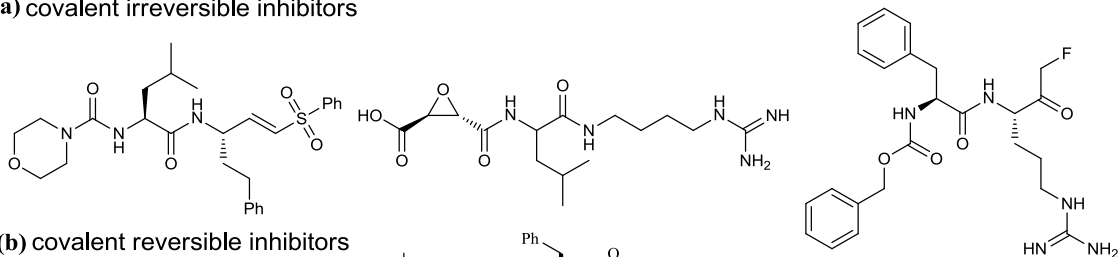
Content of this chapter is published in: Shah F., Mukherjee, P., Jiri G., Rosenthal, P., Tekwani B., Avery M. A. Identification of Novel Malarial Cysteine Protease Inhibitors Using Structure-Based Virtual Screening of a Focused Cysteine Protease Inhibitor Library, *Journal of Chemical Information and Modeling*, 2011, 51 (4), 852.

2.1. Introduction

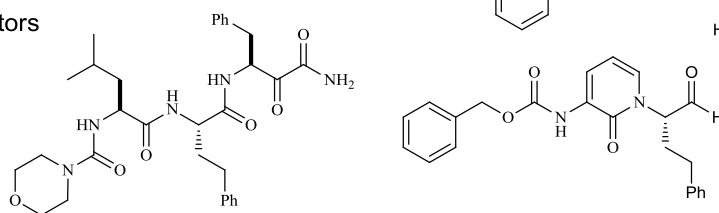
In the Chapter 1, extensive efforts conducted for the discovery of inhibitors against FPs using rational drug design approaches were described. In general, drug design efforts against FPs can be broadly classified in two main categories: peptide/peptidomimetic covalent inhibitors and non-peptidic small molecule inhibitors. The covalent irreversible inhibitors contain reactive warheads or electrophiles such as vinyl sulfones^{67, 75}, epoxy-succinate^{43, 50}, and fluoromethyl ketones (FMKs)^{43, 61} whereas compounds with electrophiles such as α -keto amide⁶⁹, aldehydes⁶⁹ and nitriles¹⁸³ function as a reversible covalent inhibitors of the FPs (Figure 2.1a and 2.1b). In particular, the covalent modifiers with reactive warheads blocked the hydrolysis of hemoglobin by erythrocytic parasites, formed hemoglobin-filled food vacuoles in trophozoites and, hampered parasites development^{43, 60-62, 64, 184}. Despite their excellent invitro potency against cultured parasites, these compounds exhibited poor selectivity towards the parasitic cysteine protease over mammalian cysteine proteases^{62, 77, 185}. In addition, potential toxicity associated with these compounds due to the formation of protein adducts restricted their further development as potential drug candidates. Moreover, peptidic nature of these inhibitors makes them susceptible to hydrolysis by host cell proteases. The other categories of FP-2 inhibitors include non-peptidic small molecule inhibitors of the enzymes (Figure 2.1c). Several structure based virtual screening attempts were made to identify non-peptidic reversible inhibitors against 3D models of FPs from the commercial databases^{80, 86, 186}. However, to date, only a few non-peptidic inhibitors of FP-2 are available with a lower micromolar (<10 μ M) affinity as well as with the ability to inhibit cultured parasites. Thus, design of non-peptidic, non-covalent inhibitors of FPs is an important goal, with the likelihood of being able to develop compounds with desirable druggable

properties, good *in vivo* potency, and biochemical selectivity.

(a) covalent irreversible inhibitors



(b) covalent reversible inhibitors



(c) non-peptidic small molecule inhibitors

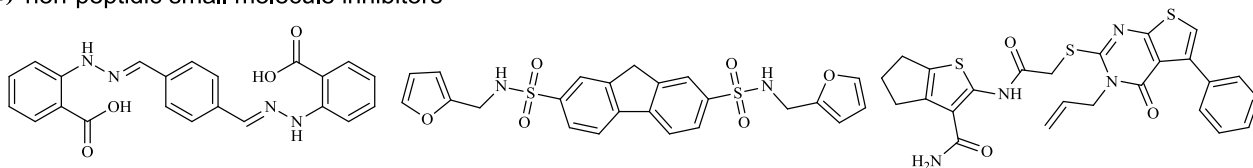


Figure 2.1. Representative FP-2 inhibitors (a) covalent irreversible inhibitors^{50, 75, 77}; (b) covalent reversible inhibitors^{69, 75} and, (c) non-peptidic small molecule inhibitors^{80, 86, 186}.

Over the past decade, structure-based virtual screening (SBVS) has gained attention and successfully identified novel scaffolds against different targets¹⁸⁷. In brief, SBVS involves the computational docking of a commercial database into the active site of the 3D structure of the target, followed by the selection of putative binders using scoring functions¹⁸⁸⁻¹⁹¹. With this approach, only a small set of compounds is selected and evaluated in a biological assay, as opposed to the high throughput screening of an entire compound library^{192, 193}. As a part of our efforts to discover non-peptidic reversible inhibitors of FP-2, we conducted SBVS against the crystal structure of FP-2 (PDB code: 2GHU)⁴⁹ using the Glide docking program. We first built a focused cysteine protease inhibitor (FCPI) library by substructure searching for compounds bearing soft electrophiles of interest from the pool of several commercial databases. The

developed FCPI library was then screened against the crystal structure of FP-2 (PDB code: 2GHU, apo form of the enzyme) to discover novel non-covalent inhibitors of the enzyme. Evaluation of the selected set of compounds identified 21 diverse non-peptidic inhibitors of FP-2 with lower micromolar potency and a potential for further structural optimizations in to the lead candidates. Four compounds also inhibited FP-3. Two compounds showed corresponding inhibition of the cultured parasites. In addition, selected compounds were evaluated against mammalian cysteine proteases (cathepsin B, K and L) of the papain-family. One compound was found to be a selective inhibitor of FP-2. Finally, an assessment of electrophilicity of electrophiles in selected target compounds was carried out by calculation of LUMO density of each atom. The results suggested the soft-nature of electrophiles in identified hits.

2.2. Methods

2.2.1. Computational Tools. All calculations were performed on a Linux workstation equipped with four parallel Intel Xeon X5460 processors (3.16 GHZ) with 8 GB total RAM. LigPrep¹⁵⁷ was used to produce low energy 3D structures of compounds. Glide docking program¹⁹⁴ was used for pose validation, enrichment study and SBVS. QikProp¹⁹⁵ was used to calculate chemical properties of the FCPI library. Phase¹⁹⁶ was used to build docking pose based pharmacophore query. Diversity analyses of the final hits were performed using the Dissimilarity module available in Sybyl¹⁹⁷. Geometry optimization and measurement of electrophilicity of identified compounds were performed using Atomic Fukui indices implemented in Jaguar¹⁹⁸. All images were created using Pymol¹⁹⁹.

2.2.2. Building the Focused Cysteine Protease Inhibitor (FCPI) Library. The selected soft-electrophiles of interest used for building of FCPI library include α -heteroatom substituted

ketones, α -heteroatom substituted amides, azetidinones, α -keto amide, α -keto acid and α -keto ester (Figure 2.2.2). Substructure searches were conducted with single line notations (SLN) of the soft electrophiles (Figure 2.2.2) as a query in Asinex platinum (<http://www.asinex.com>), Chembridge (<http://www.hit2lead.com>), Specs (<http://www.specs.net>), Enamine (<http://www.enamine.net>), IBScreen (<http://www.ibscreen.com>) and Aurora Fine Chemicals Ltd (<http://www.aurorafinechemicals.com>), using the dbsearch command from the Unity search module of Sybyl 8.1 (Tripos Inc., St. Louis, MO). The identical compounds obtained from a pool of different databases were removed. Thereafter, the database was filtered using chemical properties calculated by QikProp, such as molecular weight (MW) = 250-550, H-bond acceptor 1-10, H-bond donar = 0-5 and no. of rotatable bonds = 0-10. In addition, compounds containing problematic groups, such as metals, N-oxides, chloramines, aldehydes, and peroxides, were removed from the FCPI library. The dbslfilter utility of the Unity module of Sybyl 8.1 was used to carry out database filtration. Around 65,000 compounds with desired soft-electrophiles were thus stored in the FCPI library.

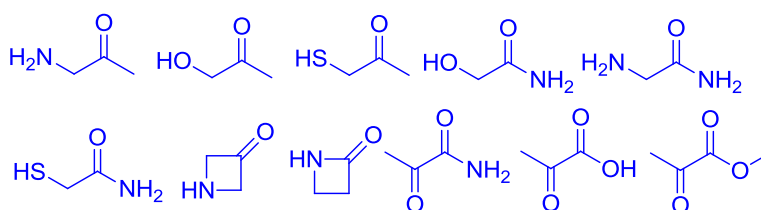


Figure 2.2.2. Soft-electrophiles used in the substructure search to build FCPI library.

2.2.3. Protein Preparation. In the present SBVS study, we utilized a recently published crystal structure of FP-2 available in the apo form (PDB code: 2GHU)⁴⁹. The residues within 14 Å of the Cys42 were included in the binding site definition. The PPREP utility of Maestro

(Schrodinger, LLC, Portland, OR) was used for protein preparation. The protein preparation was carried out at pH 5 to mimic the acidic environment of the *P.falciparum* food vacuole (pH = 4-6) where the enzyme is located. Acidic residues Asp and Glu and basic residues Arg and Lys were treated as charged unless they were surrounded by hydrophobic residues. Figure 3a shows the magnified view of the FP-2 active site with subsites S1, S1', S2 and S3 labeled. Protein with all hydrogens added was then submitted to restrained molecular mechanics refinement using the OPLS2001 force field incorporated in the IMPREF (Schrodinger, LLC, Portland, OR) protein structure refinement utility, and minimization was continued until the root mean square deviation (RMSD) reached 0.3 Å. The final refined structure was used for the docking calculations.

2.2.4. Pose Validation, Enrichment Study and Docking Pose-Based Pharmacophore Query.

In the absence of the co-crystallized structure of FP-2, a preliminary docking study was carried out in the homologous cysteine protease cruzain, the major cysteine protease of *T. cruzi*, co-crystallized with a vinyl sulfone (VS) inhibitor (PDB ID: 1F2A)²⁰⁰. The criteria for selecting cruzain as a surrogate platform to establish a docking protocol are as follows: (a) both cruzain and falcipains are papain family cysteine proteases; (b) the superimposition of FP-2 (2GHU) and cruzain (1F2A), using DaliLite server²⁰¹, matches 209 α -carbons with an RMSD of 1.6 Å, a Z score of 30.4 and, with an overall sequence identity of 38% (Figure 2.2.4.1). Both proteins were found to have 62% of sequence identity and, RMSD of 2.6 Å in the ligand binding domain as defined by 10 Å radius from catalytic cysteine (Cys42 in FP-2 and Cys25 in cruzain, see Figure S0); (c) vinyl sulfone inhibitors, for example, shown in Figure 2.1(a) is a common inhibitor of both enzymes^{67, 200}.

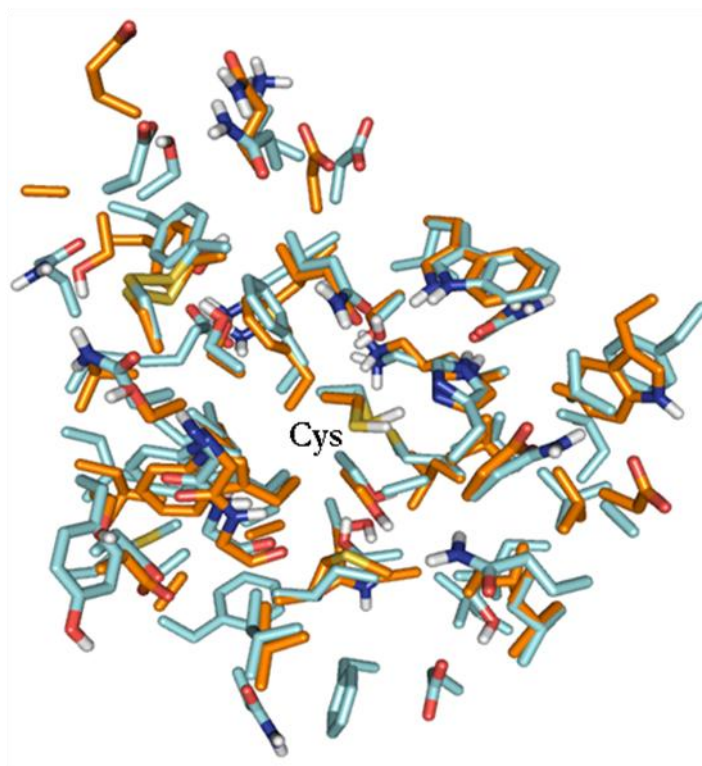


Figure 2.2.4.1. Superimposition of FP-2 (shown in cyan) and Cruzain (shown in orange) using DaliLite server.

Structural alignment of cruzain complexes, for example, with the irreversible covalent vinyl-sulfone inhibitor (PDB code: 1F2A, 2OZ2) as well as the reversible covalent inhibitors such as hydroxyl methyl ketones (PDB code: 1ME3) and a purine nitrile (PDB code: 3I06) using superimposition module of maestro showed an average RMSD of 0.45 Å. This suggests that receptor structure does not change substantially upon the binding of inhibitors from different chemotypes. Therefore, rigid receptor docking protocol implemented in glide was used throughout the study. Once initial docking protocol is established in cruzain, subsequent enrichment and docking studies were performed in the crystal structure of FP-2.

A pose validation study was performed using the Glide XP mode without any docking constraints. The top ranking pose obtained for the VS inhibitor in cruzain using the Gscore

scoring function was closest to its experimental binding mode (Figure 2.2.4.2b). The RMSD between predicted and experimental poses of VS inhibitor was found to be 0.97 Å, which was quite satisfactory considering the number of rotatable bonds of VS (structure shown in Figures 2.2.4.2b and 2.2.4.2d). Subsequent docking of VS inhibitor in FP-2 active site reproduced the arrangements of P1-P3 groups as observed in X-ray structure of cruzain. Moreover, hydrogen bonds of VS inhibitor with Trp177, Gln19, Gly66, and Asp158 of cruzain binding site were reproduced in the FP-2 active site.

The actives from the previous SBVS study^{80, 86} (having FP-2 IC₅₀ ≤ 30 μM, see Figure 2.2.4.3 and Table 2.2.4.1) were selected for the enrichment study. Active compounds were seeded with a total of 1000 decoys downloaded from the Directory of Useful Decoys (DUD) database²⁰². Decoys were selected by considering the chemical properties such as MW (250-550), no. of HBD (2-4), no. of HBA (2-8) of known actives. Selection of the diverse sets of decoys resembling properties of known actives made the enrichment study quite challenging.

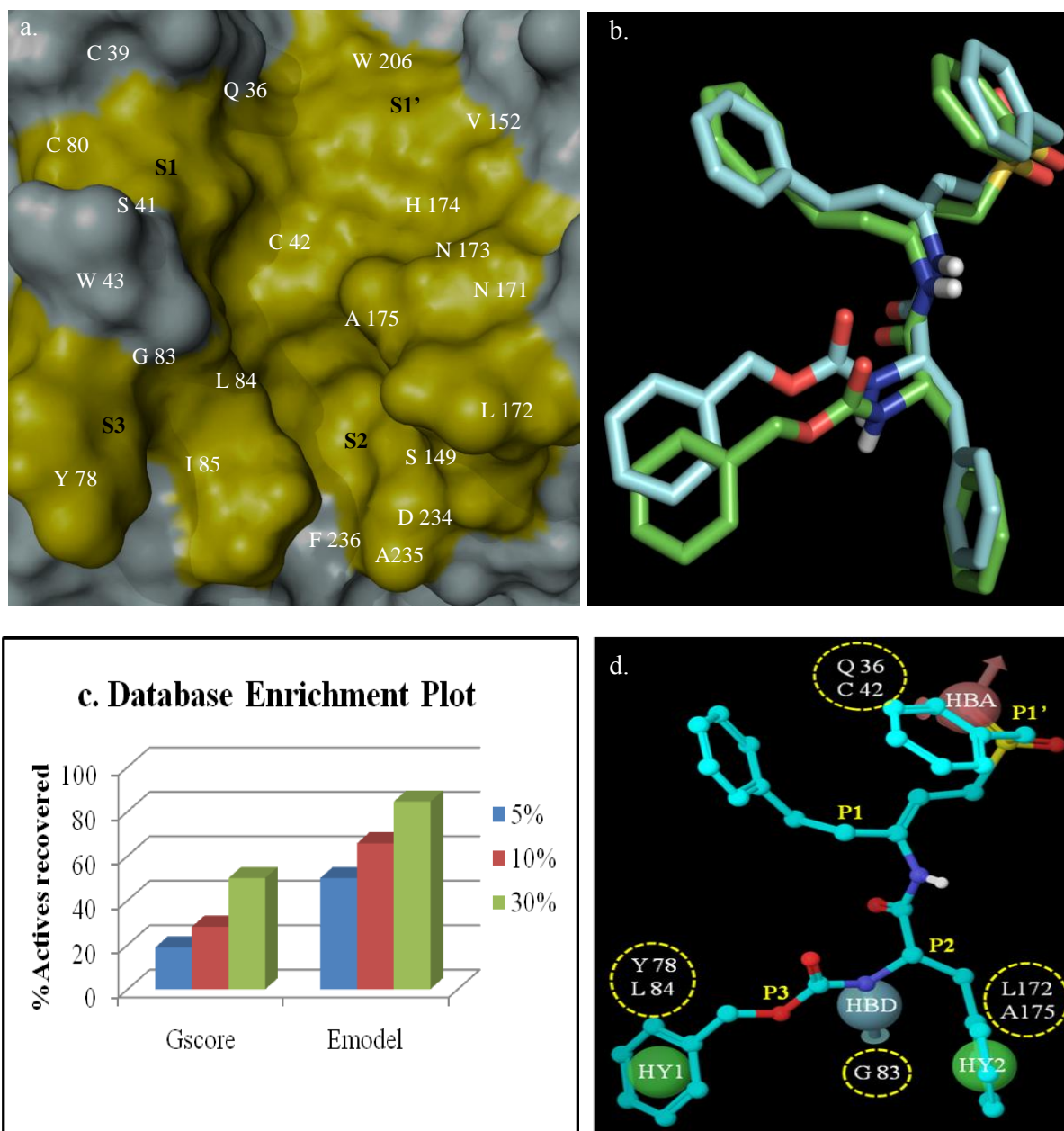


Figure 2.2.4.2 (a) Magnified view of the substrate-binding pocket of FP-2 (PDB code: 2GHU). Subsites S1, S1', S2 and S3 of the enzyme are shown; (b) Pose validation study: superimposition of co-crystallized pose (green) and the top-ranked pose (cyan) obtained through docking; (c) Enrichment study: a bar graph showing percentage of actives recovered at 5% (blue), 10% (red) and 30% (green) of the ranked database with Gscore and Emodel scoring functions; (d) Docking pose-based pharmacophore query: HBD= hydrogen bond donor; HBA=hydrogen bond acceptor; HY=hydrophobic. P1, P1', P2 and P3 side chains of VS and interacting residues Gln36, Cys42, Gly83, Tyr78, Leu84, Leu172 and Ala175 of FP-2 binding site are shown.

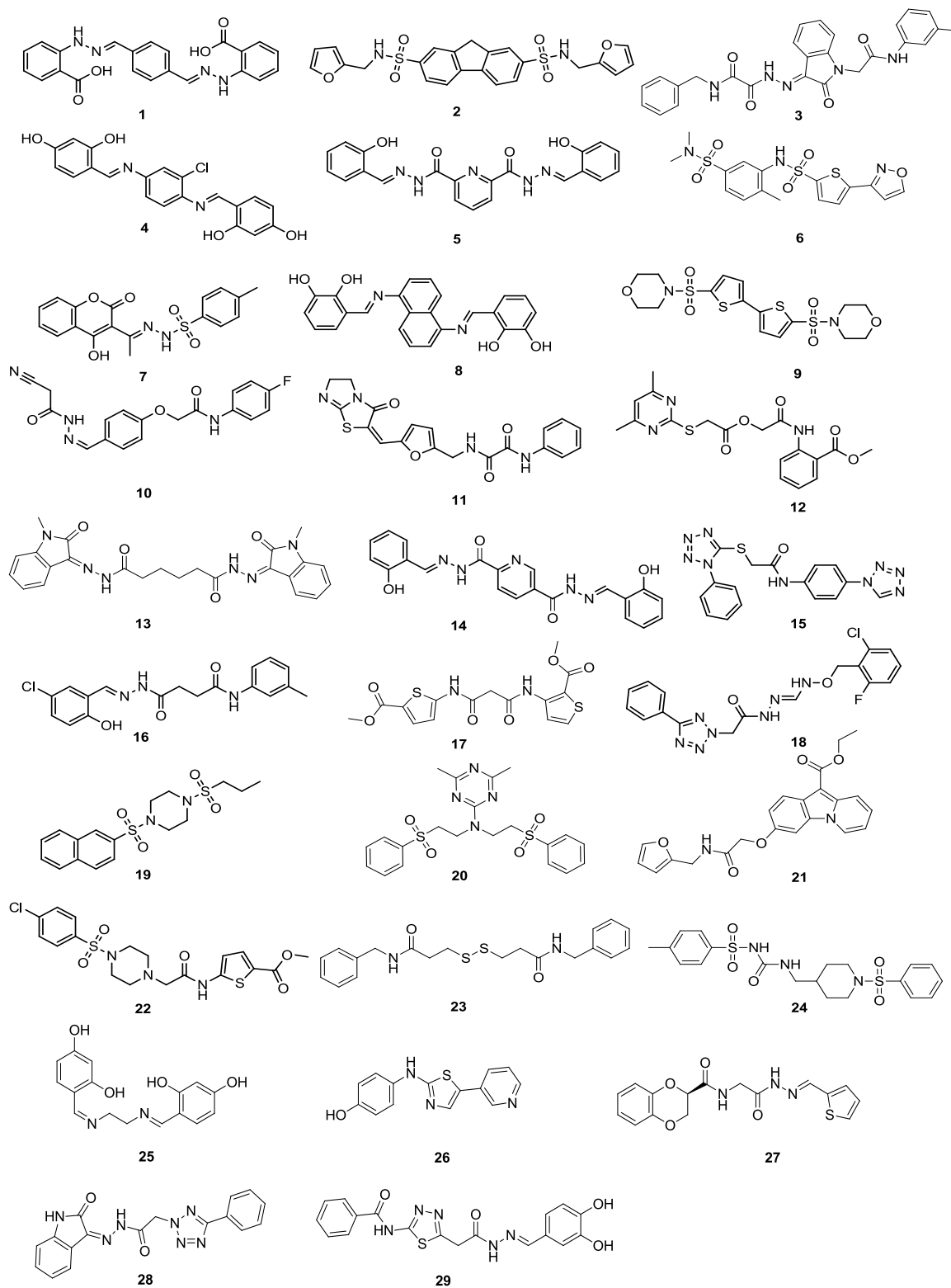


Figure 2.2.4.3. Chemical structures of previously reported FP-2 hits ($IC_{50} \leq 30 \mu M$) used in enrichment study^{80, 86}

Table 2.2.4.1 Reported IC₅₀ of previously identified FP-2 hits.

Compound Code	FP-2 ^a IC ₅₀ (μM)	Compound Code	FP-2 ^b IC ₅₀ (μM)
1	1.0	16	7.5
2	1.4	17	10.4
3	1.6	18	10.8
4	2.1	19	12.4
5	2.2	20	13.2
6	2.9	21	13.8
7	3.9	22	14.3
8	4.6	23	15.3
9	5.4	24	19.0
10	5.7	25	20.8
11	6.2	26	22.5
12	6.2	27	26.6
13	6.5	28	27.3
14	7.0	29	31.1
15	7.5		-

In addition, a receptor-based pharmacophore pre-filter was generated based on the docking pose of the peptidic vinyl sulfone in the crystal structure of FP-2 (2GHU, PDB code) encoding key receptor-ligand interactions. Phase 3.0 was used to build the pharmacophore query. The pharmacophore shown in Figure 2.2.4.3d consists of two hydrophobic features (green) mapped onto the P2 and P3 side chains of VS interacting with hydrophobic residues of S2 and S3 pockets, a hydrogen bond donor feature (cyan) mapped onto the –NH of the P3 side chain forming H-bond with Gly83 and, a hydrogen bond acceptor (pink) forming the H-bond with the residues of oxyanion hole (Cys42 and Gln36). As in the S2/S3 pockets, there is no strict requirement for the ring aromatic feature, the built-in definition for the directional ring aromatic feature was excluded for performing the search against the FCPI library, and the SMART

patterns for aromatic rings were included in the hydrophobic feature definition.

2.2.5. Structure-based virtual screening protocol. The FCPI library consisting of ~65,000 compounds was prepared for virtual screening using the ligprep module of Schrodinger (see Figure 2.2.5). Compounds were subjected to hydrogen additions, removal of salt, ionization, and generation of low energy ring conformations. The chiralities of the original compounds were preserved. Finally, the low energy 3D structures of all compounds were produced. To curtail the database prior to docking, a prepared database was prefiltered using a four-point structure-based pharmacophore query based on the docking pose of vinyl sulfone. Compounds matching a minimum 3 out of 4 features of the pharmacophore query were retained. At this stage, filtered compounds were expected to possess soft electrophiles of interest and relevant pharmacophoric features of FP-2 inhibitors, which increase the probability for docking of more relevant compounds in the active site of FP-2. Approximately 30,000 compounds thus retained were subjected to docking in the crystal structure of FP-2 using the previously validated Glide XP protocol. The top 3,000 molecules (top 10%) ranked on the basis of Emodel score were selected for eMBrAce minimization calculation (Macromodel). The eMBrAce energy minimization in the energy difference mode was carried out for the protein-ligand complexes to study the association of ligands with the receptor. The energy difference was calculated as follows:

$$\Delta E = E_{complex} - E_{ligand} - E_{protein}$$

where ΔE = energy changes upon association; $E_{complex}$ = energy of the receptor-ligand complex; E_{ligand} = energy of ligand alone, and $E_{protein}$ = energy of protein alone. Based on EMBRACE score, the top 500 molecules were selected.

Next, compounds were visually inspected for the following criteria: a) presence of the soft

electrophile in the vicinity of catalytic Cys42; b) key hydrophobic (Leu172, Leu84) or H-bonding interactions (Gln36, Gly83) between ligand atoms and protein residues; c) the internal geometry of the ligand in the binding site; d) occupancy of the S2 sub-pocket by the hydrophobic group of the ligand (hemoglobin cleavage studies demonstrated strong preference for cleavage sites with Leu at P2 position as a major determinant of the specificity against this enzyme²⁰³, therefore, hydrophobic interactions of ligand in S2 sub-site were considered essential); and, e) commercial availability of putative hits. Based on these criteria, 200 molecules were selected and subjected to diversity analysis using the Dissimilarity module of Sybyl 8.1 (Tripos Inc., St. Louis, MO). A total of 50 diverse compounds were purchased from respective vendors. The Tanimoto similarity matrix was generated to calculate an average pair-wise Tanimoto similarity index of identified hits using Similarity/Distance matrix utility implemented in Canvas, version 1.2 (Schrödinger, LLC, New York, NY, 2009), using atom-pair fingerprints. The comparisons of the obtained hits in the present study with those from previous SBVS studies were carried out using the dbcmpr utility of Sybyl 8.1 (<http://www.tripos.com>). The comparison was done by constructing a UNITY database of each set using Sybyl 8.1.

Virtual Screening Workflow

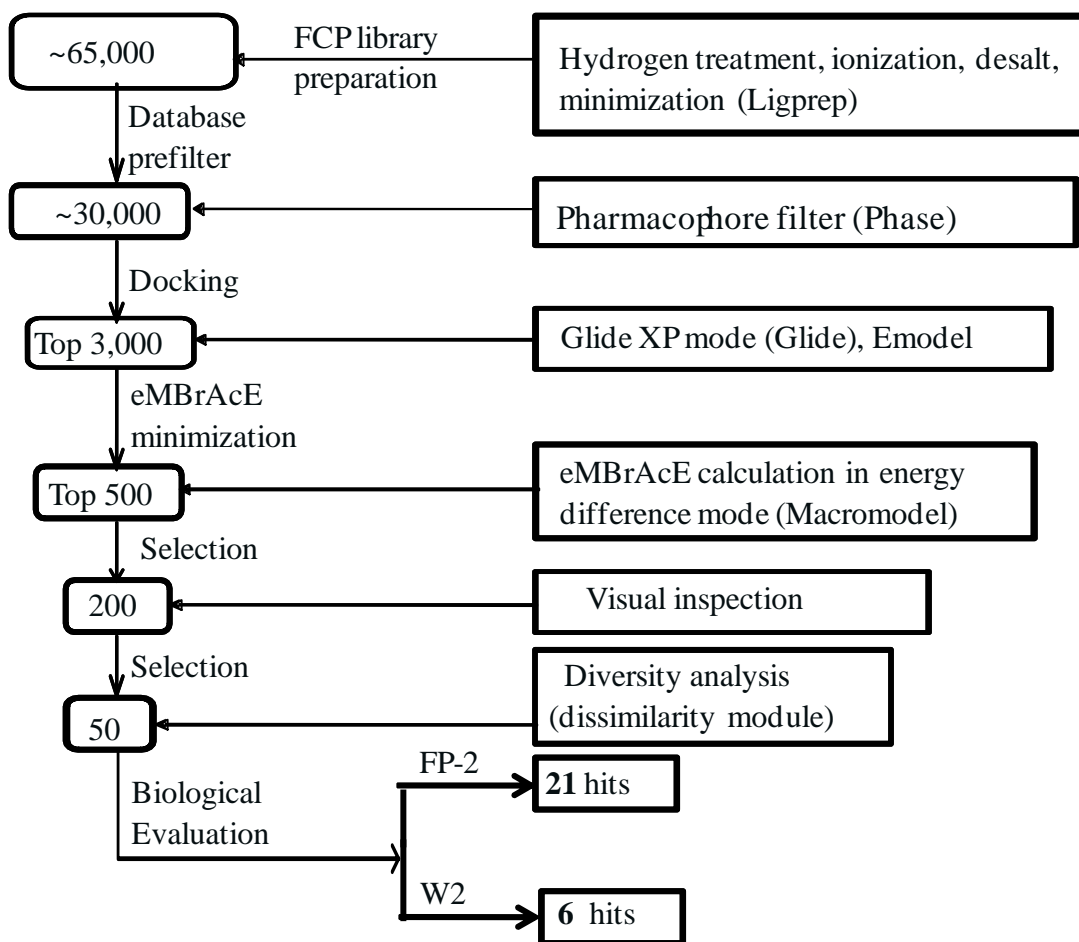


Figure 2.2.5. A flow chart depicting the virtual screening protocol utilized in the study. W2 stands for chloroquine resistance strain of *P.falciparum*.

2.2.6. Biological Evaluation. The selected 50 compounds were evaluated for inhibition of FP-2, FP-3 and against chloroquine resistant (W2 strain) *P.falciparum* parasites. To determine IC_{50} values, recombinant FP-2 and FP-3 were incubated for 30 minutes at room temperature in 100 mM sodium acetate, pH 5.5, and 10 mM dithiothreitol with different concentrations of inhibitors prepared from stocks in DMSO (maximum concentration of DMSO in the assay was 1%). After 30 minutes, the substrate Z-Leu-Arg-AMC (benzoxycarbonyl-Leu- Arg-7-amino-4-methyl-

coumarin) in the same buffer was added to a final concentration of 25 μ M. Fluorescence was monitored for 15 minutes at room temperature in a Lab systems Fluoroskan Ascent spectrofluorometer. IC₅₀ values were determined from plots of percent activity over compound concentration using GraphPad Prism software²⁰⁴.

The assays for selected human cysteine proteases, cathepsin B, K and L were conducted in a similar way as described above for FP-2. The assay concentration used for cathepsin B, K and L were 7.94 nM, 3.40 nM and, 5.38 nM, respectively. The appropriate substrates (Z-Phe-Arg-AMC for cathepsin K and L, and Z-Arg-Arg-MCA for cathepsin B) were added to a final concentration of 25 μ M.

Activity against malaria parasites was tested against the chloroquine-resistant (W2) strain of *P. falciparum*, which was cultured at 2% hematocrit of human erythrocytes in RPMI 1640 medium supplemented with 0.5% Albumax (Gibco), 2% heat inactivated human serum and 100 μ M hypoxanthine in 96-well culture plates. Parasites were synchronized with 5% sorbitol²⁰⁵. All tested compounds were prepared as 10 mM stock solutions in DMSO and diluted in medium at least 1:1000, resulting in \leq 0.1% final concentrations of DMSO. Parasites at 1% parasitemia were incubated with compounds at 37 °C under 3% O₂, 5% CO₂, 92% N₂. After 48 hours, the medium was removed, and cells were fixed with 2% formaldehyde in PBS for 24 hours. After fixation, 5 μ l aliquots were incubated for 15 min in the dark in 150 μ l of 100 mM NH₄Cl, 0.1% Triton X-100 in PBS, and 10 nM YOYO-1 (Molecular Probes). Parasitemia was determined based on counts from a FACSort flow cytometer (Beckton Dickinson) using CellQuest software (Beckton Dickinson). IC₅₀ values for growth inhibition were determined from plots of percent control parasitemia over inhibitor concentration using Prism[®] software (GraphPad). Compounds were

also evaluated for cytotoxicity on VERO (monkey kidney fibroblast) cells by the neutral red assay²⁰⁶. IC₅₀ values for each compound were computed from the growth inhibition curve.

2.2.7. Calculation of electrophilicity of soft-electrophiles in identified hits

Atomic Fukui indices, derived from Mulliken population of the Highest Occupied Molecular Orbital (HOMO) and the Lowest Unoccupied Molecular Orbital (LUMO), were used to quantify electrophilicity of a molecule at a particular atomic site. Two different set of Fukui indices calculated include the HOMO and the LUMO. Each index has two subscripts: N stands for the electron density and, S stands for the spin density. Thus, a total of four indices were calculated for both, the HOMO and the LUMO: F_{NN}, F_{NS}, F_{SN}, and F_{SS}. The first index shows which property responds to a change in the property indicated by the second index, hence, F_{NN} indicates the change in the electron density when only the electron density that is changing, and the spin density is constant. Similarly, F_{NS} represents the change in the electron density when only the spin density is changing etc. F_{NN_LUMO} was considered for measuring electrophilicity of electrophilic center in the target compounds. Structures were subjected to the geometry optimization using the hybrid density functional B3LYP^{207, 208} and a basis set 6-31G* in a gas phase using Jaguar. The default convergence criterion implemented in Jaguar were used for SCF calculations (accuracy level= Quick, convergence criteria: maximum iteration=48 and, energy change= 5×10^{-5} hartree), and, optimization (Maximum steps=100, convergence criteria=default, initial hessian=Schlegel guess²⁰⁹). The fully optimized geometry of compounds were subjected to the calculation of Atomic Fukui indices implemented in the property section of the Jaguar optimization panel, according to formalism presented elsewhere^{210, 211}. Fukui function for the electron density integrates to one, so the predicted reactivity here are scaled from zero to

one, one being the most reactive. It is worthwhile mentioning that in all cases, the SH group of the active site cysteine has been assumed as a nucleophile or HOMO component of a nucleophilic-electrophilic addition reaction.

2.3. Result and Discussions

Out of 50 compounds submitted for biological testing, a total of 21 diverse non-peptidic hits were identified. The IC₅₀ values of hits for FP-2 inhibition ranged from 1.4 μM to 49 μM (shown in Table 2.3.1). Four compounds **2**, **5**, **8** and **11** were found to be dual inhibitors of FP-2 and -3. The structures of identified inhibitors are depicted in Figure 2.3.1.

Table 2.3.1. Biological evaluation of compounds selected hits from SBVS against FP-2 and cultured parasites.

Compound Code ^c	FP-2 ^a IC ₅₀ (μM)	FP-3 ^{a1} IC ₅₀ (μM)	W2 ^b IC ₅₀ (μM)	Compound Code ^c	FP-2 ^a IC ₅₀ (μM)	FP-3 ^{a1} IC ₅₀ (μM)	W2 ^b IC ₅₀ (μM)
1	1.39	>50	NA	12	23.81	>50	NA
2	2.18	4.95	NA	13	25.25	>50	NA
3	3.26	>50	NA	14	25.96	>50	NA
4	4.59	>50	NA	15	25.99	>50	NA
5	7.51	30.27	1.90	16	28.66	>50	NA
6	7.85	>50	4.34	17	32.85	>50	NA
7	10.32	>50	NA	18	37.34	>50	1.92
8	11.58	45.64	NA	19	44.94	>50	NA
9	11.86	>50	NA	20	46.97	>50	NA
10	15.11	>50	1.57	21	49.18	>50	5.83
11	20.87	36.05	NA	22	>50	>50	7.21
E-64	0.08	0.15	-	E-64	0.08	0.15	-
Artemisinin	-	-	0.01	Artemisinin	-	-	0.01
Chloroquine	-	-	0.05	Chloroquine	-	-	0.05

NA= not active up to 50 μM; ^afalcipain-2, ^{a1}falcipain-3, ^bAntimalarial activity against chloroquine resistant (W2) strain of *P. falciparum*. ^call compounds were non-cytotoxic at least up to 50 μM in VERO (monkey kidney fibroblast) cells. E-64, artemisinin and chloroquine were used as positive controls.

A major bottleneck in SBVS is finding a suitable scoring function for the target of interest to enable the identification of potential active compounds prior to the screening of large chemical databases. Novel approaches to address this problem include the use of a target-biased scoring function which takes into consideration the nature of the target (polar vs apolar binding site) as well as the nature (pharmacophoric patterns and similarity considerations) and key interaction of known ligands with the target²¹². In the present study, two scoring functions implemented in the Glide module of Schrodinger, Gscore and Emodel were evaluated for their abilities to retrieve known FP-2 inhibitors from decoys downloaded from the DUD database. The recovery rates (Figure 2.2.4.2c) showed that among the scoring functions evaluated, Emodel performed better than Gscore in 5%, 10% and 30% database screens. At the 5% database screened mark, the percentage of actives recovered was 50% for Emodel while at the 10% and 30% database screened marks, the percentage of actives recovered was 66% and 84% by the Emodel scoring function as compared to 28% and 50%, respectively, for the Gscore. These results were quite encouraging considering the diversity of the actives used in this study and enhanced our confidence in using Emodel as a scoring function for the SBVS study. The higher enrichment rate associated with Emodel might be because of significant weighting of force field components such as the Coulomb and *van der Waals* energy in Emodel as compared to Gscore²¹³. These interactions may be a major driving factor for the ligand binding for the apolar FP-2 binding site, and, as a result, Emodel performed better with the FP-2 active site than Gscore. A similar observation was made by Li *et al.* in their virtual screening study, suggesting the importance of Dock energy score (GAsDock) over a Chemscore-based (GScore) scoring function more suitable for the apolar FP-2 binding site¹⁸⁶. The success ratio for the present study in finding novel

chemotypes was 42%, which suggest the score-based active/inactive separation ability of current docking protocol against this target. The higher hit ratio also implied the importance of scoring function based enrichment studies prior to SBVS.

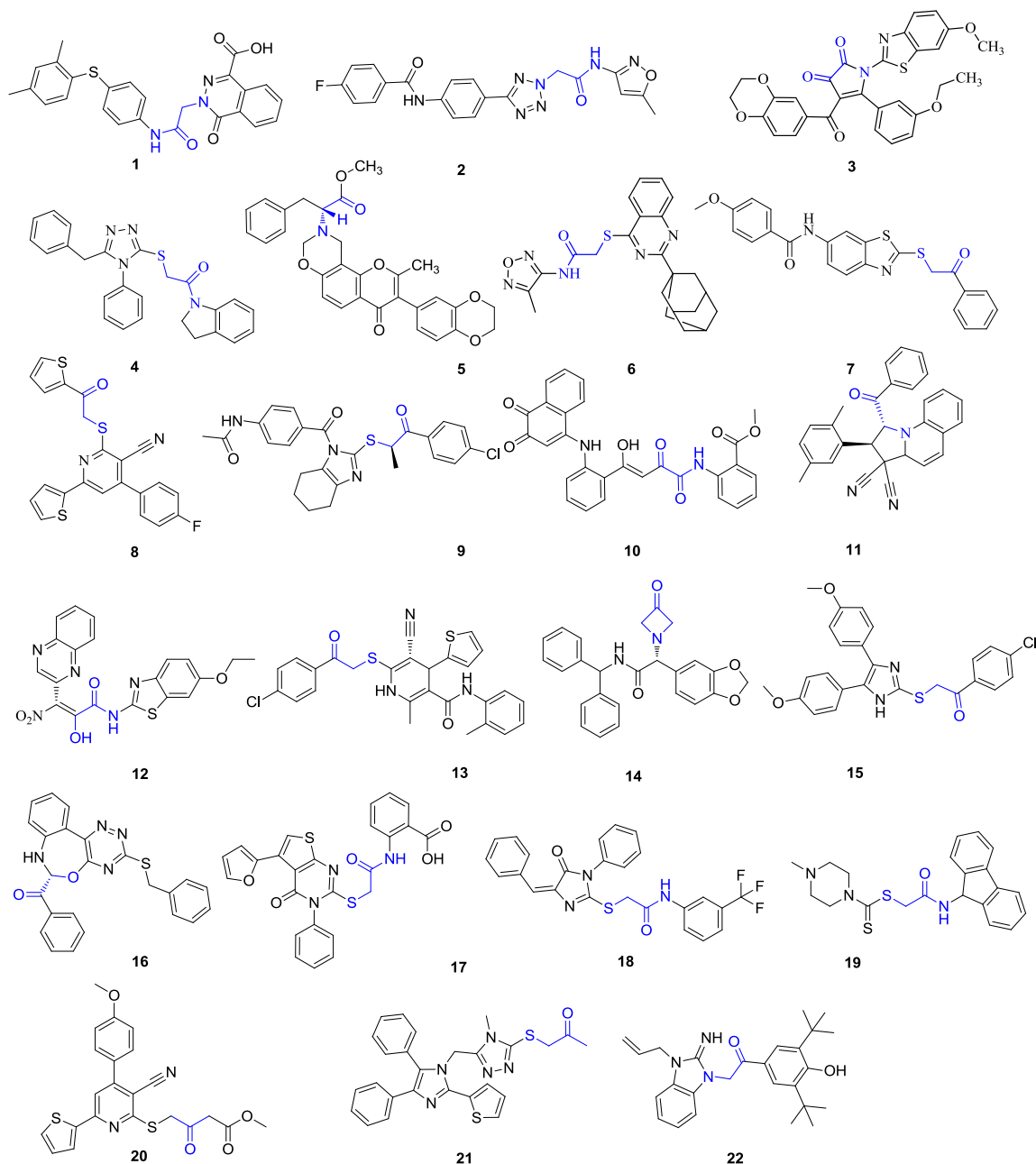


Figure 2.3.1. Structures of the FP-2 inhibitors identified through SBVS of FCPI library. Substructure containing electrophiles of interest are highlighted in blue.

Prior to virtual screening of FCPI library, the database was curtailed with a docking pose based pharmacophore query. Key interactions of vinyl sulfone inhibitor with residues of FP-2 active site such as Gly83, Cys42, Gln36, along with the hydrophobic residues of S2 and S3 pocket (shown in Figure 3d) were used to build the pharmacophore query, consisting of four descriptors: HBD, HBA, HY1, and, HY2. The selection of pharmacophoric descriptors were based on a thoughtful consideration of crucial interactions of cysteine protease inhibitors with the conserved residues of the papain-family cysteine proteases. For example, Gly83 is highly conserved in the S3 subsite of clan CA cysteine proteases^{34, 214} (Figure 2.3.2), and, the lack of potency was reported for the cysteine protease inhibitors particularly binding to the non-prime (S1-S4) site lacking hydrogen bond with the carbonyl oxygen of glycine^{11, 215}. Similarly, inhibitors of the papain-family of enzymes, are commonly shown to form hydrogen bonding interactions with residues forming oxyanion hole such as Gln36 and backbone amine of catalytic cysteine Cys42 in FP-2^{11, 216, 217}. Moreover, as with the most other papain-family of proteases, analysis of the specificity of FP-2 indicated that the amino acid at the P2 position plays a key role in mediating substrate specificity, at least with peptide substrates³⁴. In addition, substrate mapping and inhibitor profiling study against FP-2 revealed strong preference for hydrophobic residue at P2 (Leu>Phe>Val) in the S2 Pocket^{34, 185}. Finally, the recently published X-ray structure of FP-2 bound to a covalent irreversible epoxysuccinate inhibitor (E-64)⁵⁰ gives an additional validity to our selection of docking-pose based pharmacophoric descriptors (Figure 1.2, Chapter 1). The carboxyl group of E-64 occupies the oxyanion hole formed by Gln36 and Cys42, resembling HBA feature of our pharmacophore query. Similarly, peptidyl backbone amine of E-64 forms hydrogen bond with a backbone carbonyl of Gly83, similar to assigned HBD descriptor of

pharmacophore query. The P2-leucine of E-64 exhibits hydrophobic interactions in the S2 pocket with Ile85, L172 and Ala175, whereas P3 side chain of E-64 exhibited hydrophobic interaction with Gly83, Leu84, and, cation-pi interactions with the Tyr78 in S3 pocket, relating to the HY1 and HY2 descriptors of docking pose-based pharmacophore query.

		*
FP-2	-----QMNYEEVIKKYR-GEENFDHAAYDWRLHS----GVTVPKDQK	37
FP-3	TLSPVSYEANYEDVIKKYKPADAKLDRIAYDWRLHG----GVTVPKDQA	46
CAT_K	-----RAPDSVDYRKKG----YVTPVKNQG	21
CAT_L1	-----FSVDWREKG----YVTPVKNQG	18
CAT_B	YLKRLCGTFLGGPKPPQRMFTEDLKLPA SF DAREQWPQCPTIKEIRDQG	50
	*	
FP-2	NCGSCWAFSSIGSVESQYAIRK--NKLITLSEQELVDCSFK--NYGCNGG	83
FP-3	LCGSCWAFSSVGSVESQYAIRK--KALFLFSEQELVDCSVK--NNGCYGG	92
CAT_K	QCGSCWAFSSVGALEGQLKKKT--GKLLNLS PQNLVDCVSE--NDGCGGG	67
CAT_L1	QCGSCWAFSATGALEGQMFRKT--GRLISLSEQNLVDCSGPQGN EGCNGG	66
CAT_B	SCGSCWAFGAVEAISDRICIH TNAHVSVEVSAEDLLTCCGSMCGDGCNGG	100
FP-2	LINNAFEDMIELGGICP---DGDPYVSDAPNLCNIDRCTEKYGIKNYLS	130
FP-3	YITNAFDDMIDLGLCS---QDDYPYVSNLPETCNLKRCNERYTIKSYVS	139
CAT_K	YMTNAFQYVQKNRGIDS---EDAYPYVQGE-ESCMYNPTGKAAKCRGYRE	113
CAT_L1	LMDYAFQYVQDNGGLDS---EESYPYEATE-ESCKYNPKYSVANDTGFVD	112
CAT_B	YPAEAWNFWTRKGLVSGGLYESHVGC RPYSIPPCEHHVNGSRPPCTGEGD	150
FP-2	VPD-----NKLKEALRFLGPISISV	150
FP-3	IPD-----DKFKEALRYLGPISISI	159
CAT_K	IPEG-----NEKALKRAVARVGPVSVAI	136
CAT_L1	IPK-----QEKALMKAVATVGPISVAI	134
CAT_B	TPKCSKICEPGYSPTYKQDKHYGNSYSVSNSEKDIMAEIYKNGPVEGAF	200
FP-2	AVS-DDFAFYKEG-IFDGECG-DELNHAVMLVGFGMKEIVNPLTKKGEKH	197
FP-3	AAS-DDFAFYRGG-FYDGECG-AAPNHAVILVGYGMKDIYNEDTGRMEKF	206
CAT_K	DASLTSFQFYSKGVYDESCNSDNLNHAVLAVGYGIQ-----KGN	176
CAT_L1	DAGHESFLFYKEGIYFEPDCSSDMDHGVLVVG YGFESTE-----SDNN	178
CAT_B	SVY-SDFLLYKSG-VYQHVTGEMMGGHAIRILGWGVEN-----GT	238
	*	
FP-2	YYYYIKNSWGQQWGERGFINIETDESGLMRKCGLGTDAFIPLIE-----	241
FP-3	YYYYIKNSWGSDWEGGYINLETDENG YKKTCSIGTEAYVPLLE-----	250
CAT_K	KHWI IKNSWGENWGNKGYILMARNKN---NACGIANLASFPM-----	216
CAT_L1	KYWLKNSWGEEWGMGGYVKMAKDRR---NHCGIASAASYPTV-----	218
CAT_B	PYWLVANSWNTDWGDN GFFKILRGQD---HCGIES E VVAGIPRTDQYWE	284

Figure 2.3.2. Multiple sequence alignment of FP-2 with homologous cysteine proteases performed using Clustal 2.0.12¹. Sequence information is from SWISS-PROT (accession numbers: Q9NAW2, Q9NB39, P43235, P07711 and P07858). The conserved residues among all papain-family cysteine proteases are shown in green. The amino acid residues of the S2 pocket differing between FPs and mammalian cysteine proteases are highlighted in red. Catalytic residues are highlighted in *.

As mentioned before, the main drawback of covalent modifiers (for example, compounds bearing electrophiles such as α , β -unsaturated ketones, α -halo ketones, cyanamide) is lack of target specificity. In addition, non-specific binding of covalent modifiers to protein, DNA or glutathione (GSH) may lead to unfavorable toxicological events such as immunogenic response or life-threatening idiosyncratic adverse drug reactions (ADRs)^{218, 219}. Soft-electrophiles attenuate the reactive nature of hard electrophiles and might reduce the risk of toxicity associated with covalent modifiers. Thus, use of the soft-electrophiles based FCPI library for SBVS provides a strategic approach to discover novel, selective inhibitors, preferably modulating the enzyme through non-covalent interactions. The soft-electrophiles selected here had previously shown to form tetrahedral intermediates with cysteine proteases, for example, α -heteroatom substituted ketones^{217, 220-222}, α -keto amide^{56, 223}, α -keto acid⁵⁶, α -keto ester⁵⁶ and azetidinones²¹⁵. To assess the soft-nature of electrophiles present in target compounds, we performed calculations of the Atomic Fukui indices on a few selected compounds (**1**, **3-5**, **7**, **11**, **12**, **14**, and **15**) containing diverse soft-electrophiles. Fukui indices are partial second derivatives of the electron or spin density with respect to a change in the electron count or the unpaired spin count. They are regional descriptor of site reactivity predicting which atoms in a molecule are most reactive towards electrophilic or nucleophilic attack based on the analysis of Frontier Molecular Orbitals (FMOs) in a molecule²¹⁰. The greater the magnitude of an index, the greater change in electron or spin density near the atoms of interest, and thus, higher reactivity of a molecule at that atomic site for an electrophilic or a nucleophilic reaction²²⁴. Generally, a nucleophilic reaction takes place preferably in the substrate sites with the largest values for the LUMO density. Likewise, an

electrophilic reaction is most likely to occur at the largest HOMO electron density in the molecular site^{211, 225}. Since our goal was to probe the most electrophilic center in identified FP-2 hits and predict its electrophilicity, we considered calculation of Fukui indices, F_{NN_LUMO} at a constant spin density in selected compounds. The electrophilicity of electrophilic centers present in identified FP-2 hits were compared with the known cysteine protease inhibitors with an irreversible covalent (α -halo ketone, **R1**; vinyl sulfone, **R2**) and a reversible covalent (α -ketoamide, **R3**) warhead groups. The values of Fukui index F_{NN_LUMO} , of the most electrophilic atoms of the FP-2 inhibitors, and, reference compounds (**R1-R3**) are shown in Figure 2.3.3.

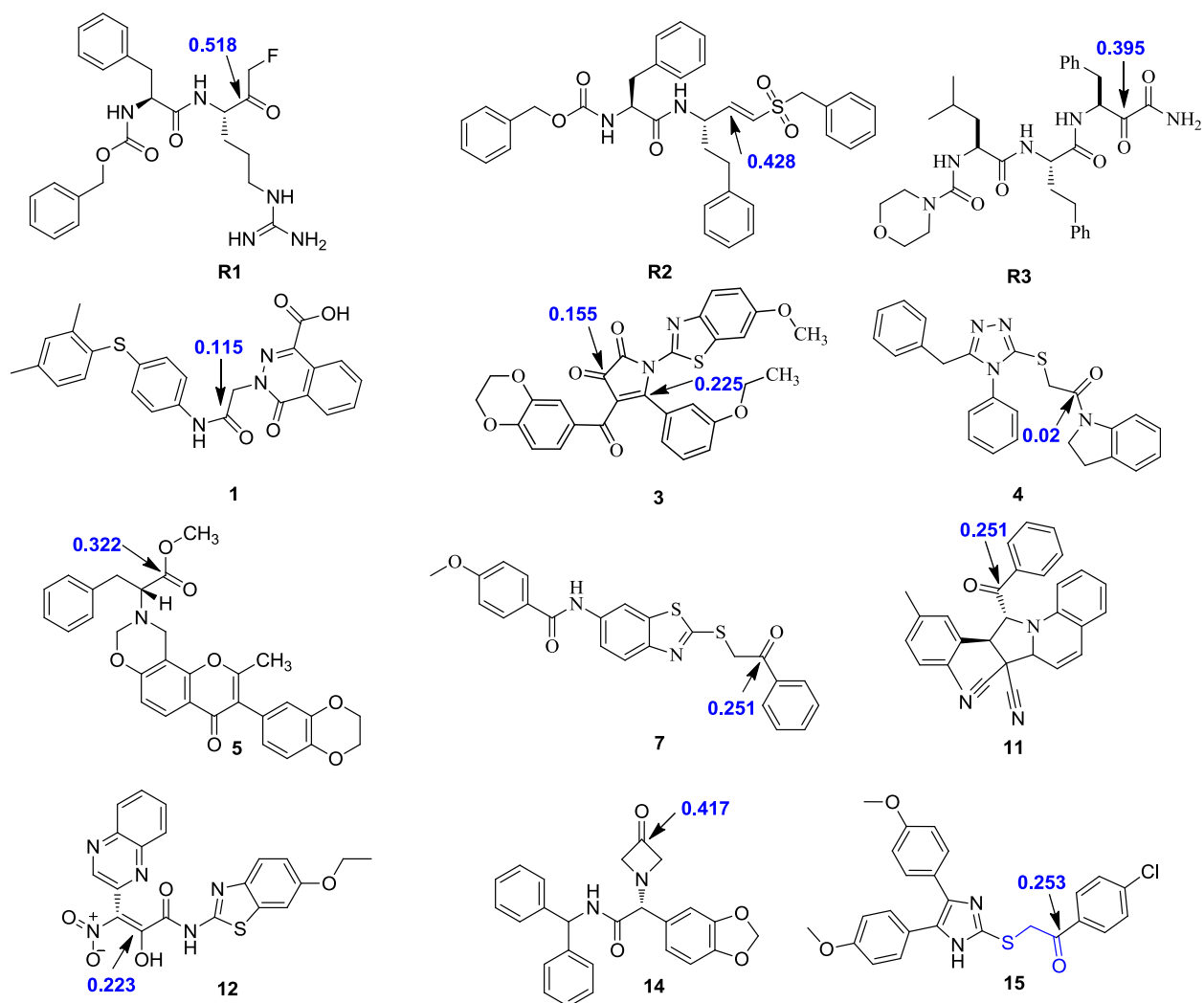


Figure 2.3.3. An evaluation of regional Fukui functions to predict the electrophilicity of atomic site of reference compounds (**R1-R3**) and selected FP-2 inhibitors is shown here. The most electrophilic center in the compounds is shown by the arrows. The corresponding value of the Fukui index, F_{NN_LUMO} , for the reactive atoms (scaled from zero to one) is shown in blue.

The result from the calculation (Figure 2.3.3) demonstrates that the site specific reactivity of the identified FP-2 inhibitors is less than the irreversible covalent inhibitors: **R1** with an α -halo ketone (LUMO = 0.518) and **R2** with a vinyl sulfone warhead group (LUMO = 0.428) and, are comparable to the known reversible covalent inhibitor, **R3** with an α -keto amide (LUMO = 0.395) moiety. These calculations suggest the soft-nature of the electrophilic atoms in the

identified FP-2 hits. However, one exception here is compound **14** with a higher than expected electrophilicity value, presumably because of the ring strain of an azetidin-3-ones moiety, which is reflected in the calculated LUMO value (0.417). On the basis of the predicted site reactivity of soft-electrophiles, compounds with an α -hetero ester (**5**, LUMO = 0.322) were found to be more electrophilic than those with an α -hetero ketone (**7**, **11**, LUMO = 0.251; **15**, LUMO = 0.253), which in turn had a higher electrophilicity than hits with an α -hetero amide functionality (**1**, LUMO = 0.115; **4**, LUMO = 0.020). One point worth mentioning here is that the calculated electrophilicity is influenced by the complex interplay of factors such as a neighboring group effects, the internal geometry of the compound, intermolecular hydrogen-bonding and the solvation effects. For example, the predicted electrophilicity of the α -keto amide **12** (LUMO = 0.223) was found to be lower than the **R3** LUMO (0.395). This difference may be due to the presence of an adjacent heteroaromatic system leading to extensive conjugation as well as intramolecular hydrogen bonding between the nitro group and the enolic OH forming a six membered chelate. Similarly, in case of hits with α -heteroatom substituted ketones, a slightly higher predicted electrophilicity of **15** than **7** or **11** might be due to the presence of the electron withdrawing *p*-chloro group, increasing the electrophilicity of the carbonyl carbon. However, a comparison of predicted site-specific reactivity of identified hits with FP-2 inhibition data suggests a lack of correlation of electrophilicity with the enzymatic inhibition. For instance, compound **14** with an azetidine-3-one, had the highest predicted LUMO based electrophilicity among the identified hits, however, it was only moderately active in the FP-2 inhibition assay (IC_{50} =25.9 μ M). By comparison, compounds with an α -hetero amide (compounds **1**, **4**) were predicted to have a negligible electrophilicity but showed relatively higher experimental

inhibition of FP-2 ($IC_{50} = 1-5 \mu M$). This observation emphasizes the obvious involvement of other factors in enzyme inhibition, such as the overall geometry and interaction patterns of the compounds in the FP-2 binding site.

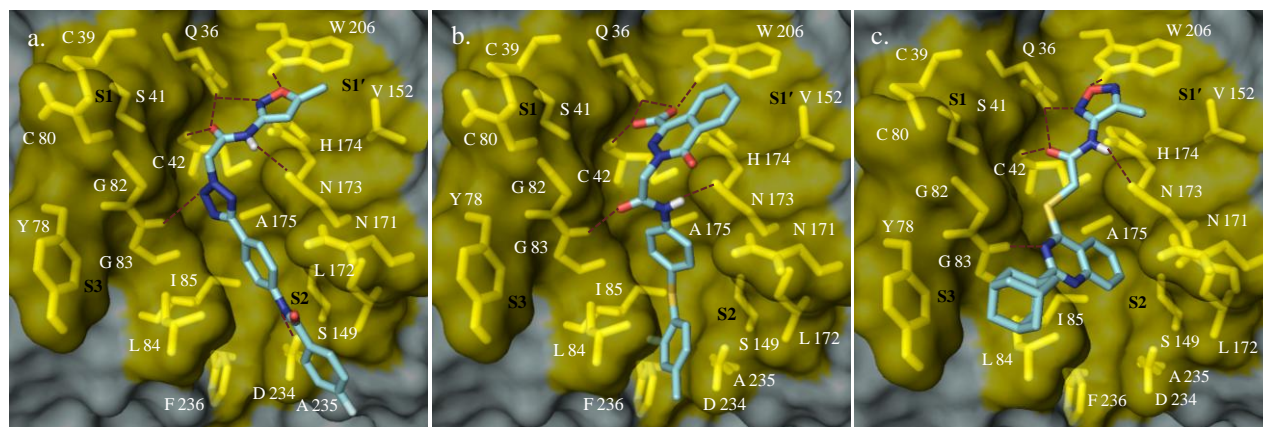


Figure 2.3.4. Predicted binding mode of virtual screening hits (a) **1** (b) **2** (c) **6** in the FP-2 binding site.

Figure 2.3.4 shows predicted interaction profiles of representative compounds **1**, **2** and **6** in FP-2 active site. The 4-oxo-phthalazine core in molecule **1** appears to protrude into the S1' pocket with the carboxylic acid group interacting with the residues of an active site-forming oxyanion hole such as the backbone $-NH$ of Cys 42 and the amide side chain of Gln 36 (see Figure 2.3.4a). Also, Trp206 and Val152 of the S1' pocket are involved in the van der Waals interactions with the phthalazine moiety wherein the indole $-NH$ of Trp206 participates in hydrogen bonding with the carboxylate moiety of **1**. The α -ketoamide soft-electrophile exhibits hydrogen bonding interaction with the backbone $-NH$ of Gly82 and the backbone carbonyl of Asn173. The phenyl ring at the terminus of the amide linker can undergo hydrophobic interaction with the side chains of Ala175. The terminal 2, 4 dimethyl phenyl moiety is involved in hydrophobic interactions with the side chains of Leu84, Ile85 at the entrance of the S3 pocket as well as with the Phe236

buried deeper in the S2 pocket.

In case of compound **2**, the 5-methyl-isooxazole moiety was well placed in the S1' pocket of FP-2, exhibiting hydrophobic interactions with the side chains of Tyr206 and Val152 (Figure 2.3.4b). Also, the isoxazolyl moiety of **2** was involved in multiple hydrogen bonding interactions with the side chain amide proton of Gln36 as well as the indole -NH of Trp206. The carbonyl group of the α -keto amide electrophile located over the catalytic Cys42 formed a critical hydrogen bonding interaction with the backbone amine of Cys42, whereas the -NH of the α -keto amide electrophile acts as a hydrogen bond donor for the backbone carbonyl of Asn 173. In addition, the α -keto amide linker was involved in hydrophobic interactions with Gly82. The nitrogen of the tetrazole moiety acted as a hydrogen bond acceptor with the backbone amine of Gly83. The phenyl ring attached to the tetrazole moiety of **2** vectors into the critical S2 pocket, forming hydrophobic interactions with Leu172 at the entrance of the S2 pocket as well as with the Leu84 side chain lining the S3 pocket. The anilinic -NH of **2** donated a hydrogen bond to the side chain carboxylic group of Asp234 buried in the hydrophobic environment of the S2 pocket. The terminal phenyl ring was found to exhibit van der Waals interactions with the side chain of Leu172, exposing the para fluoro group to the solvent.

A similar observation was made for compound **6** near the S1' pocket, especially for the 5-methyl-isooxazole moiety (Figure 2.3.4c). The thioacetamide electrophile is placed over the catalytic cysteine sulfur, and its carboxamide group participates in the hydrogen bonding interactions with Cys42 and Asn173 in a similar manner as observed for **2**. The quinazoline core of **6** is placed aptly in the S2 pocket, forming van der Waals interactions with Ala175 and Leu172. In addition, nitrogen on the quinazoline ring of **6** accepts a hydrogen bond from the

backbone of Gly83. Moreover, the adamantyl moiety of **6** protrudes into the S3 pocket and participates in the van der Waals interactions with the Tyr78 of the S3 pocket.

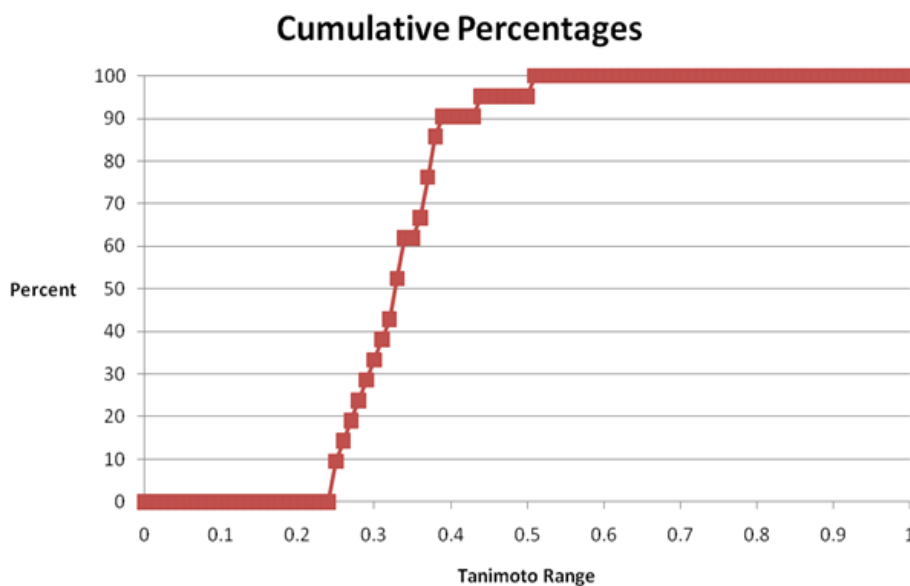


Figure 2.3.5. Comparison of similarity of hits identified in the current study with previously identified hits (structure shown in Figure 2.2.4.3) using dbcmpr utility of Sybyl 8.1. The median Tanimoto coefficient is about 0.33 (where the graph crosses the 50% threshold), and all of compounds in the current dataset have a Tanimoto of 0.5 or less with their nearest neighbour in the reference database suggesting identified hits are structurally novel.

The average pair-wise Tanimoto similarity index of the 21 FP-2 inhibitors was 0.24, which point towards the diverse nature of the identified hits (see similarity matrix, Table 2.3.2). The diversity of hits obtained in the present study with previously identified FP-2 inhibitors was analyzed using the dbcmpr module of Sybyl. dbcmpr utility calculates the Tanimoto similarities using UNITY 2D fingerprints between all compounds identified in current study (test set) with their nearest neighbors in the previous study (reference set). The result of this comparison showed that 95% of compounds from the test set has the Tanimoto similarity of less than or equal to 0.5 to its nearest neighbor in the reference set (see database comparison, Figure S3 in the supporting information). The results clearly suggest novelty of the hits identified in the

present study.

Table 2.3.2. The pair-wise Tanimoto similarity matrix table generated using Canvas, version 1.2 (Schrödinger, LLC, New York, NY, 2009) for 21 identified hits. The maximum similarity was 0.5 between compound pairs **4** and **21** which share same core.

Comp	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	20	21	19
1	1.00	0.32	0.25	0.29	0.22	0.19	0.33	0.20	0.33	0.36	0.17	0.26	0.27	0.19	0.22	0.23	0.38	0.29	0.20	0.28	0.17
2	0.32	1.00	0.24	0.27	0.21	0.26	0.32	0.20	0.25	0.31	0.13	0.24	0.20	0.19	0.17	0.23	0.30	0.24	0.20	0.26	0.15
3	0.25	0.24	1.00	0.30	0.36	0.24	0.31	0.20	0.29	0.26	0.19	0.32	0.25	0.29	0.23	0.27	0.31	0.23	0.29	0.28	0.19
4	0.29	0.27	0.30	1.00	0.21	0.26	0.32	0.31	0.33	0.23	0.16	0.23	0.25	0.26	0.21	0.41	0.42	0.46	0.26	0.50	0.29
5	0.22	0.21	0.36	0.21	1.00	0.21	0.26	0.14	0.24	0.37	0.17	0.24	0.23	0.32	0.21	0.25	0.23	0.17	0.26	0.19	0.18
6	0.19	0.26	0.24	0.26	0.21	1.00	0.25	0.20	0.24	0.19	0.15	0.24	0.21	0.23	0.18	0.27	0.25	0.26	0.28	0.24	0.19
7	0.33	0.32	0.31	0.32	0.26	0.25	1.00	0.32	0.27	0.38	0.15	0.32	0.36	0.27	0.36	0.40	0.38	0.33	0.32	0.32	0.21
8	0.20	0.20	0.20	0.31	0.14	0.20	0.32	1.00	0.21	0.18	0.21	0.19	0.33	0.18	0.24	0.31	0.27	0.28	0.42	0.28	0.17
9	0.33	0.25	0.29	0.33	0.24	0.24	0.27	0.21	1.00	0.27	0.17	0.24	0.33	0.20	0.27	0.26	0.31	0.28	0.23	0.33	0.22
10	0.36	0.31	0.26	0.23	0.37	0.19	0.38	0.18	0.27	1.00	0.14	0.28	0.27	0.34	0.30	0.29	0.35	0.28	0.25	0.21	0.18
11	0.17	0.13	0.19	0.16	0.17	0.15	0.15	0.21	0.17	0.14	1.00	0.14	0.24	0.21	0.15	0.16	0.13	0.17	0.18	0.18	0.15
12	0.26	0.24	0.32	0.23	0.24	0.24	0.32	0.19	0.24	0.28	0.14	1.00	0.26	0.27	0.22	0.30	0.28	0.23	0.20	0.20	0.16
13	0.27	0.20	0.25	0.25	0.23	0.21	0.36	0.33	0.33	0.27	0.24	0.26	1.00	0.25	0.29	0.34	0.29	0.27	0.31	0.24	0.19
14	0.19	0.19	0.29	0.26	0.32	0.23	0.27	0.18	0.20	0.34	0.21	0.27	0.25	1.00	0.25	0.34	0.24	0.22	0.22	0.17	0.27
15	0.22	0.17	0.23	0.21	0.21	0.18	0.36	0.24	0.27	0.30	0.15	0.22	0.29	0.25	1.00	0.26	0.25	0.23	0.25	0.21	0.20
16	0.23	0.23	0.27	0.41	0.25	0.27	0.40	0.31	0.26	0.29	0.16	0.30	0.34	0.34	0.26	1.00	0.38	0.35	0.25	0.31	0.22
17	0.38	0.30	0.31	0.42	0.23	0.25	0.38	0.27	0.31	0.35	0.13	0.28	0.29	0.24	0.25	0.38	1.00	0.49	0.26	0.34	0.21
18	0.29	0.24	0.23	0.46	0.17	0.26	0.33	0.28	0.28	0.28	0.17	0.23	0.27	0.22	0.23	0.35	0.49	1.00	0.22	0.31	0.24
20	0.20	0.20	0.29	0.26	0.26	0.28	0.32	0.42	0.23	0.25	0.18	0.20	0.31	0.22	0.25	0.25	0.26	0.22	1.00	0.28	0.17
21	0.28	0.26	0.28	0.50	0.19	0.24	0.32	0.28	0.33	0.21	0.18	0.20	0.24	0.17	0.21	0.31	0.34	0.31	0.28	1.00	0.27
19	0.17	0.15	0.19	0.29	0.18	0.19	0.21	0.17	0.22	0.18	0.15	0.16	0.19	0.27	0.20	0.22	0.21	0.24	0.17	0.27	1.00

In addition to inhibition of FP-2, compounds **5**, **6**, **10**, **18**, **21** and **22** exhibited in vitro anti-malarial activity. Compounds **5** and **6** showed the same log order IC_{50} values in whole-cell assays as in the enzyme assay. Other compounds (**10**, **18**, **21** and **22**) displayed greater activity against cultured parasites than FP-2. These observations might be explained in a number of ways: First, the compounds might be acting both against FP-2 and non-protease targets, as suggested by results with phenothiazines²²⁶. For example, compound **10** contains naphthoquinone scaffold conferring its redox potential²²⁷ and may exert its activity via protein crosslink in the

cultured parasites in addition to inhibition of FP-2. Second, the compounds may be acting against targets other than FP-2. This possibility is suggested by the fact that compounds **18**, **21** and **22** did not form a swollen, hemoglobin-filled food vacuole, as seen with highly potent covalent falcipain inhibitors^{64, 183}. Third, the compounds might concentrate in the food vacuole, allowing activity at culture concentrations below those that inhibit the protease. In any event, our screens identified a number of compounds with low micromolar antimalarial activity. Of note, all the compounds shown in Figure 2.3.1 were non-toxic in cytotoxicity assay, suggesting potential for development of related compounds as drugs.

Next, we screened identified FP-2 hits (compound **1-7**, FP-2 IC₅₀ ≤ 10 μM) against selected human papain-family cysteine proteases (cathepsin B, cathepsin K and cathepsin L), for possible off-target activity (Table 2.3.3). As expected, the evaluated hits (**1-7**) were found to inhibit selected cathepsins that are closely related to the parasitic cysteine proteases. Although, homology exists between the parasitic cysteine protease and its human orthologues, there are significant differences in the binding site residues, especially near the S2 pocket of these enzymes. For instance, sequence comparison (Figure 2.3.2) suggested Ser149 in the S2 pocket of FPs is replaced by Ala in human orthologues. Ser149A mutation in FP-2 was reported to decrease the proteolytic activity by 91%⁴⁸ and, can be targeted for design of selective protozoal cysteine protease inhibitors. In addition, Ile85 of the S2 pocket of FP-2 and FP-3 is replaced by methionine in cathepsin K and L and, proline in cathepsin B. Moreover, the polar residues such as Asp234 and Glu243 in the S2 pocket of the FP-2 and FP-3 binding sites, respectively are replaced by hydrophobic residues such as Leu and Ala in cathepsin K and L²²⁸. These differences between the proteases can be utilized for the design of selective inhibitors of FPs. To our

surprise, compound **4** did not show any inhibition up to 50 μM against the evaluated mammalian cysteine proteases. Further docking studies of **4** in the selected mammalian cysteine protease panel might provide valuable insights for the binding mode and selectivity of this compound in FP-2.

Table 2.3.3. The IC_{50} values of FP-2 hits against selected human cathepsin peptidases

Comp No.	Cathepsin K IC_{50} (μM)	Cathepsin L IC_{50} (μM)	Cathepsin B IC_{50} (μM)
1	20.00	9.30	10.72
2	3.83	24.39	37.14
3	25.60	41.57	14.74
4	>50	>50	>50
5	26.70	5.86	49.08
6	29.66	16.52	21.44
7	6.12	31.63	17.66
E-64	0.004	0.016	0.007

2.4. Conclusions

The focused cysteine protease inhibitor library consisting of ~65,000 compounds was screened against the X-ray crystal structure of FP-2. The attractive components of the virtual screening approach were the FCPI library based on soft-electrophiles of interest, the enrichment study to select the best scoring function, the docking pose-based pharmacophore query as a filter in the SBVS protocol, and, consideration of binding energy of the receptor-ligand complex for the selection of putative hits. The study reported a high success ratio with novel antimalarial

scaffolds. Twenty one compounds were active against FP-2 and four were also active against FP-3. Two FP-2 hits, compound **5** and **6**, showed corresponding inhibition of the cultured parasites, showing promise for further structural optimization. Moreover, the predicted site-specific electrophilicity, as measured by calculating the LUMO electron density of each atom in the form of the atomic Fukui Indices, suggested a soft-nature of the electrophiles present in identified hits. Thus, any potential risk of toxicity associated with the irreversible inhibition of the cysteine protease is attenuated. Assessment of the selectivity of identified hits against related mammalian cysteine proteases (cathepsin B, K and L) revealed compound **4** as a selective inhibitor of FP-2 ($IC_{50} > 50 \mu\text{M}$ against tested mammalian proteases). The hits obtained through the present screening efforts can be optimized via substructure/similarity search approaches to identify structural analogs of the active hits taking into consideration the key residues of the S2 pocket for the biochemical selectivity as discussed above. In addition, combinatorial library synthesis for selected hits (for example, compounds **2**, **4**, **6**, **7**) using commercially available building blocks can be pursued taking insights from the recently published co-crystallized structure of FP-2⁵⁰, for enhanced binding site occupancy and interaction profiles, to establish an SAR, and, will be addressed in follow up work.

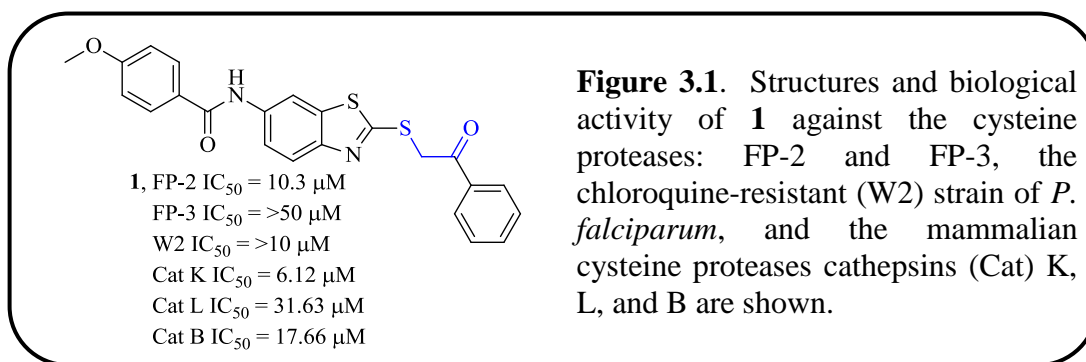
Chapter 3

Design, Synthesis and Biological Evaluation of Benzothiazole Analogs as Falcipain-2 Inhibitors

The content of this chapter is published in: Shah, F., Wu, Y., Jiri G., Pedurri, Y., Rosenthal, P., Avery M. A. Design, synthesis and biological evaluation of novel benzthiazole and triazole analogs as inhibitors of falcipain, cysteine protease of malaria parasite Plasmodium falciparum, *MedChemComm*, 2011_DOI: 10.1039/C1MD00129A

3.1. Introduction

As discussed in previous chapters, most of the potent FP-2 inhibitors described in the literature were peptide analogs, displaying a covalent irreversible binding mode with the active site cysteine, and were relatively non-selective. It is highly desirable to develop non-peptidic, non-covalent inhibitors of FPs to minimize toxicity risk while retaining *in vivo* activity and selectivity. In our previous study, we carried out the structure-based virtual screening (SBVS) of the FCPI library built based on soft-electrophiles against the X-ray structure of FP-2²²⁹. The study identified 21 diverse micromolar inhibitors of FP-2. In the present work, we initiated efforts to optimize the potency of lower micromolar FP-2 hit **1** with a benzothiazole core (Figure 1). Compound **1** was selected for further optimization based on a) ease to synthesize combinatorial library around commercially available benzthiazole core and to establish an SAR and b) to test selectivity hypothesis of these series of compounds to FPs over other homologous mammalian cysteine proteases of the papain-family. The optimization of benzothiazole series was carried out with mutation of the benzothiazole core with other complementary groups followed by combinatorial library design and synthesis around the benzothiazole core. A total of 23 analogs were made in this series and evaluated against the enzyme FP-2, FP-3 and for the activity against cultured *P. falciparum* parasites. Twelve compounds from the benzothiazole series showed moderate activity against FP-2. Two compounds from the benzothiazole series showed dual inhibition of FP-2 and FP-3. One particular compound from benzothiazole series showed corresponding inhibition of the cultured parasites. Docking of these inhibitors in the FP-2 active site provided insights about SAR and interactions of the inhibitors with the active site of FP-2.



3.2. Optimization strategies: Core hopping, Structure-based design

Core or scaffold hopping is a technique frequently used by medicinal chemists to discover novel compounds by mutating the molecule core or scaffold starting from known active molecules²³⁰. Core hopping gives medicinal chemists an opportunity to pursue drug design in a different chemical space to improve potency, to improve ADMET liabilities of a given scaffold, to circumvent intellectual property rights associated with a parent scaffold or to speed up the lead optimization process²³⁰.

Our primary goal was to improve the potency of virtual screening hits **1**. For the initial optimization of **1** with core-hopping strategy, we selected analogs which were commercially available, affordable and had an α -thio ketone to act as a soft-electrophile prior to initiating extensive synthetic efforts. From the docking studies of **1**, we envisioned that the replacement of the benzothiazole core with benzimidazole and thioacetamide moieties may improve the overall affinity of **1** by forming strong H-bond networks with the backbone Gly83 and Asn173. Therefore, our selection was limited to commercially available compounds (**3-10**) containing these cores with the goal of improving potency. While selecting compounds in these series, compounds with electron withdrawing groups *para* to the phenyl ring (for example, **3**, **4**, **7**) were preferred to enhance the reactivity of the carbonyl carbon with the catalytic Cys42 of the FP-2 active site. It is worthwhile to mention that compounds **9** and **10** can be reasonably enolic and

may predominantly behave as α , β -unsaturated ketones in the FP-2 active site. This enolic behavior is in equilibrium, thus availability of the carbonyl tautomer should not be problematical. Selected compounds were docked in the crystal structure of FP-2 using Glide extra precision (XP) mode and a previously validated docking protocol²²⁹. Although thioacetamide and benzimidazole compounds showed improved interaction profiles in the binding site (Figure 3.2.1), they were inactive against FP-2 at concentration up to 50 μ M. These findings, although surprising, suggested the importance of the benzothiazole core for FP-2 inhibition and prompted us to design a library around this core.

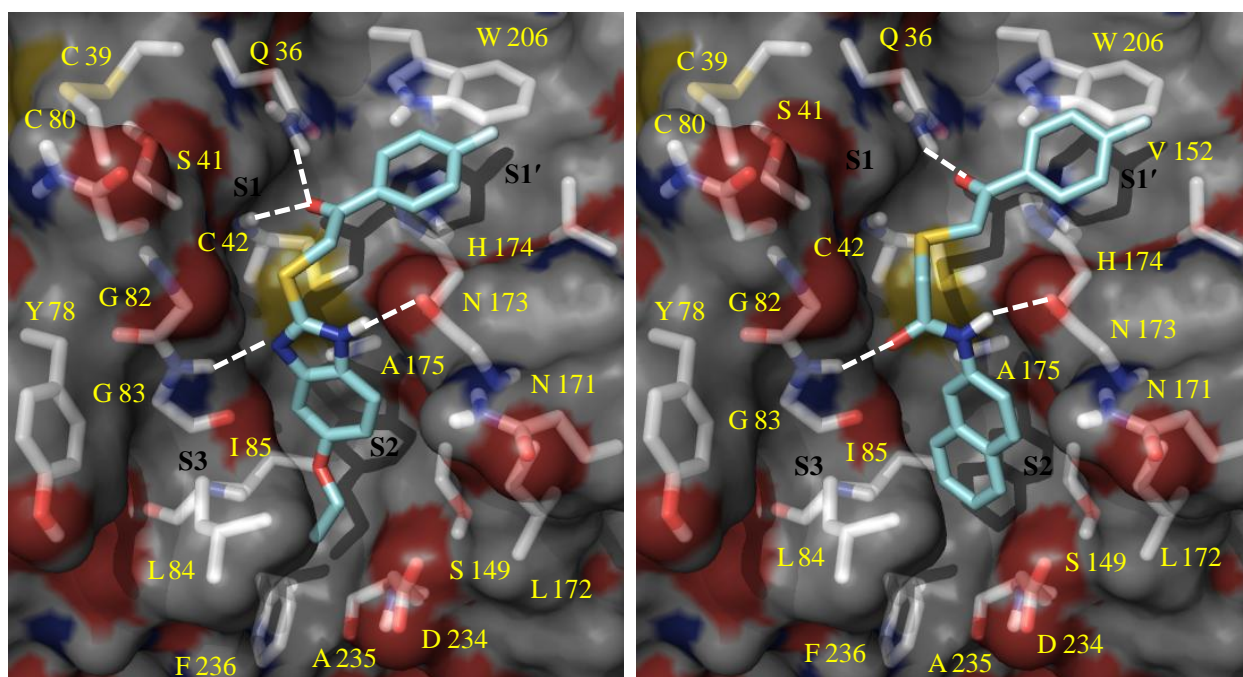


Figure 3.2.1. Docking pose of representative compounds from benzimidazole (**3**, Left) and thioacetamide (**7**, Right) series.

From the docking pose of compound **1** shown in Figure 3.2.2, it was evident that **1** could be modified to better occupy the S1' and, S2 pockets of FP-2. For example, hydrophobic substituents can be placed on the 6-aminobenzothiazole scaffold to interact with hydrophobic

residues such as Ala235, Ile85 and Leu172 lining the border of the S2 pocket. Also, the flat S1' pocket of FP-2 is composed of hydrophobic residues such as Trp206 and Val152. Therefore, two regions of the benzothiazole core spanning the S1' and S2 sites were selected to perform chemical transformations and, thereby to develop SAR information. In addition, α -thioketones of **1** were modified to α -thioesters and α -thioacetamide electrophiles to investigate the chemical reactivity of these electrophiles with the active site cysteine sulfur and determine its effects on FP-2 inhibition. Thus, two series were designed (Table 3.5.1).

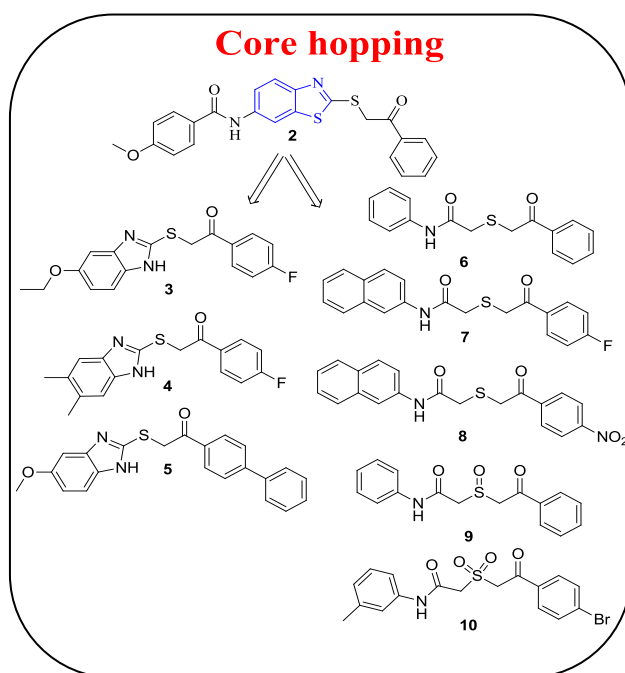
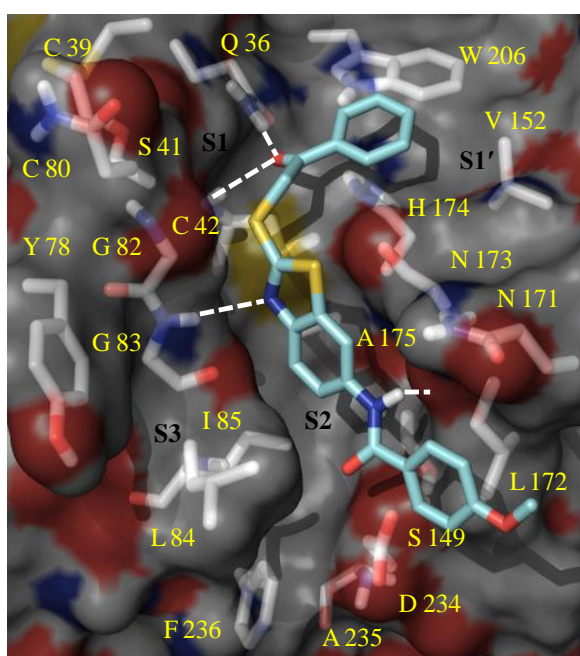
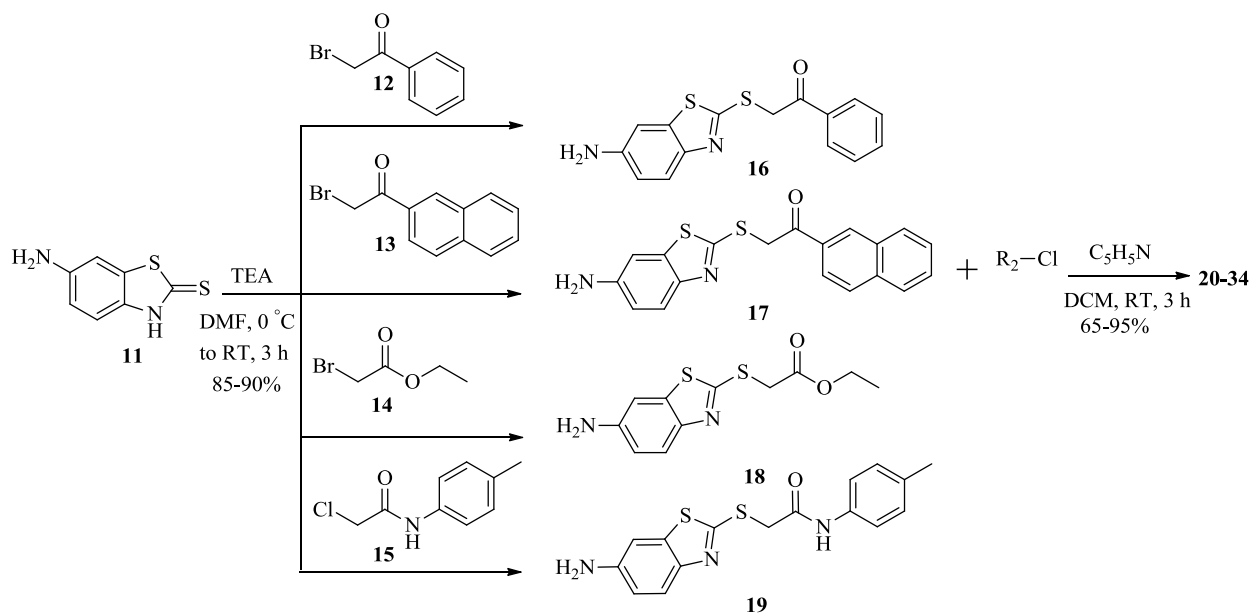


Figure 3.2.2. (top) docking pose of compounds **1** in the FP-2 binding site; (bottom) core hopping of compound **1** and **2** with tetrazole and benzothiazole cores (highlighted in blue), respectively are shown. The tetrazole core of **1** was modified to triazole core to take an advantage of click chemistry. The benzothiazole core of **2** was modified to commercially available benzimidazole and thioacetamide core containing compounds (**3-10**) with anticipated gain in potency.

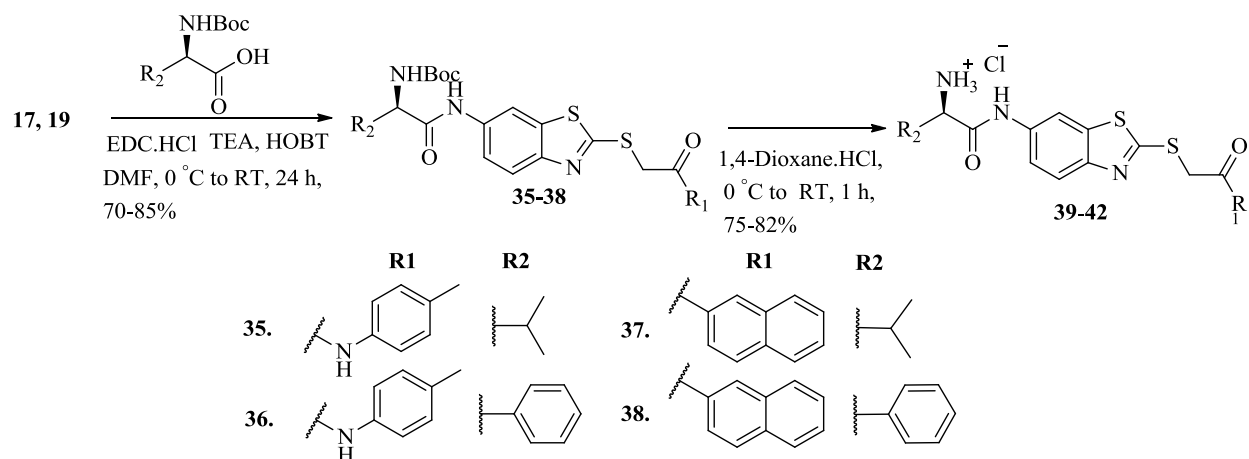
The deep hydrophobic S2 pocket in FPs harbors polar residues such as Asp234 and, Ser149 in FP-2 and Glu243 and, Ser158 in FP-3. Electrostatic interactions of compounds with the polar residues of the S2 pocket might lead to an overall boost in potency. In addition, sequence

alignments of FPs with homologous mammalian protease of papain family (Figure 2.3.2) suggested that polar residues of the S2 pocket are replaced by hydrophobic residues in cathepsin K, and L and can be targeted to design selective inhibitors of FPs. Therefore, in the α -thioketones and α -thioacetamide series, compounds with protonated amino groups were designed (compounds **39-42**) to exhibit ionic interactions with Asp234 of the S2 subsite. The preliminary docking studies suggested a preference for R over S isomers of designed compounds, and, hence, Boc-(D) phenyl glycine or Boc-(D) valine were used in the synthesis of **39-42**.

3.3. Synthetic Schemes. The synthetic strategies adopted to obtain various benzothiazole derivatives are highlighted in Schemes 1 and 2. Scheme 1 depicts condensation of appropriate α -haloketones (**12, 13**), α -haloacetate (**14**), and α -haloamides (**15**) with the commercially available 6-aminobenzothiazole-2-thione (**11**) in the presence of triethylamine to give respective *S*-alkylated α -thioketones (**16, 17**), α -thioesters (**18**) or α -thioacetamide (**19**). In the next step of *N*-acylation, intermediate **16-19** provided corresponding *N*-acylated derivatives **20-34** by treatment with various acid chlorides. Synthesis of designed analogs targeting polar residues of the S2 pocket was accomplished as depicted in Scheme 2. Accordingly, compounds **17** and **19** were coupled with (D)-*N*-Boc phenyl glycine and (D)-*N*-Boc valine using standard 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide/ hydroxybenzotriazole (EDCI/HOBT) peptide coupling conditions and triethylamine as a base to yield Boc-protected compounds **35-38**. Deprotection of Boc with 4 M HCl in 1, 4-dioxane provided the target compounds **39-42** as HCl salts.



Scheme 3.1



Scheme 3.2

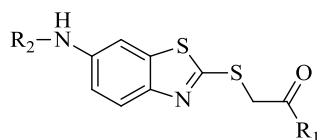
3.4. Biological Evaluation. The synthesized compounds were evaluated for inhibition of recombinant FP-2 and, FP-3 and, the *in vitro* growth of the chloroquine resistant W2 strain of *P. falciparum* according to previously described procedures²²⁹. The epoxysuccinate E-64 was used as a positive control for the enzyme assays whereas artemisinin and chloroquine were used as positive controls for testing against W2 strain of *P. falciparum*.

3.5. Results and Discussion

The structure and biological evaluations of benzthiazole analogs are shown in Table 3.5.1. The SAR of compounds in this series is discussed below.

In the benzothiazole series, compounds lacking R₂ substituents (for example, compounds **16-19**) were devoid of activity against FP-2 (Table 3.5.1). A few analogs showed similar log order activity as that of parent compound **2**, whereas others were inactive. Interestingly, no direct correlation was observed between the chemical reactivity of electrophiles such as α -thioketones (**20-25**, and **41, 42**), α -thioesters (**26-29**) and α -thioacetamide (**30-34**, and **39, 40**) towards Cys-42 and inhibition of FP-2. Overall, the compounds with α -thioacetamide and α -thioketones electrophiles showed moderate activity against FP-2 compared to those possessing α -thioesters electrophiles. Moreover, compounds **41** and **42** with protonated amino groups were the only compounds from the benzothiazole series displaying dual inhibition of FP-2 and FP-3.

Table 3.5. 1. Biological activity of benzothiazole analogs.



Sr. No.	R ₁	R ₂	FP-2 IC ₅₀ (μM)	Sr. No.	R ₁	R ₂	FP-2 IC ₅₀ (μM)
16		H	>50	27			>50
17		H	>50	28			>50

18		H	>50	29			31.34
19		H	>50	30			24.33
20			11.14	31			13.89
21			35.70	32			13.96
22			>50	33			16.48
23			13.18	34			16.87
24			>50	39			>50
25			43.25	40			>50
26			>50	≠41			12.22
E-64	-	-	0.05	≠42			12.75

[≠] IC₅₀ of compound **41** and **42** against FP-3 were 13.77 μM and 14.94 μM, respectively; compound **24** & **41** also inhibited the growth of W2 strain of *P. falciparum* with IC₅₀ = 2.08 μM and 4.65 μM, respectively. IC₅₀ of positive controls, artemisinin and chloroquine, against W2 strain were 0.012 μM and 0.062 μM, respectively.

Among the evaluated compounds, compound **24** and **41** showed activity against the W2

strain of *P. falciparum*. However, compound **24** was inactive against FP-2, suggesting action against targets other than FP-2. The majority of identified compounds specifically inhibited FP-2 but not FP-3, which may account for their inability to block parasite development, as deletion of only FP-2 by gene knockout was not lethal to parasites²³¹. Activity of compound **41** can be attributed to its dual inhibition against FP-2 and FP-3. However, compound **42** was not active up to 10 μ M concentration in culture parasites, despite of having same log order of activity against FPs as **41**. This suggests the obvious involvement of other physicochemical properties for inactivity of **42** in cultured parasites development such as the solubility, lipophilicity, polar-surface area etc.

To gain insight into the binding mode of test compounds, docking of compounds into the FP-2 binding site using the Glide docking program was performed. Previously, we compared abilities of two scoring functions implemented in Glide: Gscore and Emodel, to enrich known FP-2 actives from decoys downloaded from the DUD database²²⁹. In our study, Emodel outperformed Gscore in retrieving actives from decoys and proved more appropriate for the apolar FP-2 binding site²²⁹. Thus, we applied our validated Glide XP protocol with an Emodel scoring function to rank order the analogs from both series. Here, the distance of the electrophilic carbonyl carbon present in the putative hits to the cysteine sulfur (typically <3.5 Å) was used as a measure of covalent adduct formation. Docking studies revealed some important trends which are discussed below.

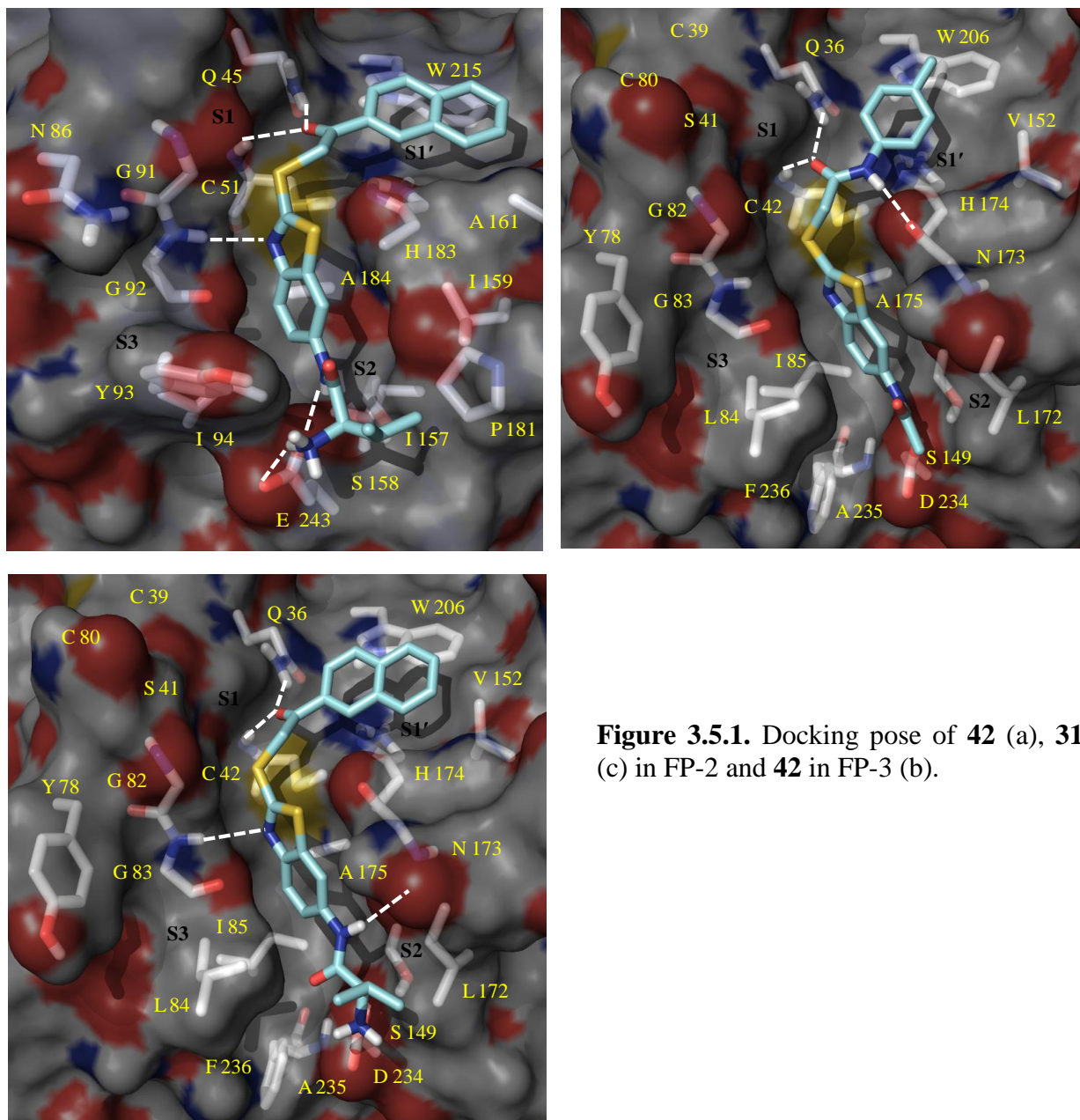


Figure 3.5.1. Docking pose of **42** (a), **31** (c) in FP-2 and **42** in FP-3 (b).

Among the benzothiazole series, compounds with α -thioacetamide and α -thioketones electrophiles were ranked higher (Emodel scores from -61 to -84), whereas compounds with α -thioesters showed Emodel scores of greater than -56, except compound **29** (Emodel score -61). This trend was consistent with the experimental affinity of compounds from the benzothiazole series. For example, most of the compounds from the α -thioesters were inactive up to 50 μ M

concentration, and, thus, ranked lower, except **29**, which showed moderate activity ($IC_{50} = 31.34 \mu\text{M}$) against FP-2. In addition, in α -thioesters containing compounds, the distance of reactive carbonyl carbon from the catalytic Cys 42 of the FP-2 active site was always $>4.0\text{\AA}$. On the other hand, the majority of compounds in the α -thioacetamide series showed a similar log order of activity against FP-2, and they were always more active than corresponding α -thioesters. These observations suggest the predictive nature of the scoring scheme used here to separate actives from inactives in congeneric series of compounds. However, correlation of docking scores with activity was poor, perhaps due to similar log orders of activity for identified hits and the parent compound **2**. In the benzothiazole series, bulkier R1 and R2 substituents affected the placement of the reactive carbonyl carbon in the vicinity of the catalytic Cys42 of the FP-2 active site. For example, compounds **24** and **25** with naphthalene in the S1' pocket and quinoline or substituted phenyl in the S2 pocket, showed modest or no activity against FP-2.

The compounds with the α -thioacetamide electrophile (**30-39**) formed an additional H-bond with the Asn173 of FP-2, placing the reactive carbonyl carbon in an appropriate position for nucleophilic attack by Cys42. The docking pose of representative compound **31** from thioacetamide series is shown in Figure 3. The *p*-toluene moiety of **31** showed hydrophobic interactions with Tyr 206 of the S1' pocket. The thioacetamide linker exhibited interactions with key catalytic residues such as Cys 42 and Gln36 in addition to H-bond formation with Asn173. The anilinic -NH of **31** formed an H-bond with Asp234 of the S2 site.

FP-2 and FP-3 share 65% sequence identity⁵⁰. Both of these enzymes have a preference for substrate with a hydrophobic residue at the P2 position²⁰³. The major difference between these two enzymes is in their S2 pockets. The Asp234, Leu84 and Leu172 of the S2 pocket in FP-2 is

replaced by the bulkier Glu243, Tyr93 and Pro181 in FP-3, making the S2 pocket more narrow in FP-3⁵⁰. These differences are likely major contributing factors for distinct biochemical specificity and ligand recognition among these two homologous hemoglobins. In the present study, compounds **41** and **42** showed inhibition of FP-2 and FP-3. The binding modes of **41** in FP-2 and FP-3 explain their inhibition, as follows (Figure 3.5.1). First, as expected, the protonated amino functionality of **41** forms charge-charge interactions with Asp234 and Glu243 in the S2 pockets of FP-2 and FP-3, respectively. Second, the isopropyl side chain of **41** forms van der Waals interactions with Leu172 in FP-2 and Pro181 in FP-3. Third, the thiazole nitrogen of **41** interacts with the main chains of glycine residue (Gly83 in FP-2 and Gly92 in FP-3) that is highly conserved in papain-like family of cysteine proteases and considered important for inhibition of these enzymes. Fourth, the carbonyl carbon of α -thioketone linker interacts with the residues of an oxyanion hole formed by the backbone of Cys42 and side chain of Gln36. Finally, the naphthalene moiety exhibits van der Waals interactions with Tyr206 and Val152 in S1' pocket of FP-2 and Tyr215 and Ala161 in S1' pocket of FP-3. Collectively, the above interaction profiles of **41** contribute to its activity against both parasitic cysteine proteases.

Next, we screened compound **41** and **42** with protonated amino group against selected human papain-family of cysteine proteases, cathepsins K, L1 and B (Table 3.5.2). Although, cathepsins K and L1 lack corresponding polar residues in the S2 pocket (Leu in Cat K and Ala in Cat L1, see Figure 2.3.2 in Chapter 2) as present in FPs (Asp in FP-2, Glu in FP-3), compound **41** and **42** also inhibited these enzymes, whereas both compounds were inactive against cathepsin B having polar glutamate residue at the same position. These suggest the polar residues of S2 pocket, although important for improving potency, may not be crucial for design of selective inhibitors

against falcipains.

Table 3.5.2. The IC₅₀ values of FP-2 hits against selected human cathepsin (Cat) peptidases, K, L1 and B.

Compound No.	Cat K IC ₅₀ (μM)	Cat L1 IC ₅₀ (μM)	Cat B IC ₅₀ (μM)
41	14.52	5.87	>50
42	5.62	4.39	>50
E-64	0.004	0.016	0.005

The compounds reported in this study are ketone-based inhibitors and are likely to bind in a covalent but reversible fashion with the active site cysteine. However, extensive efforts would be required to actually prove that these compounds undergo covalent adduct formation with the catalytic cysteine moiety. The proposed mechanism of action of these compounds (shown in Figure 4) which involves formation of a tetrahedral intermediate (hemithioacetal) after nucleophilic attack by the catalytic cysteine (stabilized by the Gln 36, Cys42 residues of the oxyanion hole) is supported by SAR and recent studies by Ellman and co-workers^{232, 233}.

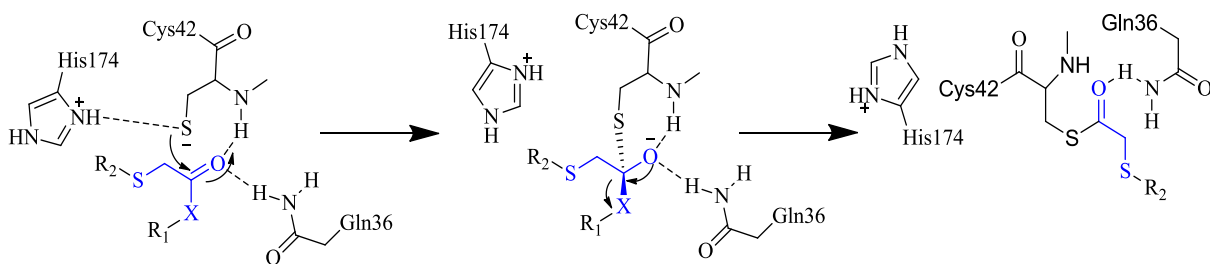


Figure 3.5.2. The proposed mechanisms of action for the α -thioketone soft-electrophiles containing compounds.

3.6. Conclusions

The present study reports novel, non-peptidic active analogs of the benzothiazole and triazole series. Fifteen compounds showed moderate activity against FP-2 and, three were also active against FP-3. In the benzothiazole series, compounds **41** and **42**, with protonated amines inhibited both hemoglobinases. However, these compounds also displayed activity against homologous mammalian cysteine proteases lacking corresponding polar residues (Asp 234 in FP-2 and Glu243 in FP-3), suggesting less significance of these residues in the design of selective inhibitors against FPs. The relative inactivity of synthesized analogs against FP-3 and cultured *P. falciparum* needs to be addressed in the subsequent lead optimization process.

Chapter 4

Analogs of FP-2 Hits: Virtual Screening, Structure-Activity Relationships, Solvent Thermodynamics and Reactivity Analysis

Content of this chapter is submitted for publication: Shah F., Jiri G., Legac, J., Shivakumar, D., Sherman W., Rosenthal, P., Avery M. A. Computer-Aided Drug Design of Falcipain Inhibitors: Virtual Screening, Structure-Activity Relationships, Solvent Thermodynamics and Reactivity Analysis.

4.1. Introduction

Previously, we reported 21 diverse low-micromolar non-peptidic hits against FP-2, originating from novel scaffolds²²⁹. Compound **2** with a tetrazole scaffold was identified as a selective inhibitor of FP-2. In the first part of our current study, we describe efforts to explore the chemical space around the core group of virtual screening hits **1-3** with low-micromolar affinity to FP-2 (Figure 4.1). A combined ligand- and structure-based virtual screening approach was adopted that involved mining of compounds bearing substructures of **1-3** from the Asinex and Chembridge databases, followed by docking of putative hits into the active site of FP-2. A number of active analogs were identified, some of which also showed activity against FP-3 and against development of cultured *P. falciparum*. Most of the active analogs showed the same order of potency as the parent compounds. However, the resulting SAR in these congeneric series of compounds was complex and could not be deciphered by only considering the steric, electrostatic, or van der Waals interactions of inhibitors in the protein active site. Therefore, the second part of this work considers additional factors involved in the binding process, such as the displacement of explicit water molecules and changes in the reactivity of electrophiles based on different chemical substitutions.

The solvent thermodynamics of FP-2 and FP-3 ligand binding domain were computed using the recently published method referred to as WaterMap^{234, 235}. WaterMap predicts the location and thermodynamic properties of the active site solvent using a combination of explicit solvent molecular dynamics, spatial clustering, and inhomogeneous solvation theory. The result of this procedure is a map containing water locations, occupancy, and determinations of whether the water clusters (referred to as “hydration sites”) are favorable or unfavorable in the binding site

relative to bulk solvent. WaterMap has been successfully applied to number of targets in a variety of contexts, such as assessing binding affinity of potent inhibitors^{236, 237}, understanding the SAR in congeneric series of compounds,^{238, 239} and investigating selectivity of compounds within gene family targets such as kinases²⁴⁰. Recently, the method was used to rationalize accurate isothermal titration calorimetry data showing that the addition of lipophilic groups to a ligand resulted in increased potency that was gained via enthalpy from released water molecules, rather than the more traditional view that entropy gain from released water is the only driver of the hydrophobic effect.²⁴¹

The water displacement patterns of FP-2 and FP-3 inhibitors aided in our understanding of key experimental observations of known FP-2 inhibitors, perplexing trends in the SAR of identified hits. However, the relatively modest affinity of some inhibitors (also displacing unstable hydration sites from FP-2 LBD in a few cases) was not evident from the WaterMap analysis alone. We hypothesized that poor reactivity of soft-electrophiles were responsible for the low affinity of some inhibitors towards FP-2. To further support our hypothesis, virtual screening of a small compound library enriched in diverse aryl and aliphatic nitriles with a moderate reactivity profile (based on the reactivity index of nitrile-containing compounds proposed by Obella *et al.*²⁴²) was carried out against FP-2. The biological results showed good agreement between chemical reactivity and the inhibitory potency of these compounds against FP-2.

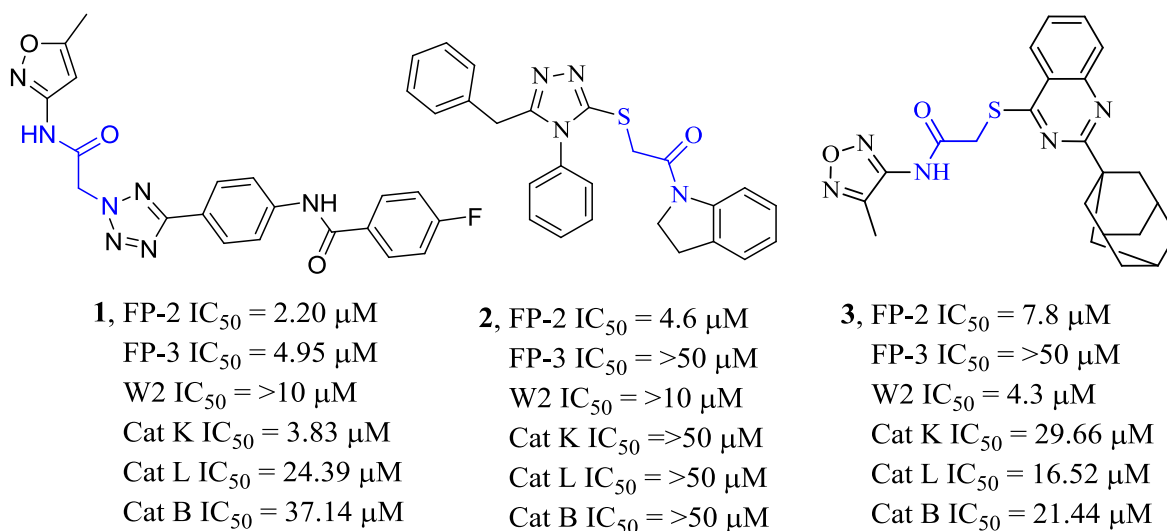


Figure 4.1. FP-2 inhibitors identified by structure-based virtual screening of the Focused Cysteine Protease Inhibitor library ($IC_{50} < 10 \mu M$). W2 stands for the chloroquine resistant strain of *P. falciparum* parasites. Cathepsin (Cat) K, L and B are human cysteine proteases of the papain-family. α -heteroacetamide soft-electrophile is highlighted in blue.

4.2. METHODS

4.2.1. Virtual Screening. The protocol used for virtual screening is shown in Figure 4.2. The substructure search on the highlighted query atoms for compounds **1-3** was conducted in the Chembridge (<http://www.hit2lead.com>) and Asinex (<http://www.asinex.com>) 2009 collections using the built-in substructure search utilities of these databases. The substructure hits were prepared for docking using the LigPrep program (Schrodinger LLC, NY). Approximately 1500 compounds matching sub-graphs of **1-3** were subjected to docking using our previously validated Glide XP protocol²²⁹. The top 20% (~300) compounds were selected based on the top Emodel score. Docking poses were visually inspected for the following criteria prior to submitting them for biological evaluation: (a) placement of a soft-electrophile within 3.5 Å of the sulfur atom of the catalytic cysteine; (b) formation of key hydrogen bonds and hydrophobic interactions by ligand atoms with the key residues of the FP-2 active site such as Gly82, Gly83, Cys42, and

Gln36; (c) energetically reasonable ligand geometry without obvious strain; and (d) commercial availability and diversity of putative hits. Based on these criteria, 65 analogs with diverse R₁-R₃ groups were purchased for analysis.

Virtual Screening Workflow

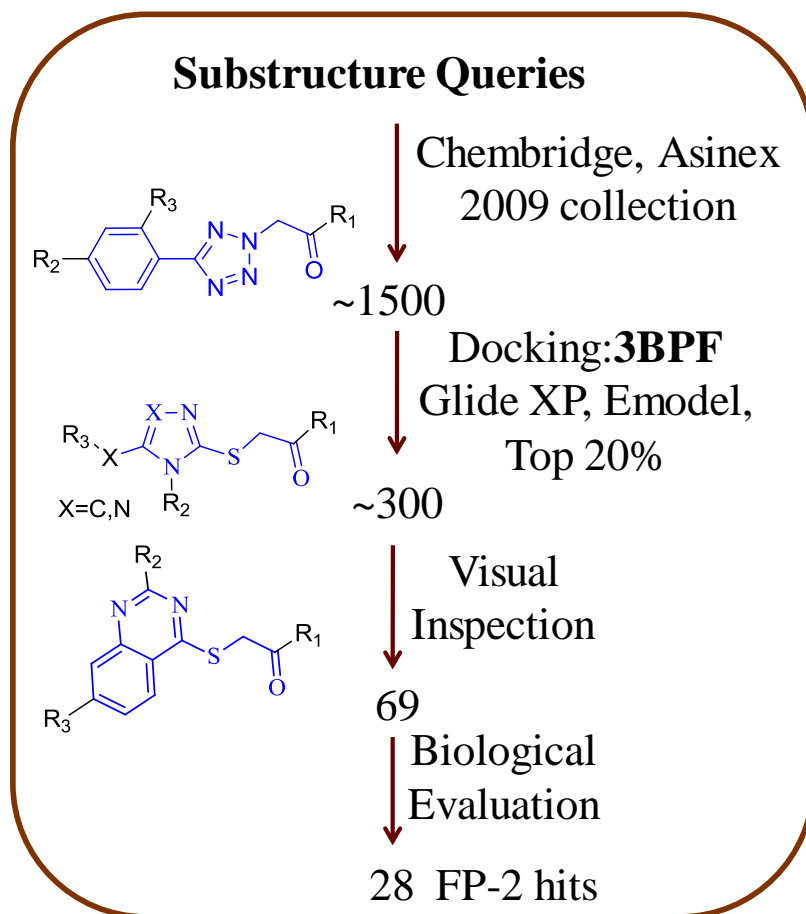


Figure 4.2. Virtual screening workflow to find hits related to compounds **1-3**. Substructure queries for **1-3** are shown in blue.

An additional substructure search was performed for compounds with aliphatic or aromatic nitriles, followed by a similar virtual screening protocol to that shown in Figure 2. Compounds with a match to the query “CC#N” (i.e. carbon connected to carbon connected to nitrogen via a

triple bond) were retrieved from the Chembridge database (2009 collection). The matching set of ~5000 compounds was then docked into the FP-2 binding site. The top 10% (~500) of compounds was visually inspected for the criteria described in the preceding paragraph. Finally, a total of 14 compounds with aliphatic or aromatic nitriles were prioritized for biological evaluation based on two criteria a) compound should have similar interactions in the binding site as predicted for pyrimidine nitrile compounds, a potent class of FP inhibitors¹⁸³ and b) they should displace the key unstable hydration sites from the FP-2 LBD

4.2.2. Biological Evaluation. The prioritized hits were evaluated against FP-2, FP-3, and cultured W2 strain (chloroquine-resistant) *P. falciparum*. A selected set of compounds was also evaluated against mammalian cysteine proteases cathepsin K (Cat K), cathepsin B (Cat B), and cathepsin L1 (Cat L1). These biological assays were conducted using the same protocols as described previously²²⁹.

4.3.3. Thermodynamic Characterization of Waters in the Binding Site. WaterMap predicts the location of water molecules around a protein and provides a quantitative estimation of enthalpy (ΔH), entropy ($-T\Delta S$), and free energy (ΔG) associated with the waters^{234, 235}. The approach combines explicit solvent molecular dynamics, solvent clustering, and statistical thermodynamics. In short, a grand canonical Monte Carlo simulation is run to obtain the initial placement of the waters, followed by a series of equilibration stages and a 2.0 ns production run of explicit solvent molecular dynamics. The protein is constrained in the simulations, allowing for convergence in the sampling of water configurations around the protein conformation of interest. The waters from approximately 2000 equally spaced frames are clustered to determine regions of high solvent density, termed “hydration sites”²³⁴. The entropy for each hydration site is

computed using inhomogeneous solvation theory.^{243, 244} The enthalpy of each hydration site is calculated using the non-bonded molecular mechanics interaction energy with the OPLS_2005 force field.²⁴⁵ These energy terms are calculated relative to a hydration site in bulk water.

WaterMap profiles were generated for FP-2 in complex with an epoxy succinate (E-64) inhibitor (3BPF, PDB code)⁵⁰ and for FP-3 in complex with a vinyl sulfone (VS) inhibitor (3BWK, PDB code)²¹⁶. Proteins were prepared using the Protein Preparation Wizard in Maestro (Schrodinger, LLC, Portland, OR). The protonation states of the catalytic His174 in FP-2 and His183 in FP-3 were adjusted to HIE (histidine with hydrogen on the epsilon nitrogen). Finally, the proteins were minimized with a harmonic potential restraint as implemented in the Impref module from Impact to prepare the final systems for the WaterMap calculations. All waters within 5 Å of the co-crystallized inhibitor were considered for the WaterMap calculations.

4.3. Result and Discussion

The work presented here was initiated to establish structure-activity relationships (SAR) around the active core of compounds **1-3** (Figure 4.1). The combined sub-structure search and rigid-receptor docking identified 28 active analogs of **1-3**. The activity of these compounds against FP-2, FP-3, and cultured parasites is shown in Table 1. The SAR in each series is discussed below.

4.3.1. SAR in 1, 2, 3, 4-tetrazole series. In the tetrazole series, linear compounds showed greater inhibitory activity against FP-2 as compared to the branched compounds. For example, compounds with R₃ substitutions showed either a decrease (**8**, **9**, and **11**) in activity or no activity (**12**), whereas compounds with R₂ substitutions (**4-6**) showed moderate inhibition of FP-2 (Figure 3 and Table 1). In addition, compounds lacking R₂ substitutions (**13**) or containing smaller R₂

groups (**14**) were inactive. This trend was consistent with the SAR of compounds in this series recently reported by us using the combinatorial library synthesis approach²⁴⁶. Compounds **1**, **4** and **9**, with heteroaromatic moieties at R₁, seem to be well tolerated in the S1-S1' pocket, possibly because of additional interactions with residues of an oxyanion hole, such as Gln36 and Cys 42 in FP-2 (Figure 4.3.8). Although linear compounds such as **1** and **4** showed decent potency against both falcipains, they also appeared to inhibit the mammalian cysteine proteases cathepsin B, K, and L1, and were non-selective (Table 4.3.4). In addition, none of the compounds in this series showed activity against cultured parasites at concentrations up to 10 μM (Table 4.3.1).

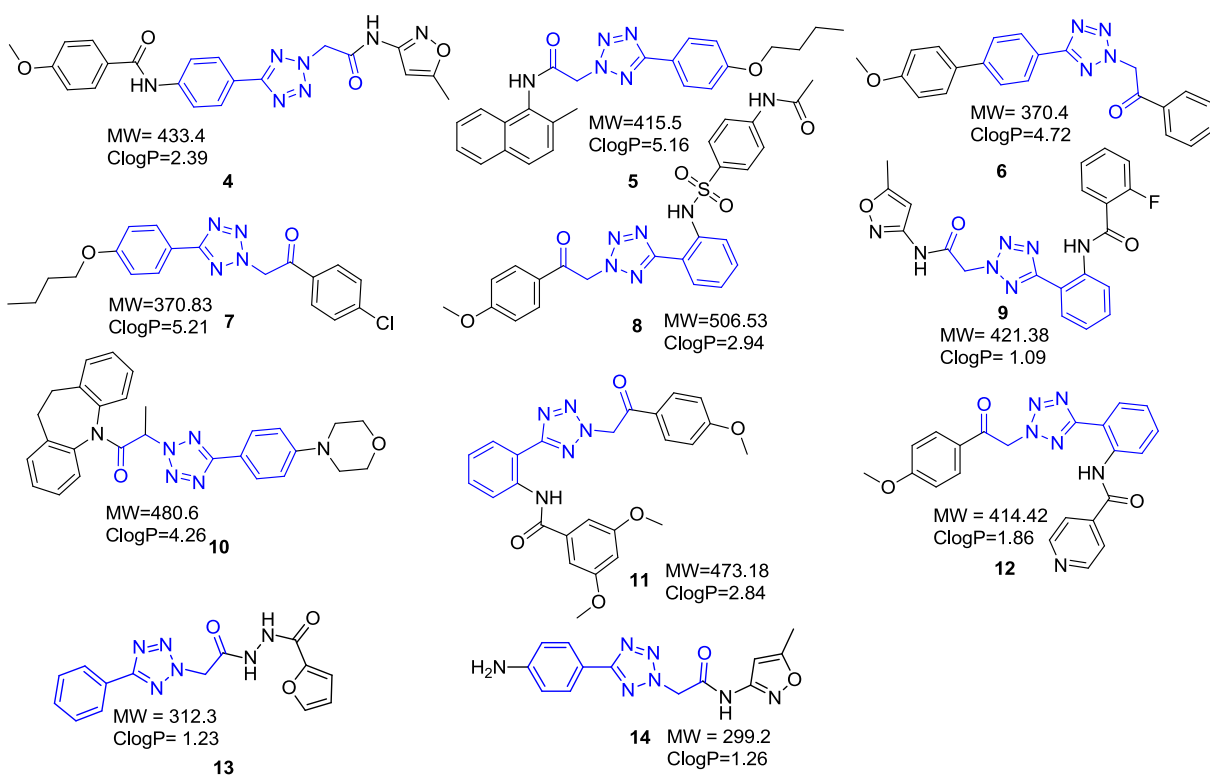


Figure 4.3.1. Chemical structures of compounds with 1, 2, 3, 4-tetrazole core. Original subgraph is highlighted in blue.

Table 4.3.1. Biological evaluation of analogs of virtual screening hit **1** with 1, 2, 3, 4-tetrazole

core

Compound Code	FP-2 IC ₅₀ (μM)	FP-3 IC ₅₀ (μM)	W2 IC ₅₀ (μM)	Compound Code	FP-2 IC ₅₀ (μM)	FP-3 IC ₅₀ (μM)	W2 IC ₅₀ (μM)
4	6.67	19.4	>10	11	38.13	>50	>10
5	11.35	43.9	>10	12	38.13	>50	>10
6	19.15	>50	>10	13	38.13	>50	>10
7	31.46	>50	>10	14	38.13	>50	>10
8	33.04	>50	>10	E-64	0.03	0.16	NT
9	35.92	>50	>10	Artemisinin	-	-	0.092
10	37.81	>50	>10				

NT= not tested, W2= chloroquine-resistant strain of *P. falciparum*

4.3.2. SAR in 1, 2, 4-triazole series. Substituents at R₁ and R₂ had greater impact on the activity in the 1, 2, 4-triazole series. For example, compounds lacking R₂ substituents (**27-34**) or with shorter R₂ substituents (**35-38**) were inactive (Figure 4.3.3). The majority of the active analogs had phenyl at the R₂ position. Compounds **43-48**, with substitutions on the R₂-phenyl moiety, were not tolerated either (Figure 4.3.3). At the R₁ position, the amino-substituted compounds (forming thioacetamide) were well tolerated (Figure 4.3.2 & Table 4.3.2). Analogs with the fused nitrogen heterocycles (**40-45**) or lacking polar -NH (**49, 50**) at that position were inactive. This could be due to lack of ability of these compounds to form an H-bond with Asn173 (Figure 4.3.9c).

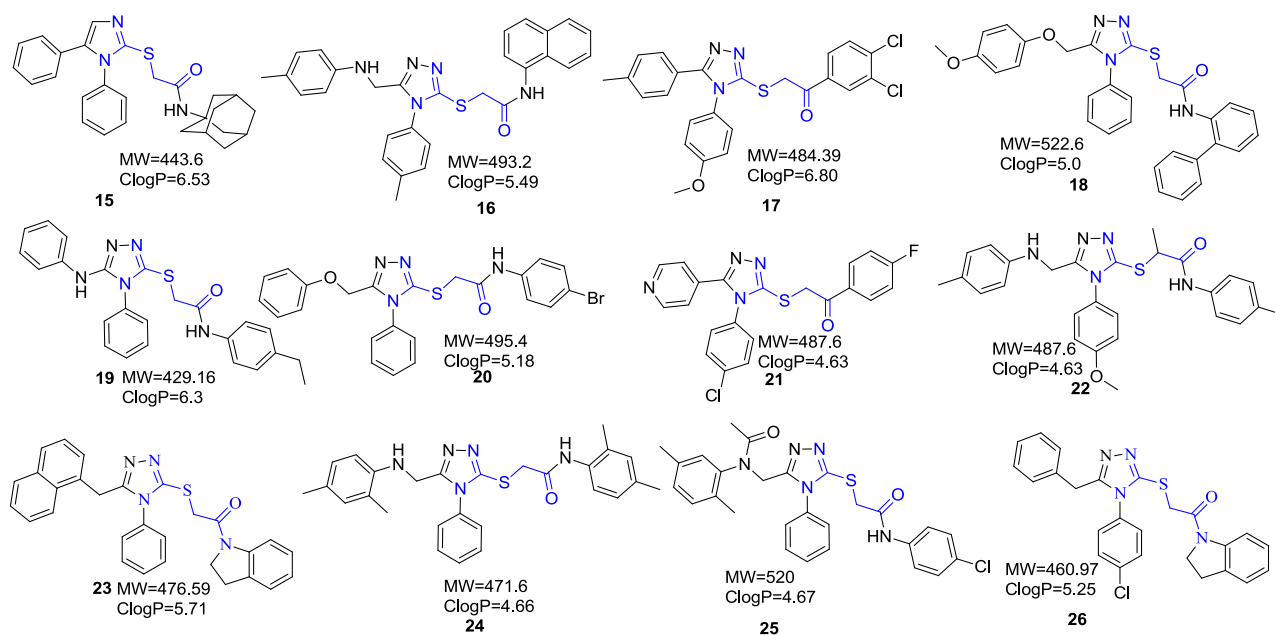


Figure 4.3.2. Chemical structures of compounds with 1, 2, 4-triazole core. Original subgraph is highlighted in blue.

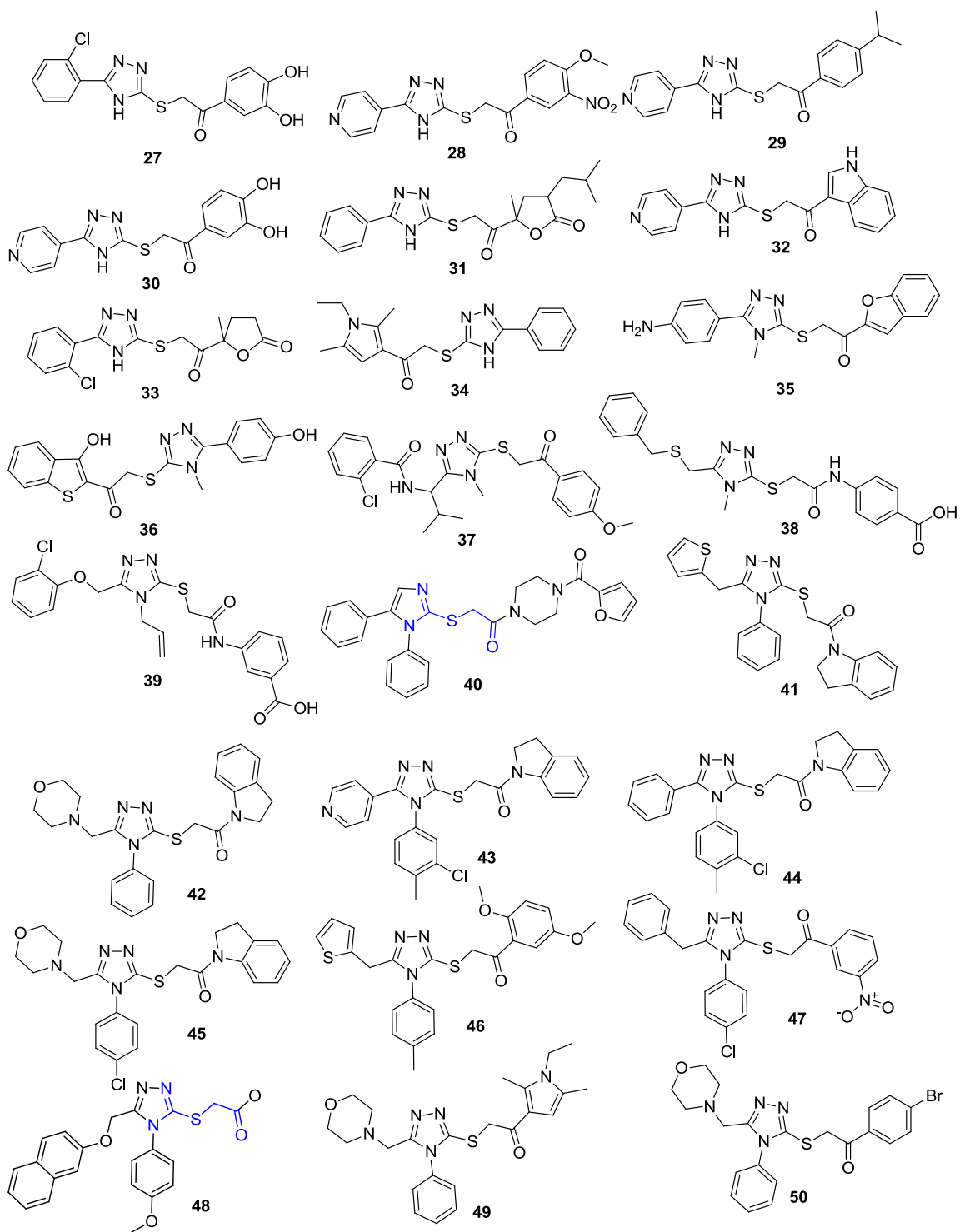


Figure 4.3.3. Inactive compounds in 1, 2, 4-triazole series

Table 4.3.2. Biological evaluation of analogs of virtual screening hit **2** with 1, 2, 4-triazole core

Compound Code	FP-2 IC ₅₀ (μM)	FP-3 IC ₅₀ (μM)	W2 IC ₅₀ (μM)	Compound Code	FP-2 IC ₅₀ (μM)	FP-3 IC ₅₀ (μM)	W2 IC ₅₀ (μM)
15	15.32	>50	1.61	34	>50	>50	NT
16	18.61	>50	5.91	35	>50	>50	NT
17	20.66	>50	NT	36	>50	>50	NT
18	23.21	>50	6.42	37	>50	>50	NT
19	26.45	>50	1.91	38	>50	>50	NT
20	27.43	>50	4.65	39	>50	>50	NT
21	28.61	31.33	NT	40	>50	>50	NT
22	29.40	>50	8.91	41	>50	>50	NT
23	36.95	>50	NT	42	>50	>50	NT
24	>50	>50	NT	43	>50	>50	NT
25	>50	>50	NT	44	>50	>50	NT
26	>50	>50	NT	45	>50	>50	NT
27	>50	>50	NT	46	>50	>50	NT
28	>50	>50	NT	47	>50	>50	NT
29	>50	>50	NT	48	>50	>50	NT
30	>50	>50	NT	49	>50	>50	NT
31	>50	>50	NT	50	>50	>50	NT
32	>50	>50	NT	E-64	0.03	0.16	NT
33	>50	>50	NT	Artemisinin	NT	NT	0.092

NT= not tested, W2= chloroquine-resistant strain of *P. falciparum*

Substitution at R₃ seemed to be crucial for modulating selectivity against FP-2 for this series of compounds. From the multiple sequence alignment of falcipains and other homologous mammalian cysteine proteases of the papain family (cathepsin B, K, and L) it was evident that the composition of the S2 and S3 pocket differs critically in these cysteine proteases^{229, 246}. More importantly, the gate-keeper Leu at the entrance of the S3 pocket in FP-2 and in cathepsin L is substituted by the bulkier Tyr in FP-3, Cat K, and Cat B. To further understand the selectivity of **2** against FP-2, preliminary docking studies were carried out in FP-2 and the cathepsins (Figure 4.3.4). As anticipated, the R₃ methyl phenyl group of **2** was aptly placed in the S3 pocket of FP-2 and Cat L, whereas this moiety was directed into the S2 pocket in FP-3, Cat K, and Cat B,

leading to unfavorable placement of **2** (Figure 4.3.5). This might explain the exclusive selectivity of **2** for FP-2.

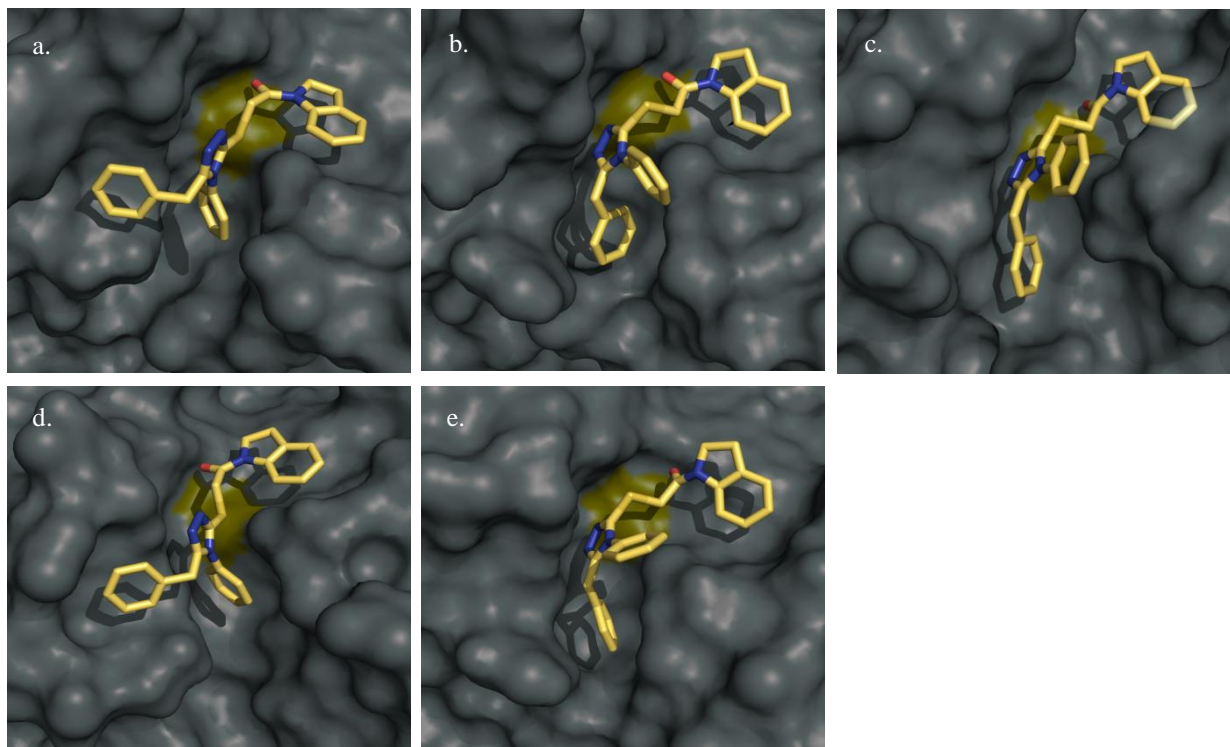


Figure 4.3.4. The proposed binding mode of **2** in a) FP-2; b) FP-3; c) Cat K; d) Cat L1 and e) Cat B.

To further analyze the effects of R_3 substituents on FP-3 selectivity, we evaluated selected active compounds from this series against FP-3 and cathepsins B, K, and L1 (Figure 4.3.5). As expected, compounds with smaller R_3 substituents (e.g. pyridine in compound **21**) tended to be non-selective, showing activity against all the cysteine proteases (except Cat B). As the length of R_3 substituents increased, activity against FP-3, Cat K, and Cat B decreased, probably due to steric clash with Tyr at the entrance of the S3 pocket (for example, compounds **17**, **19**, and **2**). Interestingly, compounds with more flexible R_3 substituents (**16** and **18**) showed activity against FP-2 as well as cathepsins, probably due to the abilities of flexible R_3 substituents to enter into

the S3 pocket. Most of the active compounds from this series were also active against cultured parasites, with an IC_{50} of $<10 \mu M$. Moreover, a few compounds from this series, although inactive against FP-2, were active against the cultured parasites, suggesting involvement of other targets in addition to FP-2 for the mode of action of these compounds (Table 4.3.4).

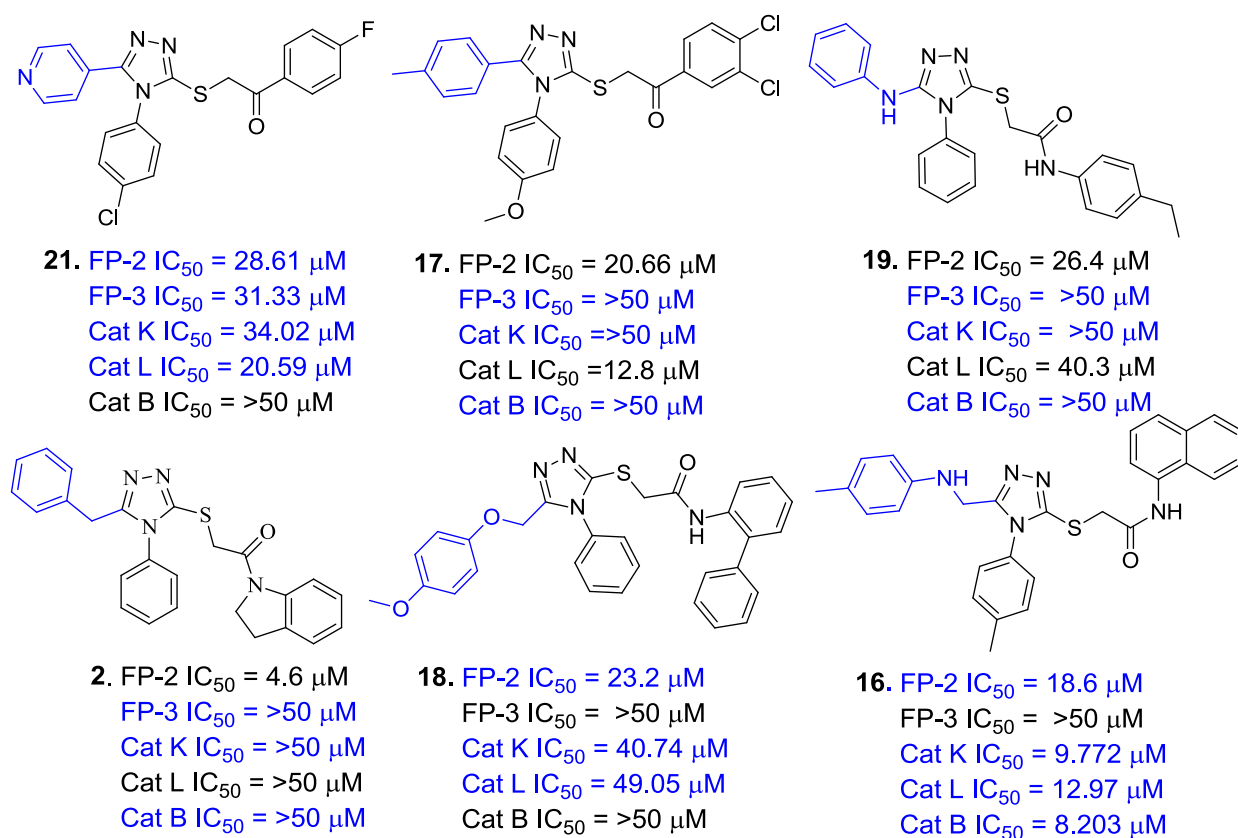


Figure 4.3.5. Effect of R_3 substituents of 1, 2, 4-triazole series on selectivity to FPs and homologous mammalian cysteine proteases of the papain family: Cat K, Cat L1, and Cat B.

4.3.3. SAR in quinazoline series

In the quinazoline series, activity was retained when aliphatic (t-butyl, isopropyl) or aryl (phenyl) substituents at the R_3 position were directly attached to the quinazoline core (e.g. compounds **52-54**). Alicyclic substitutions such as alkyl cyclopentane resulted in diminished activity (for example, compound **59**) or no activity (compounds **60, 61**, Figure 4.3.6, Table 4.3.3)

against FP-2. Substitutions on the quinazoline core at the R₂ position resulted in decreased activity (compound **58**) or inactive analogs (compound **62**), probably due to the steric clash in the narrow S2 pocket of FP-2. At the R₁ position, there was preference for amphiphilic aryl groups such as 1, 2, 5-oxadiazole (**3**), thiazole (**52**, **55**), 1, 3, 4-thiadiazole (**53**), isoxazole (**56**, **59**), and furan (**57**). Also, the heterocyclic moiety with three heteroatoms (**3**, **53**) at R₁ was more potent than a two (**52**, **55**, **56**) or single (**57**) heteroatom containing ring. Substitutions at R₁ either with only hydrophobic groups, such as compounds **63** and **64** or polar groups at R₁ (such as compound **65**), were found to be inactive. From the docking studies, it was suggested that these heterocycles exhibit hydrogen bonding interactions with the side chains of Gln 36 and Trp 206 in addition to hydrophobic interactions with Tyr206 (Figure 4.3.10a). The inactivity of **66** versus **54** might be explained by unfavorable placement of carboxylic acid in the S1' pocket, which is composed predominantly of hydrophobic residues, whereas the inactivity of **67-69** is not clearly understood based on visual inspection of the interactions.

Among the active analogs, compound **54** showed an IC₅₀ value of 8.5 μM against FP-2 and was also active against cultured parasites (W2 IC₅₀ = 6 μM). Compound **54** moderately inhibited Cat K (IC₅₀ = 39.2 μM) and Cat L (IC₅₀ = 14.6 μM) and was inactive against Cat B at concentration up to 50 μM. In addition, as reported previously by our laboratory¹⁹, compound **54** was a low-micromolar inhibitor of SARS severe acute respiratory syndrome 3CL^{pro} (IC₅₀ of 5.8 μM), which is the main protease of SARS, playing an important role in viral replication²⁴⁷. To the best of our knowledge, **54** is the first common inhibitor of malarial (protozoal) and SARS (viral) cysteine proteases reported to date. The discovery of **54** holds merit as these two cysteine proteases are quite different from each other, in terms of their protease family (FP-2 is a member of the papain-like family whereas SARS-3CL^{pro} is a chymotrypsin-like cysteine protease), and in

their active site geometry.

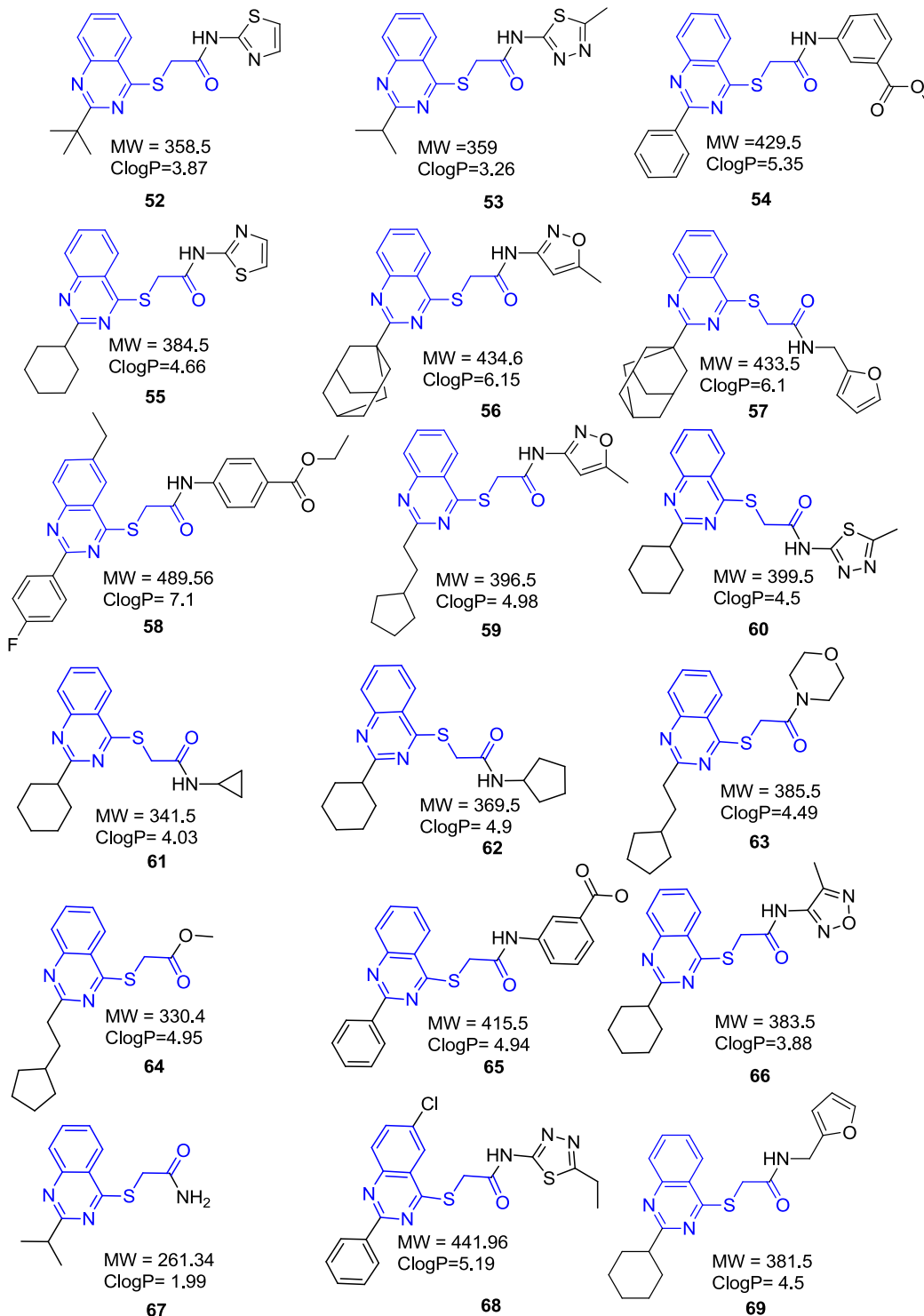


Figure 4.3.6. Chemical structures of compounds from quinazoline core. Original subgraph is highlighted in blue.

Table 4.3.3. Biological evaluation of analogs of virtual screening hit **3** with quinazoline core

Compound Code	FP-2 IC ₅₀ (μM)	FP-3 IC ₅₀ (μM)	W2 IC ₅₀ (μM)	Compound Code	FP-2 IC ₅₀ (μM)	FP-3 IC ₅₀ (μM)	W2 IC ₅₀ (μM)
52	4.64	>50	NT	62	>50	>50	NT
53	5.93	>50	NT	63	>50	>50	NT
54	8.57	>50	6.0	64	>50	>50	NT
55	9.19	>50	NT	65	>50	>50	NT
56	14.00	49.50	NT	66	>50	>50	NT
57	15.81	>50	NT	67	>50	>50	NT
58	18.82	>50	NT	68	>50	>50	NT
59	>50	>50	NT	69	>50	>50	NT
60	>50	>50	NT	E-64	0.03	0.16	-
61	>50	>50	NT	Artemisinin	-	-	0.092

Table 4.3.4. Evaluation of selected FP-2 hits against the mammalian cysteine proteases, cathepsins K, L1, and B

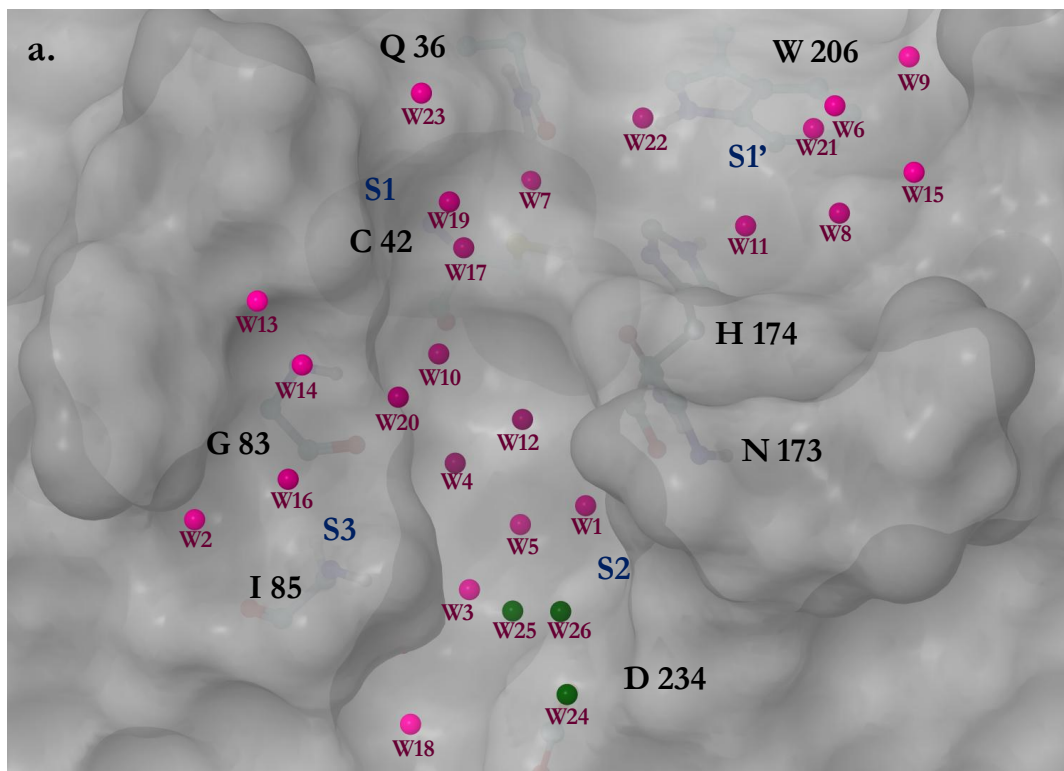
Compound No.	Cat K IC ₅₀ (μM)	Cat L1 IC ₅₀ (μM)	Cat B IC ₅₀ (μM)
4	4.22	3.12	10.07
17	>50	12.80	>50
18	40.74	49.05	>50
19	>50	40.3	>50
21	34.02	20.59	>50
54	39.21	14.55	>50
E-64	3.99	16.15	5.01

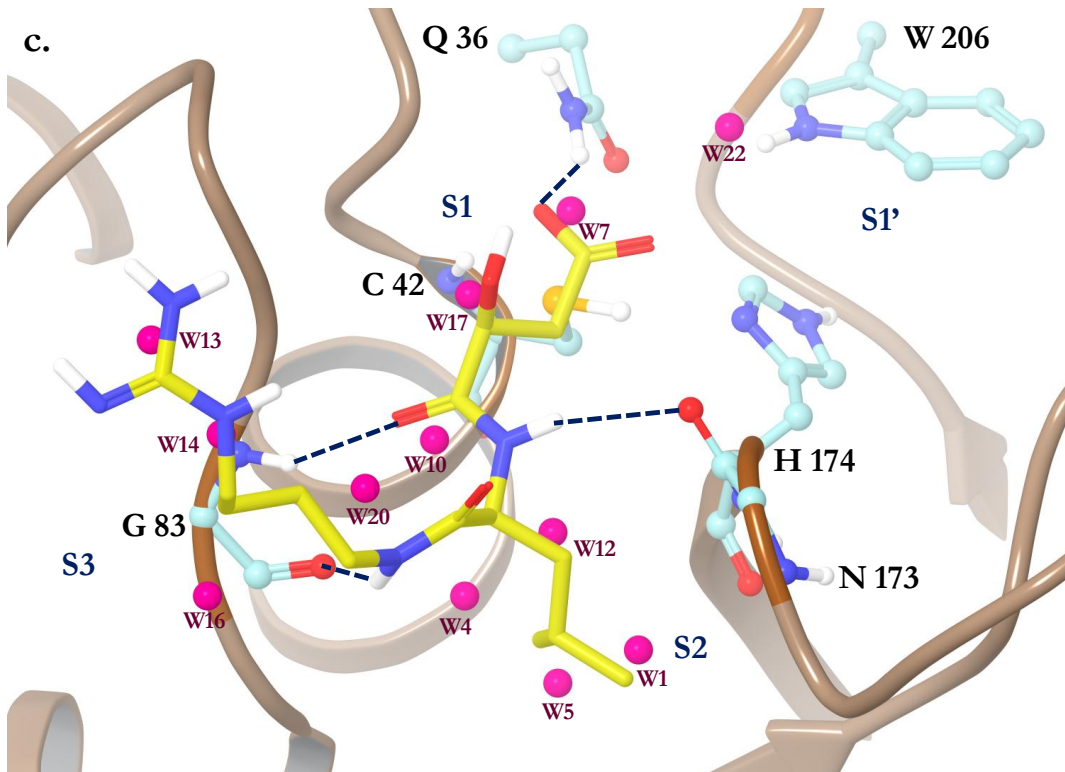
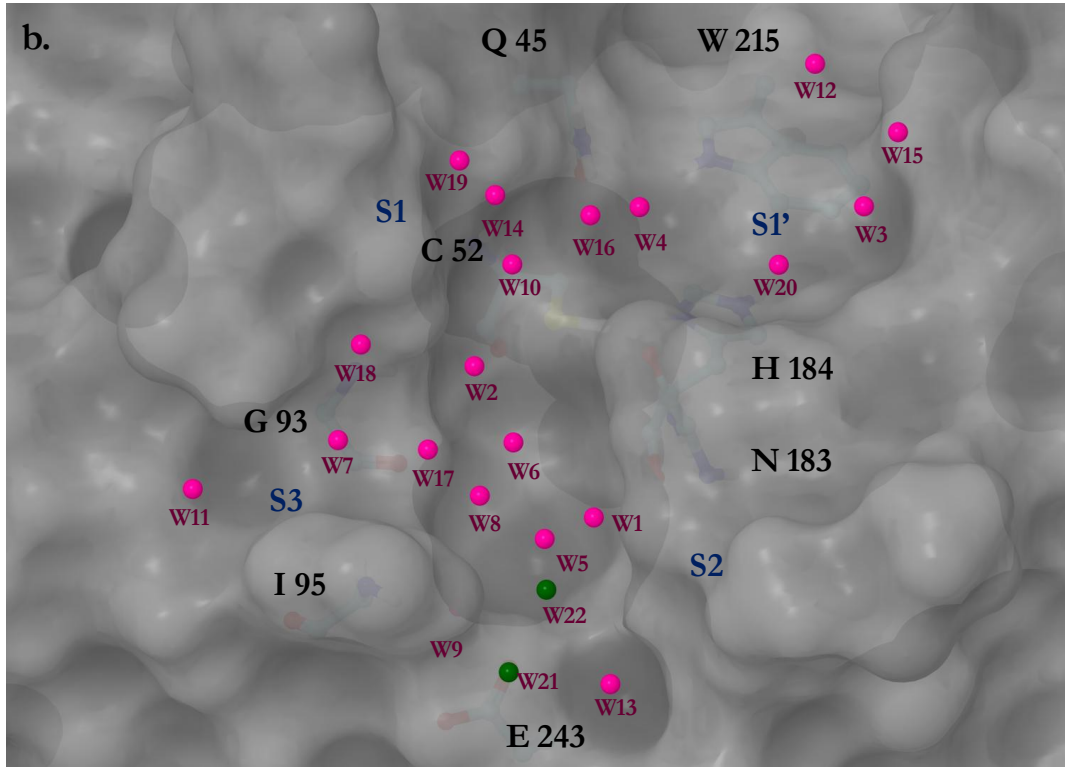
Several trends in the SAR could not be understood from the molecular docking studies with FP-2. For example, the experimental SAR of identified hits was very sensitive for compounds with R₂ substituents pointing towards the S2 pocket of FP-2, although no specific interactions in this pocket suggested such a significant change in potency. Also, there was a general preference for amphiphilic substituents in the S1-S1' pocket that could not be explained based on molecular interactions. Moreover, the identified analogs showed the same log order of activity, in spite of

enhanced interaction profiles, compared to the parent compounds **1-3** in the ligand binding domain of FPs. The steric and electrostatic interactions alone in addition to other terms of the Emodel scoring function (e.g. van der Waals, H-bond, lipophilic, ligand strain) were not sufficient to understand the steep SAR or moderate affinities of these compounds. Therefore, we anticipated a potentially important role of explicit solvent in the binding site and/or reactivity of the electrophiles in these compounds, both of which are generally not estimated by typical scoring functions.

Water molecules in protein binding sites have unfavorable entropy compared with those in bulk solvent due to their restrained movement in the active site, with the amount of entropy loss related to the degree of translational and rotational restriction^{234, 235}. They may have favorable or unfavorable enthalpy relative to bulk water, depending on the exact nature of the binding pocket and interaction with neighboring waters. Therefore, the thermodynamic properties of waters cannot be estimated by simple visual inspection. Accurately assessing the thermodynamics of water molecules is important for drug design, since displacement of high-energy water molecules from the protein binding site by complementary groups of ligands is considered to be a principle source of binding affinity²³⁶. A number of methods can account for solvation effects in the protein binding site, although many methods, such as MM-GBSA²⁴⁸ and MM-PBSA²⁴⁹, are based on continuum electrostatics and do not account for the explicit nature of water. One method that accounts for waters explicitly, WaterMap, uses a combination of explicit solvent MD simulation and statistical mechanics calculations to provide a quantitative description of enthalpy, entropy, and free energies of water molecules at the surface of proteins (see methods). From the design perspective, water molecules with positive values of ΔG should be displaced by

complementary groups of the ligand, with expected gains in binding affinity resulting from an increase in entropy when water is released into bulk solvent. Conversely, water molecules with negative enthalpy (ΔH) should be considered carefully, as displacement will result in a loss of enthalpy. Therefore, such waters should be avoided, bridged, or judiciously displaced with polar groups to replace polar interactions of water molecules²³⁶.





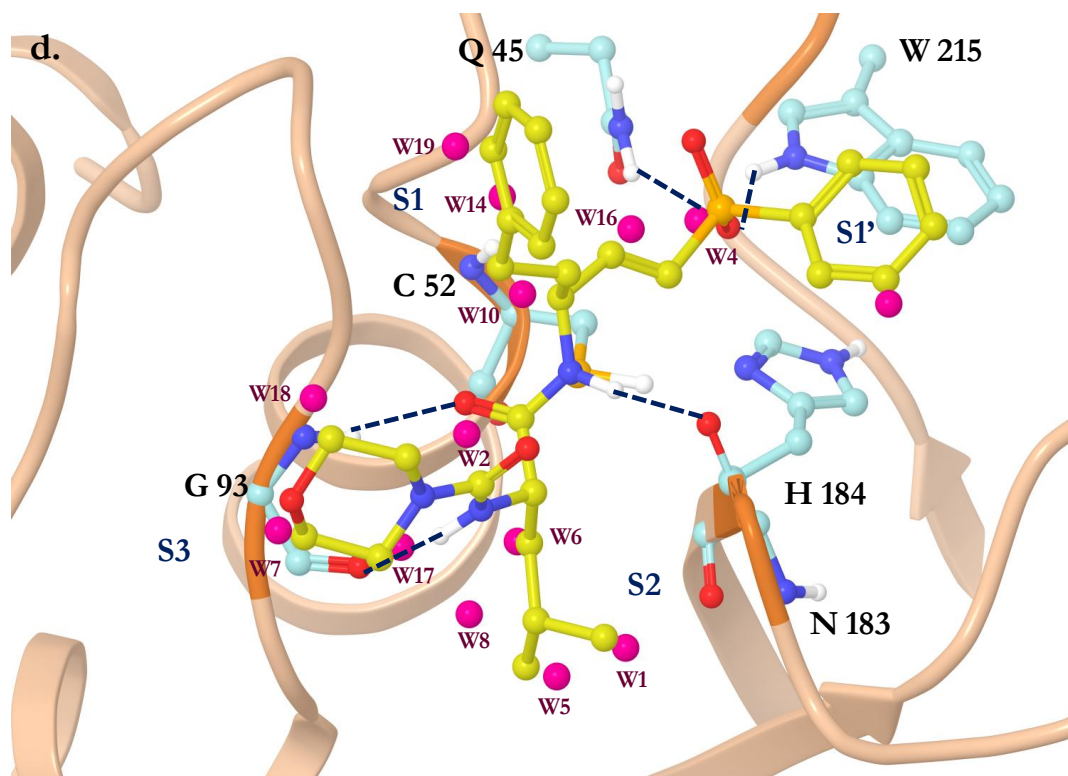


Figure 4.3.7. The WaterMap profile of (a) FP-2, (b) FP-3, (c) FP-2 in complex with E-64, and (d) vinyl sulfone in complex with FP-3. The thermodynamically interesting hydration sites important for the SAR are shown in spheres. Stable hydration sites ($\Delta G < 0$ kcal mol⁻¹) are colored in green, whereas significantly unstable hydration sites ($\Delta G > 1$ kcal mol⁻¹) are shown in purple. The water sites are labeled based on decreasing values of predicted ΔG . Key hydrogen bonding interactions of inhibitors (shown in yellow) with the residues of FP binding sites (shown in grey) are displayed as dotted lines.

The WaterMap profiles of FP-2 and FP-3 with and without co-crystallized inhibitors (E-64 in FP-2 and a vinyl sulfone inhibitor k11017 in FP-3) are shown in Figure 4.3.7a-d. The thermodynamic breakdown of predicted hydration sites are shown in Tables 4.3.5. and 4.3.6. The location and energetics values of waters in these homologous proteases are consistent. The solvent analysis identifies the triangle of enthalpically unstable waters (W1, W5, W12 in FP-2 and W1, W5, W6 in FP-3), solvating the hydrophobic S2 pocket in both FPs. The hydration sites in the narrow hydrophobically enclosed region of the S2 pocket are enthalpically unfavorable

(positive ΔH), as they are unable to form a full complement of hydrogen bonds with protein residues ($\Delta G = >2.3$ kcal mol⁻¹ for all three in FP-2 and $\Delta G = >3.4$ kcal mol⁻¹ for all three in FP-3).

WaterMap predicts two unfavorable waters (W10 = 2.6 kcal mol⁻¹ and W19 = 1.7 kcal mol⁻¹ in FP-2; W2 = 4.22 kcal mol⁻¹ and W6 = 3.45 kcal mol⁻¹ in FP-3) near Gly83 in FP-2 and Gly92 in FP-3. The interaction of a ligand with the conserved glycine near the entrance of the S3 pocket is considered essential for inhibitors of papain-family cysteine proteases binding to non-prime sites (S1-S3)^{11, 50, 215}. Compounds lacking these predicted interactions were void of activity. When co-crystallized inhibitors were mapped onto the generated WaterMap profile, they can be seen to selectively displace these unfavorable hydration sites. For example, the shape of the P2-Leu of E-64 in FP-2 and of VS k11017 in FP-3 complements the arrangement of the unstable waters of the S2 pocket, expelling them to the solvent (Figure 4.3.7c and 4.3.7d). This result is consistent with the observation that FP-2 and FP-3 prefer substrates with leucine in the P2 position^{203, 250} and is consistent with previous findings with other targets that show strong preference for specific amino acids being driven by water displacement²⁵¹. The amide carbonyl near the P2 side chain displaces the unfavorable water (W10 in FP-2, $\Delta G = 2.6$ kcal mol⁻¹; W2 in FP-3, $\Delta G = 4.2$ kcal mol⁻¹) and replaces the H-bond with Gly83 in FP-2 and Gly92 in FP-3, thereby presumably imparting the observed gain in binding affinity.

Table 4.3.5. Predicted occupancy and thermodynamic properties of selected WaterMap waters in FP-2

Water Site	Occupancy	ΔH	$-T\Delta S$	ΔG
W1	0.68	2.55	2.01	4.56
W2	0.51	2.73	1.78	4.51
W3	0.79	0.63	3.03	3.66
W4	0.59	1.67	1.86	3.53
W5	0.40	2.22	1.27	3.49
W6	0.86	0.42	2.8	3.22
W7	0.81	0.13	2.83	2.96
W8	0.46	1.32	1.55	2.87
W9	0.97	-1.63	4.35	2.72
W10	0.46	1.28	1.33	2.61
W11	0.35	1.3	1.03	2.33
W12	0.43	0.98	1.33	2.31
W13	0.38	1.18	1.06	2.23
W14	0.33	1.32	0.9	2.22
W15	0.63	0.28	1.78	2.06
W16	0.52	0.56	1.46	2.02
W17	0.54	0.47	1.51	1.98
W18	0.61	0.15	1.73	1.88
W19	0.37	0.67	1.03	1.70
W20	0.50	-0.09	1.43	1.34
W21	0.47	-0.3	1.46	1.16
W22	0.95	-2.25	3.36	1.11
W23	0.50	-0.52	1.55	1.03
W24	0.58	-3.88	1.94	-1.94
W25	0.58	-4.10	1.97	-2.13
W26	0.36	-3.95	1.19	-2.76

Table 4.3.6. Predicted occupancy and thermodynamic properties of selected WaterMap waters in FP-3

Water Site	Occupancy	ΔH	$-T\Delta S$	ΔG
W1	0.43	3.00	1.25	4.25
W2	0.86	1.45	2.77	4.22
W3	0.71	1.62	2.32	3.94
W4	0.96	-0.09	3.91	3.82
W5	0.91	2.50	1.10	3.60
W6	0.45	2.21	1.24	3.45
W7	0.42	2.30	1.15	3.45
W8	0.37	2.27	1.14	3.41
W9	0.84	-0.79	3.28	2.49
W10	0.57	0.62	1.71	2.33
W11	0.51	0.80	1.45	2.25
W12	0.40	1.09	1.09	2.18
W13	0.88	-0.93	2.98	2.05
W14	0.29	0.98	0.84	1.82
W15	0.39	0.45	1.13	1.58
W16	0.43	0.15	1.29	1.44
W17	0.44	0.19	1.24	1.43
W18	0.40	0.28	1.14	1.42
W19	0.31	0.47	0.83	1.30
W20	0.52	-0.36	1.6	1.24
W21	0.61	-2.55	1.89	-0.66
W22	0.48	-2.28	1.37	-0.91

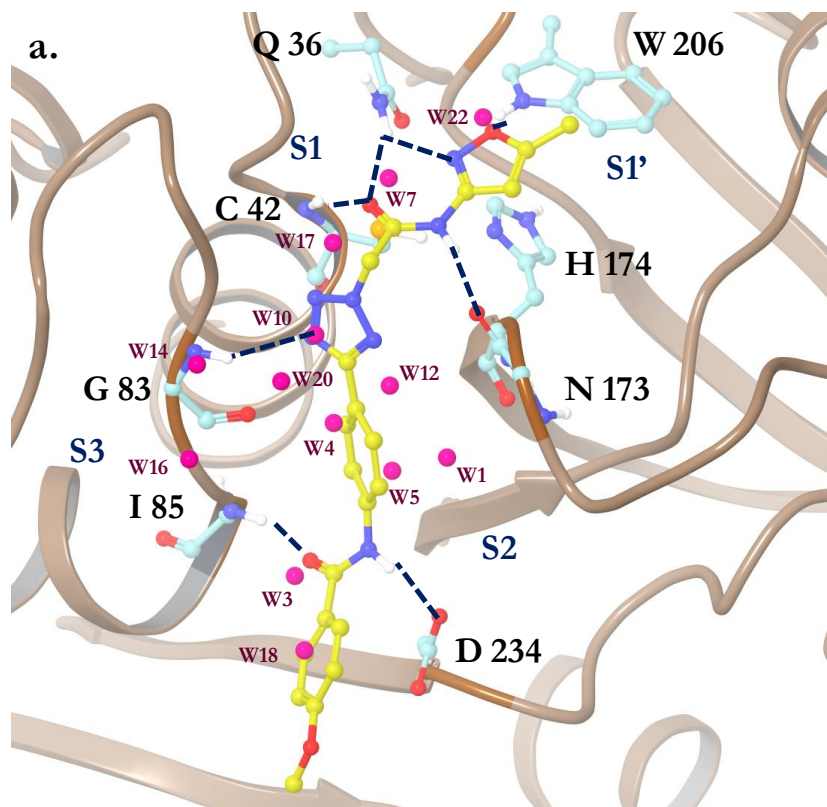
WaterMap identifies unfavorable waters (W7, $\Delta G = 3.0 \text{ kcal mol}^{-1}$ and W22, $\Delta G = 1.1 \text{ kcal mol}^{-1}$ in FP-2; W4, $\Delta G = 3.8 \text{ kcal mol}^{-1}$ and W16, $\Delta G = 1.4 \text{ kcal mol}^{-1}$ in FP-3) in the S1 pocket near the residues of an oxyanion hole (Cys42/51 and Gln36/45). W7 in FP-2 and both unstable waters (W4 and W16) in FP-3 are displaced by corresponding atoms of the co-crystallized inhibitors. In addition, the guanidine moiety of E-64 and morpholine side chain of the vinyl sulfone inhibitor displaces several unfavorable hydration sites from the S3 pocket of FP-2 and FP-3, respectively. In addition, WaterMap predicts other unfavorable hydration sites near the S3

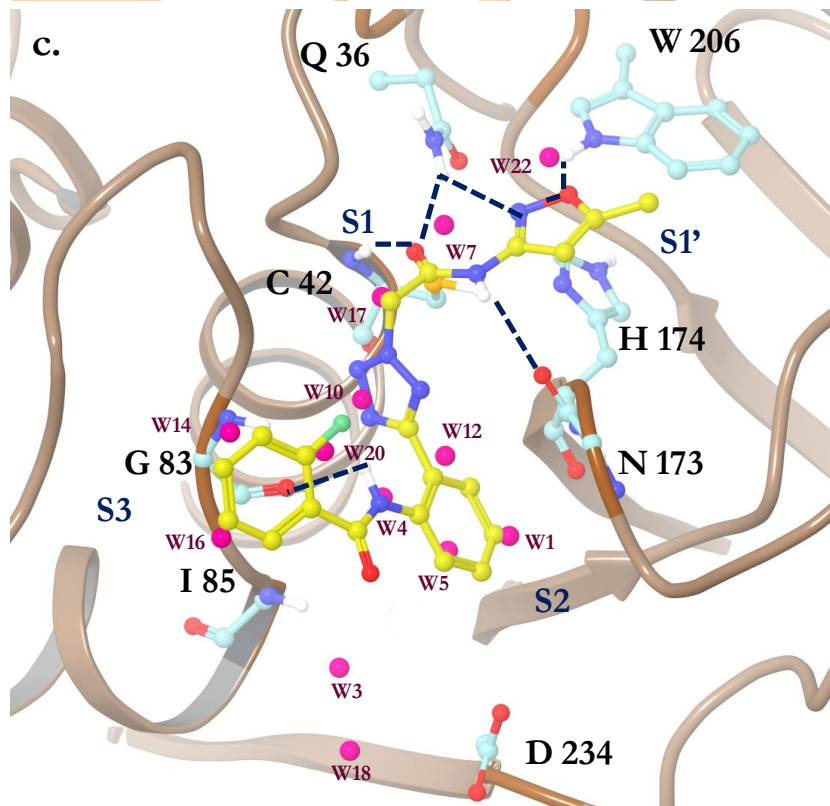
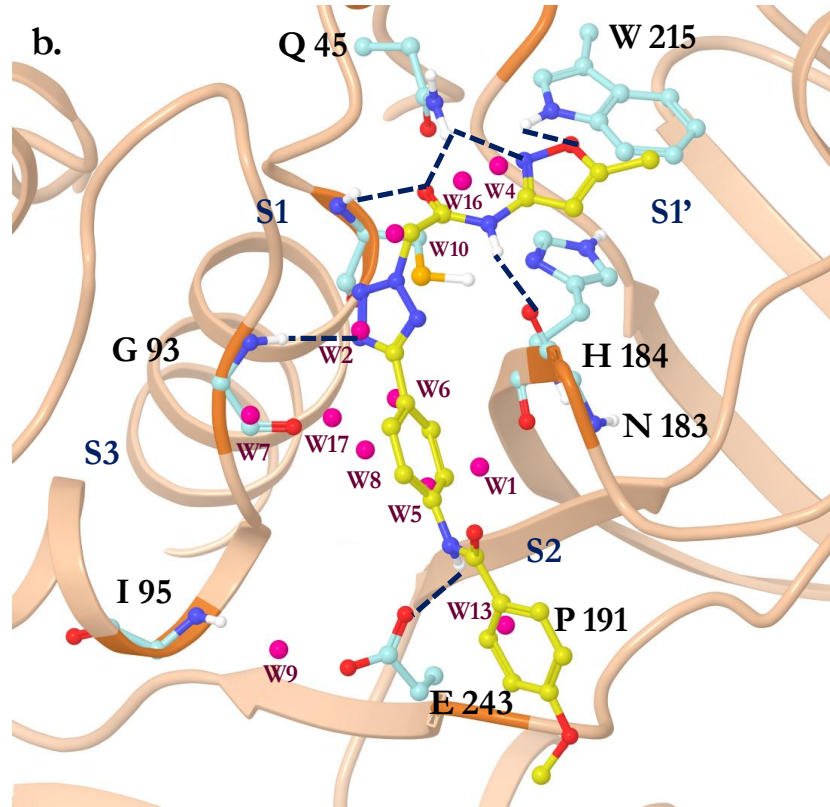
pocket (W2, W3 in FP-2 and W9, W11 in FP-3) and in the S1' pocket (W6, W8, W9, W11, W15, and W21 in FP-2; W3, W12, W15, and W20 in FP-3) that are out of the plane of these peptidic inhibitors. These sites could potentially be targeted, with anticipated gain in binding affinity, with a different ligand scaffold that is able to place functional groups to displace these waters. Finally, WaterMap predicts stable clusters of waters (W24-W26 in FP-2; and W21, W22 in FP-3, ΔG and $\Delta H < 0 \text{ kcal mol}^{-1}$) in the vicinity of Asp234 and Ser149 of FP-2 and Glu243 and Ser158 of FP-3 of the S2 pocket that can potentially be bridged for water-mediated hydrogen bonding to gain binding affinity.

Mapping of the newly identified hits in this work on the WaterMap-generated thermodynamic hydration site profiles of FPs elucidated salient features of the SAR. For example, compound **4**, a linear analog from the tetrazole series, has a complementary shape to the binding site of FPs and it displaces six unfavorable water molecules with $\Delta G > 2.0 \text{ kcal mol}^{-1}$ of unfavorable hydration sites (Figure 4.3.8a and b). In particular, it displaces the unfavorable water W3 bound to Ile 85 in FP-2 ($\Delta G = 3.7 \text{ kcal mol}^{-1}$) and W18 ($\Delta G = 1.9 \text{ kcal mol}^{-1}$). Interestingly, compound **4** binds in a similar fashion to FP-3 and FP-2, but it did not displace the unfavorable water W9 ($\Delta G = 2.5 \text{ kcal mol}^{-1}$) near Ile 95 of FP-3, potentially explaining the log difference in potency of **4** to FP-3 over FP-2. This difference in hydration site displacement is possibly due to the narrow S2 pocket of FP-3 (Glu243 in FP-3 versus Asp234 in FP-2, respectively), directing the methoxy phenyl moiety of **4** close to Pro181.

Likewise, compounds **9** and **14** are unable to displace W3 and W18 from FP-2, and as a result, these compounds display either diminished activity or inactivity against FP-2 (Figure 4.3.8c and d). In FP-3 these compounds also miss key unfavorable hydration sites and hence are devoid of

activity. This may explain why linear compounds in the tetrazole series were more active than the branched compounds. The oxazole moiety of **4** and **9** displaces the unfavorable water (W22 in FP-2, $\Delta G= 1.1 \text{ kcal mol}^{-1}$, W4 in FP-3, $\Delta G= 3.8 \text{ kcal mol}^{-1}$) from the S1 pocket and forms an additional H-bond with Tyr 206, potentially explaining why heteroaromatic substituents were preferred in the S1' pocket. These interactions and water displacements profiles together likely contribute to the higher affinity of **1** and **4** for both falcipains compared to other analogs in the series.





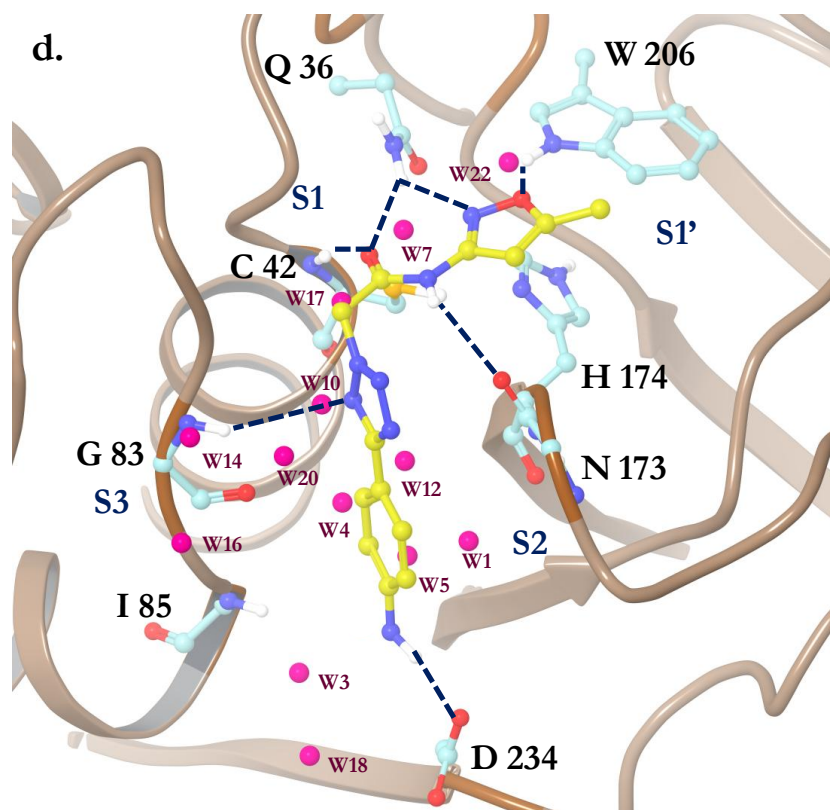
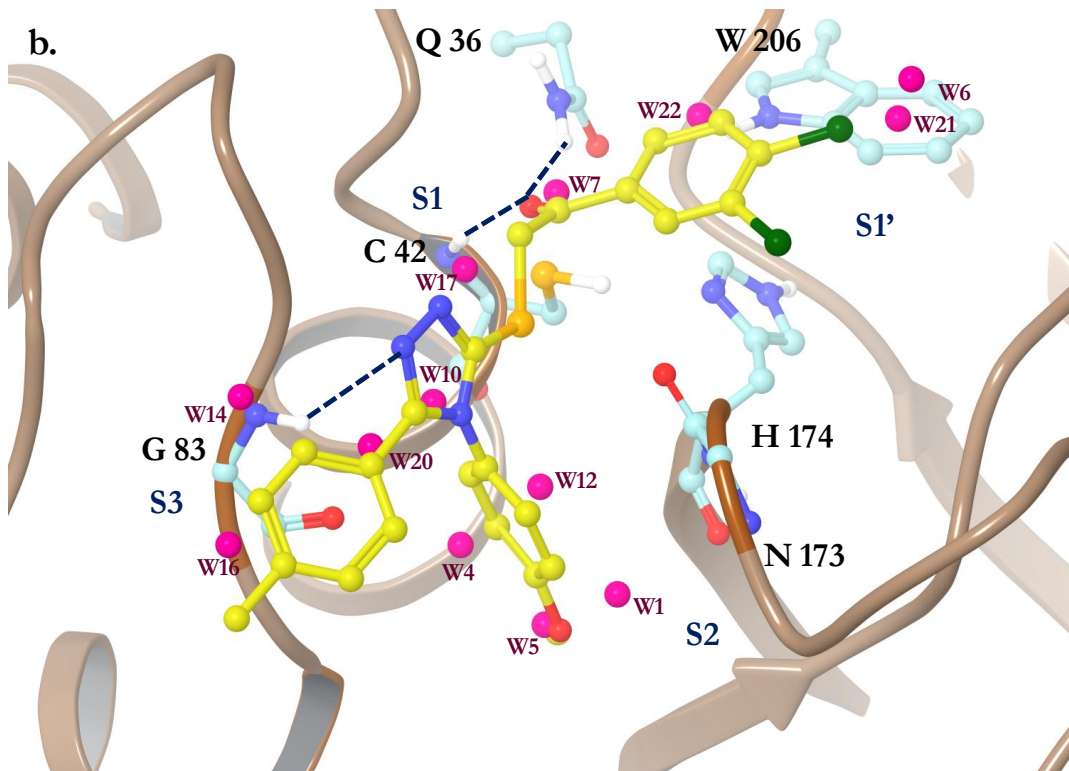
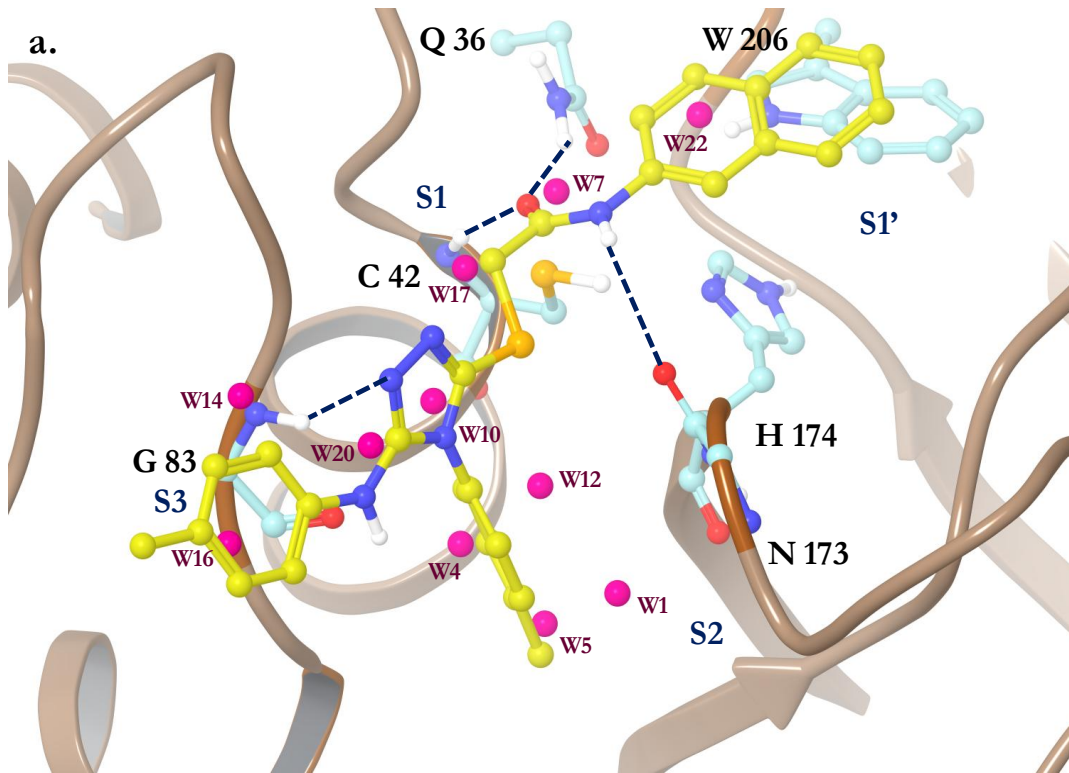


Figure 4.3.8. The predicted binding poses and thermodynamic hydration site profile of **4** in complex with (a) FP-2 and (b) FP-3; the predicted binding poses and thermodynamic profile of (c) **9** and (d) **14** in complex with FP-2. Water site labeling and color codes are the same as shown in Figure 4.3.7.

In the triazole series, compounds lacking R_2 substituents (**29**) or with shorter R_2 substituents (**36**) were unable to displace the hydrophobically enclosed unstable waters from the S2 pocket, and hence they are inactive (Figure 4.3.9c and 4.3.9d). Moreover, R_2 substituents of this series projected differently than the P2-Leu of E-64. For example, compounds with a phenyl (**16**) or substituted phenyl (**17**) at R_2 could not displace the least favorable hydration sites from the S2 pocket, and hence they showed reduced affinity for FP-2 (Figures 4.3.9a and 4.3.9b).



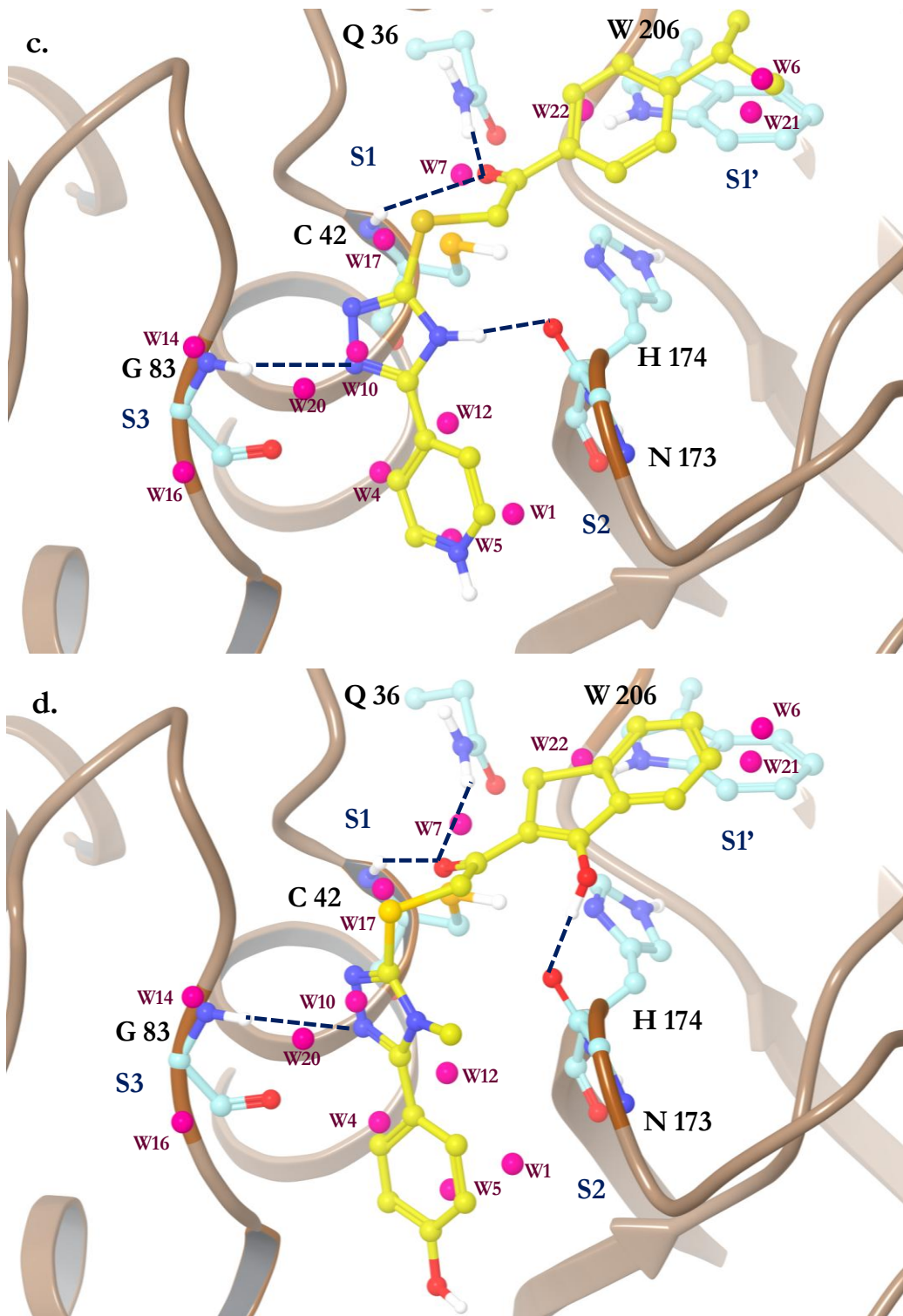
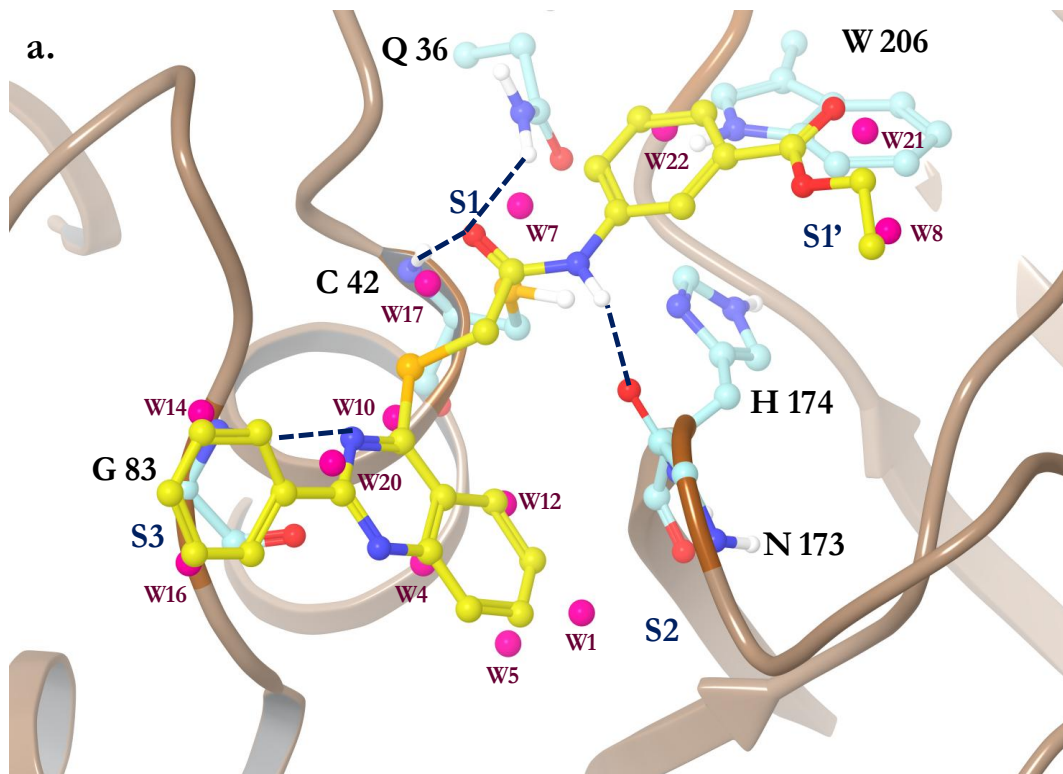
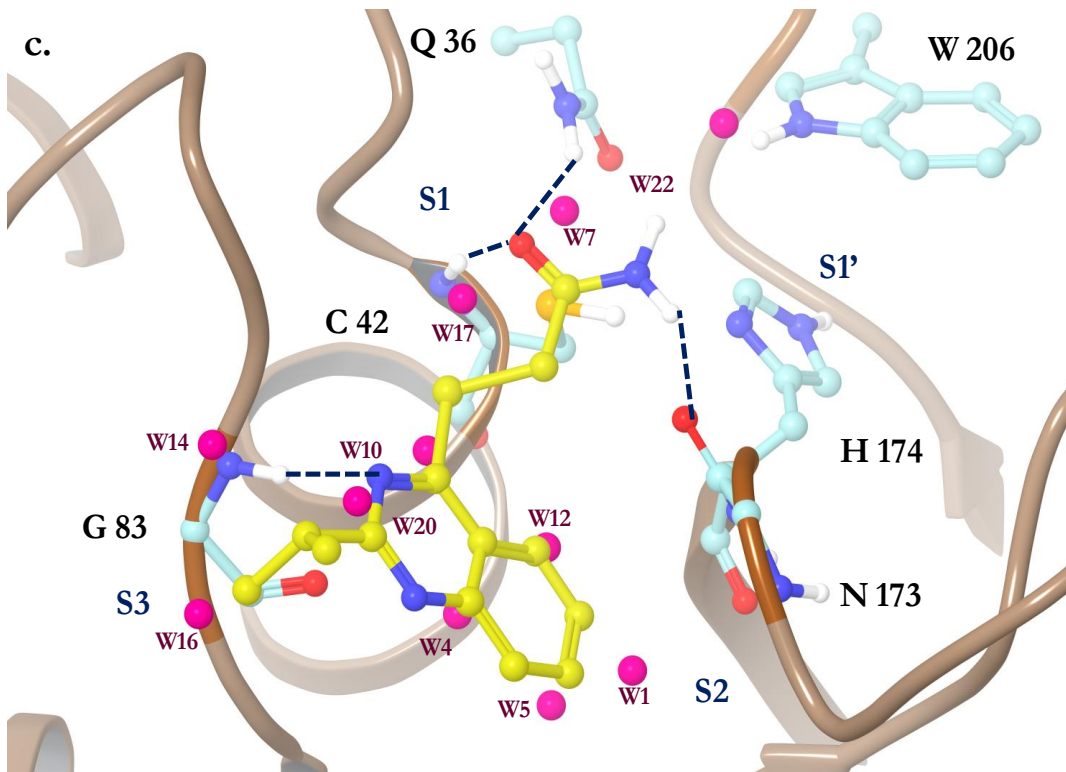
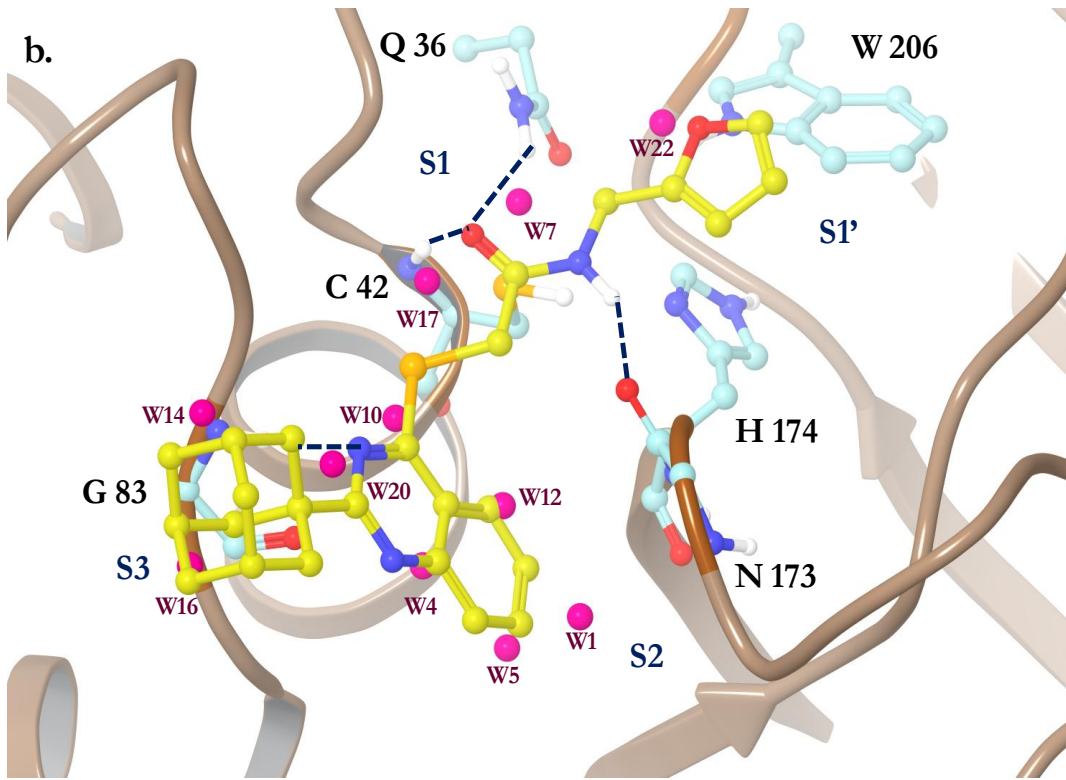


Figure 4.3.9. The predicted binding poses and thermodynamic hydration site profile of (a) **16** (b) **17**; (c) **29** and (d) **36** in complex with FP-2. Water site labeling and color codes are the same as shown in Figure 4.3.7.

Compounds **3** and **57** from the quinazoline series with an oxazole moiety are predicted to displace the unfavorable waters (W7 and W22) from the S1-S1' pocket as well as form an H-bond with the residues of the oxyanion hole such as Gln36, Cys42 and Tyr206 (Figure 4.3.10b). Compounds such as **32** and **38**, lacking corresponding heteroaromatic substitutions or unable to displace the unfavorable waters of the S1' pocket, were inactive (Figure 4.3.10c). Compound **54** additionally displaces the unfavorable waters (W8, $\Delta G = 2.9 \text{ kcal mol}^{-1}$, W21, $\Delta G = 1.2 \text{ kcal mol}^{-1}$) from the S1' pocket. The quinazoline moiety of **54** displaces the unstable waters from the S2 pocket (except the least favorable W1) as well as from the S3 pocket (W14, $\Delta G = 2.2 \text{ kcal mol}^{-1}$ and W16, $\Delta G = 2.0 \text{ kcal mol}^{-1}$), which likely accounts for moderate affinity of this compound for FP-2.





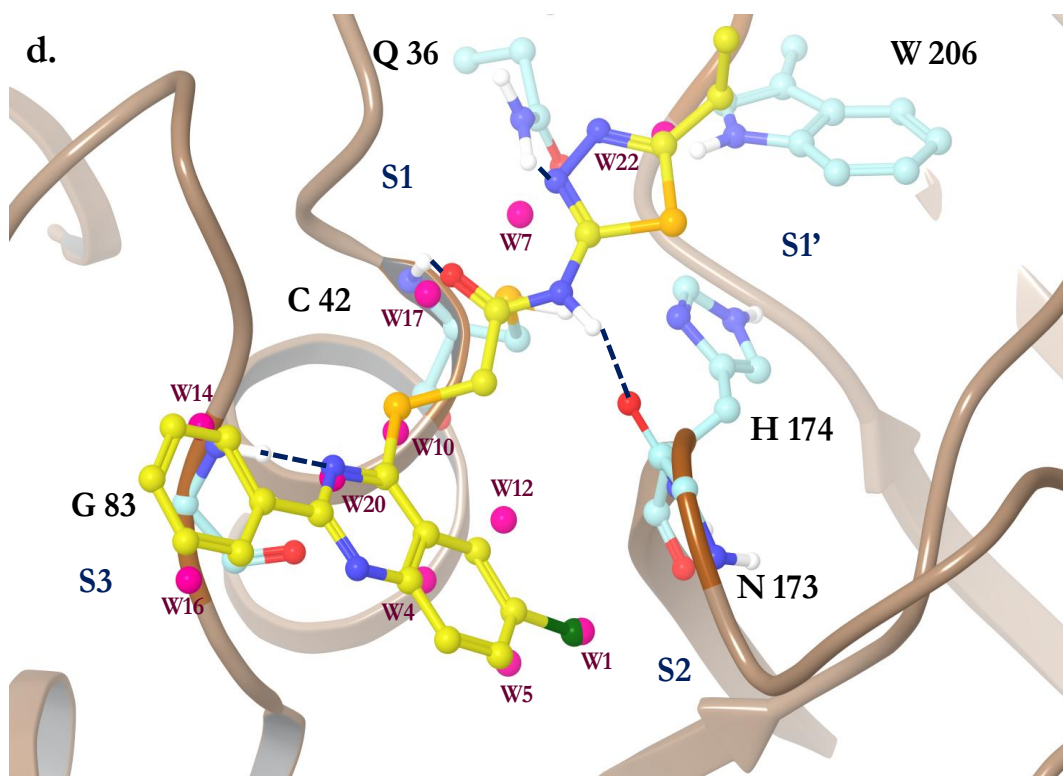


Figure 4.3.10. The predicted binding poses and thermodynamic profiles of (a) **54** (b) **57** (c) **67** and (d) **68** in complex with FP-2. Water site labeling and color codes are the same as shown in Figure 4.3.7.

A few analogs with the ability to displace the high-energy waters of the S2 pocket, such as compound **68** from the quinazoline series (Figure 4.3.10d) or compound **30** (Figure 4.3.13d) from the triazole series, did not show additional gains in binding affinity. Also, none of the analogs of this series were more potent than compounds **1-3**, suggesting that there was an additional factor reducing the binding affinity of these compounds. Based on prior work in our group, we reasoned that the soft-nature of electrophiles present in these hits could account for their moderate affinity against FPs. In our previous study²²⁹, we utilized the LUMO density variant of atomic Fukui indices to locate the most electrophilic center and to estimate the reactivity of electrophiles present in the virtual screening hits. The predicted LUMO density of α -

hetero amide electrophiles were in the range of 0.02-0.32 (on the scale of zero to one, one being the most electrophilic), lower than the LUMO density of compounds containing hard electrophiles, such as α -fluoro ketone (0.52) or vinyl sulfone (0.43). In a similar reactivity analysis, Obella and co-workers assessed the reactivity of diverse nitrile containing compounds using DFT calculations²⁴². Their method involved approximating the free energy of formation of thioimidate by calculating the difference between the energy of thioimidate adduct [E(Adduct)], the nitrile molecule [E(Nitrile compound)], and the precursor methanethiol [E(Methanethiol)]. The generated reactivity index was scaled from 0 to -10 kcal/mol. As anticipated, the heteroaromatic nitriles, in particular triazine and pyrimidine nitriles, were excellent electrophiles (predicted reactivity of -10 kcal mol⁻¹), whereas the aliphatic and the aryl nitriles were predicted as poor electrophiles (0 to -2 kcal mol⁻¹). Furthermore, the authors validated their findings by measuring the ability of representative aliphatic, aromatic, and heteroaromatic nitriles to form irreversible thiazoline adducts upon incubation with cysteine. The calculations showed that poorly electrophilic aryl nitriles (such as naphthalene or pyridine nitrile) have a reduced ability to form cysteine adducts, whereas more electrophilic pyrimidine nitriles or cyanamide have enhanced cysteine adduct formation. Their results showed excellent correlation between predicted reactivity and the ability to form irreversible cysteine adducts²⁴².

Recently, Coteron et al. published data for a 2-pyrimidinecarbonitrile series, showing sub-nanomolar activity against FPs as well as cultured parasites.¹⁸³ The representative compounds from this series are shown in Figure 4.3.11. As discussed above, the strong electrophile pyrimidine nitrile could be a major contributor in the high affinity of these compounds against FPs. This motivated us to further assess the importance of reactivity of electrophiles in the

inhibition of cysteine proteases by evaluating a subset of aliphatic or aryl nitriles (purchased from Chembridge Corporation) with moderate electrophilicity (as proposed by Obella et al.) against FP-2 (Figure 4.3.12). If reactivity of electrophiles plays an important role in the affinity of FP inhibitors, then these compounds were expected to show moderate or less binding affinity as compared to **CN-1** to **3**.

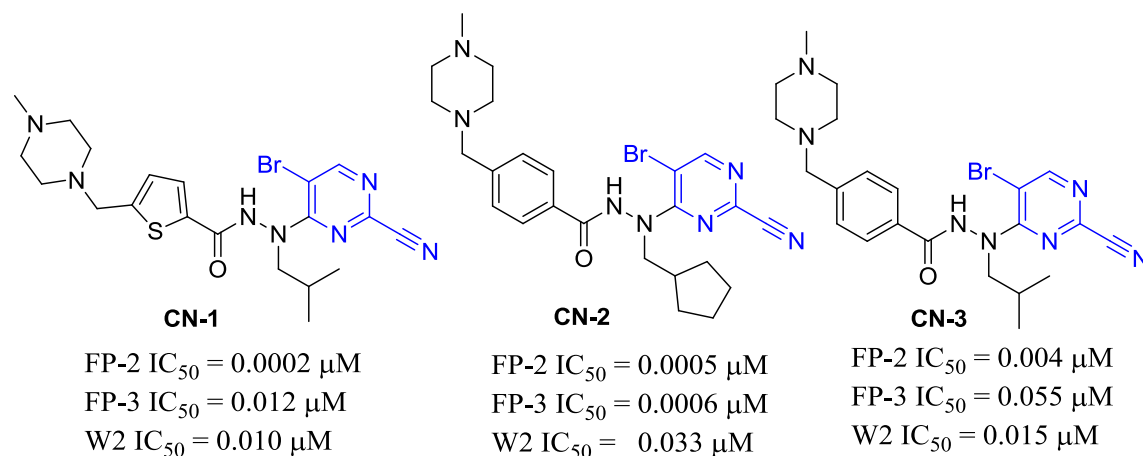


Figure 4.3.11. Potent inhibitors of FP-2, FP-3, and W2 from the 2-pyrimidinecarbonitrile series reported by Obella et al.¹⁸³

The identified hits **70-73**, although having similar binding modes and water displacement profiles as **CN-1** (Figure 4.3.13a-c), displayed moderate affinity towards FP-2 (IC_{50} = 10-20 μ M). In addition, out of fourteen compounds evaluated, only four compounds (**70-73**) were active against FP-2 (Figure 4.3.12). Our data clearly demonstrates that it can be difficult to achieve potent inhibition of FPs with compounds lacking reactive electrophilic centers.

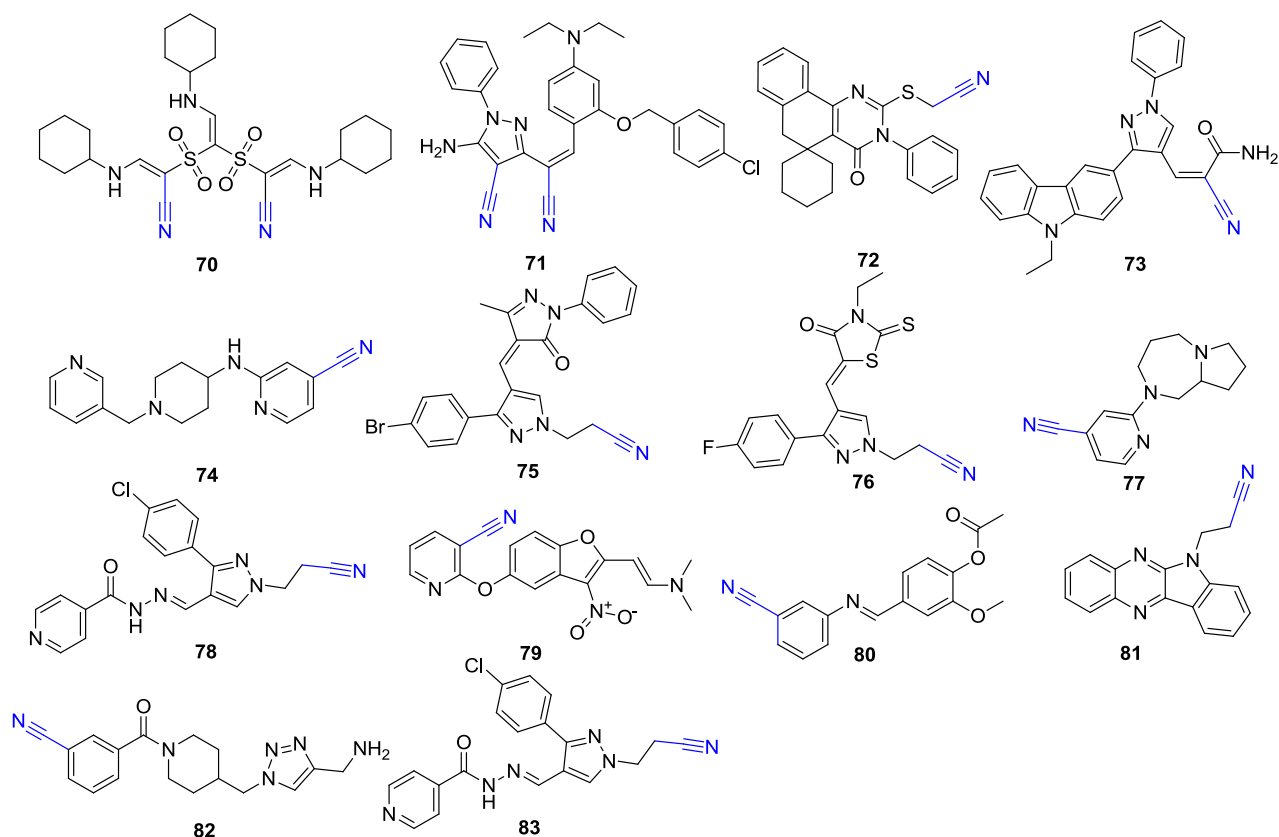


Figure 4.3.12. Aryl- or aliphatic nitrile-containing compounds discovered by substructure search in the Chembridge database. Alkyl nitrile electrophiles are highlighted in blue.

Table 4.3.5. Biological evaluation of analogs of virtual screening hit **3** with quinazoline core

Compound Code	FP-2 IC ₅₀ (μM)	FP-3 IC ₅₀ (μM)	W2 IC ₅₀ (μM)	Compound Code	FP-2 IC ₅₀ (μM)	FP-3 IC ₅₀ (μM)	W2 IC ₅₀ (μM)
70	18.96	>50	7.89	78	>50	NT	>10
71	10.33	>50	5.92	79	>50	NT	>10
72	18.39	>50	>10	80	>50	NT	>10
73	11.21	>50	6.75	81	>50	NT	>10
74	>50	NT	>10	82	>50	NT	>10
75	>50	NT	>10	83	>50	NT	>10
76	>50	NT	8.53	E-64	0.03	0.16	-
77	>50	NT	>10	Artemisinin	-	-	0.092

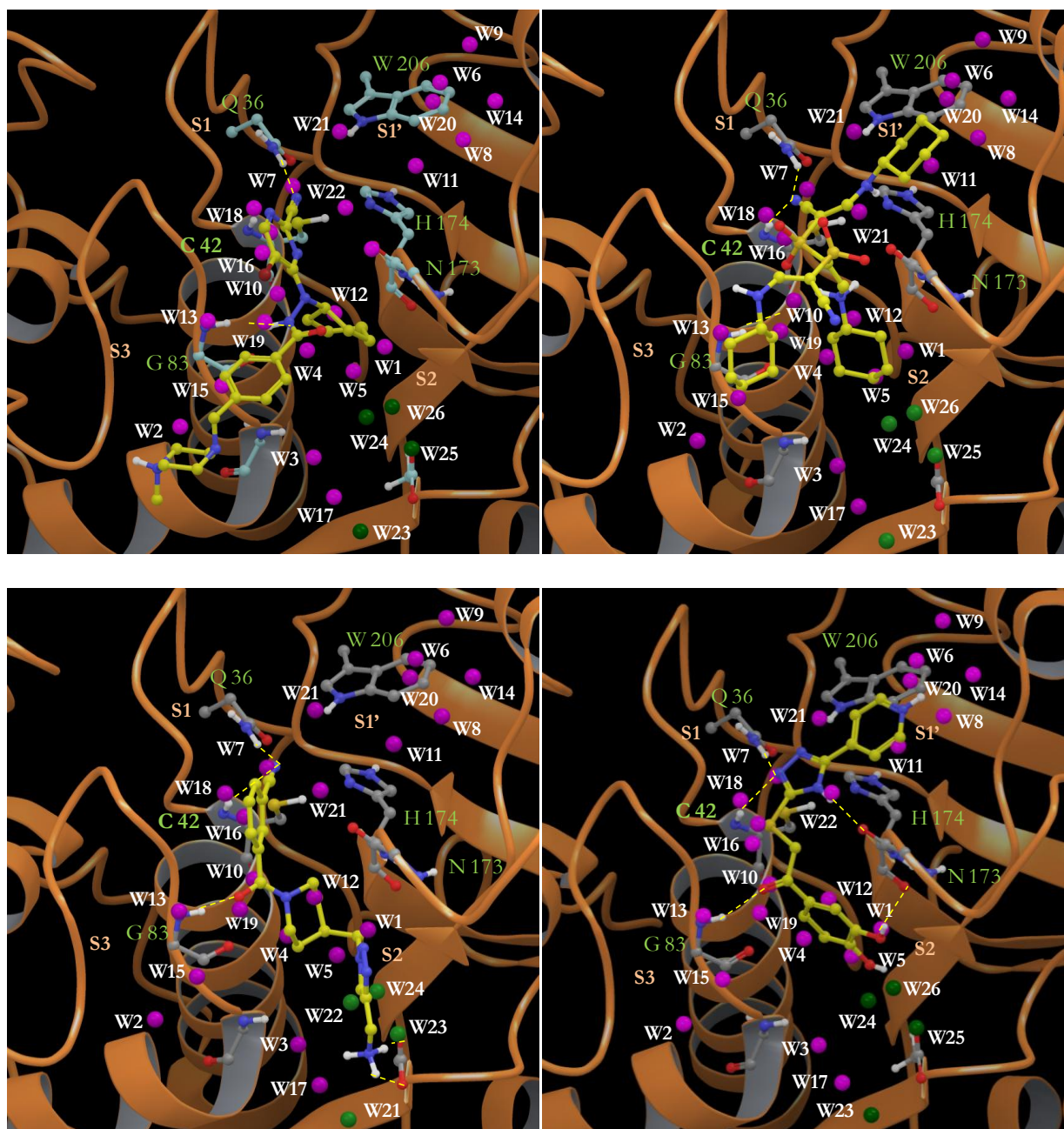


Figure 4.3.13. The predicted binding poses and thermodynamic hydration site profiles of (a) CN-1 (b) CN-2; (c) CN-3, and (d) 30 in complex with FP-2. Water site labeling and color codes are identical to those shown in Figure 3.

4.4. Conclusions

A combined ligand- and structure-based virtual screening was performed against the Asinex and Chembridge databases to discover non-peptidic inhibitors of FP-2. The core structures of previously discovered virtual screenings hits were used as a query to mine commercial databases for structural analogs. A total of 69 putative hits with diverse R-groups, prioritized based on docking calculations, were evaluated against FP-2. As a result, 28 compounds inhibited FP-2, with IC_{50} s of 5-48 μ M. Some of these compounds were also active against cultured malaria parasites, with IC_{50} s <10 μ M.

The complex SAR of these congeneric series of compounds was difficult to explain using traditional computational descriptors like hydrogen bonds, van der Waals energy, electrostatic interactions, and ligand strain. However, we found that accounting for the thermodynamic characteristics of explicit solvent in the ligand binding domain of FP-2 and FP-3 provided an explanation for most of the SAR. The hydration site thermodynamics revealed multiple unfavorable hydration sites, most notably in the S2 pocket of FP-2 and FP-3. Modeling predicted that the leucine moiety at the P2 position of the co-crystallized inhibitors E-64 and vinyl sulfone k11017 in the S2 pockets of FP-2 and FP-3 complemented the arrangement of these hydrophobically enclosed waters and, as a result, displaced them from the S2 pocket of both FPs for increased binding affinity. This also reflected the preference for a P2-Leu as a recognition element for FP inhibition. The most potent hits generally displaced the least favorable hydration sites; however, a majority of the identified hits were unable to displace the least favorable water of the S2 pocket (in a similar pattern to leucine), which accounts for the moderate affinity observed for most of these compounds.

In a few cases, displacement of the least favorable hydration sites did not correlate with the

binding affinity. This could be explained based on the poor reactivity of soft electrophiles toward the catalytic cysteine in these weak hits. We evaluated moderately electrophilic alkyl and aryl nitrile inhibitors against FP-2 to further highlight the effect of the reactivity on cysteine protease inhibition. The aryl and alkyl nitriles only showed moderate affinity (10-20 μM) towards FP-2 as opposed to compounds with strong electrophiles such as pyrimidine nitriles, which have sub-nanomolar affinity.

The present study highlights three important factors that appear to maximize inhibition of FP: a) strong hydrogen bond networks with the key residues of the FP-2 active site such as Gly83, Cys42, Gln36 and Asn173; b) displacement of unfavorable hydration sites (in particularly the triangle of unstable waters from the S2 pocket) by complementary groups of the ligands, and c) chemical reactivity of electrophiles to the catalytic cysteine of the active site. In general, computing binding energies is a complex process with many factors involved. The work presented here shows that terms that are often neglected from scoring functions (i.e. explicit water energies and chemical reactivity) can play an important role in binding. While the chemical reactivity of a ligand is likely to be most important in cases of covalent bond formation, as was the case for the FPs studied here, the hydration site thermodynamics are expected to be important in most ligand binding processes. More work is needed to develop a unified scoring function that can simultaneously account for all of these terms.

Chapter 5

Understanding Binding Affinity of α -Keto Substituted Peptidomimetics for the Inhibition of *Plasmodium falciparum*

Part of this chapter will be submitted for publication: Weldon, D., **Shah F.**, Chittiboyina, A., Jiri, G., Rosenthal, P., Shivakumar, D., Sherman W., Desai, P., Jung J. C., Avery M. A. Design, Synthesis, and Biological Results of α -Keto Substituted Peptidomimetics for the Inhibition of *Plasmodium falciparum*.

5.1. Introduction

In our parallel efforts for drug discovery against FP-2, several peptidomimetics were designed and synthesized taking an advantage of a published crystal structure for cruzain, a related cysteine protease from *Trypanosoma cruzi*. Docking of compound **1** into a homology model of falcipain-2 developed in our laboratory²⁵²

revealed that inhibitors of cruzain would be good templates for the design of inhibitors against falcipains. Therefore, a congeneric series of compounds utilizing **1** as a template were designed. The initial design **2** and its

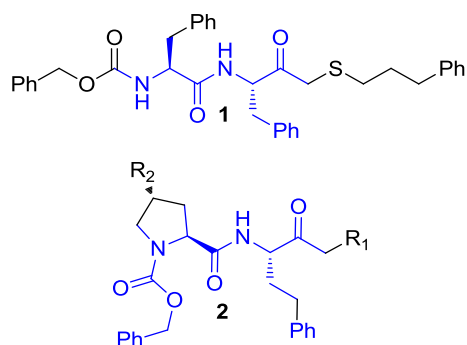


Figure 5.1.1. Overlap of cruzain inhibitor **1** with design template **2**.

overlap with **1** are shown in blue in Figure 5.1.1. Of particular note, one of the phenylalanine residues in **1** has been replaced with a homophenylalanine in **2**. In addition, the other Cbz-protected phenylalanine has been replaced with a Cbz-protected proline. These changes from natural to unnatural amino acids were incorporated to prevent peptidases from cleaving the compound, thereby increasing the compound half-life, and also to increase the rigidity of the falcipain inhibitor in order to improve its cell permeability. Docking studies of these compounds in a homology model of falcipain-2 suggested that they could closely mimic the binding modes of non-covalent inhibitor **1** and might form important hydrogen bonds, similar to those seen in the co-crystal structures of **1**-like compounds in complex with cruzain¹⁰. Thirty-six peptidomimetics were designed and synthesized (Figure 5.1.2, compounds **15-50**) in our laboratory considering the prime and non-prime substrate sites of falcipain-2 to elucidate the

SAR and evaluated for activity against falcipain-2, falcipain-3, and cultured malaria parasites (Table 5.1).

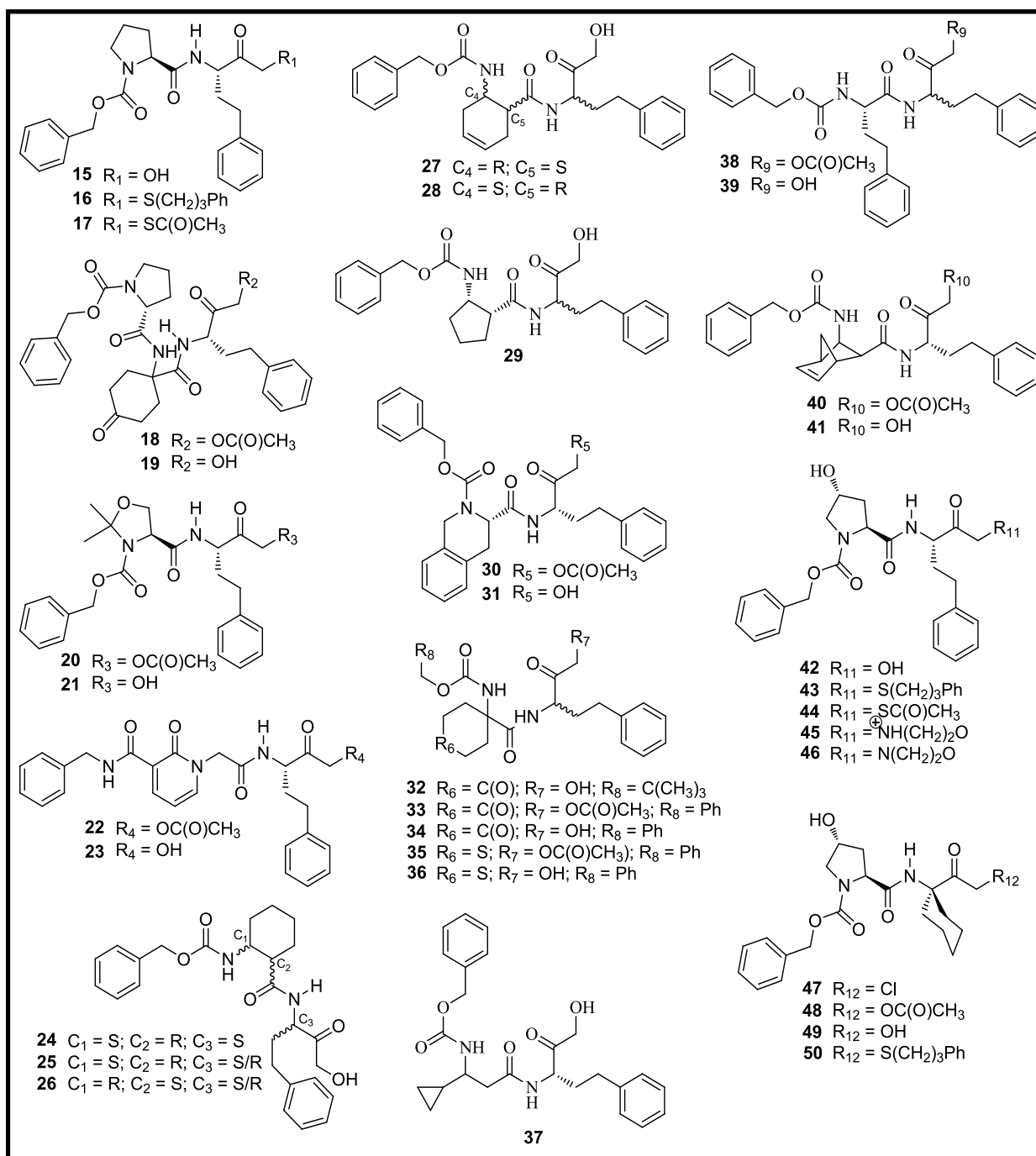


Figure 5.1.2. Chemical structures of synthesized peptidomimetics 15-50.

Table 5.1. Activity of compounds **15-50** against Falcipain-2 (FP-2), Falcipain -3 (FP-3) and cultured malaria parasites. All values shown are IC₅₀s in μ M.

Series	Compound	FP-2	FP-3	W2
1	15	112.22	357.06	>50
	16	48.13	149.01	13.77
	17	154.23	36.36	18.12
	18	>250	>250	>50
	19	72.81	5.43	21.57
2	20	235	>250	>50
	21	170	>250	>50
	22	192	>250	>50
	23	168	47.46	>50
	24	5.52	0.54	>50
	25	7.32	1.86	32
	26	4.42	0.91	47.6
3	27	33.74	5.88	>50
	28	61.64	20.18	>50
	29	80.71	23.90	>50
	30	>250	>250	>50
	31	>250	>250	>50
	32	36.63	37.47	27.5
	33	>250,000	>250000	>50
	34	69.17	47.23	23.29
	35	33.86	73.98	>50
	36	3.25	8.22	28.16
	37	16.54	11.53	>50
	38	29.67	131.13	>50
	39	30.08	11.91	37.21
	40	87.23	>250	>50
	41	33.64	25.44	47.21
4	42	0.08	0.06	7.70
	43	1.10	0.52	19.64
	44	65.3	3.56	18.64
	45	232	223	10.16
	46	>250	>250	41.51
	47	>250	>250	19.48
	48	>250	>250	>50
	49	>250	>250	>50
	50	>250	>250	19.92

From the analysis of experimental SAR, it was evident that hydroxyl proline in the S2 pocket, an α -hydroxy ketone electrophile in the S1-S1' pocket, and homophenylalanine (hPhe) in the S1 pocket were suitable substituents for high binding affinity peptidomimetics. Efforts to understand the SAR of this series by means of steric and electrostatic interactions derived from the docking calculations in falcipains did not provide useful insights. Van der Waals energy, electrostatics, hydrogen bonds, or DockingScore from Glide SP^{253, 254} docking calculations did not reveal any correlation with experimental affinity. Furthermore, implicit solvent binding energy estimations from MM-GBSA (as implemented in the program Prime^{255, 256}) did not show a significant correlation with experimental affinity. In light of the above findings, we anticipated the involvement of explicit water molecules in the binding of these inhibitors and therefore generated thermodynamic profiles of waters present in the LBD of falcipain-2 and falcipain-3 using WaterMap.

5.2. Methods

5.2.1. WaterMap Calculations: Thermodynamic characterization of water molecules in the falcipain ligand binding domain were carried out with WaterMap, which computes the location and energetics of water molecules around protein using explicit solvent MD simulation, solvent clustering, and statistical thermodynamics.^{234, 235} In short, the simulation consists of an initial equilibration followed by 2 ns of production molecular dynamics. Water molecules from a collection of snapshots taken every 1.2 ps were clustered into “hydration sites”. The enthalpy of the waters within each hydration site was computed using the non-bonded interaction energy with the OPLS_2005 force field.²⁴⁵ The entropy was computed using a local expansion of the spatial and orientational correlation functions as described in inhomogeneous solvation theory^{243,}

²⁴⁴. WaterMap has been effectively applied to a broad range of pharmaceutically relevant targets, including PDZ domains²⁵¹, kinases²⁴⁰, G-protein coupled receptors (GPCRs)²³⁹, protein-protein interaction interfaces²⁵⁷, and serine proteases²³⁸.

WaterMap analysis was carried out for falcipain-2 in complex with the epoxy succinate inhibitor E-64 (3BPF, PDB code)⁵⁰ and with falcipain -3 in complex with the vinyl sulfone inhibitor k11017 (3BWK, PDB code)²¹⁶. Structures were prepared using Protein Preparation Wizard of Maestro (Schrodinger, LLC, Portland, OR). Protonation states of catalytic His 174 in falcipain -2 and His 183 in falcipain -3 were adjusted to HIE (proton on the epsilon nitrogen). Finally, proteins were minimized with a harmonic positional restraint using the Impref module from Impact to prepare the final systems for the explicit MD simulation. All waters within 5 Å of the co-crystallized inhibitor were considered for the WaterMap calculations.

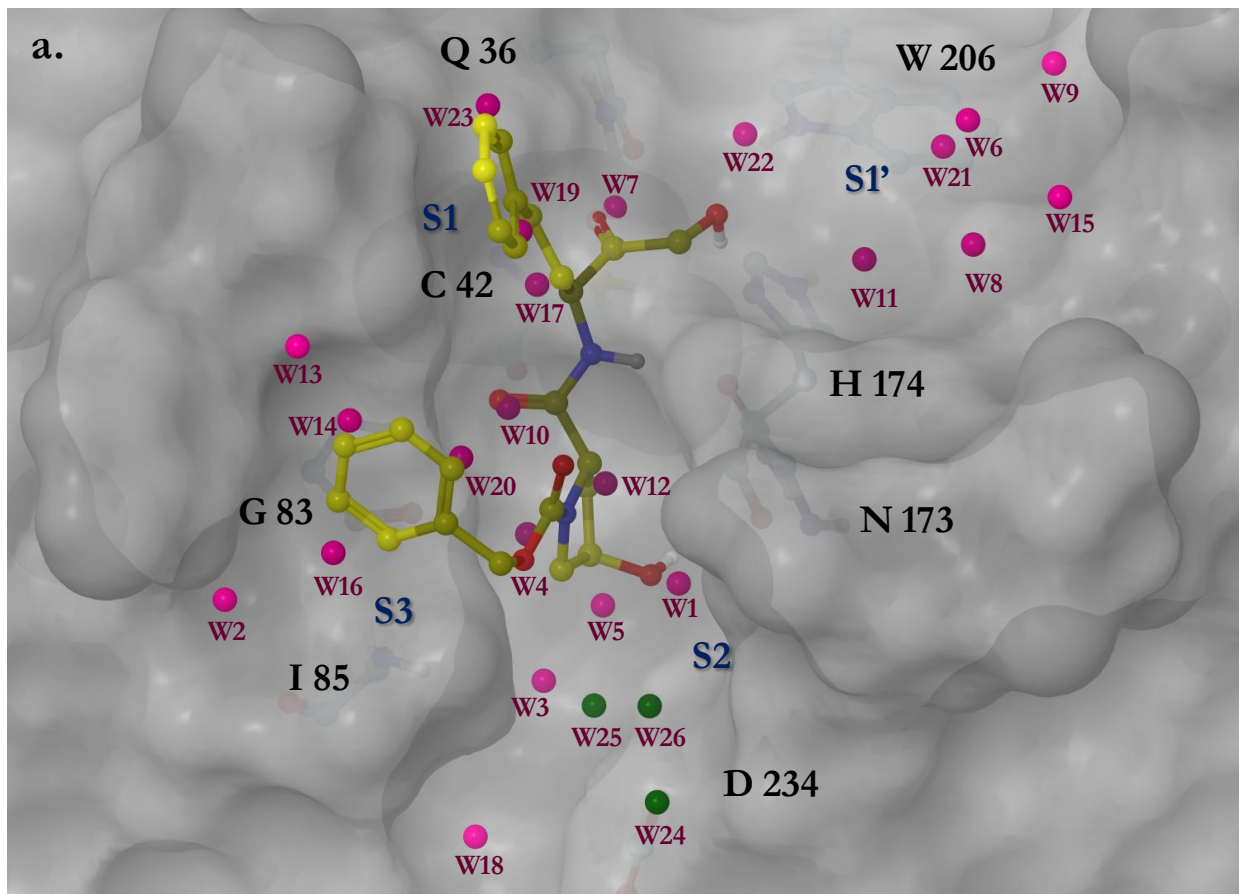
5.2.2. Calculation of LUMO density using Atomic Fukui Indices. Quantum mechanics calculations were run to determine the effect of electronegative substituents on the reactivity of a series of compounds, with the aim of correlating the activation of a bond with the potential to gain binding potency through enhanced interactions. Atomic Fukui indices are regional descriptors of site-specific reactivity and can be used to predict which atoms in a molecule are most reactive toward electrophilic or nucleophilic attack by analyzing frontier molecular orbitals (FMOs).²¹⁰ We used the program Jaguar¹⁹⁸ to compute Atomic Fukui indices for several compounds in a similar manner as described in Chapter 2. The values were normalized on a scale from 0-1, with 1 being the most reactive.

5.3. Results and Discussion

WaterMap computes the location and thermodynamic properties (enthalpy, entropy, and free

energy) of water clusters (“hydration sites”) using explicit solvent simulations.^{234, 235} WaterMap analysis was performed on the recently published crystal structures of falcipain-2 (3BPF, PDB code) and falcipain -3 (3BWK, PDB code) using default settings. The WaterMap calculations revealed several highly unstable water molecules in the binding domain of falcipains. The thermodynamic values of selected hydration sites in falcipain -2 and falcipain-3 as calculated by WaterMap are shown in Tables 5.3.1 and 5.3.2.

Typically, water molecules with unfavorable free energy ($\Delta G > 1 \text{ kcal mol}^{-1}$) should be displaced by ligand functional groups complementary to the protein active site to gain in the overall binding free energy through release of the water molecule into bulk solvent. Conversely, water molecules with negative free energy values should be considered as structural water molecules that can either be bridged or judiciously displaced with the polar atoms of ligands.²⁵⁸ Waters with negative ΔG also have negative enthalpy (ΔH), because the entropy contribution ($-T\Delta S$) is, by definition, always unfavorable for waters in the binding site relative to bulk solution. Given the above considerations, the emerging SAR trends of peptidomimetics from series 1-4 are explained here in the context of the most active compound **42**.



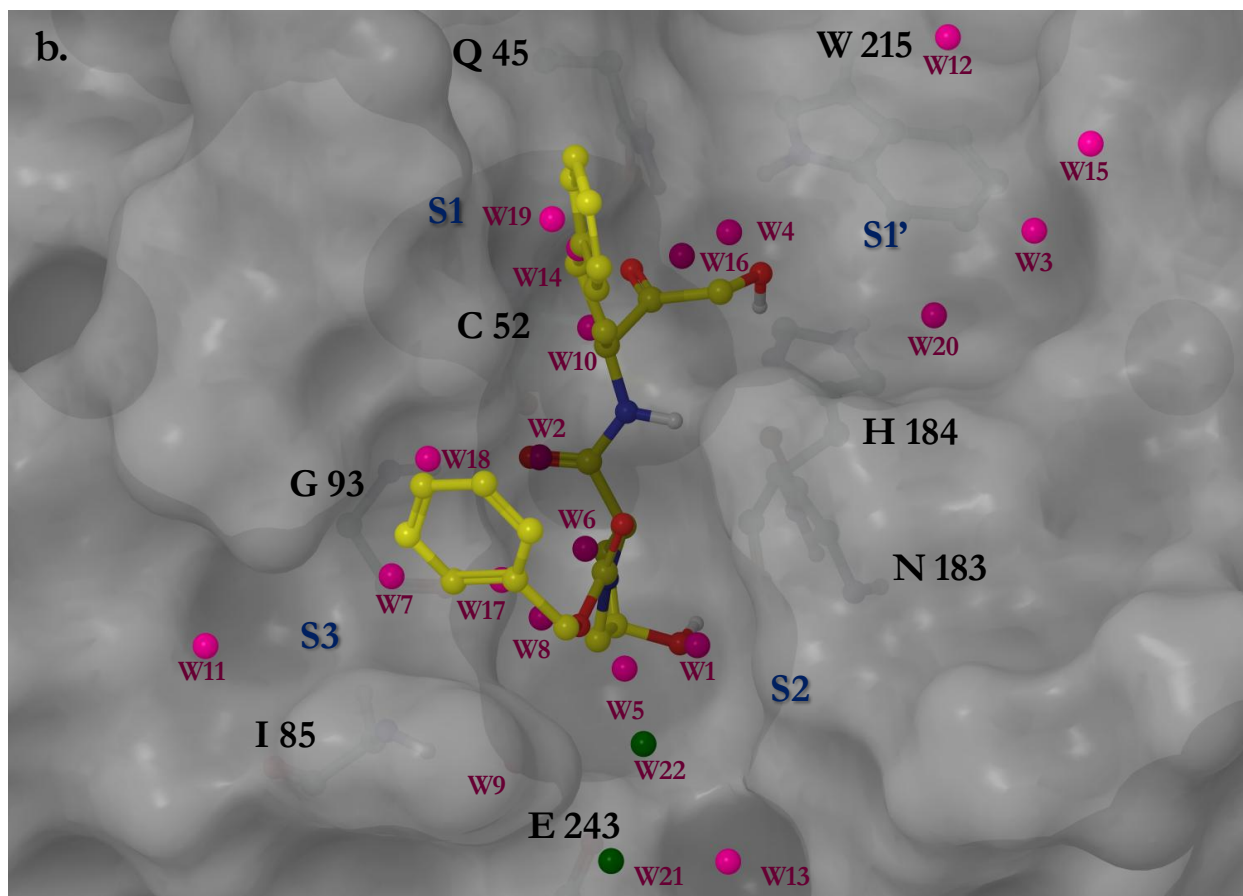


Figure 5.3.1. The WaterMap profile of (a) FP-2 and (b) FP-3 LBD in complex with **42** is shown. The thermodynamically interesting hydration sites important for the SAR are shown in spheres. Stable hydration sites ($\Delta G < 0 \text{ kcal mol}^{-1}$) are colored in green, whereas significantly unstable hydration sites ($\Delta G > 1 \text{ kcal mol}^{-1}$) are shown in purple. The water sites are labeled based on decreasing values of predicted ΔG .

The full thermodynamic profile of falcipain -2 and falcipain-3 LBD in complex with **42** are shown in Figure 5.3.1. The WaterMap hydration sites in falcipain -2 and falcipain-3 binding sites along with compound **42** are shown in Figure 5.3.2. The proline hydroxy group of **42** successfully displaces the most unstable hydration site (W1, $\Delta G = 4.6 \text{ kcal mol}^{-1}$ in falcipain -2; W1, $\Delta G = 4.3 \text{ kcal mol}^{-1}$ in falcipain-3) in the S2 pocket of both falcipains and forms an additional H-bond with the backbone of His174/183 in the active site of the proteases.

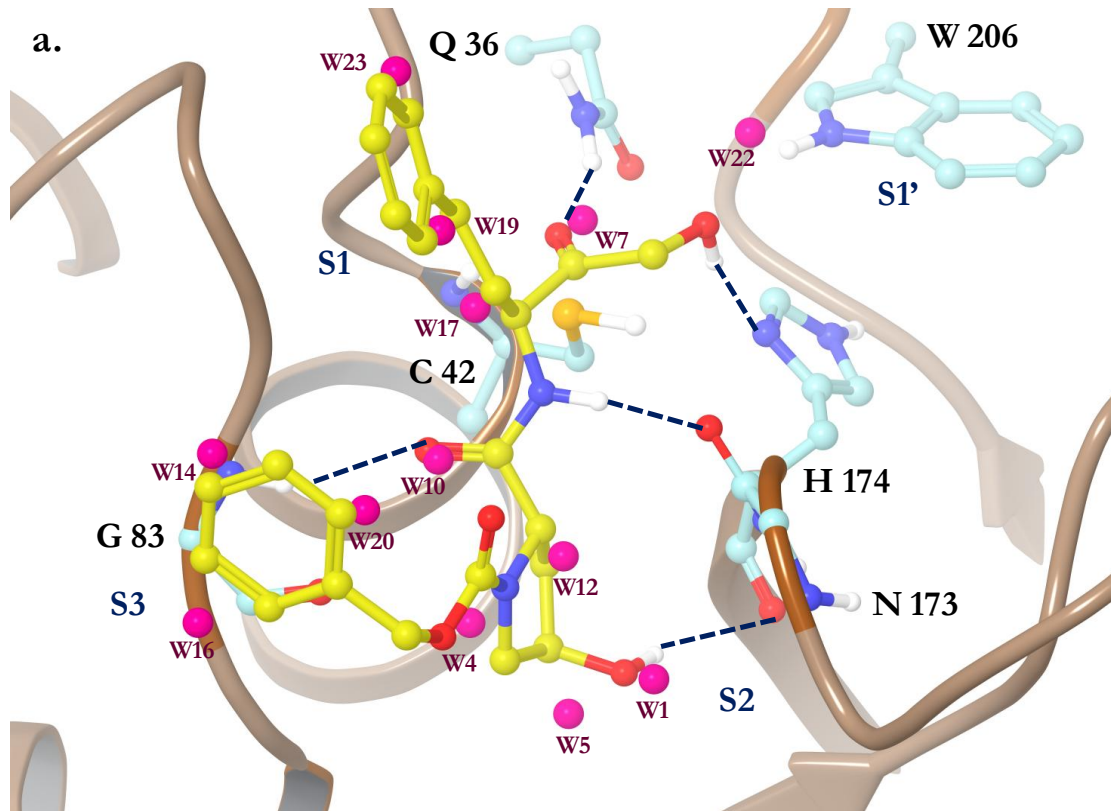
Furthermore, the hydroxy proline moiety displaces two other unfavorable hydration sites (W5 and W12 in falcipain -2; W5 and W6 in falcipain -3) with $\Delta G \geq 2.3 \text{ kcal mol}^{-1}$ from the S2 pockets of both enzymes. The boost in the potency of **42**, as compared to other analogs in this series, is mainly attributed to its proline hydroxyl substituent displacing highly unstable water sites of the S2 pocket. The inactivity of **15** and **23** (Figures 5.3.3a and 5.3.3b) from series 1 and 2, respectively, can be explained by their inability to displace the unfavorable water sites of the S2 pocket. Compound **42** also displaces W10 in falcipain -2 ($\Delta G = 2.6 \text{ kcal mol}^{-1}$) and W2 in falcipain -3 ($\Delta G = 4.2 \text{ kcal mol}^{-1}$), with the replacement of a critical hydrogen bond with Gly83, which is considered important for inhibitors of papain family cysteine proteases binding to the non-prime site²²⁹. Moderate activities of compound **24** (Figure 5.3.3c), **36** **37**, **39**, and **41**(not shown) are due to the formation of a strong H-bond network with Gly83 in addition to the displacement of a few unfavorable water sites ($\Delta G \geq 2.3 \text{ kcal mol}^{-1}$) of the S2 pocket of falcipain-2. The Cbz moiety of **42** also displaces several unfavorable water sites from the S3 pocket of falcipains (W4, W14, W16 and W20 in falcipain -2; W7, W8, W17 and W18 in falcipain -3) into the solvent, and is a suitable substituent for the S3 subsite.

Table 5.3.1. Predicted occupancy and thermodynamic properties of selected WaterMap waters in FP-2

Water Site	Occupancy	ΔH	$-T\Delta S$	ΔG
W1	0.68	2.55	2.01	4.56
W2	0.51	2.73	1.78	4.51
W3	0.79	0.63	3.03	3.66
W4	0.59	1.67	1.86	3.53
W5	0.40	2.22	1.27	3.49
W6	0.86	0.42	2.8	3.22
W7	0.81	0.13	2.83	2.96
W8	0.46	1.32	1.55	2.87
W9	0.97	-1.63	4.35	2.72
W10	0.46	1.28	1.33	2.61
W11	0.35	1.3	1.03	2.33
W12	0.43	0.98	1.33	2.31
W13	0.38	1.18	1.06	2.23
W14	0.33	1.32	0.9	2.22
W15	0.63	0.28	1.78	2.06
W16	0.52	0.56	1.46	2.02
W17	0.54	0.47	1.51	1.98
W18	0.61	0.15	1.73	1.88
W19	0.37	0.67	1.03	1.70
W20	0.50	-0.09	1.43	1.34
W21	0.47	-0.3	1.46	1.16
W22	0.95	-2.25	3.36	1.11
W23	0.50	-0.52	1.55	1.03
W24	0.58	-3.88	1.94	-1.94
W25	0.58	-4.10	1.97	-2.13
W26	0.36	-3.95	1.19	-2.76

Table 5.3.2. Predicted occupancy and thermodynamic properties of selected WaterMap waters in FP-3

Water Site	Occupancy	ΔH	$-T\Delta S$	ΔG
W1	0.43	3.00	1.25	4.25
W2	0.86	1.45	2.77	4.22
W3	0.71	1.62	2.32	3.94
W4	0.96	-0.09	3.91	3.82
W5	0.91	2.50	1.10	3.60
W6	0.45	2.21	1.24	3.45
W7	0.42	2.30	1.15	3.45
W8	0.37	2.27	1.14	3.41
W9	0.84	-0.79	3.28	2.49
W10	0.57	0.62	1.71	2.33
W11	0.51	0.80	1.45	2.25
W12	0.40	1.09	1.09	2.18
W13	0.88	-0.93	2.98	2.05
W14	0.29	0.98	0.84	1.82
W15	0.39	0.45	1.13	1.58
W16	0.43	0.15	1.29	1.44
W17	0.44	0.19	1.24	1.43
W18	0.40	0.28	1.14	1.42
W19	0.31	0.47	0.83	1.30
W20	0.52	-0.36	1.6	1.24
W21	0.61	-2.55	1.89	-0.66
W22	0.48	-2.28	1.37	-0.91



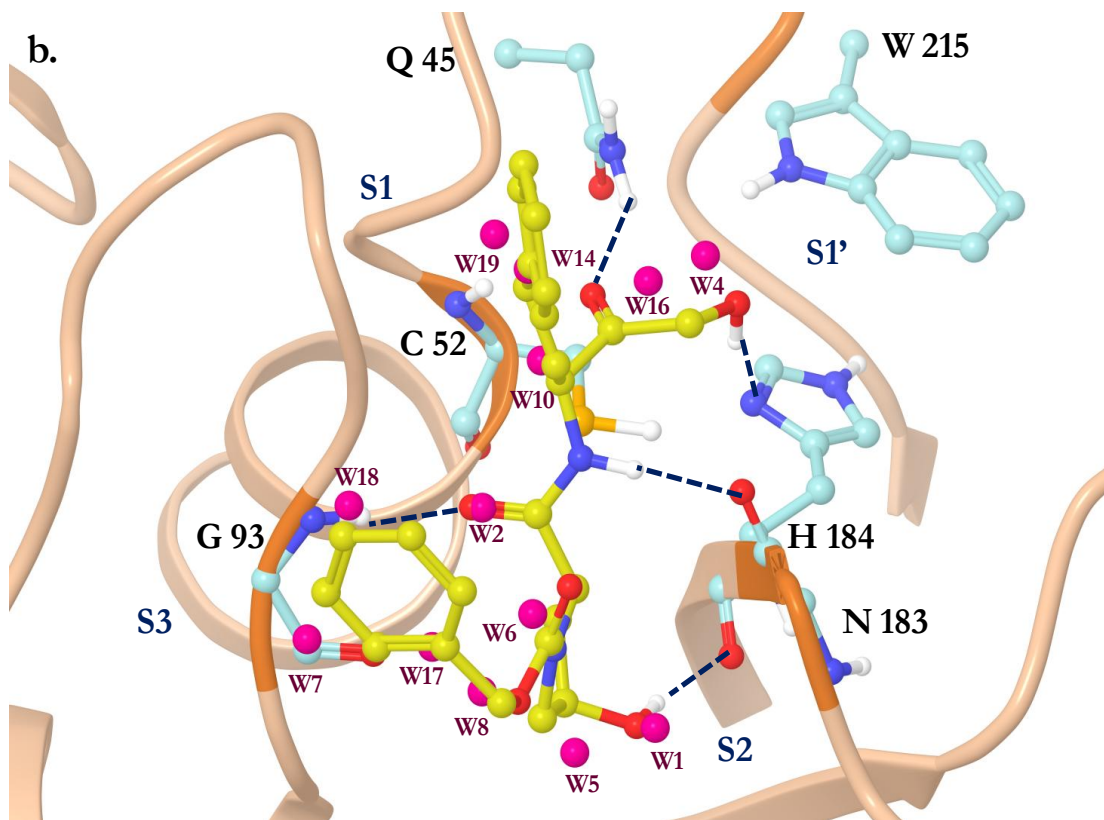
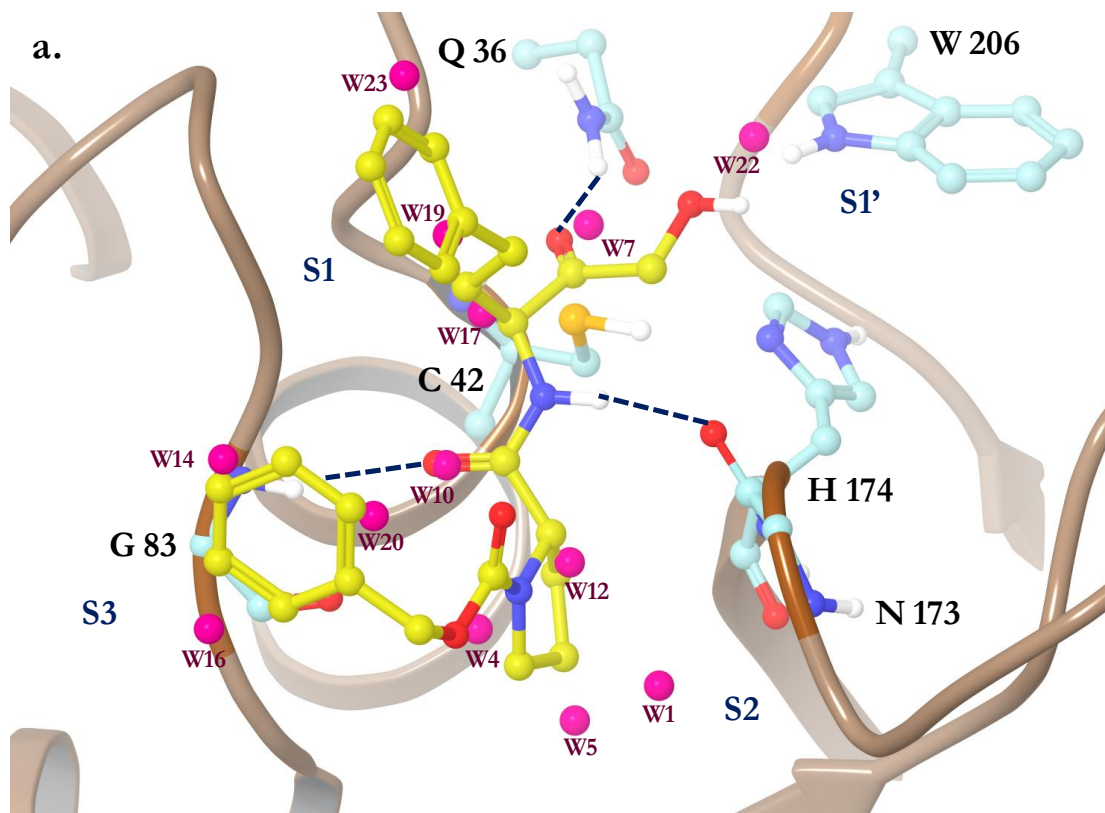
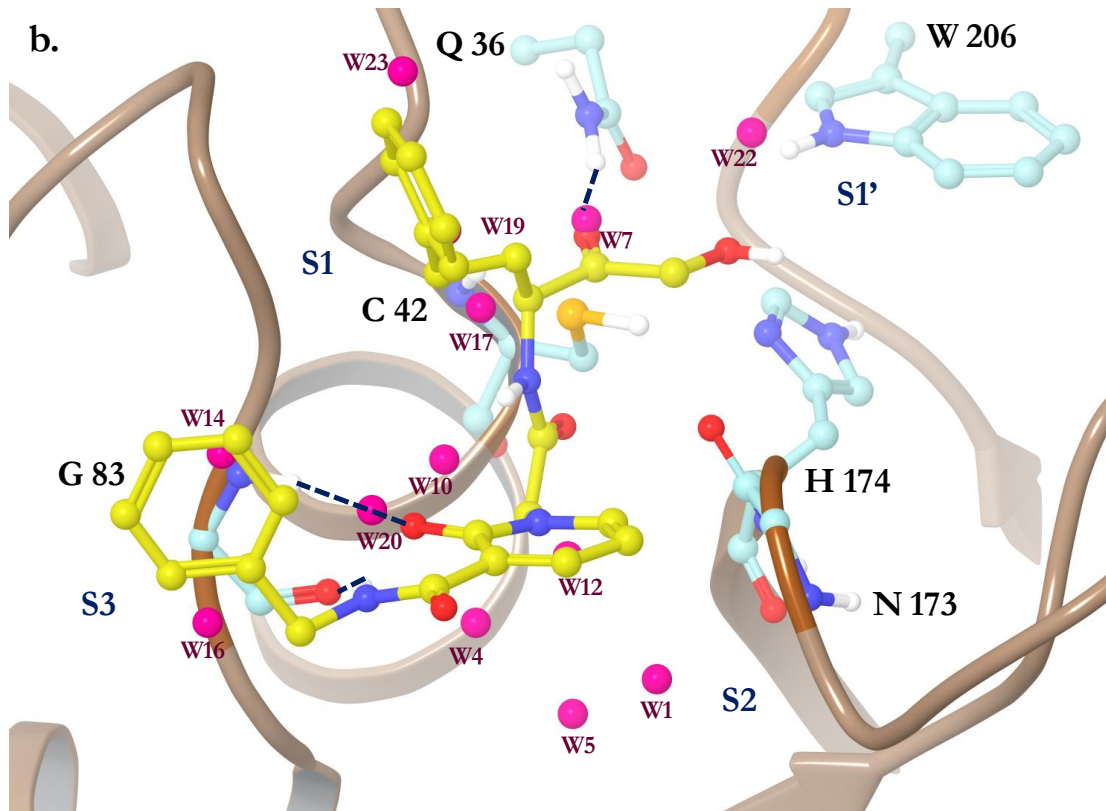


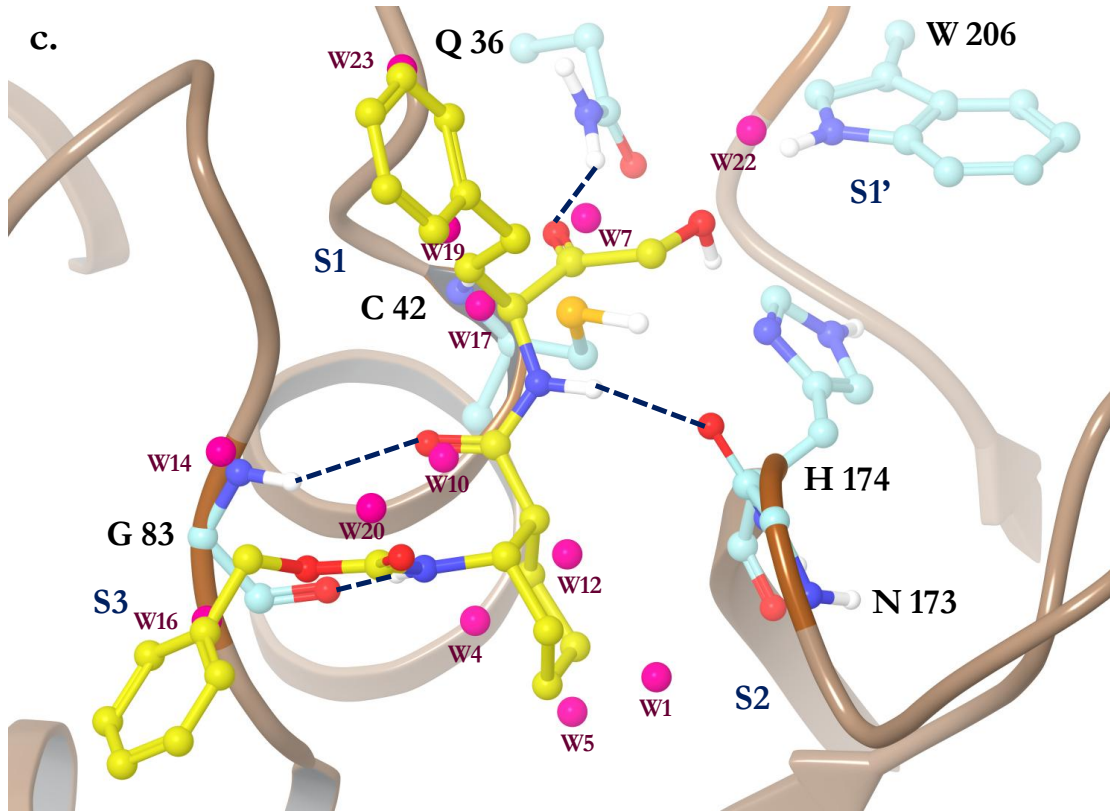
Figure 5.3.2. The WaterMap profile of **42** in (a) falcipain -2 and (b) falcipain -3 is shown. The thermodynamically interesting hydration sites important for SAR are shown in spheres. The unstable hydration sites ($\Delta G > 1$ kcal mol⁻¹) are shown in purple. The water sites are labeled based on decreasing value of predicted ΔG . Key hydrogen bonding interactions of **42** (shown in yellow) with the residues of falcipain binding (shown in cyan) sites are displayed as dotted blue lines.

The carbonyl moiety of the α -hydroxy ketone in **42** displaces the proximal unstable water (W7 in FP-2, $\Delta G = 3.0$ kcal mol⁻¹, W16 in FP-3, $\Delta G = 1.4$ kcal mol⁻¹) from the S1 pocket and replaces its H-bond interaction with the side chain of Gln36/45 and backbone of Cys42/51 in FP-2/3, respectively. The hydroxy group of the α -hydroxy ketone forms an H-bond with His 174/283 in FP-2/3. Additionally, the hydroxy ketone also stabilizes the unfavorable water molecule of the S1' pocket (W22 in FP-2 with $\Delta G = 1.1$ kcal mol⁻¹, W4 in FP-3, $\Delta G = 3.6$ kcal mol⁻¹) by forming a water-mediated H-bond with Trp (206/215) in both enzymes. However, WaterMap did not

provide any additional insights about the preference of α -hydroxy ketone in the S1-S1' pocket over other electrophiles. This might be expected, given that WaterMap only accounts for solvent displacement and not direct interactions made with the receptor. We anticipate that the reactivity of the carbonyl carbon in the vicinity of catalytic cysteine (42/51) with different α -substituents might play an important role in the activity of these compounds (discussed below).







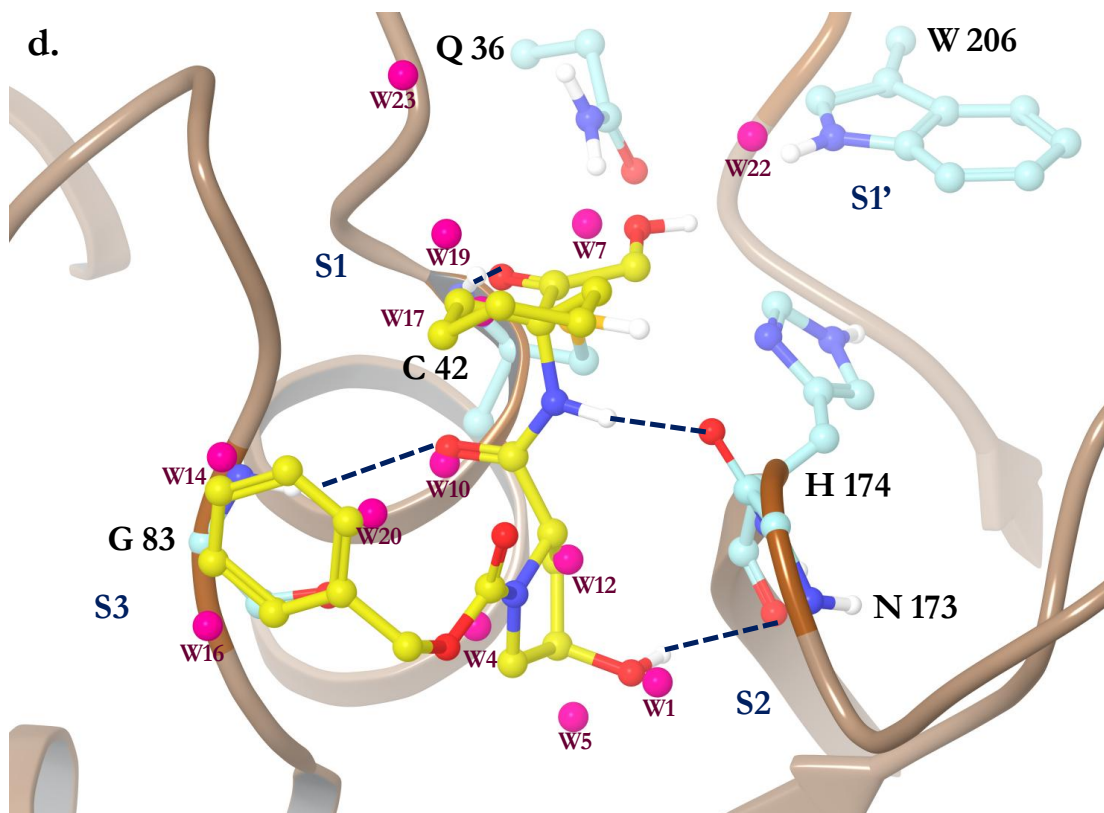


Figure 5.3.3. The predicted binding poses and interesting hydration sites for (a) **15** (b) **23** (c) **24** and (d) **49** in FP-2 are shown. Hydration sites labeling and color codes are same as shown in Figure 5.3.2.

The HPhe in the S1 pocket displaces unfavorable water sites (W19 and W23 in FP-2; W14 and W19 in FP-3) in both enzymes. The replacement of HPhe with a phenyl moiety, a modification that was expected to reduce the flexibility of these peptidomimetics, resulted in the inactive analogs **47-50**. As shown in Figure 5.3.3d, compound **49** is not able to displace unfavorable waters (W19 and W23 in FP-2) into the bulk of the solvent, as observed with HPhe in **42**. In addition, the increased rigidity resulted in unfavorable placement of the α -hydroxy ketone electrophile near the catalytic cysteine in the S1-S1' pocket (Figure 5.3.3d). This observation is further supported by the inactivity of compound **47**, with an α -chloro ketone electrophile,

anticipated to show higher reactivity with the catalytic cysteine.

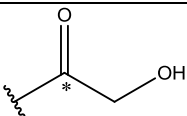
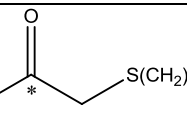
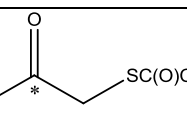
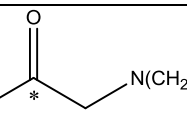
In addition, WaterMap predicts other unfavorable waters in the LBD of FPs that were not assessed in this work. For example, WaterMap predicts unstable water (W3 in FP-2 with $\Delta G = 3.66 \text{ kcal mol}^{-1}$; W9 in FP-3 with $\Delta G = 2.5 \text{ kcal mol}^{-1}$) interacting with the backbone carbonyl of ILE 85/87 lining the S2 pocket of the proteases. W9 is also present in the X-ray structure of FP-3²¹⁶. Moreover, WaterMap shows a stable water (W24-W26 in FP-2 with $\Delta G = <-2 \text{ kcal mol}^{-1}$; W21, W22 in FP-3 with $\Delta G = <-0.7 \text{ kcal mol}^{-1}$) at the bottom of the S2 pocket forming an H-bond with Asp234 in FP-2 and Glu243 in FP-3. These water sites can be either targeted (displacing unstable waters) or interacted with (water-mediated H-bonds) in future work to gain additional binding affinity.

To understand the reactivity of electrophiles with the catalytic cysteine residue and its effect on the inhibitory activity of **42-45**, we calculated the LUMO variant of Atomic Fukui indices at constant spin density using the quantum mechanics package Jaguar, as described previously.²⁵⁹ The congeneric series of compounds **42-45** with different electrophilic groups were considered for these calculations. The LUMO variant of Fukui indices was considered for predicting electrophilicity of the carbonyl carbon. The predicted LUMO values were scaled from zero to one, one being the most reactive, and are shown in Table 2.

The predicted reactivity of electrophiles was significantly higher for the compound with an α -hydroxy ketone (**42**, LUMO = 0.52), than the compound with an α -thio ketone (**43**, LUMO = 0.46; **44**, LUMO = 0.26), which in turn had a higher electrophilicity than the α -amino ketone (**45**, LUMO = 0.01). The activity of **43** with the α -thio ketone might be due to its elongated alkyl

phenyl chain displacing unfavorable waters from the S1' pocket in FP-2 over **42**. The calculated LUMO density values predict correct ordering of the inhibition of FPs for compounds **42-45**.

Table 5.3.3. Predicted LUMO density of the most electrophilic center (shown by asterisk *) in compound **42-45**.

Compound No.	Electrophiles	F_NN_LUMO values
42		0.52
43		0.46
44		0.26
45		0.01

5.4. Conclusions

WaterMap highlighted the putative role played by water molecules in the binding of peptidomimetics in falcipains and provided important insights about the SAR. In particular, WaterMap analysis explained the role played by the hydroxyproline moiety in the potency of **42** based on the displacement of the most unstable predicted waters in the falcipain binding site and the reason a homophenylalanine (HPhe) binds significantly tighter than phenylalanine. Furthermore, the quantum mechanics calculations of Atomic Fukui indices correctly predicted the trend in electronegative substitution effects on potency. The information gained from

WaterMap analysis and Atomic Fukui indices can be used to design the next generation of inhibitors of falcipains. Specifically, we plan to synthesize compounds that are able to displace unstable hydration sites that were not displaced by the compounds presented in this work. This may require developing new chemical scaffolds, as some of the unstable hydration sites are not accessible with the R-group positions in the chemotypes studied here. Further design efforts against FPs will consider the reactivity of electrophiles in addition to the displacement of unfavorable waters, in particular from the S2 pocket of falcipain-2 and falcipain-3.

Chapter 6

Understanding the Binding Affinity of Inhibitors of SARS-3CL^{pro} by Explicit Molecular Dynamic Simulation

Content of this chapter is partly adapted from: Mukherjee, P., Shah, F., Desai, P., Avery M. A. SARS-3CL^{pro}: Virtual Screening, Biological Evaluation and Molecular Dynamics Simulation Studies, *Journal of Chemical Information and Modeling*, **2011**, 51 (6), 1376.

6.1. Introduction

SARS-CoV (Severe acute respiratory syndrome-coronavirus) is a previously unidentified virus belonging to the coronaviridae family which has been recognized as the etiological agent for SARS, a highly infective upper respiratory tract disease. The disease was first diagnosed in the winter of 2002 amongst patients in the remote Guangdong province of southern China. This highly infectious disease with ~10% mortality rate, quickly reached pandemic status, spreading to over 37 countries worldwide and causing over ~9000 infections^{96, 97, 260-262}. Although the initial outbreak of the virus was stymied about an year from its first discovery, recent findings on the isolation of the virus from animals including the Chinese horse shoe bats²⁶³⁻²⁶⁵ which act as natural reservoirs for this virus are suggestive of the possibility of a future animal to human transmission of the SARS-CoV or other closely related corona viruses in the near future. The imminent threat from this disease led to the discovery of a number of viral proteins²⁶⁶ which could be utilized as possible targets for the development of antiviral therapy. Among these targets, SARS-3CL^{pro}, a cysteine protease indispensable to the viral life cycle, has been identified as one of the key therapeutic target against SARS. Chapter 1 describes the detailed overview of SARS-3CL^{pro} functions and various classes of inhibitors developed so far. As a part of our efforts to discover inhibitors against SARS-3CL^{pro}, we previously reported a cascading virtual screening study, involving similarity/substructure searching, docking and ranking based on docking pose based descriptors was carried out against the Asinex Platinum collection in our laboratory^{19, 267}. As a result, numbers of lower micromolar inhibitors of SARS-3CL^{pro} were identified (Table 6.1), the structures of which are shown in Figure 7.1.1 and their docking pose in the binding site of SARS-3CL^{pro} are shown in Figure 6.1. 2.

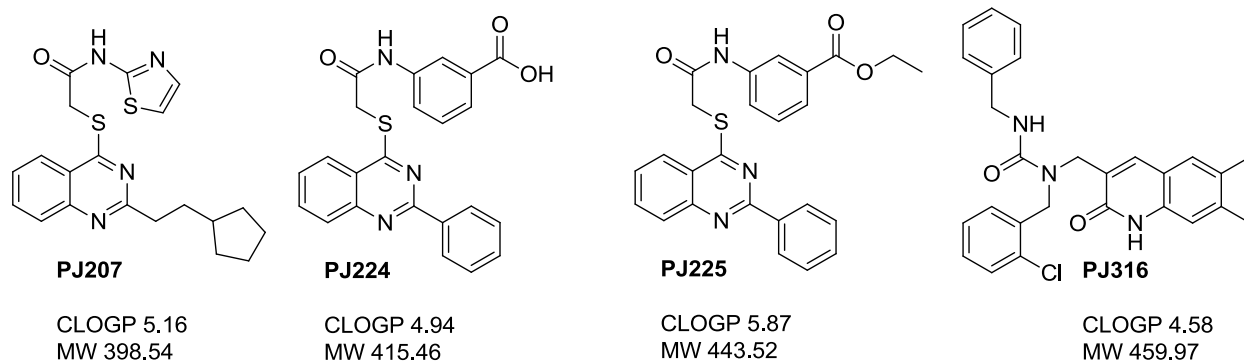


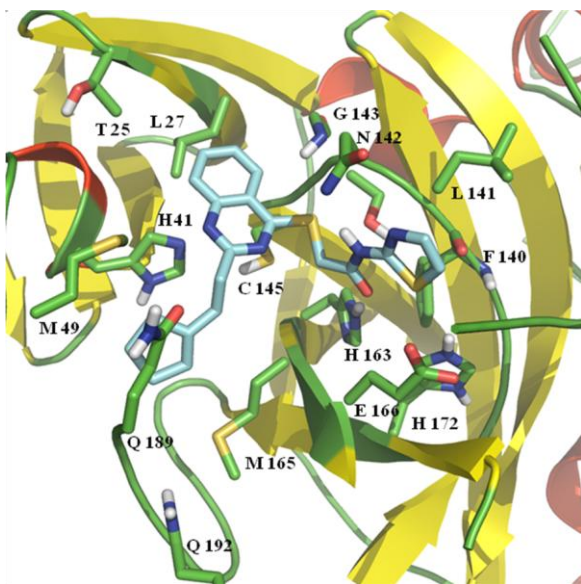
Figure 6.1.1. SARS-3CL^{pro} inhibitors identified by combined structure/ligand-based drug design approach.

Table 6.1. Results from the SARS-3CL^{pro} enzyme inhibition assay

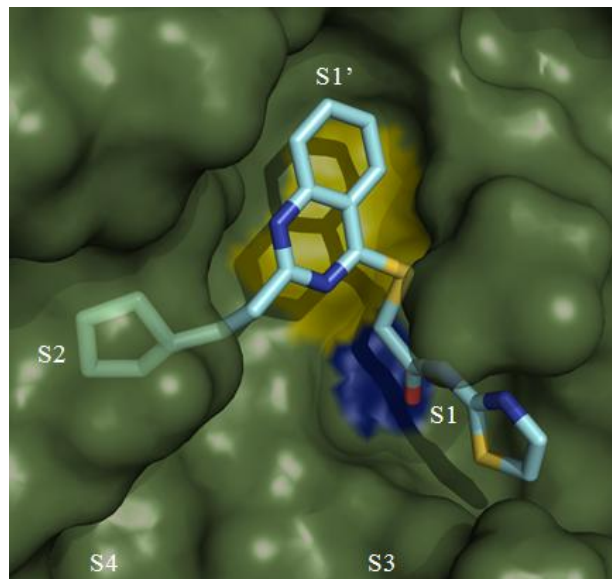
ID	%Inh (10 μ M)	SARS-3CL ^{pro} IC ₅₀ (μ M)
PJ207	25.03	39.4
PJ224	44.93	15
PJ225	61.26	5.8
PJ316	55.69	10.3

The present efforts were initiated to understand the binding affinity of the identified hits and to highlight trends in the SAR of these compounds. Our aim was to incorporate the information gained from binding affinity analysis into future structure-based design of SARS-3CL^{pro} inhibitors. Therefore, we carried out multi nanosecond explicit MD simulation of SARS-3CL^{pro} in complex with **PJ207**, **PJ224**, **PJ225**, **PJ316** to study the overall stability of the interaction profiles predicted from rigid docking studies as well as novel interactions that were not observed during the docking study. The studies revealed several novel interactions that can be considered for future structure-based drug design of SARS-3CL^{pro} inhibitors.

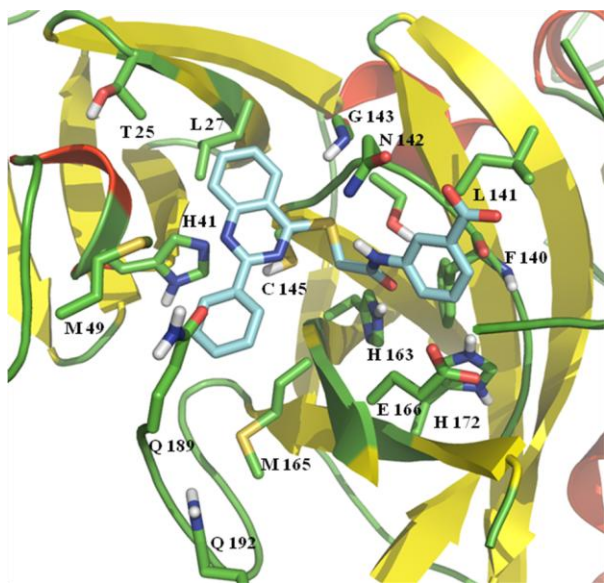
a)



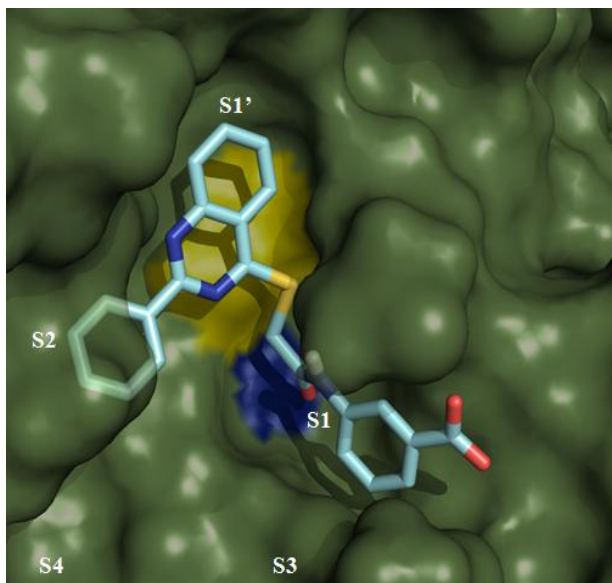
b)



c)



d)



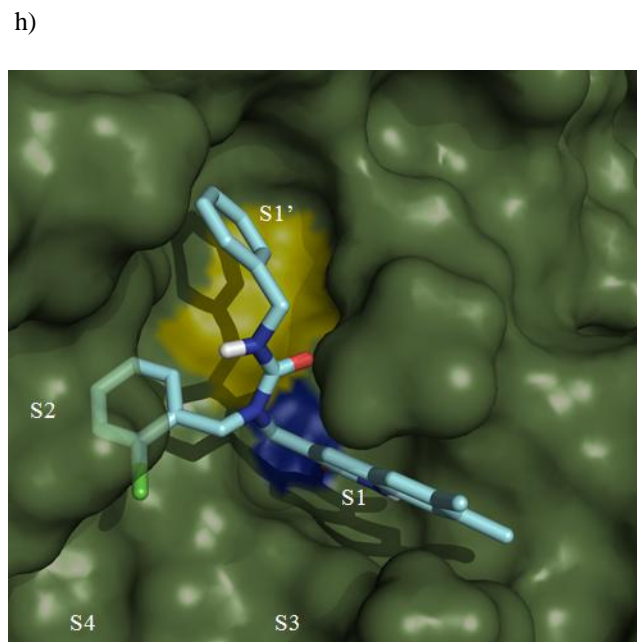
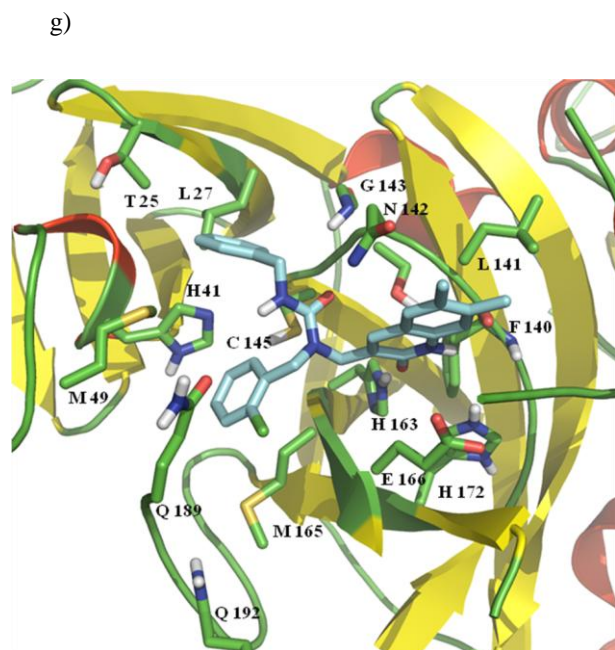
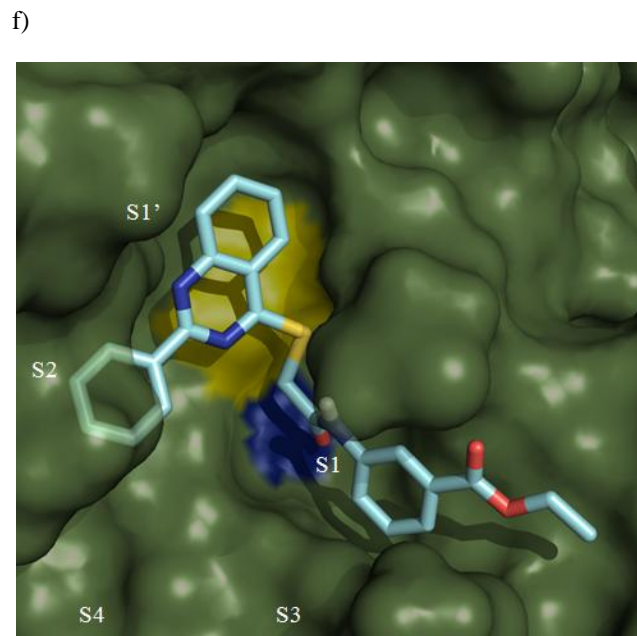
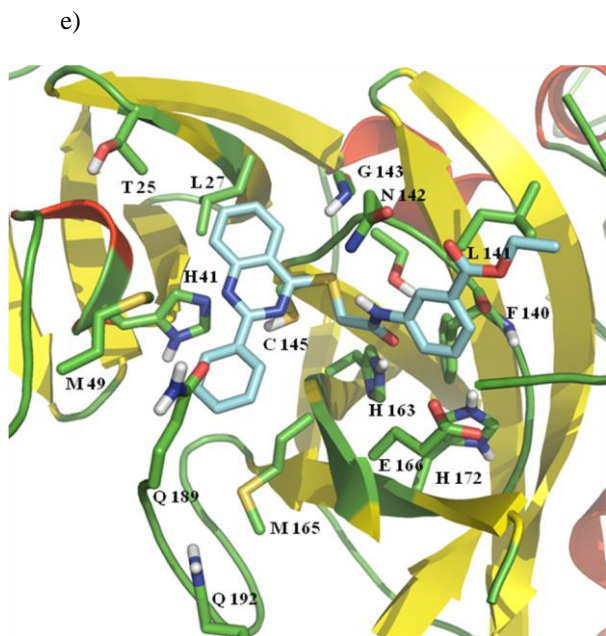


Figure 6.1.2. Interaction profile and surface representation of the poses of the identified hits (a)(b) PJ207, (c)(d) PJ224, (e)(f) PJ225 and (g)(h) PJ316.

6.2. Methods

The docked ligand/SARS-3CL^{pro} complexes of the biological hits (Figure 6.1.2) were used to perform MD simulations with the parallelized Desmond Molecular Dynamics System v2.2 (D. E. Shaw Research, New York, NY) within the maestro environment. The model system for simulation was prepared using the ‘system builder’, a graphics tool in maestro. The OPLS-AA/2005 force field parameters were assigned for all the simulation systems. Each inhibitor-enzyme complex was placed in the center of a cubic box with a dimension of 94.4 Å ensuring 10 Å solvent buffers (~23,400 water molecules) between the solute and the boundary of simulation box in each direction. The SPC model was used to describe the water molecules. Na⁺ and Cl⁻ ions were added at physiological concentration of 0.15 M to ensure the overall neutrality of the systems. Any solvent molecules that overlapped with the solute structures were removed. The final size of the solvated system was close to ~33,000 atoms. The crystallographic waters were retained as part of the simulation system.

The model systems were relaxed using a six step default protocol implemented in Maestro, utilized to prepare systems for production-quality simulation. The relaxation protocol implemented in Maestro involved a series of minimizations followed by a series of short MD runs to relax the model systems. The solvated systems were first minimized for 10 steps of steepest-descent followed by the limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) minimization over a maximum of 2000 steps, keeping all the protein and ligand atoms fixed. The convergence criterion and force constant on the solute was set to 50.0 kcal mol⁻¹ Å⁻¹. In the second step all positional restraints were removed and the system was minimized to a convergence criterion of 5 kcal mol⁻¹ Å⁻¹ using the same sequence as the last step. In the third

step, a 12 ps simulation was performed using the NVT ensemble using a Berendsen thermostat²⁶⁸, a temperature of 10K, a fast temperature relaxation constant of 0.1 ps, velocity rescaling at every 1 ps and restraints on all non-hydrogen solute atoms with a force constant of 50.0 kcal mol⁻¹ Å⁻¹. Following these, two short simulations of 12 and 24 ps were performed using the NPT ensemble using a Berendsen barostat²⁶⁸, maintaining pressure at 1 atm and a slow pressure relaxation constant of 50 ps with all other parameters preserved from the previous NVT simulation. The restraint set from the third step was preserved through the fourth and fifth steps. Temperature was maintained at 10k for the fourth step and then elevated to 300k for the fifth step. In the final stage of the relaxation protocol, all atoms were set free and the simulation was performed using the NPT ensemble for a simulation time of 24 ps using a pressure relaxation constant of 2.0 ps. This final step resembled the conditions of the actual production runs.

The relaxed systems were subjected to a production phase, using the NPT ensemble and periodic boundary conditions for a period of 5 ns. The Martyna-Tobias-Klein²⁶⁹ pressure control method was used to maintain the pressure of the cell at 1.013 bar using the isotropic coupling method. The Nose-Hoover thermostat method was used to control the temperature at 300K. Heavy atom-hydrogen covalent bonds were constrained using the SHAKE algorithm²⁷⁰ which allowed a 2-fs integration step to be used. The integration of the equation of motion as implemented according to the RESPA²⁷¹ multiple time step scheme was used in the simulations. Long-range electrostatic forces were computed using the smooth Particle-Mesh Ewald (PME) method²⁷². The cutoff distance for calculating short-range electrostatics and Lennard-Jones interaction was set to 9.0 Å. The trajectories and the energies were recorded at every 25 and 6 ps respectively. The glue option was turned on for proper display of recorded trajectories and

convenience in structure analysis. The simulation quality analysis tool was used for the analysis of the energetics of the simulations whereas the simulation event analysis tool provided the graphical interface for calculating the root mean square deviations (RMSDs), as well as the H-bond distance, angles and Van der Waals interaction monitors.

6.3 Result and Discussions

Molecular Dynamics simulations were carried out to investigate the overall stability of the predicted docking poses as well as conformational changes of the inhibitors relative to the active site resulting in altered interaction profiles as well as movements of amino acid side chains lining the binding site. The overall stability of the individual simulations was evaluated by plotting the total energy of the system as a function of time (Figure 6.3.1) and was suggestive of energetically stable complexes with maximum fluctuations at < 1% of the total energy of the system.

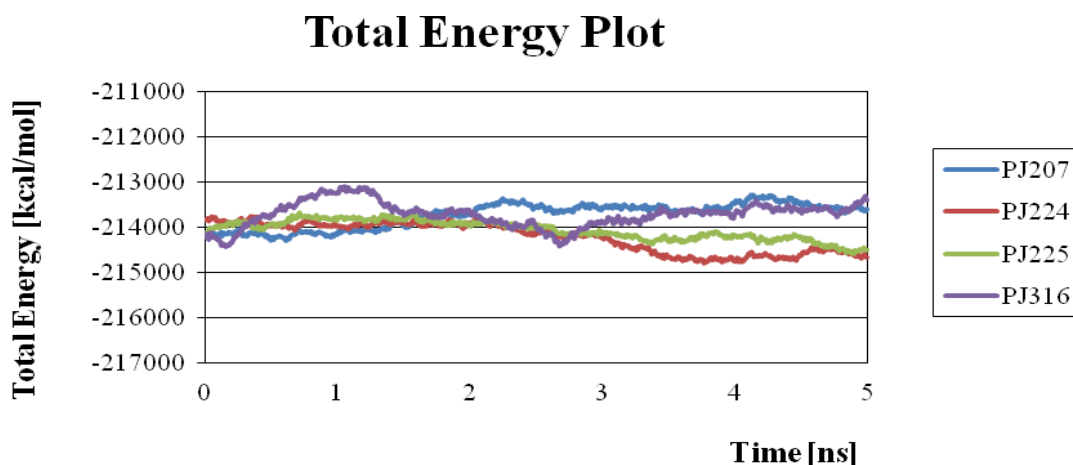


Figure 6.3.1. Potential energy plot for the production phase of dynamic simulation of ligand/SARS-3CL^{pro} complex.

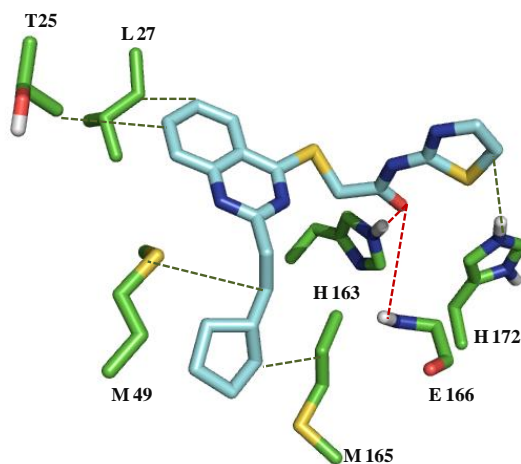
The superposition of the trajectories onto the system conformations at the start of the

production phase helped in the analysis of the RMSD progression over the time of the simulation. The average heavy atom RMSD of the binding site (BS) side chain residues (Thr25, Leu27, His41, Met49, Phe140, Leu141, Asn142, Cys145, His 163, Met 165, Glu166, His172, Asp187, Gln189 and Gln 192) of SARS-3CL^{pro} in complex with **PJ207**, **PJ224**, **PJ225** and **PJ316** were found to be 1.14 Å, 1.41 Å, 1.44 Å and 1.49Å respectively for 5 ns of simulation time and were suggestive of a stable binding site. The average RMSD of the ligands **PJ207**, **PJ224**, **PJ225** and **PJ316** were 0.72 Å, 1.05 Å, 1.0 Å, and 1.48 Å respectively (Figure 7b, 8b, 9b, and 10b) and were suggestive of stable binding poses. The average RMSD of the inhibitors were at the same level (**PJ316**) or lower (**PJ207**, **PJ224**, **PJ225**) than the corresponding average heavy atom RMSD of the binding site (BS) side chain residues, suggesting that the movement of inhibitors in the active site should be relatively smaller as compared to the overall conformational changes of the SARS-3CL^{pro} binding site. The most noticeable side chain movements were shown by Gln189. This residue are solvent exposed and are located in the opening of the catalytic cleft located between domain I and II²⁷³ might explain their positional variability due to inter domain movements.

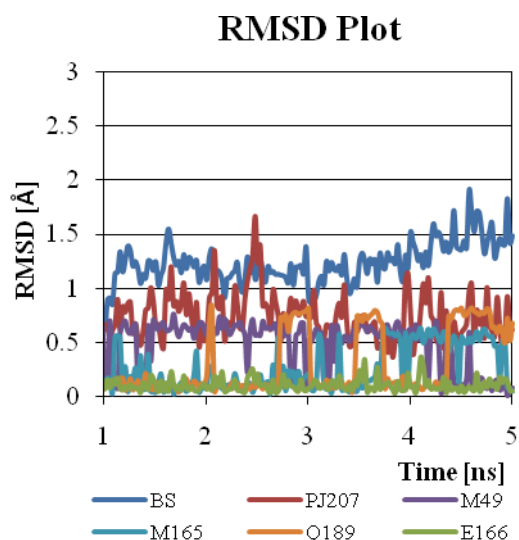
In addition, the MD trajectories of the representative inhibitors were analyzed for the presence of H-bonds as well as hydrophobic interactions between the inhibitors and important residues of the enzyme binding site predicted by the docking studies. Analysis of the MD trajectories indicated the presence of several H-bonds and hydrophobic interactions with moderate to high frequency of occurrence. MD studies also helped to identify new interactions formed over the period of the simulation and suggested the possibility of considering these interactions for the rational designing of the inhibitors against SARS-3CL^{pro}. A D-A (Donor-

Acceptor) distance cut-off of ≤ 3.5 Å and angle (D-H-A) of $\geq 120^\circ$ was used for the identification of H-bonds. For hydrophobic interactions, a cut-off distance of ≤ 4.5 Å between corresponding hydrophobic heavy atoms from the protein and ligand was considered. The interactions of each inhibitor with the active site residues of SARS-3CL^{PRO} are discussed below in detail. The plots of the MD simulation of all inhibitors are graphically shown in Figure 6.3.2.-6.3.5. The distance monitors (Figure 6.3.2a, 6.3.3a, 6.3.4a and 6.3.5a) used for H-bonds and hydrophobic interactions are shown using red and green dotted lines respectively.

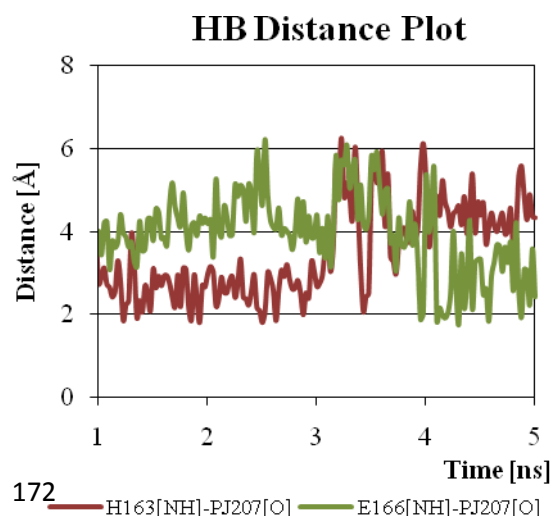
a)



b)



c)



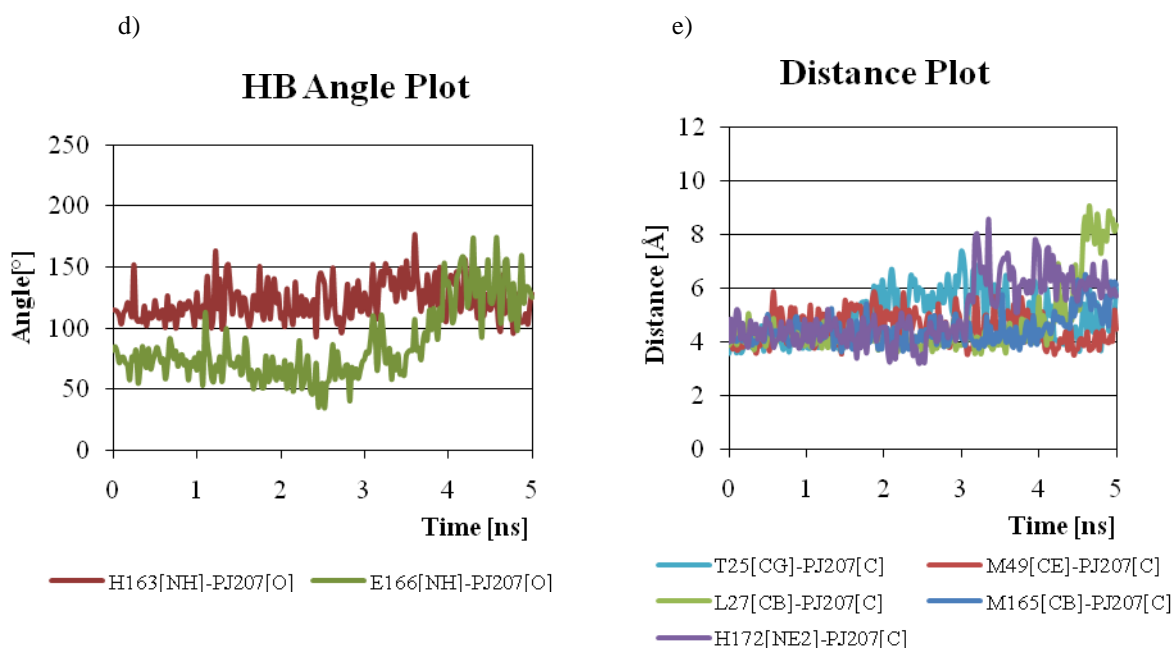
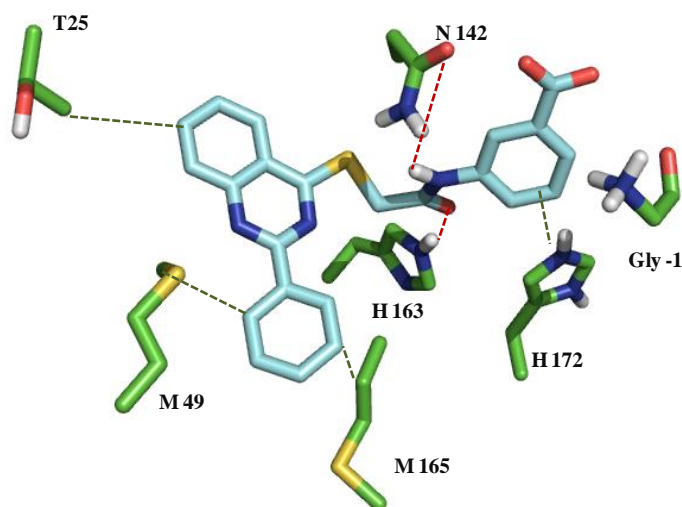


Figure 6.3.2. (a) Docking pose of PJ207 showing the H-bond (red) and hydrophobic (green) monitors used during the simulation; (b) Heavy atom RMSD of the side chain residues of enzyme's binding site (BS), PJ207 and side chain of four residues of BS that showed maximum fluctuations during the simulation; (c) H-bond (H-A) distances and (d) D-H-A angle of His163[-NH]-[O]PJ207 and Glu166[-NH]-[O]PJ207; (e) Hydrophobic [CH-CH] interactions of PJ207 with Thr25, Leu27, Met49 and M165 and a cation(pi)-pi interaction with His172.

The results from the simulation of the **PJ207/SARS-3CL^{Pro}** complex are shown in Figure. 6.3.2. All the H-bonds (Figure 6.3.2 c, d) formed between **PJ207** and residues of the BS were transient in nature. A very interesting observation was made in respect to the critical H-bonding interaction in the S_1 sub pocket. In the 0-3ns of the simulation the H-bond interaction between the NE2 of His 163 from the S_1 subsite of the enzyme and the amide side chain carbonyl of **PJ207** was retained an average D-A distance of 2.69Å and D-H-A angle of 120.1°. Following this the H-bond is lost and was compensated by the formation of a new H-bond between the carbonyl side chain of **PJ207** and backbone -NH of Glu 166 located close to the S_1 pocket with an average D-A distance of 2.85 Å and D-H-A angle of 135° for 4-5 ns of the simulation time (Figure 6.3.2 c, d). The hydrophobic interaction profile (Figure 6.3.2e) showed interactions of

PJ207 with Thr25, Leu27 of the S_1 ' site as well as Met49 from S_2 and Met165 where the corresponding heavy atom distances were retained at ≤ 4.5 Å for 42%, 70.5%, 50%, and 64.5% of the simulation time respectively. A cation(pi)-pi interaction could also be observed between the thiazole group and the His172 in the 0-3ns of the simulation. It's likely, that the relatively polar nature of the aromatic group (compared to the other active analogues) at the end of the S_1 substitution could also contribute to the observed higher movement of the amide side chain. It is also interesting to note that the RMSD fluctuation of Met165 shows a biphasic behavior between 0-3 and 3-5 ns of the simulation showing larger fluctuations in the second phase which corresponds with the H-bond compensation phenomenon observed between the ligand's amide side chain carbonyl and the His163 and Glu166 residues (discussed above). Overall, the more transient nature of the interactions in case of **PJ207** (compared to the other hits) may be indicative of its relatively weaker affinity.

a)



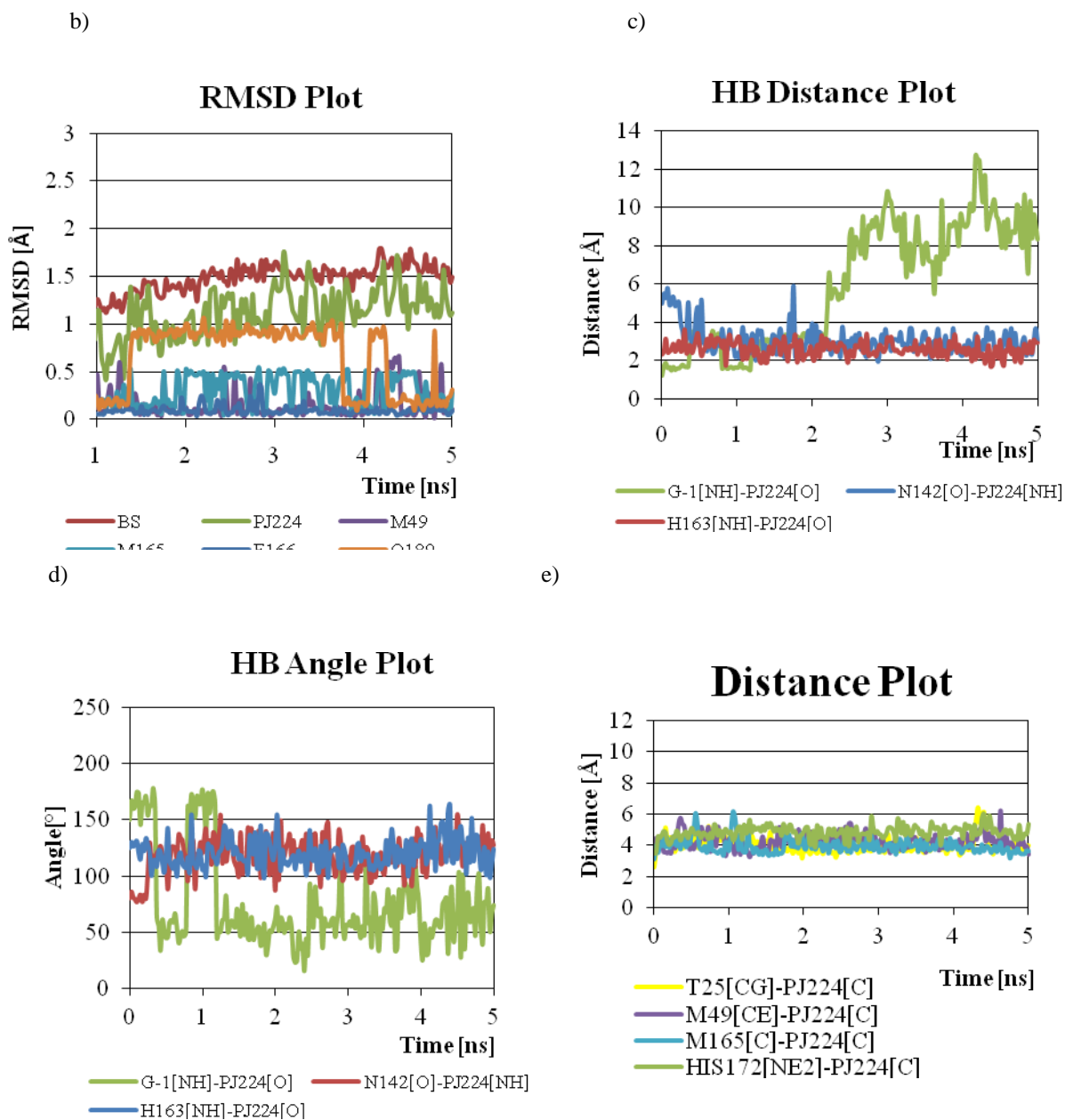
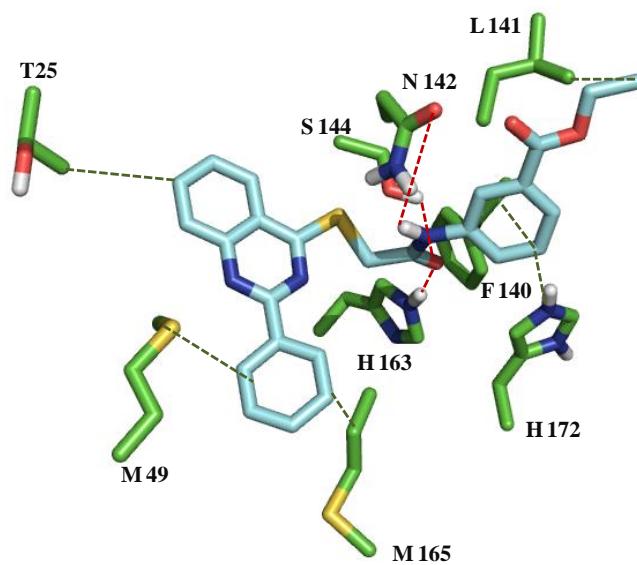


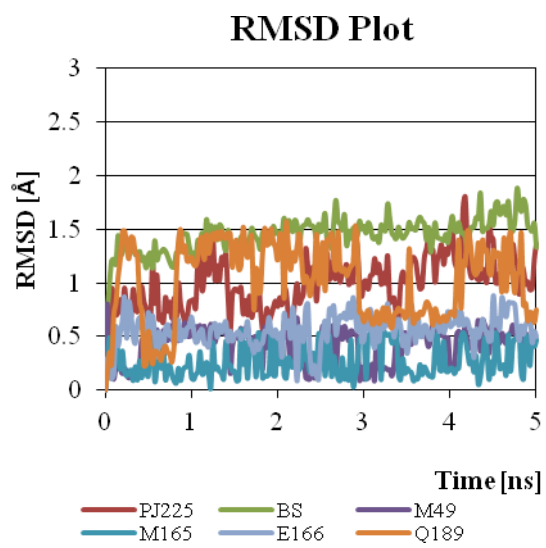
Figure 6.3.3. (a) Docking pose of PJ224 showing the H-bond (red) and hydrophobic (green) monitors used during the simulation; (b) Heavy atom RMSD of the side chain residues of enzyme's binding site (BS), PJ224 and side chain of four residues of BS that showed maximum fluctuations during the simulation; (c) H-bond (H-A) distances and (d) D-H-A angle of Asn142[O]-PJ224[NH], His163[-NH]-[O]PJ224 and Gly-1[NH]-PJ224[O]; (e) Hydrophobic [CH-CH] interactions of PJ224 with Thr25, Met49, and Met165 and a cation(pi)-pi interaction with His172.

The results from the simulation of the **PJ224**/SARS-3CL^{pro} complex are shown in Figure. 6.3.3. As predicted from the docking studies, the amide carbonyl side chain of **PJ224** appeared to form a strong H-bond with the NE2 of His163 with an average D-A distance of 2.61 Å and average D-H-A angle of 124.2° (Figure 6.3.3 c, d). Another H-bond was observed between the side chain -NH of Asn142 and the side chain carbonyl oxygen of **PJ224** with an average D-A distance of 2.9 Å and average D-H-A angle of 120.9° for 0.55-5.0 ns of the simulation time. This interaction was not observed in the static docking pose and pointed towards the importance of the MD simulation in detecting a possible additional interaction. MD simulation also revealed a new H-bond between the carboxylate moiety of **PJ224** and N-terminal Gly-1 (of subunit B) with an average D-A distance of 2.69 Å for the first 2.2 ns of the simulation time which was not observed in the docking pose. The additional H-bond interactions in **PJ224** vis-a-vis **PJ207** may contribute to its higher affinity. Thr25 from the S₁' sub site exhibited strong hydrophobic interactions with the benzpyrimidine scaffold of **PJ224** with the corresponding hydrophobic heavy atom distance at ≤4.5 Å for 88% of the simulation time with an average distance of 4.03 Å. Also, Met165 and Met49 from the S₂ sub site were found to interact with the phenyl side chain of **PJ224** with the corresponding hydrophobic heavy atom distance at ≤4.5 Å for 87% and 71% of the simulation time respectively (Figure 6.3.3 e). A transient cation (pi)-pi interaction was observed between the phenyl group and His172 in the S₁ pocket.

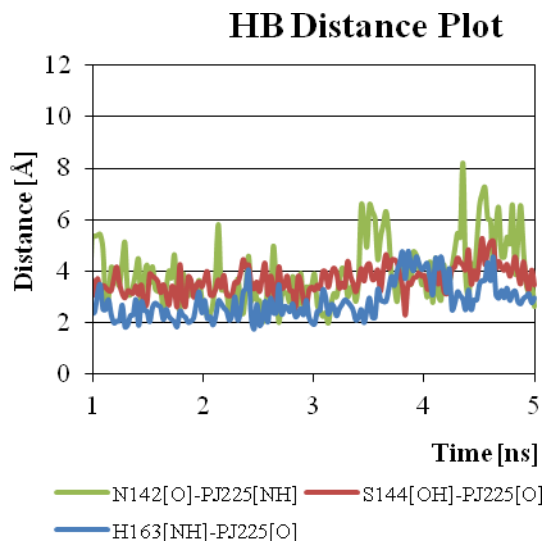
a)



b)



c)



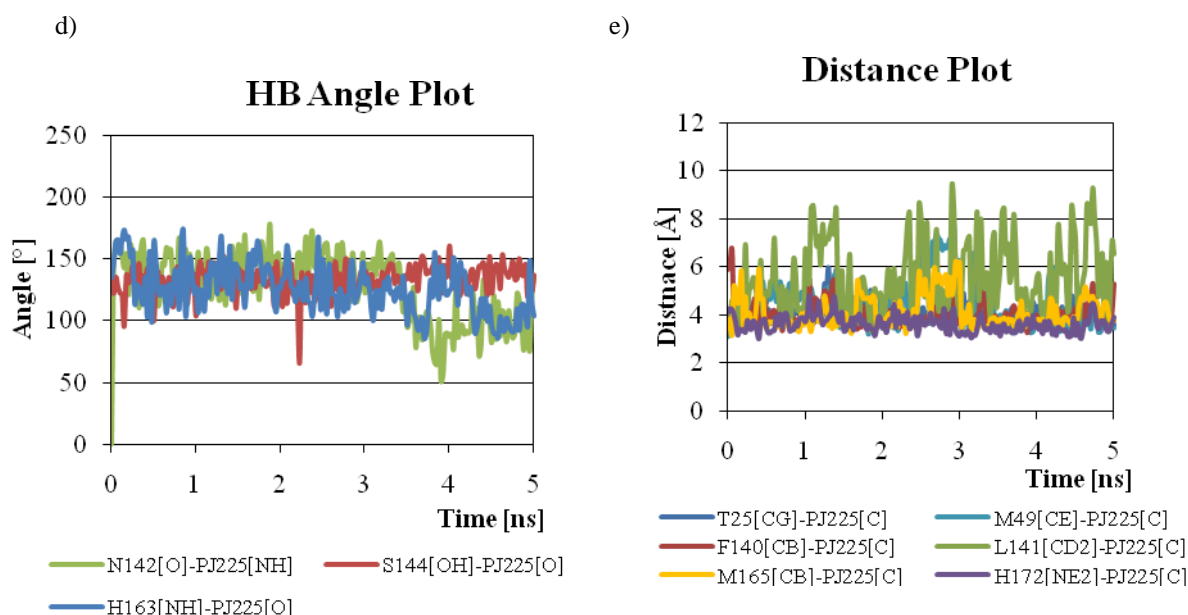
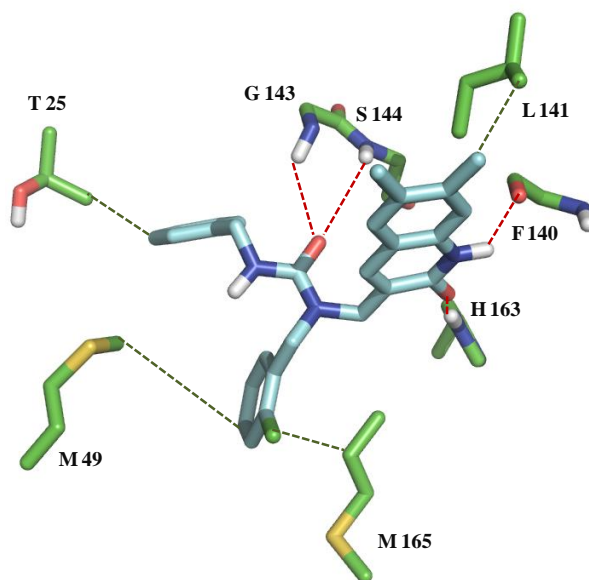


Figure 6.3.4. (a) Docking pose of PJ225 showing the H-bond (red) and hydrophobic (green) monitors used during the simulation; (b) Heavy atom RMSD of the side chain residues of enzyme's binding site (BS), PJ225 and side chain of four residues of BS that showed maximum fluctuations during the simulation; (c) H-bond (H-A) distances and (d) D-H-A angle of Asn142[NH]-PJ225[O], Ser144[OH]-PJ225[O] and His163[NH]-[O]PJ225; (e) Hydrophobic [CH-CH] interactions of PJ225 with Thr25, Met49, Phe140, Leu141 and Met165 and a cation(pi)-pi interaction with His172.

The results of the MD studies for SARS-3CL^{pro} in complex with **PJ225** are shown in Figure 7.3.4. Similar to **PJ224**, a stable H-bond was observed between the side chain amide carbonyl oxygen of **PJ225** with the NE2 of His 163 with an average D-A distance of 2.69 Å and average D-H-A angle of 125.6° (Figure 9 c, d). Also, a transient H-bond was observed between the side chain OH of Ser144 and amide carbonyl oxygen of **PJ225** (D-A distance ≤ 3.5 43.5% of the simulation time). This residue is located on the loop (residues 140-146) involved in the oxyanion hole formation and stabilization of the covalent tetrahedral intermediate generated during substrate hydrolysis by the enzyme. Coincidentally, this interaction was not observed in the

docking pose of **PJ225**. However, the Gly-1 H-bond is not possible in case of **PJ225** and the Asn142 interaction had a lower residence time in comparison to **PJ224**. During the course of the simulation, several strong hydrophobic interactions were observed. These included interactions of **PJ225** with the side chains of Thr25, Leu141, Phe140, Met165 and Met49 where the corresponding hydrophobic heavy atoms were retained at ≤ 4.5 Å for 82.5%, 30.5%, 85.5%, 72% and 59% of the simulation time respectively. As predicted from the docking study, **PJ225** in comparison to **PJ224** and **PJ207** demonstrated additional hydrophobic interactions (Figure 7.3.4 e) between the ethyl moiety of the ligand with Phe140 and Leu141. A highly stable cation(pi)-pi interaction was observed between the phenyl ring and His 172 with the distance ≤ 4.5 Å 100% of the time. These additional and/or stronger interactions in the S_1 pocket observed for **PJ225** could account for its higher inhibitory activity ($IC_{50}=5.8$ μ M) against SARS-3CL^{pro} in comparison to **PJ224** ($IC_{50}=15$ μ M) and **PJ207** ($IC_{50}=39.4$ μ M).

a)



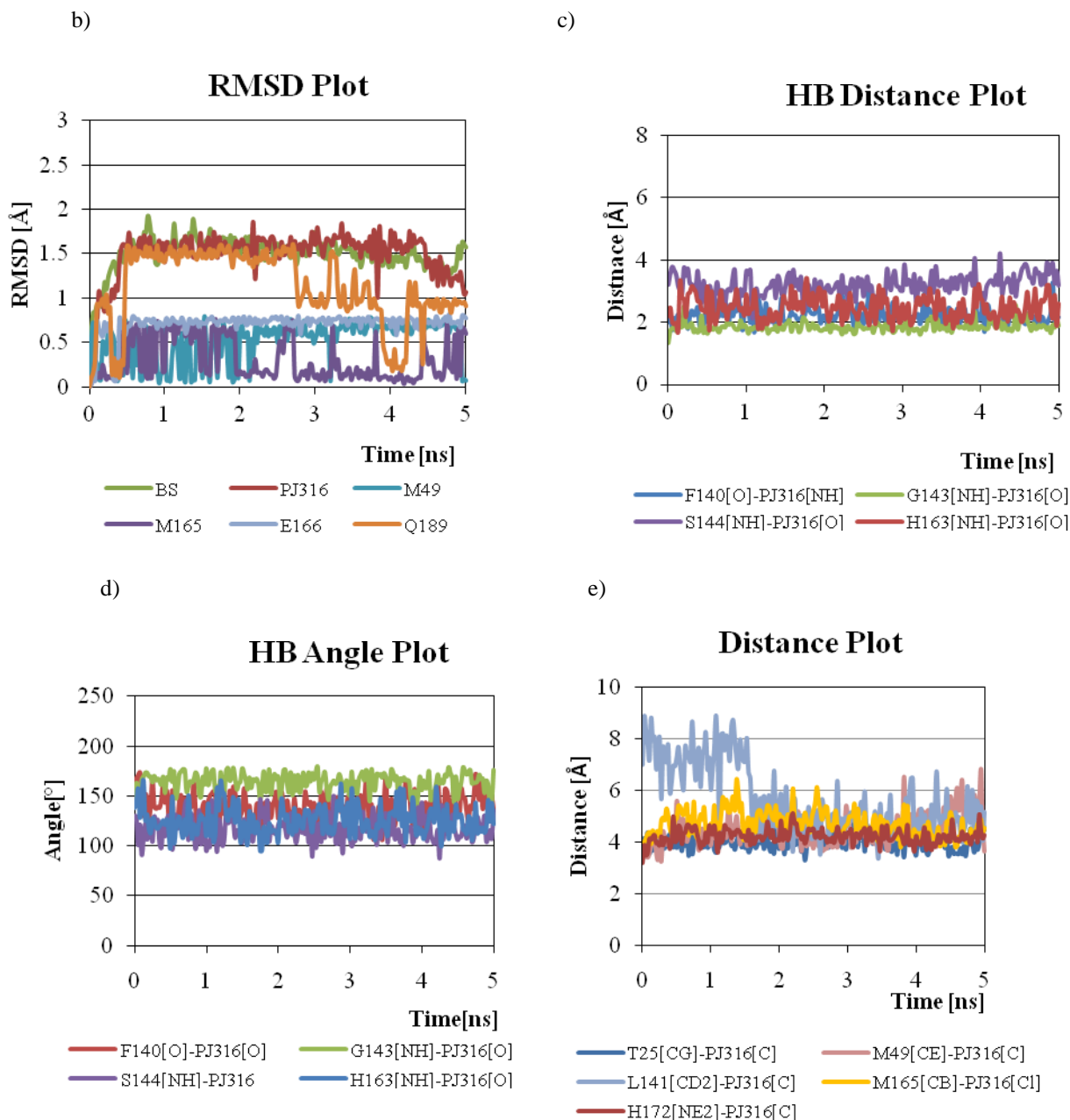


Figure 6.3.5. (a) Docking pose of PJ316 showing the H-bond (red) and hydrophobic (green) monitors used during the simulation; (b) Heavy atom RMSD of the side chain residues of enzyme's binding site (BS), PJ316 and side chain of four residues of BS that showed maximum fluctuations during the simulation; (c) H-bond (H-A) distances and (d) D-H-A angle of Phe140 [O]-[NH]PJ316, G143[-NH]-[CO]PJ316, Ser144 [NH]-[O]PJ316 and H163[-NH]-[O]PJ316; (e) Hydrophobic [CH-CH] interactions of PJ316 with Thr25, Met49, Leu141 and Met165 and a cation(pi)-pi interaction with His172.

The simulation plots for the **PJ316**/SARS-3CL^{pro} complex are shown in Figure 7.3.5. Multiple H-bonds were predicted for **PJ316** based on the docking simulations (Figure 7.3.5 c, d). The amide carbonyl oxygen of the benzpyridone moiety of **PJ316** appeared to form a stable H-bond with NE2 of His 163 (Avg. D-A distance= 2.47Å and Avg. D-H-A angle=125°) of the S₁ pocket. The amide nitrogen of the bicyclic ring of **PJ316** also formed a H-bond with the backbone carbonyl oxygen of Phe140 with an average D-A distance of 2.22 Å and D-H-A angle of 139°. In addition, the amide side chain of **PJ316** formed a bifurcated H-bond with the backbone amide of Gly143 (Avg. D-A distance= 1.83Å and Avg. D-H-A angle=164°) and Ser144 which forms the oxyanion hole to stabilize the covalent tetrahedral intermediate resulting from substrate hydrolysis by the enzyme. The hydrophobic interactions profile for **PJ316**-enzyme complex is shown in Figure 7.3.5e. Also, Thr25 exhibited a strong interaction with the benzyl side chain of **PJ316** with a distance ≤ 4.5 Å for 91.5% of the simulation time and an average distance of 4.48 Å. Met49 exhibited hydrophobic interaction with **PJ316** for 57.5% of the simulation time. In addition, the hydrophobic interactions between the side chain of Met165 with the phenyl side chain of **PJ316**, and Leu141 with the benzpyridone moiety of the ligand falling within the cutoff distance of ≤ 4.5 Å for 40% and 51% of the simulation time respectively. Amongst the current hits only **PJ316** may explore the S₄ pocket through the chlorine substitution on the phenyl ring. Also of interest is the RMSD fluctuation of the Met49 and Met165 residues which show a biphasic behavior. The residues appear to undergo induced-fit in the 0-2 ns of the simulation following which they preserve their conformational states for the remainder of the simulation. The hydrophobic distance progression (Figure 7.3.5e) for Leu141 also showed a biphasic behavior (0-1.5ns and 1.5-5ns) suggesting the possible stabilization of the interaction

with the dimethyl moiety of the ligand in the second phase which would also stabilize the two H-bonds in the S₁ pocket.

6.4 Guidelines for future structure-based drug design of SARS-3CL^{pro} inhibitors.

Based on the analysis of SAR trends as well as MD simulation, certain guidelines could be formed for structure-based modification schemes that could be used for the synthesis of analogues. First, the S₂ sub site appears to be highly specific to the hydrophobic groups that may be utilized to occupy this subsite. While the smaller/bulkier aliphatic groups didn't provide adequate hydrophobic interaction/caused steric clash, longer groups would require a corresponding S₁ group of compatible size for proper fit. A phenyl group at the S₂ position is advantageous because it allows a vector into the S₄ pocket which could be occupied with a hydrophobic group potentially leading to an increase in overall affinity. In case of **PJ316**, we observed the hydrophobic chlorine group vectored into the S₄ cavity adds to the overall stability of the binding pose and suggest that such structure-based modifications in both active chemotypes could be beneficial towards potency gain. Furthermore, additional smaller hydrophobic substituents may be useful in probing the depths of the S₂ pocket. The S₁ interaction with His163 was found to be highly stable in all the simulations except **PJ207** and is one of the most critical binding site requirements. The alternate H-bonding demonstrated by **PJ207** with the backbone of Glu166 in lieu of His163 and the Ser144 (from the oxyanion hole) interaction observed in case of **PJ225** suggests that possible modifications vectoring two hydrogen acceptors (e.g. sulfonamide for Ser144) could be utilized to harness additional binding affinity. Also, of interest would be the interaction with Asn142 as observed in the MD simulations of **PJ224** and **PJ225**. It would also appear that, the interaction with Leu141 (observed in the

docking pose and MD simulations of the two more potent hits **PJ225** and **PJ316**) through aliphatic hydrophobic groups may add to the overall stability by forming additional hydrophobic interactions which in turn helps in the stabilization of the H-bonds formed in the S₁ pocket. Additionally, as shown in case of **PJ316** the H-bonding interactions with the backbone of Gly143 and Ser144 residue forming the oxyanion hole could form a strong anchor point for ligand stability. The Gly-1 residue from the second monomer was found to participate in a transient H-bonding interaction during the MD simulation and may be pursued as part of the structure-based design strategy. The binding site appears to prefer an amphiphilic group extending into the S₁ pocket and is critical towards the stability of the interactions formed in this region.

6.5. Conclusions

A multi nanosecond explicit solvent simulation were carried out using the docking poses of the identified hits to study the overall stability of the binding site interactions as well as identify important changes in the interaction profile that may not be apparent from a docking study.

Additionally, the docking poses from the identified hits were utilized for a series of multi nanosecond simulations to study the overall stability of the interaction profiles and well as novel interactions not observed during the docking study. The information from these analyses would be utilized for further structure guided synthesis of analogues that might help in the identification of potent antiviral leads targeted against SARS-CoV. Cumulative analysis of the evaluated compounds and the simulation studies led to the identification of certain SAR traits which would be useful in further structure based design efforts.

Appendices

^1H and ^{13}C spectra of intermediates and final compounds **16-41 (Chapter 3)** were recorded on Bruker J[500MHz, 400MHz, or Bruker 400MHz Ultra Shield and the analyses conducted in CDCl_3 , MeOD, or DMSO- d_6 .

2-((6-aminobenzo[d]thiazol-2-yl)thio)-1-phenylethanone (16)

^1H NMR (500 MHz, DMSO- d_6) δ 8.06 (d, $J = 6.84$ Hz, 2H), 7.64 - 7.78 (m, 1H), 7.58 (d, $J = 6.41$ Hz, 2H), 7.43 (d, $J = 8.55$ Hz, 1H), 6.99 (br. s., 1H), 6.69 (d, $J = 7.27$ Hz, 1H), 5.34 (br. s., 2H), 5.03 (br. s., 2H); ^{13}C NMR (126 MHz, DMSO- d_6) δ 193.2, 157.4, 146.6, 144.1, 136.6, 135.4, 133.7, 128.8, 128.4, 121.5, 114.4, 103.9, 40.9; HRMS (m/z): [M + H] calcd for $\text{C}_{15}\text{H}_{13}\text{N}_2\text{OS}_2$ 301.0469, found 301.0448

2-((6-aminobenzo[d]thiazol-2-yl)thio)-1-(naphthalen-2-yl)ethanone (17)

^1H NMR (500 MHz, DMSO- d_6) δ 8.83 (br. s., 1H), 8.11 - 8.42 (m, 2H), 8.04 (br. s., 2H), 7.58 - 7.78 (m, 2H), 7.44 (br. s., 1H), 7.01 (br. s., 1H), 6.70 (br. s., 1H), 5.27 - 5.47 (m, 2H), 5.17 (br. s., 2H); ^{13}C NMR (126 MHz, DMSO- d_6) δ 193.0, 157.3, 146.5, 144.0, 136.5, 135.1, 132.5, 131.9, 130.5, 129.5, 128.8, 128.3, 127.6, 126.9, 123.6, 121.3, 114.2, 103.7
HRMS (m/z): [M + H] calcd for $\text{C}_{19}\text{H}_{15}\text{N}_2\text{OS}_2$ 351.0626, found 351.0626

ethyl 2-((6-aminobenzo[d]thiazol-2-yl)thio)acetate (18)

^1H NMR (400 MHz, DMSO- d_6) δ 7.48 (dd, $J = 3.14, 8.41$ Hz, 1H), 7.00 (br. s., 1H), 6.66 - 6.86 (m, 1H), 5.36 (br. s., 2H), 4.16 - 4.29 (m, 2H), 4.13 (dd, $J = 3.01, 7.03$ Hz, 2H), 1.09 - 1.30 (m, 3H); ^{13}C NMR (101 MHz, DMSO- d_6) δ 168.2, 146.7, 144.1, 136.6, 135.9, 121.6, 114.4, 103.9, 61.2, 34.8, 14.0; HRMS (m/z): [M + H] calcd for $\text{C}_{11}\text{H}_{13}\text{N}_2\text{O}_2\text{S}_2$ 269.0418, found 269.0389

2-((6-aminobenzo[d]thiazol-2-yl)thio)-*N*-(*p*-tolyl)acetamide (19)

^1H NMR (400 MHz, DMSO- d_6) δ 10.27 (br. s., 1H), 7.42 - 7.66 (m, 2H), 7.10 (br. s., 2H), 6.74 (br. s., 1H), 5.34 (br. s., 2H), 4.24 (br. s., 2H), 2.14 - 2.39 (m, 3H); ^{13}C NMR (101 MHz, DMSO- d_6) δ 165.3, 157.8, 146.7, 144.3, 136.7, 136.3, 132.6, 129.2, 121.6, 119.3, 114.5, 104.0, 37.9, 20.5; HRMS (m/z): [M + H] calcd for $\text{C}_{16}\text{H}_{16}\text{N}_3\text{OS}_2$ 330.0735, found 330.0700

2-nitro-*N*-(2-((2-oxo-2-phenylethyl)thio)benzo[d]thiazol-6-yl)benzenesulfonamide (20)

^1H NMR (400 MHz, CHLOROFORM- d) δ 8.07 (d, $J = 7.28$ Hz, 2H), 7.87 (d, $J = 7.78$ Hz, 1H), 7.76 - 7.80 (m, 1H), 7.67 - 7.72 (m, 2H), 7.62 - 7.67 (m, 2H), 7.55 - 7.59 (m, 1H), 7.50 - 7.55 (m, 2H), 7.35 (s, 1H), 7.17 (dd, $J = 2.13, 8.66$ Hz, 1H), 4.93 (s, 2H); ^{13}C NMR (126 MHz, CHLOROFORM- d) δ 193.8, 166.9, 150.6, 148.4, 136.3, 135.8, 135.5, 134.7, 133.6, 133.3, 131.6, 130.8, 129.7, 129.1, 125.2, 122.1, 121.6, 115.2, 41.3; HRMS (m/z): [M - H] calcd for $\text{C}_{21}\text{H}_{14}\text{N}_3\text{O}_5\text{S}_3$ 484.0095, found 484.0066

***N*-(2-((2-oxo-2-phenylethyl)thio)benzo[d]thiazol-6-yl)-4-(trifluoromethoxy)benzene sulfonamide (21)**

^1H NMR (400 MHz, CHLOROFORM- d) δ 8.06 - 8.12 (m, 2H), 7.77 (d, $J = 8.78$ Hz, 2H), 7.62 - 7.67 (m, 2H), 7.58 (d, $J = 2.01$ Hz, 1H), 7.50 - 7.56 (m, 2H), 7.25 (s, 1H), 6.99 (dd, $J = 2.26, 8.53$ Hz, 1H), 6.64 (br. s., 1H), 4.94 (s, 2H); ^{13}C NMR (101 MHz, DMSO- d_6) δ 192.8, 179.1, 165.4, 151.1, 149.7, 138.1, 135.7, 135.3, 133.9, 133.8, 129.3, 128.9, 128.4, 121.5, 121.4, 120.2,

113.8, 40.9; HRMS (m/z): [M-H] calcd for C₂₂H₁₃F₃N₂O₄S₃ 523.0067, found 523.9977

***N*-(2-((2-(naphthalen-2-yl)-2-oxoethyl)thio)benzo[d]thiazol-6-yl)-4-(trifluoromethoxy)benzenesulfonamide (22)**

¹H NMR (400 MHz, DMSO-d₆) δ 8.86 (br. s., 1H), 8.16 (br. s., 2H), 7.99 - 8.12 (m, 2H), 7.86 (br. s., 1H), 7.60 - 7.81 (m, 3H), 7.27 (br. s., 1H), 5.29 (br. s., 2H), 4.56 (d, J = 9.54 Hz, 2H), 4.13 (br. s., 1H), 1.27 (br. s., 3H), 0.86 (br. s., 2H); ¹³C NMR (126 MHz, CHLOROFORM-d) δ 192.7, 165.2, 149.7, 135.8, 135.3, 133.7, 132.7, 132.1, 130.7, 129.7, 129.0, 128.5, 127.7, 127.1, 123.7, 121.5, 120.2, 113.3, 41.0; HRMS (m/z): calcd for C₂₁H₁₄F₃N₂O₃S₃ 495.0041, found 495.0118

***N*-(2-((2-(naphthalen-2-yl)-2-oxoethyl)thio)benzo[d]thiazol-6-yl)acetamide (23)**

¹H NMR (500 MHz, CHLOROFORM-d) δ 8.65 (s, 1H), 8.27 (br. s., 1H), 8.10 (dd, J = 1.71, 8.55 Hz, 1H), 7.99 (d, J = 7.69 Hz, 1H), 7.93 (dd, J = 8.33, 17.31 Hz, 2H), 7.72 (d, J = 8.55 Hz, 1H), 7.65 (t, J = 7.05 Hz, 1H), 7.59 (t, J = 7.05 Hz, 1H), 7.34 (br. s., 1H), 7.20 - 7.26 (m, 1H), 5.07 (s, 2H), 2.21 (s, 3H); ¹³C NMR (126 MHz, DMSO-d₆) δ 192.9, 181.9, 168.4, 163.8, 136.2, 135.4, 135.3, 132.7, 132.1, 130.7, 129.7, 129.0, 128.5, 127.7, 127.1, 123.7, 121.0, 118.4, 111.1, 79.2, 41.0, 24.0; HRMS (m/z): [M + Na] calcd for C₂₁H₁₆N₂NaO₂S₂ 415.0550, found 415.0566

***N*-(2-((2-(naphthalen-2-yl)-2-oxoethyl)thio)benzo[d]thiazol-6-yl)quinoline-8-sulfonamide (24)**

¹H NMR (400 MHz, CHLOROFORM-d) δ 9.15 - 9.22 (m, 1H), 8.58 (s, 1H), 8.55 (br. s., 1H), 8.32 (br. s., 1H), 8.30 (d, J = 2.35 Hz, 1H), 8.00 - 8.07 (m, 2H), 7.96 (d, J = 8.02 Hz, 1H), 7.86 - 7.94 (m, 2H), 7.64 (dt, J = 4.18, 8.27 Hz, 2H), 7.53 - 7.60 (m, 2H), 7.50 - 7.53 (m, 1H), 7.48 (d, J = 8.80 Hz, 1H), 6.92 (dd, J = 1.96, 8.61 Hz, 1H), 4.99 (s, 2H); ¹³C NMR (126 MHz, DMSO-d₆) δ 192.7, 164.6, 151.5, 149.1, 142.7, 137.0, 135.4, 135.2, 135.0, 134.6, 134.3, 132.6, 132.2, 132.1, 130.7, 129.6, 129.0, 128.4, 128.4, 127.7, 127.1, 125.6, 123.7, 122.6, 121.1, 119.6, 112.5, 41.0; HRMS (m/z): [M + Na] calcd for C₂₈H₁₉N₃O₃S₃Na, 564.0486; found, 564.0499

4-fluoro-*N*-(2-((2-(naphthalen-2-yl)-2-oxoethyl)thio)benzo[d]thiazol-6-yl)benzenesulfonamide (25)

¹H NMR (500 MHz, CHLOROFORM-d) δ 8.63 (s, 1H), 8.08 (s, 1H), 7.99 (d, J = 8.12 Hz, 1H), 7.92 (dd, J = 8.33, 16.03 Hz, 2H), 7.75 (dd, J = 4.92, 8.76 Hz, 2H), 7.58 - 7.69 (m, 3H), 7.55 (d, J = 1.71 Hz, 1H), 7.09 (t, J = 8.55 Hz, 1H), 6.98 - 7.04 (m, 1H), 5.05 (s, 3H); ¹³C NMR (126 MHz, DMSO-d₆) δ 192.7, 165.3, 149.6, 135.7, 135.2, 134.1, 132.6, 132.1, 130.7, 129.7, 129.7, 129.0, 128.5, 128.4, 127.7, 127.1, 127.0, 123.7, 121.4, 120.2, 116.5, 116.3, 113.6, 40.9; HRMS (m/z): [M + Na] calcd for C₂₅H₁₇FN₂O₃S₃Na, 531.0283; found, 531.0308

ethyl 2-((6-(2,2,2-trifluoroethylsulfonamido)benzo[d]thiazol-2-yl)thio)acetate (26)

¹H NMR (500 MHz, CHLOROFORM-d) δ 7.74 (d, J = 8.55 Hz, 1H), 7.69 (s, 1H), 7.40 (br. s., 1H), 7.22 (dd, J = 2.14, 8.55 Hz, 1H), 4.28 (d, J = 7.27 Hz, 4H), 4.16 (s, 2H), 3.80 (d, J = 8.98 Hz, 2H), 1.32 (t, J = 7.05 Hz, 3H); ¹³C NMR (126 MHz, DMSO-d₆) δ 168.0, 164.8, 149.7, 135.9, 133.7, 123.4, 121.7, 120.3, 113.4, 61.4, 52.5, 34.7, 14.0; HRMS (m/z): [M + H] calcd for C₁₃H₁₄F₃N₂O₄S₃ 415.0068, found 415.0059

ethyl 2-((6-(4-fluorophenylsulfonamido)benzo[d]thiazol-2-yl)thio)acetate (27)

¹H NMR (400 MHz, CHLOROFORM-d) δ 7.72 - 7.81 (m, 2H), 7.63 (d, *J* = 8.61 Hz, 1H), 7.56 (s, 1H), 7.17 (br. s., 1H), 7.10 (t, *J* = 8.51 Hz, 2H), 7.01 (dd, *J* = 1.96, 8.61 Hz, 1H), 4.26 (q, *J* = 7.17 Hz, 2H), 1.30 (t, *J* = 7.14 Hz, 3H); ¹³C NMR (126 MHz, DMSO-d₆) δ 168.0, 165.3, 164.8, 163.3, 149.5, 135.8, 135.6, 134.2, 129.8, 129.7, 121.5, 120.2, 116.5, 116.3, 113.6, 61.3, 34.7, 13.9; HRMS (*m/z*): [M + H] calcd for C₁₇H₁₆FN₂O₄S₃ 427.0256 found 427.02562

ethyl 2-((6-(2-nitrophenylsulfonamido)benzo[d]thiazol-2-yl)thio)acetate (28)

¹H NMR (500 MHz, CHLOROFORM-d) δ 7.85 - 7.93 (m, *J* = 7.69 Hz, 1H), 7.80 (d, *J* = 7.69 Hz, 1H), 7.71 (br. s., 2H), 7.54 - 7.61 (m, 1H), 7.38 (br. s., 1H), 7.29 (s, 1H), 7.16 - 7.23 (m, *J* = 8.12 Hz, 1H), 4.23 - 4.31 (m, 2H), 4.16 (s, 2H), 1.31 (t, *J* = 6.84 Hz, 3H); ¹³C NMR (101 MHz, DMSO-d₆) δ 168.0, 165.3, 149.8, 147.9, 135.8, 134.7, 133.3, 132.6, 131.2, 130.1, 124.6, 121.6, 120.6, 114.2, 61.4, 34.7, 14.0; HRMS (*m/z*): [M + Na] calcd for C₁₇H₁₅N₃NaO₆S₃ 476.0021, found 476.0068

ethyl 2-((6-(quinoline-7-sulfonamido)benzo[d]thiazol-2-yl)thio)acetate (29)

¹H NMR (500 MHz, CHLOROFORM-d) δ 9.20 (d, *J* = 4.27 Hz, 1H), 8.58 (br. s., 1H), 8.33 (t, *J* = 6.62 Hz, 2H), 8.04 (d, *J* = 8.12 Hz, 1H), 7.65 (dd, *J* = 4.27, 8.55 Hz, 1H), 7.55 - 7.61 (m, 2H), 7.53 (d, *J* = 8.55 Hz, 1H), 6.93 (dd, *J* = 1.92, 8.76 Hz, 1H), 4.21 (q, *J* = 7.12 Hz, 2H), 4.09 (s, 2H), 1.26 (t, *J* = 7.05 Hz, 3H); ¹³C NMR (126 MHz, DMSO-d₆) δ 167.9, 164.1, 151.5, 149.0, 142.7, 136.9, 135.4, 135.0, 134.7, 134.3, 132.2, 128.3, 125.6, 122.6, 121.2, 119.6, 112.6, 61.3, 34.6, 13.9; HRMS (*m/z*): [M+H] calcd for C₂₀H₁₈N₃O₄S₃ 460.0459, found 460.0483

***N*-(*p*-tolyl)-2-((6-(2,2,2-trifluoroethylsulfonamido)benzo[d]thiazol-2-yl)thio) acetamide (30)**

¹H NMR (500 MHz, DMSO-d₆) δ 10.29 - 10.34 (m, 1H), 9.77 (br. s., 1H), 7.69 - 7.75 (m, 1H), 7.40 - 7.46 (m, 1H), 7.32 - 7.37 (m, 2H), 7.31 (br. s., 1H), 7.01 (d, *J* = 8.12 Hz, 2H), 4.09 (br. s., 2H), 3.72 - 3.89 (m, 2H), 2.21 (s, 3H); ¹³C NMR (101 MHz, DMSO-d₆) δ 165.2, 164.6, 149.7, 143.7, 135.7, 133.3, 132.4, 128.8, 121.2, 119.8, 119.1, 112.9, 37.5, 20.3; HRMS (*m/z*): [M - H] calcd for C₁₈H₁₅ F₃N₃ O₃ S₃ 474.0306, found 474.0145

2-((6-acetamidobenzo[d]thiazol-2-yl)thio)-*N*-(*p*-tolyl)acetamide (31)

¹H NMR (500 MHz, DMSO-d₆) δ 9.88 (br. s., 1H), 9.67 (br. s., 1H), 8.28 (s, 1H), 7.66 (d, *J* = 8.98 Hz, 1H), 7.48 - 7.54 (m, 1H), 7.39 - 7.44 (m, 1H), 7.35 (d, *J* = 8.12 Hz, 2H), 7.00 (d, *J* = 8.12 Hz, 2H), 4.08 (br. s., 2H), 2.21 (s, 3H), 2.08 (s, 3H); ¹³C NMR (101 MHz, DMSO-d₆) δ 168.5, 165.0, 164.1, 148.5, 136.3, 135.4, 132.6, 129.2, 121.0, 119.2, 118.5, 111.2, 37.8, 24.0, 20.4; HRMS (*m/z*): [M + H] calcd for C₁₈H₁₈ N₃ O₂ S₂ 372.0840, found 372.0797

2-((6-(quinoline-7-sulfonamido)benzo[d]thiazol-2-yl)thio)-*N*-(*p*-tolyl)acetamide (32)

¹H NMR (500 MHz, CHLOROFORM-d) δ 9.59 (br. s., 1H), 9.22 (d, *J* = 2.99 Hz, 1H), 8.69 (br. s., 1H), 8.35 (dd, *J* = 8.12, 9.83 Hz, 2H), 8.06 (d, *J* = 7.69 Hz, 1H), 7.68 (dd, *J* = 4.06, 8.33 Hz, 1H), 7.64 (d, *J* = 8.55 Hz, 1H), 7.60 (d, *J* = 7.69 Hz, 1H), 7.56 - 7.59 (m, 1H), 7.33 (d, *J* = 8.12 Hz, 2H), 7.29 (s, 1H), 7.07 (d, *J* = 8.12 Hz, 2H), 3.99 (s, 2H), 2.26 - 2.36 (m, 3H); ¹³C NMR (101 MHz, DMSO-d₆) δ 164.8, 151.5, 149.2, 142.7, 137.0, 136.2, 135.4, 135.0, 134.6, 134.3,

132.5, 132.2, 129.2, 128.4, 125.6, 122.6, 121.1, 119.6, 119.2, 112.6, 52.9, 37.7, 20.4; HRMS (*m/z*): [M + Na] calcd for C₂₅H₂₀N₄O₃S₃Na, 543.0595; found 543.0584

2-((6-(4-fluorophenylsulfonamido)benzo[d]thiazol-2-yl)thio)-N-(*p*-tolyl)acetamide (33)

¹H NMR (500 MHz, CHLOROFORM-*d*) δ 9.55 (br. s., 1H), 7.78 (dd, *J* = 5.77, 8.33 Hz, 3H), 7.62 (d, *J* = 1.71 Hz, 1H), 7.36 (d, *J* = 8.55 Hz, 2H), 7.08 - 7.17 (m, 4H), 6.78 (br. s., 1H), 4.06 (s, 2H), 2.30 (s, 3H); ¹³C NMR (101 MHz, DMSO-*d*₆) δ 166.0, 165.4, 163.5, 150.1, 136.7, 136.2, 136.0, 134.6, 133.0, 130.2, 129.6, 121.9, 120.7, 119.7, 117.0, 116.8, 114.1, 63.8, 38.2, 20.8; HRMS (*m/z*): [M + H] calcd for C₂₂H₁₉ FN₃ O₃ S₃ 488.0573, found 488.0602

N-(*p*-tolyl)-2-((6-(4-(trifluoromethoxy)phenylsulfonamido)benzo[d]thiazol-2-yl)thio)acetamide (34)

¹H NMR (400 MHz, CHLOROFORM-*d*) δ 9.60 (br. s., 1H), 7.80 (dd, *J* = 2.35, 8.80 Hz, 2H), 7.64 (d, *J* = 1.96 Hz, 1H), 7.36 (d, *J* = 8.22 Hz, 2H), 7.28 (br. s., 1H), 7.14 (d, *J* = 1.76 Hz, 1H), 7.10 (d, *J* = 8.41 Hz, 2H), 6.81 (br. s., 1H), 4.06 (s, 2H), 2.30 (s, 3H); ¹³C NMR (101 MHz, DMSO-*d*₆) □ 164.9, 151.0, 149.6, 136.2, 136.0, 135.8, 134.5, 132.6, 129.3, 129.2, 121.4, 120.4, 119.2, 113.7, 63.2, 37.7, 20.4; HRMS (*m/z*): [M +H] calcd for C₂₃H₁₉ F₃N₃ O₄ S₃ 553.0412, found 553.0545

(*R*)-tert-butyl (3-methyl-1-oxo-1-((2-((2-oxo-2-(*p*-tolylamino)ethyl)thio)benzo [d]thiazol-6-yl)amino)butan-2-yl)Carbamate (35)

¹H NMR (400 MHz, CHLOROFORM-*d*) δ 9.81 (s, 1H), 9.36 (br. s., 1H), 8.13 (br. s., 1H), 7.56 (d, *J* = 7.78 Hz, 1H), 7.36 - 7.43 (m, *J* = 8.28 Hz, 2H), 7.14 (d, *J* = 8.03 Hz, 1H), 7.04 - 7.11 (m, *J* = 8.03 Hz, 2H), 5.54 (d, *J* = 8.03 Hz, 1H), 4.24 (t, *J* = 6.65 Hz, 1H), 3.90 - 4.09 (m, 2H), 2.26 (s, 3H), 2.14 (d, *J* = 19.07 Hz, 1H), 1.50 (s, 9H), 1.08 (d, *J* = 5.77 Hz, 6H); ¹³C NMR (101 MHz, CDCl₃) δ 171.11, 166.42, 165.92, 156.85, 148.23, 135.84, 135.51, 135.19, 133.92, 129.49, 120.47, 119.64, 118.59, 111.58, 80.55, 77.38, 77.07, 76.75, 61.14, 37.44, 30.91, 28.44, 20.86, 19.33, 18.64.

(*R*)-tert-butyl (2-oxo-2-((2-((2-oxo-2-(*p*-tolylamino)ethyl)thio)benzo[d]thiazol-6-yl)amino)-1-phenylethyl)carbamate (36)

¹H NMR (400 MHz, DMSO-*d*₆) δ 10.07 (br. s., 0H), 9.88 (br. s., 1H), 8.19 - 8.32 (m, 1H), 7.65 (d, *J* = 8.78 Hz, 1H), 7.38 - 7.48 (m, 3H), 7.34 (d, *J* = 8.53 Hz, 3H), 7.23 - 7.29 (m, 2H), 7.21 (d, *J* = 6.53 Hz, 1H), 6.99 (d, *J* = 8.53 Hz, 2H), 6.20 (br. s., 1H), 5.43 (br. s., 1H), 4.06 (s, 3H), 3.42 (br. s., 1H), 2.93 - 2.99 (m, 1H), 2.20 (s, 3H), 1.36 (br. s., 9H); ¹³C NMR (101 MHz, DMSO-*d*₆) δ 165.1, 164.5, 148.2, 137.7, 135.3, 135.1, 135.0, 132.8, 128.6, 128.0, 127.3, 126.4, 120.2, 118.9, 118.8, 118.4, 111.3, 37.0, 27.6, 20.1

(*R*)-tert-butyl (3-methyl-1-((2-((2-(naphthalen-2-yl)-2-oxoethyl)thio)benzo[d]thiazol-6-yl)amino)-1-oxobutan-2-yl)carbamate (37)

¹H NMR (400 MHz, CHLOROFORM-*d*) δ 8.90 (br. s., 1H), 8.61 (s, 1H), 8.16 (br. s., 1H), 8.06 (d, *J* = 8.53 Hz, 1H), 7.96 (d, *J* = 8.03 Hz, 1H), 7.88 (t, *J* = 8.66 Hz, 2H), 7.59 - 7.65 (m, 1H), 7.56 (d, *J* = 7.53 Hz, 2H), 7.15 (d, *J* = 7.28 Hz, 1H), 5.37 (d, *J* = 7.78 Hz, 1H), 5.05 (d, *J* = 2.26 Hz, 2H), 4.16 (br. s., 1H), 2.19 (br. s., 1H), 1.47 (s, 9H), 1.05 (t, *J* = 5.77 Hz, 6H); ¹³C NMR

(101 MHz, CHLOROFORM-d) δ 193.0, 170.7, 156.6, 149.4, 135.8, 134.5, 132.8, 132.4, 130.6, 129.7, 128.9, 128.7, 127.8, 127.0, 123.9, 121.1, 118.4, 111.8, 77.3, 77.2, 77.0, 76.7, 71.5, 41.2, 30.7, 28.4, 19.4

(R)-tert-butyl (2-((2-((2-(naphthalen-2-yl)-2-oxoethyl)thio)benzo[d]thiazol-6-yl)amino)-2-oxo-1-phenylethyl)carbamate (38)

^1H NMR (400 MHz, CHLOROFORM-d) δ 8.63 (s, 1H), 8.27 (br. s., 1H), 8.14 (br. s., 1H), 8.09 (d, J = 8.28 Hz, 1H), 7.99 (d, J = 8.03 Hz, 1H), 7.90 (d, J = 8.53 Hz, 1H), 7.93 (d, J = 8.78 Hz, 1H), 7.63 (d, J = 8.03 Hz, 2H), 7.55 - 7.61 (m, 1H), 7.48 (d, J = 7.03 Hz, 2H), 7.37 (d, J = 7.78 Hz, 3H), 7.15 (d, J = 8.53 Hz, 1H), 5.77 (br. s., 1H), 5.42 (br. s., 1H), 5.06 (s, 2H), 1.45 (s, 9H)

(R)-2-amino-3-methyl-N-(2-((2-oxo-2-(p-tolylamino)ethyl)thio)benzo[d]thiazol-6-yl)butanamide hydrochloride (39)

^1H NMR (400 MHz, MeOD) δ 8.39 (s, 1H), 7.77 (d, J = 8.78 Hz, 1H), 7.61 (d, J = 8.53 Hz, 1H), 7.44 (d, J = 8.03 Hz, 2H), 7.09 (d, J = 8.03 Hz, 2H), 4.32 (s, 2H), 3.97 (br. s., 1H), 2.27 (s, 3H), 1.13 (dd, J = 6.78, 12.80 Hz, 6H); ^{13}C NMR (101 MHz, MeOD) δ 167.8, 166.8, 166.2, 147.6, 135.5, 135.1, 134.0, 129.0, 120.2, 119.9, 119.5, 112.3, 60.9, 59.0, 37.6, 30.4, 19.6, 17.7, 16.6; HRMS (m/z): [M + Na] calcd for $\text{C}_{21}\text{H}_{24}\text{N}_4\text{O}_2\text{S}_2$ Na 451.5601, found 451.1238

(R)-2-amino-N-(2-((2-oxo-2-(p-tolylamino)ethyl)thio)benzo[d]thiazol-6-yl)-2-phenylacetamide hydrochloride (40)

^1H NMR (400 MHz, MeOD) δ 8.26 - 8.37 (m, 1H), 7.69 (d, J = 8.78 Hz, 1H), 7.65 - 7.67 (m, 1H), 7.64 (br. s., 1H), 7.51 (d, J = 1.76 Hz, 4H), 7.42 (d, J = 8.28 Hz, 2H), 7.09 (d, J = 8.03 Hz, 2H), 5.24 (s, 1H), 4.27 (s, 2H), 2.26 (s, 3H); ^{13}C NMR (101 MHz, MeOD) δ 168.8, 167.8, 167.4, 149.5, 137.0, 136.8, 136.7, 135.5, 134.2, 131.4, 130.7, 130.5, 129.6, 121.8, 121.8, 120.7, 113.6, 58.6, 38.9, 21.1; HRMS (m/z): [M + Na] calcd for $\text{C}_{24}\text{H}_{22}\text{N}_4\text{ClO}_2\text{S}_2\text{Na}$ 521.0298, found 521.0345

(R)-3-methyl-1-((2-((2-(naphthalen-2-yl)-2-oxoethyl)thio)benzo[d]thiazol-6-yl)amino)-1-oxobutan-2-aminium chloride (41)

^1H NMR (400 MHz, DMSO- d_6) δ 11.36 (br. s., 1H), 8.86 (s, 1H), 8.48 (br. s., 2H), 8.42 (s, 1H), 8.15 (d, J = 8.03 Hz, 1H), 7.98 - 8.09 (m, 3H), 7.60 - 7.76 (m, 3H), 5.23 - 5.36 (m, 2H), 3.99 (br. s., 1H), 2.15 - 2.33 (m, 1H), 1.01 (d, J = 6.53 Hz, 6H); ^{13}C NMR (101 MHz, DMSO- d_6) δ 192.9, 167.0, 164.7, 149.0, 135.5, 135.3, 135.1, 132.7, 132.1, 130.8, 129.7, 129.0, 128.5, 127.8, 127.2, 123.7, 121.2, 118.9, 111.9, 66.4, 58.0, 41.1, 30.0, 18.5, 18.0; HRMS (m/z): [M + Na] calcd for $\text{C}_{24}\text{H}_{24}\text{N}_3\text{ClO}_2\text{S}_2\text{Na}$ 508.0311, found 508.0896

(R)-2-((2-((2-(naphthalen-2-yl)-2-oxoethyl)thio)benzo[d]thiazol-6-yl)amino)-2-oxo-1-phenylethanaminium chloride (42)

^1H NMR (400 MHz, DMSO- d_6) δ 11.50 (br. s., 1H), 8.93 (br. s., 3H), 8.86 (br. s., 1H), 8.37 (s, 1H), 8.16 (d, J = 7.78 Hz, 1H), 7.98 - 8.09 (m, 3H), 7.67 - 7.79 (m, 5H), 7.59 - 7.67 (m, 2H), 7.45 (d, J = 7.53 Hz, 3H), 5.28 (s, 2H); ^{13}C NMR (101 MHz, DMSO- d_6) δ 192.8, 166.1, 149.1, 145.5, 135.5, 135.3, 135.0, 133.7, 132.6, 132.1, 130.7, 129.7, 129.3, 129.0, 128.9, 128.5, 127.9, 127.8, 127.2, 123.7, 121.2, 118.7, 111.8, 66.4, 41.0; HRMS (m/z): [M + Na] calcd for $\text{C}_{27}\text{H}_{22}\text{N}_3\text{O}_2\text{S}_2\text{Na}$ 507.6017, found 507.1051

Bibliography

1. Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947.
2. Lecaille, F.; Kaleta, J.; Bromme, D. Human and parasitic papain-like cysteine proteases: their role in physiology and pathology and recent developments in inhibitor design. *Chem. Rev.* **2002**, *102*, 4459-4488.
3. Powers, J. C.; Asgian, J. L.; Ekici, O. D.; James, K. E. Irreversible inhibitors of serine, cysteine, and threonine proteases. *Chem. Rev.* **2002**, *102*, 4639-4750.
4. Barrett, A. J.; Rawlings, N. D. Evolutionary lines of cysteine peptidases. *Biol. Chem.* **2001**, *382*, 727-733.
5. Leung-Toung, R.; Zhao, Y.; Li, W.; Tam, T. F.; Karimian, K.; Spino, M. Thiol proteases: inhibitors and potential therapeutic targets. *Curr. Med. Chem.* **2006**, *13*, 547-581.
6. Tong, L. Viral proteases. *Chem. Rev.* **2002**, *102*, 4609-4626.
7. Rota, P. A.; Oberste, M. S.; Monroe, S. S.; Nix, W. A.; Campagnoli, R.; Icenogle, J. P.; Penaranda, S.; Bankamp, B.; Maher, K.; Chen, M. H.; Tong, S.; Tamin, A.; Lowe, L.; Frace, M.; DeRisi, J. L.; Chen, Q.; Wang, D.; Erdman, D. D.; Peret, T. C.; Burns, C.; Ksiazek, T. G.; Rollin, P. E.; Sanchez, A.; Liffick, S.; Holloway, B.; Limor, J.; McCaustland, K.; Olsen-Rasmussen, M.; Fouchier, R.; Gunther, S.; Osterhaus, A. D.; Drosten, C.; Pallansch, M. A.; Anderson, L. J.; Bellini, W. J. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* **2003**, *300*, 1394-1399.
8. Anand, K.; Ziebuhr, J.; Wadhwani, P.; Mesters, J. R.; Hilgenfeld, R. Coronavirus Main Proteinase (3CLpro) Structure: Basis for Design of Anti-SARS Drugs. *Science (Washington, DC, U. S.)* **2003**, *300*, 1763-1767.
9. Dragovich, P. S.; Prins, T. J.; Zhou, R.; Johnson, T. O.; Hua, Y.; Luu, H. T.; Sakata, S. K.; Brown, E. L.; Maldonado, F. C.; Tuntland, T.; Lee, C. A.; Fuhrman, S. A.; Zalman, L. S.; Patick, A. K.; Matthews, D. A.; Wu, E. Y.; Guo, M.; Borer, B. C.; Nayyar, N. K.; Moran, T.; Chen, L.; Rejto, P. A.; Rose, P. W.; Guzman, M. C.; Doval Santos, E. Z.; Lee, S.; McGee, K.; Mohajeri, M.; Liese, A.; Tao, J.; Kosa, M. B.; Liu, B.; Batugo, M. R.; Gleeson, J. P.; Wu, Z. P.; Liu, J.; Meador, J. W., 3rd; Ferre, R. A. Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 8. Pharmacological optimization of orally bioavailable 2-pyridone-containing peptidomimetics. *J. Med. Chem.* **2003**, *46*, 4572-4585.
10. Huang, L.; Brinen, L. S.; Ellman, J. A. Crystal structures of reversible ketone-Based inhibitors of the cysteine protease cruzain. *Bioorg Med Chem* **2003**, *11*, 21-29.
11. Marquis, R. W.; Ru, Y.; Zeng, J.; Trout, R. E.; LoCastro, S. M.; Gribble, A. D.; Witherington, J.; Fenwick, A. E.; Garnier, B.; Tomaszek, T.; Tew, D.; Hemling, M. E.; Quinn, C. J.; Smith, W. W.; Zhao, B.; McQueney, M. S.; Janson, C. A.; D'Alessio, K.; Veber, D. F. Cyclic ketone inhibitors of the cysteine protease cathepsin K. *J. Med. Chem.* **2001**, *44*, 725-736.
12. Batra, S.; Sabnis, Y. A.; Rosenthal, P. J.; Avery, M. A. Structure-based approach to falcipain-2 inhibitors: synthesis and biological evaluation of 1,6,7-Trisubstituted dihydroisoquinolines and isoquinolines. *Bioorg. Med. Chem.* **2003**, *11*, 2293-2299.
13. Sabnis, Y.; Rosenthal, P. J.; Desai, P.; Avery, M. A. Homology modeling of falcipain-2:

- validation, de novo ligand design and synthesis of novel inhibitors. *J. Biomol. Struct. Dyn.* **2002**, *19*, 765-774.
14. Dragovich, P. S.; Webber, S. E.; Babine, R. E.; Fuhrman, S. A.; Patick, A. K.; Matthews, D. A.; Lee, C. A.; Reich, S. H.; Prins, T. J.; Marakovits, J. T.; Littlefield, E. S.; Zhou, R.; Tikhe, J.; Ford, C. E.; Wallace, M. B.; Meador, J. W., 3rd; Ferre, R. A.; Brown, E. L.; Binford, S. L.; Harr, J. E.; DeLisle, D. M.; Worland, S. T. Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 1. Michael acceptor structure-activity studies. *J. Med. Chem.* **1998**, *41*, 2806-2818.
 15. Ettari, R.; Nizi, E.; Di Francesco, M. E.; Dude, M. A.; Pradel, G.; Vicik, R.; Schirmeister, T.; Micale, N.; Grasso, S.; Zappala, M. Development of peptidomimetics with a vinyl sulfone warhead as irreversible falcipain-2 inhibitors. *J. Med. Chem.* **2008**, *51*, 988-996.
 16. Donkor, I. O. A survey of calpain inhibitors. *Curr. Med. Chem.* **2000**, *7*, 1171-1188.
 17. Zhu, S.; Hudson, T. H.; Kyle, D. E.; Lin, A. J. Synthesis and in vitro studies of novel pyrimidinyl peptidomimetics as potential antimalarial therapeutic agents. *J. Med. Chem.* **2002**, *45*, 3491-3496.
 18. Chen, L.; Chen, S.; Gui, C.; Shen, J.; Shen, X.; Jiang, H. Discovering severe acute respiratory syndrome coronavirus 3CL protease inhibitors: virtual screening, surface plasmon resonance, and fluorescence resonance energy transfer assays. *J. Biomol. Screening* **2006**, *11*, 915-921.
 19. Mukherjee, P.; Desai, P.; Ross, L.; White, E. L.; Avery, M. A. Structure-based virtual screening against SARS-3CL(pro) to identify novel non-peptidic hits. *Bioorg. Med. Chem.* **2008**, *16*, 4138-4149.
 20. <http://malaria.who.int/wmr2008/>. (Accessed August 8,, 2009).
 21. Katzung, B. G., *Basic and Clinical Pharmacology*. 8th ed.; Lange Medical Books/McGraw-Hill: New York, 2001.
 22. Sachs, J.; Malaney, P. The economic and social burden of malaria. *Nature* **2002**, *415*, 680-685.
 23. *World Malaria Report 2010*. WHO Press: 2010.
 24. Fidock, D. A. Drug discovery: Priming the antimalarial pipeline. *Nature* **2010**, *465*, 297-298.
 25. Toovey, S. *The Miraculous Fever-Tree. The Cure that Changed the World* Fiametta Rocco; Harper Collins, San Francisco, 2004, 348 pages, Paperback, ISBN 0-00-6532357. *Travel Med Infect Dis* **2004**, *2*, 109-110.
 26. Dondorp, A. M.; Nosten, F.; Yi, P.; Das, D.; Phyo, A. P.; Tarning, J.; Lwin, K. M.; Ariey, F.; Hanpithakpong, W.; Lee, S. J.; Ringwald, P.; Silamut, K.; Imwong, M.; Chotivanich, K.; Lim, P.; Herdman, T.; An, S. S.; Yeung, S.; Singhasivanon, P.; Day, N. P.; Lindergardh, N.; Socheat, D.; White, N. J. Artemisinin resistance in *Plasmodium falciparum* malaria. *N Engl J Med* **2009**, *361*, 455-467.
 27. Price, R. N.; Uhlemann, A. C.; Brockman, A.; McGready, R.; Ashley, E.; Phaipun, L.; Patel, R.; Laing, K.; Looareesuwan, S.; White, N. J.; Nosten, F.; Krishna, S. Mefloquine resistance in *Plasmodium falciparum* and increased pfmdr1 gene copy number. *Lancet* **2004**, *364*, 438-447.
 28. Wellems, T. E.; Plowe, C. V. Chloroquine-resistant malaria. *J Infect Dis* **2001**, *184*, 770-

- 776.
29. Nosten, F.; White, N. J. Artemisinin-based combination treatment of falciparum malaria. *The American journal of tropical medicine and hygiene* **2007**, *77*, 181.
 30. Klemba, M.; Goldberg, D. E. Biological roles of proteases in parasitic protozoa. *Annu. Rev. Biochem.* **2002**, *71*, 275-305.
 31. Rosenthal, P. J. Hydrolysis of erythrocyte proteins by proteases of malaria parasites. *Curr. Opin. Hematol.* **2002**, *9*, 140-145.
 32. Brady, R. L.; Cameron, A. Structure-based approaches to the development of novel anti-malarials. *Curr. Drug Targets* **2004**, *5*, 137-149.
 33. Sajid, M.; McKerrow, J. H. Cysteine proteases of parasitic organisms. *Mol. Biochem. Parasitol.* **2002**, *120*, 1-21.
 34. Shenai, B. R.; Sijwali, P. S.; Singh, A.; Rosenthal, P. J. Characterization of native and recombinant falcipain-2, a principal trophozoite cysteine protease and essential hemoglobinase of *Plasmodium falciparum*. *J. Biol. Chem.* **2000**, *275*, 29000-29010.
 35. Dahl, E. L.; Rosenthal, P. J. Biosynthesis, localization, and processing of falcipain cysteine proteases of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **2005**, *139*, 205-212.
 36. Sijwali, P. S.; Shenai, B. R.; Gut, J.; Singh, A.; Rosenthal, P. J. Expression and characterization of the *Plasmodium falciparum* haemoglobinase falcipain-3. *Biochem. J.* **2001**, *360*, 481-489.
 37. Sijwali, P. S.; Koo, J.; Singh, N.; Rosenthal, P. J. Gene disruptions demonstrate independent roles for the four falcipain cysteine proteases of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **2006**, *150*, 96-106.
 38. Rosenthal, P. J.; Sijwali, P. S.; Singh, A.; Shenai, B. R. Cysteine proteases of malaria parasites: targets for chemotherapy. *Curr. Pharm. Des.* **2002**, *8*, 1659-1672.
 39. Sijwali, P. S.; Rosenthal, P. J. Gene disruption confirms a critical role for the cysteine protease falcipain-2 in hemoglobin hydrolysis by *Plasmodium falciparum*. *Proc Natl Acad Sci U S A* **2004**, *101*, 4384-4389.
 40. Greenbaum, D. C.; Baruch, A.; Grainger, M.; Bozdech, Z.; Medzihradzky, K. F.; Engel, J.; DeRisi, J.; Holder, A. A.; Bogyo, M. A role for the protease falcipain 1 in host cell invasion by the human malaria parasite. *Science* **2002**, *298*, 2002-2006.
 41. Sijwali, P. S.; Kato, K.; Seydel, K. B.; Gut, J.; Lehman, J.; Klemba, M.; Goldberg, D. E.; Miller, L. H.; Rosenthal, P. J. *Plasmodium falciparum* cysteine protease falcipain-1 is not essential in erythrocytic stage malaria parasites. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 8721-8726.
 42. Eksi, S.; Czesny, B.; Greenbaum, D. C.; Bogyo, M.; Williamson, K. C. Targeted disruption of *Plasmodium falciparum* cysteine protease, falcipain 1, reduces oocyst production, not erythrocytic stage growth. *Mol. Microbiol.* **2004**, *53*, 243-250.
 43. Rosenthal, P. J.; Wollish, W. S.; Palmer, J. T.; Rasnick, D. Antimalarial effects of peptide inhibitors of a *Plasmodium falciparum* cysteine proteinase. *J. Clin. Invest.* **1991**, *88*, 1467-1472.
 44. Rosenthal, P. J. Cysteine proteases of malaria parasites. *Int. J. Parasitol.* **2004**, *34*, 1489-1499.
 45. Rosenthal, P. J. *Plasmodium falciparum*: effects of proteinase inhibitors on globin

- hydrolysis by cultured malaria parasites. *Exp. Parasitol.* **1995**, *80*, 272-281.
46. Sabnis, Y. A.; Desai, P. V.; Rosenthal, P. J.; Avery, M. A. Probing the structure of falcipain-3, a cysteine protease from *Plasmodium falciparum*: Comparative protein modeling and docking studies. *Protein Sci.* **2003**, *12*, 501-509.
 47. Joachimiak, M. P.; Chang, C.; Rosenthal, P. J.; Cohen, F. E. The impact of whole genome sequence data on drug discovery--a malaria case study. *Mol. Med.* **2001**, *7*, 698-710.
 48. Goh, L. L.; Sim, T. S. Homology modeling and mutagenesis analyses of *Plasmodium falciparum* falcipain 2A: implications for rational drug design. *Biochem. Biophys. Res. Commun.* **2004**, *323*, 565-572.
 49. Hogg, T.; Nagarajan, K.; Herzberg, S.; Chen, L.; Shen, X.; Jiang, H.; Wecke, M.; Blohmke, C.; Hilgenfeld, R.; Schmidt, C. L. Structural and functional characterization of Falcipain-2, a hemoglobinase from the malarial parasite *Plasmodium falciparum*. *J. Biol. Chem.* **2006**, *281*, 25425-25437.
 50. Kerr, I. D.; Lee, J. H.; Pandey, K. C.; Harrison, A.; Sajid, M.; Rosenthal, P. J.; Brinen, L. S. Structures of Falcipain-2 and Falcipain-3 Bound to Small Molecule Inhibitors: Implications for Substrate Specificity. *J. Med. Chem.* **2009**, *52*, 852-857.
 51. Wang, S. X.; Pandey, K. C.; Scharfstein, J.; Whisstock, J.; Huang, R. K.; Jacobelli, J.; Fletterick, R. J.; Rosenthal, P. J.; Abrahamson, M.; Brinen, L. S.; Rossi, A.; Sali, A.; McKerrow, J. H. The structure of chagasin in complex with a cysteine protease clarifies the binding mode and evolution of an inhibitor family. *Structure* **2007**, *15*, 535-543.
 52. Wang, S. X.; Pandey, K. C.; Somoza, J. R.; Sijwali, P. S.; Kortemme, T.; Brinen, L. S.; Fletterick, R. J.; Rosenthal, P. J.; McKerrow, J. H. Structural basis for unique mechanisms of folding and hemoglobin binding by a malarial protease. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 11503-11508.
 53. Singh, N.; Sijwali, P. S.; Pandey, K. C.; Rosenthal, P. J. *Plasmodium falciparum*: biochemical characterization of the cysteine protease falcipain-2'. *Exp. Parasitol.* **2006**, *112*, 187-192.
 54. Hanzlik, R. P.; Jacober, S. P.; Zygmunt, J. Reversible binding of peptide aldehydes to papain. Structure-activity relationships. *Biochim. Biophys. Acta* **1991**, *1073*, 33-42.
 55. Hu, L. Y.; Abeles, R. H. Inhibition of cathepsin B and papain by peptidyl alpha-keto esters, alpha-keto amides, alpha-diketones, and alpha-keto acids. *Arch. Biochem. Biophys.* **1990**, *281*, 271-274.
 56. Li, Z.; Patil, G. S.; Golubski, Z. E.; Hori, H.; Tehrani, K.; Foreman, J. E.; Eveleth, D. D.; Bartus, R. T.; Powers, J. C. Peptide alpha-keto ester, alpha-keto amide, and alpha-keto acid inhibitors of calpains and other cysteine proteases. *J. Med. Chem.* **1993**, *36*, 3472-3480.
 57. Hanzlik, R. P.; Zygmunt, J.; Moon, J. B. Reversible covalent binding of peptide nitriles to papain. *Biochim. Biophys. Acta* **1990**, *1035*, 62-70.
 58. Rasnick, D. Synthesis of peptide fluoromethyl ketones and the inhibition of human cathepsin B. *Anal. Biochem.* **1985**, *149*, 461-465.
 59. Gour-Salin, B. J.; Lachance, P.; Magny, M. C.; Plouffe, C.; Menard, R.; Storer, A. C. E64 [trans-epoxysuccinyl-L-leucylamido-(4-guanidino)butane] analogues as inhibitors of cysteine proteinases: investigation of S2 subsite interactions. *Biochem. J.* **1994**, *299* (Pt

- 2), 389-392.
60. Palmer, J. T.; Rasnick, D.; Klaus, J. L.; Bromme, D. Vinyl Sulfones as Mechanism-Based Cysteine Protease Inhibitors. *J. Med. Chem.* **1995**, *38*, 3193-3196.
 61. Rosenthal, P. J.; Lee, G. K.; Smith, R. E. Inhibition of a Plasmodium vinckei cysteine proteinase cures murine malaria. *J. Clin. Invest.* **1993**, *91*, 1052-1056.
 62. Rosenthal, P. J.; Olson, J. E.; Lee, G. K.; Palmer, J. T.; Klaus, J. L.; Rasnick, D. Antimalarial effects of vinyl sulfone cysteine proteinase inhibitors. *Antimicrob. Agents Chemother.* **1996**, *40*, 1600-1603.
 63. Schirmeister, T.; Klockow, A. Cysteine protease inhibitors containing small rings. *Mini-Rev. Med. Chem.* **2003**, *3*, 585-596.
 64. Rosenthal, P. J.; McKerrow, J. H.; Aikawa, M.; Nagasawa, H.; Leech, J. H. A malarial cysteine proteinase is necessary for hemoglobin degradation by Plasmodium falciparum. *J. Clin. Invest.* **1988**, *82*, 1560-1566.
 65. Wickham, M. E.; Culvenor, J. G.; Cowman, A. F. Selective inhibition of a two-step egress of malaria parasites from the host erythrocyte. *J. Biol. Chem.* **2003**, *278*, 37658-37663.
 66. Olson, J. E.; Lee, G. K.; Semenov, A.; Rosenthal, P. J. Antimalarial effects in mice of orally administered peptidyl cysteine protease inhibitors. *Bioorg. Med. Chem.* **1999**, *7*, 633-638.
 67. Shenai, B. R.; Lee, B. J.; Alvarez-Hernandez, A.; Chong, P. Y.; Emal, C. D.; Neitz, R. J.; Roush, W. R.; Rosenthal, P. J. Structure-activity relationships for inhibition of cysteine protease activity and development of Plasmodium falciparum by peptidyl vinyl sulfones. *Antimicrob. Agents Chemother.* **2003**, *47*, 154-160.
 68. Valente, C.; Guedes, R. C.; Moreira, R.; Iley, J.; Gut, J.; Rosenthal, P. J. Dipeptide vinyl sultams: synthesis via the Wittig-Horner reaction and activity against papain, falcipain-2 and Plasmodium falciparum. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 4115-4119.
 69. Lee, B. J.; Singh, A.; Chiang, P.; Kemp, S. J.; Goldman, E. A.; Weinhouse, M. I.; Vlasuk, G. P.; Rosenthal, P. J. Antimalarial activities of novel synthetic cysteine protease inhibitors. *Antimicrob. Agents Chemother.* **2003**, *47*, 3810-3814.
 70. Chipeleme, A.; Gut, J.; Rosenthal, P. J.; Chibale, K. Synthesis and biological evaluation of phenolic Mannich bases of benzaldehyde and (thio)semicarbazone derivatives against the cysteine protease falcipain-2 and a chloroquine resistant strain of Plasmodium falciparum. *Bioorg. Med. Chem.* **2007**, *15*, 273-282.
 71. Martichonok, V.; Plouffe, C.; Storer, A. C.; Menard, R.; Jones, J. B. Aziridine analogs of [[trans-(epoxysuccinyl)-L-leucyl]amino]-4-guanidinobutane (E-64) as inhibitors of cysteine proteases. *J. Med. Chem.* **1995**, *38*, 3078-3085.
 72. Schulz, F.; Gelhaus, C.; Degel, B.; Vicik, R.; Heppner, S.; Breuning, A.; Leippe, M.; Gut, J.; Rosenthal, P. J.; Schirmeister, T. Screening of protease inhibitors as antiplasmodial agents. Part I: Aziridines and epoxides. *ChemMedChem* **2007**, *2*, 1214-1224.
 73. Capela, R.; Oliveira, R.; Goncalves, L. M.; Domingos, A.; Gut, J.; Rosenthal, P. J.; Lopes, F.; Moreira, R. Artemisinin-dipeptidyl vinyl sulfone hybrid molecules: design, synthesis and preliminary SAR for antiplasmodial activity and falcipain-2 inhibition. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3229-3232.
 74. Goud, P. M.; Sheri, A.; Desai, P. V.; Watkins, E. B.; Tekwani, B.; Sabnis, Y.; Gut, J.;

- Rosenthal, P. J.; Avery, M. A. Design, synthesis and evaluation of trisubstituted thiazoles targeting Plasmodium falciparum cysteine proteases. *Med. Chem. Res.* **2005**, *14*, 74-105.
75. Verissimo, E.; Berry, N.; Gibbons, P.; Cristiano, M. L.; Rosenthal, P. J.; Gut, J.; Ward, S. A.; O'Neill, P. M. Design and synthesis of novel 2-pyridone peptidomimetic falcipain 2/3 inhibitors. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 4210-4214.
76. Micale, N.; Kozikowski, A. P.; Ettari, R.; Grasso, S.; Zappala, M.; Jeong, J. J.; Kumar, A.; Hanspal, M.; Chishti, A. H. Novel peptidomimetic cysteine protease inhibitors as potential antimalarial agents. *J. Med. Chem.* **2006**, *49*, 3064-3067.
77. Ettari, R.; Nizi, E.; Di Francesco, M. E.; Micale, N.; Grasso, S.; Zappala, M.; Vicik, R.; Schirmeister, T. Nonpeptidic vinyl and allyl phosphonates as falcipain-2 inhibitors. *ChemMedChem* **2008**, *3*, 1030-1033.
78. Ring, C. S.; Sun, E.; McKerrow, J. H.; Lee, G. K.; Rosenthal, P. J.; Kuntz, I. D.; Cohen, F. E. Structure-based inhibitor design by using protein models for the development of antiparasitic agents. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 3583-3587.
79. Li, R.; Chen, X.; Gong, B.; Dominguez, J. N.; Davidson, E.; Kurzban, G.; Miller, R. E.; Nuzum, E. O.; Rosenthal, P. J.; et al. In Vitro Antimalarial Activity of Chalcones and Their Derivatives. *J. Med. Chem.* **1995**, *38*, 5031-5037.
80. Desai, P. V.; Patny, A.; Sabnis, Y.; Tekwani, B.; Gut, J.; Rosenthal, P.; Srivastava, A.; Avery, M. Identification of Novel Parasitic Cysteine Protease Inhibitors Using Virtual Screening. 1. The ChemBridge Database. *J. Med. Chem.* **2004**, *47*, 6609-6615.
81. Desai, P. V.; Patny, A.; Gut, J.; Rosenthal, P. J.; Tekwani, B.; Srivastava, A.; Avery, M. Identification of Novel Parasitic Cysteine Protease Inhibitors by Use of Virtual Screening. 2. The Available Chemical Directory. *J. Med. Chem.* **2006**, *49*, 1576-1584.
82. Li, H.; Huang, J.; Chen, L.; Liu, X.; Chen, T.; Zhu, J.; Lu, W.; Shen, X.; Li, J.; Hilgenfeld, R.; Jiang, H. Identification of Novel Falcipain-2 Inhibitors as Potential Antimalarial Agents through Structure-Based Virtual Screening. *J. Med. Chem., ACS ASAP*.
83. Zheng, S.; Luo, X.; Chen, G.; Zhu, W.; Shen, J.; Chen, K.; Jiang, H. A new rapid and effective chemistry space filter in recognizing a druglike database. *J. Chem. Inf. Model.* **2005**, *45*, 856-862.
84. Zhu, J.; Chen, T.; Liu, J.; Ma, R.; Lu, W.; Huang, J.; Li, H.; Li, J.; Jiang, H. 2-(3,4-dihydro-4-oxothieno[2,3-d]pyrimidin-2-ylthio) acetamides as a new class of falcipain-2 Inhibitors. 3. Design, synthesis and biological evaluation. *Molecules* **2009**, *14*, 785-797.
85. Zhu, J.; Chen, T.; Chen, L.; Lu, W.; Che, P.; Huang, J.; Li, H.; Li, J.; Jiang, H. 2-amido-3-(1H-indol-3-yl)-N-substitued-propanamides as a new class of falcipain-2 inhibitors. 1. Design, synthesis, biological evaluation and binding model studies. *Molecules* **2009**, *14*, 494-508.
86. Desai, P. V.; Patny, A.; Gut, J.; Rosenthal, P. J.; Tekwani, B.; Srivastava, A.; Avery, M. Identification of Novel Parasitic Cysteine Protease Inhibitors by Use of Virtual Screening. 2. The Available Chemical Directory. *J. Med. Chem.* **2006**, *49*, 1576-1584.
87. Chen, M.; Theander, T. G.; Christensen, S. B.; Hviid, L.; Zhai, L.; Kharazmi, A. Licochalcone A, a new antimalarial agent, inhibits in vitro growth of the human malaria parasite Plasmodium falciparum and protects mice from P. yoelii infection. *Antimicrob. Agents Chemother.* **1994**, *38*, 1470-1475.

88. Dominguez, J. N.; Charris, J. E.; Lobo, G.; Gamboa de Dominguez, N.; Moreno, M. M.; Riggione, F.; Sanchez, E.; Olson, J.; Rosenthal, P. J. Synthesis of quinolinyl chalcones and evaluation of their antimalarial activity. *Eur. J. Med. Chem.* **2001**, *36*, 555-560.
89. Liu, M.; Wilairat, P.; Go, M. L. Antimalarial alkoxylated and hydroxylated chalcones [corrected]: structure-activity relationship analysis. *J. Med. Chem.* **2001**, *44*, 4443-4452.
90. Dominguez, J. N.; Leon, C.; Rodrigues, J.; Gamboa de Dominguez, N.; Gut, J.; Rosenthal, P. J. Synthesis and evaluation of new antimalarial phenylurenyl chalcone derivatives. *J. Med. Chem.* **2005**, *48*, 3654-3658.
91. Shuttleworth, S. J.; Nasturica, D.; Gervais, C.; Siddiqui, M. A.; Rando, R. F.; Lee, N. Parallel synthesis of isatin-based serine protease inhibitors. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 2501-2504.
92. Webber, S. E.; Tikhe, J.; Worland, S. T.; Fuhrman, S. A.; Hendrickson, T. F.; Matthews, D. A.; Love, R. A.; Patick, A. K.; Meador, J. W.; Ferre, R. A.; Brown, E. L.; DeLisle, D. M.; Ford, C. E.; Binford, S. L. Design, synthesis, and evaluation of nonpeptidic inhibitors of human rhinovirus 3C protease. *J. Med. Chem.* **1996**, *39*, 5072-5082.
93. Chiyanzu, I.; Hansell, E.; Gut, J.; Rosenthal, P. J.; McKerrow, J. H.; Chibale, K. Synthesis and evaluation of isatins and thiosemicarbazone derivatives against cruzain, falcipain-2 and rhodesain. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3527-3530.
94. Fujii, N.; Mallari, J. P.; Hansell, E. J.; Mackey, Z.; Doyle, P.; Zhou, Y. M.; Gut, J.; Rosenthal, P. J.; McKerrow, J. H.; Guy, R. K. Discovery of potent thiosemicarbazone inhibitors of rhodesain and cruzain. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 121-123.
95. Noedl, H.; Se, Y.; Schaecher, K.; Smith, B. L.; Socheat, D.; Fukuda, M. M. Evidence of artemisinin-resistant malaria in western Cambodia. *N. Engl. J. Med.* **2008**, *359*, 2619-2620.
96. Lee, N.; Hui, D.; Wu, A.; Chan, P.; Cameron, P.; Joynt, G. M.; Ahuja, A.; Yung, M. Y.; Leung, C. B.; To, K. F.; Lui, S. F.; Szeto, C. C.; Chung, S.; Sung, J. J. A major outbreak of severe acute respiratory syndrome in Hong Kong. *N. Engl. J. Med.* **2003**, *348*, 1986-1994.
97. Poutanen, S. M.; Low, D. E.; Henry, B.; Finkelstein, S.; Rose, D.; Green, K.; Tellier, R.; Draker, R.; Adachi, D.; Ayers, M.; Chan, A. K.; Skowronski, D. M.; Salit, I.; Simor, A. E.; Slutsky, A. S.; Doyle, P. W.; Krajden, M.; Petric, M.; Brunham, R. C.; McGeer, A. J. Identification of severe acute respiratory syndrome in Canada. *N Engl J Med* **2003**, *348*, 1995-2005.
98. Tsang, K. W.; Ho, P. L.; Ooi, G. C.; Yee, W. K.; Wang, T.; Chan-Yeung, M.; Lam, W. K.; Seto, W. H.; Yam, L. Y.; Cheung, T. M.; Wong, P. C.; Lam, B.; Ip, M. S.; Chan, J.; Yuen, K. Y.; Lai, K. N. A cluster of cases of severe acute respiratory syndrome in Hong Kong. *N. Engl. J. Med.* **2003**, *348*, 1977-1985.
99. McMasters, D. R.; Garcia-Calvo, M.; Maiorov, V.; McCann, M. E.; Meurer, R. D.; Bull, H. G.; Lisnock, J.; Howell, K. L.; Devita, R. J. Spiroimidazolidinone NPC1L1 inhibitors. 1: Discovery by 3D-similarity-based virtual screening. *Bioorg Med Chem Lett* **2009**, *19*, 2965-2968.
100. Kuiken, T.; Fouchier, R. A. M.; Schutten, M.; Rimmelzwaan, G. F.; van Amerongen, G.; van Riel, D.; Laman, J. D.; de Jong, T.; van Doornum, G.; Lim, W.; Ling, A. E.; Chan, P. K. S.; Tam, J. S.; Zambon, M. C.; Gopal, R.; Drosten, C.; van der Werf, S.; Escriou, N.;

- Manuguerra, J.-C.; Stohr, K.; Peiris, J. S. M.; Osterhaus, A. D. M. E. Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. *Lancet* **2003**, *362*, 263-270.
101. Ksiazek, T. G.; Erdman, D.; Goldsmith, C. S.; Zaki, S. R.; Peret, T.; Emery, S.; Tong, S.; Urbani, C.; Comer, J. A.; Lim, W.; Rollin, P. E.; Dowell, S. F.; Ling, A.-E.; Humphrey, C. D.; Shieh, W.-J.; Guarner, J.; Paddock, C. D.; Rota, P.; Fields, B.; DeRisi, J.; Yang, J.-Y.; Cox, N.; Hughes, J. M.; LeDuc, J. W.; Bellini, W. J.; Anderson, L. J.; Cannon, A. D. L.; Curtis, M.; Farrar, B.; Morgan, L.; Pezzanite, L.; Sanchez, A. J.; Slaughter, K. A.; Stevens, T. L.; Stockton, P. C.; Wagoner, K. D.; Sanchez, A.; Nichol, S.; Vincent, M.; Osborne, J.; Honig, J.; Brickson, B. R.; Holloway, B.; McCaustland, K.; Lingappa, J.; Lowe, L.; Scott, S.; Lu, X.; Villamarzo, Y.; Cook, B.; Chen, Q.; Birge, C.; Shu, B.; Pallansch, M.; Tatti, K. M.; Morken, T.; Smith, C.; Greer, P.; White, E.; McGlothen, T.; Bhatnagar, J.; Patel, M.; Bartlett, J.; Montague, J.; Lee, W.; Packard, M.; Thompson, H. A.; Moen, A.; Fukuda, K.; Uyeki, T.; Harper, S.; Klimov, A.; Lindstrom, S.; Benson, R.; Carlone, G.; Facklam, R.; Fields, P.; Levett, P.; Mayer, L.; Talkington, D.; Thacker, W. L.; Tondella, M. L. C.; Whitney, C.; Robertson, B.; Warnock, D.; Brooks, T.; Schrag, S.; Rosenstein, N.; Arthur, R.; Ganem, D.; Poutanen, S. M.; Chen, T. J.; Hsiao, C. H.; Wai-Fu, N. G.; Ho, M.; Keung, T. K.; Nghiem, K. H.; Nguyen, H. K. L.; Le, M. Q.; Nguyen, H. H. T.; Hoang, L. T.; Vu, T. H.; Vu, H. Q.; Chunsuttiwat, S. A novel coronavirus associated with severe acute respiratory syndrome. *N. Engl. J. Med.* **2003**, *348*, 1953-1966.
 102. Drosten, C.; Preiser, W.; Gunther, S.; Schmitz, H.; Doerr, H. W. Severe acute respiratory syndrome: identification of the etiological agent. *Trends Mol. Med.* **2003**, *9*, 325-327.
 103. Thiel, V.; Ivanov, K. A.; Putics, A.; Hertzog, T.; Schelle, B.; Bayer, S.; Weissbrich, B.; Snijder, E. J.; Rabenau, H.; Doerr, H. W.; Gorbalenya, A. E.; Ziebuhr, J. Mechanisms and enzymes involved in SARS coronavirus genome expression. *J. Gen. Virol.* **2003**, *84*, 2305-2315.
 104. Hsu, M.-F.; Kuo, C.-J.; Chang, K.-T.; Chang, H.-C.; Chou, C.-C.; Ko, T.-P.; Shr, H.-L.; Chang, G.-G.; Wang, A. H. J.; Liang, P.-H. Mechanism of the Maturation Process of SARS-CoV 3CL Protease. *J. Biol. Chem.* **2005**, *280*, 31257-31266.
 105. Ziebuhr, J.; Snijder, E. J.; Gorbalenya, A. E. Virus-encoded proteinases and proteolytic processing in the nidovirales. *J. Gen. Virol.* **2000**, *81*, 853-879.
 106. Ziebuhr, J.; Heusipp, G.; Siddell, S. G. Biosynthesis, purification, and characterization of the human coronavirus 229E 3C-like proteinase. *J. Virol.* **1997**, *71*, 3992-3997.
 107. Xiong, B.; Gui, C.-S.; Xu, X.-Y.; Luo, C.; Chen, J.; Luo, H.-B.; Chen, L.-L.; Li, G.-W.; Sun, T.; Yu, C.-Y.; Yue, L.-D.; Duan, W.-H.; Shen, J.-K.; Qin, L.; Shi, T.-L.; Li, Y.-X.; Chen, K.-X.; Luo, X.-M.; Shen, X.; Shen, J.-H.; Jiang, H.-L. A 3D model of SARS_CoV 3CL proteinase and its inhibitors design by virtual screening. *Acta Pharmacol. Sin.* **2003**, *24*, 497-504.
 108. Bacha, U.; Barrila, J.; Gabelli, S. B.; Kiso, Y.; Mario Amzel, L.; Freire, E. Development of broad-spectrum halomethyl ketone inhibitors against coronavirus main protease 3CL(pro). *Chem. Biol. Drug Des.* **2008**, *72*, 34-49.
 109. Ghosh, A. K.; Xi, K.; Ratia, K.; Santarsiero, B. D.; Fu, W.; Harcourt, B. H.; Rota, P. A.; Baker, S. C.; Johnson, M. E.; Mesecar, A. D. Design and Synthesis of Peptidomimetic

- Severe Acute Respiratory Syndrome Chymotrypsin-like Protease Inhibitors. *J. Med. Chem.* **2005**, *48*, 6767-6771.
110. Lee, T. W.; Cherney, M. M.; Liu, J.; James, K. E.; Powers, J. C.; Eltis, L. D.; James, M. N. Crystal structures reveal an induced-fit binding of a substrate-like Aza-peptide epoxide to SARS coronavirus main peptidase. *J. Mol. Biol.* **2007**, *366*, 916-932.
 111. Tan, J.; Verschueren, K. H. G.; Anand, K.; Shen, J.; Yang, M.; Xu, Y.; Rao, Z.; Bigalke, J.; Heisen, B.; Mesters, J. R.; Chen, K.; Shen, X.; Jiang, H.; Hilgenfeld, R. pH-dependent Conformational Flexibility of the SARS-CoV Main Proteinase (Mpro) Dimer: Molecular Dynamics Simulations and Multiple X-ray Structure Analyses. *J. Mol. Biol.* **2005**, *354*, 25-40.
 112. Verschueren, K. H.; Pumpor, K.; Anemuller, S.; Chen, S.; Mesters, J. R.; Hilgenfeld, R. A structural view of the inactivation of the SARS coronavirus main proteinase by benzotriazole esters. *Chem. Biol.* **2008**, *15*, 597-606.
 113. Xu, T.; Ooi, A.; Lee, H. C.; Wilmouth, R.; Liu, D. X.; Lescar, J. Structure of the SARS coronavirus main proteinase as an active C2 crystallographic dimer. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **2005**, *61*, 964-966.
 114. Xue, X.; Yu, H.; Yang, H.; Xue, F.; Wu, Z.; Shen, W.; Li, J.; Zhou, Z.; Ding, Y.; Zhao, Q.; Zhang, X. C.; Liao, M.; Bartlam, M.; Rao, Z. Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. *J. Virol.* **2008**, *82*, 2515-2527.
 115. Yang, H.; Xie, W.; Xue, X.; Yang, K.; Ma, J.; Liang, W.; Zhao, Q.; Zhou, Z.; Pei, D.; Ziebuhr, J.; Hilgenfeld, R.; Yuen, K. Y.; Wong, L.; Gao, G.; Chen, S.; Chen, Z.; Ma, D.; Bartlam, M.; Rao, Z. Design of wide-spectrum inhibitors targeting Coronavirus main proteases. *PLoS Biol.* **2005**, *3*, 1742-1752.
 116. Yang, H.; Yang, M.; Ding, Y.; Liu, Y.; Lou, Z.; Zhou, Z.; Sun, L.; Mo, L.; Ye, S.; Pang, H.; Gao, G. F.; Anand, K.; Bartlam, M.; Hilgenfeld, R.; Rao, Z. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 13190-13195.
 117. Yin, J.; Niu, C.; Cherney, M. M.; Zhang, J.; Huitema, C.; Eltis, L. D.; Vederas, J. C.; James, M. N. A mechanistic view of enzyme inhibition and peptide hydrolysis in the active site of the SARS-CoV 3C-like peptidase. *J. Mol. Biol.* **2007**, *371*, 1060-1074.
 118. Anand, K.; Palm, G. J.; Mesters, J. R.; Siddell, S. G.; Ziebuhr, J.; Hilgenfeld, R. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *EMBO J.* **2002**, *21*, 3213-3224.
 119. Chen, S.; Luo, H.; Chen, L.; Chen, J.; Shen, J.; Zhu, W.; Chen, K.; Shen, X.; Jiang, H. An overall picture of SARS coronavirus (SARS-CoV) genome-encoded major proteins: structures, functions and drug development. *Curr. Pharm. Des.* **2006**, *12*, 4539-4553.
 120. Shi, J.; Wei, Z.; Song, J. Dissection Study on the Severe Acute Respiratory Syndrome 3C-like Protease Reveals the Critical Role of the Extra Domain in Dimerization of the Enzyme: Defining the extra domain as a new target for design of highly specific protease inhibitors. *J. Biol. Chem.* **2004**, *279*, 24765-24773.
 121. Kuang, W. F.; Chow, L. P.; Wu, M. H.; Hwang, L. H. Mutational and inhibitive analysis of SARS coronavirus 3C-like protease by fluorescence resonance energy transfer-based assays. *Biochem. Biophys. Res. Commun.* **2005**, *331*, 1554-1559.

122. Chen, S.; Chen, L.; Tan, J.; Chen, J.; Du, L.; Sun, T.; Shen, J.; Chen, K.; Jiang, H.; Shen, X. Severe Acute Respiratory Syndrome Coronavirus 3C-like Proteinase N Terminus Is Indispensable for Proteolytic Activity but Not for Enzyme Dimerization: Biochemical and thermodynamic investigation in conjunction with molecular dynamics simulations. *J. Biol. Chem.* **2005**, *280*, 164-173.
123. Hsu, W.-C.; Chang, H.-C.; Chou, C.-Y.; Tsai, P.-J.; Lin, P.-I.; Chang, G.-G. Critical Assessment of Important Regions in the Subunit Association and Catalytic Action of the Severe Acute Respiratory Syndrome Coronavirus Main Protease. *J. Biol. Chem.* **2005**, *280*, 22741-22748.
124. Yang, H.; Xie, W.; Xue, X.; Yang, K.; Ma, J.; Liang, W.; Zhao, Q.; Zhou, Z.; Pei, D.; Ziebuhr, J.; Hilgenfeld, R.; Yuen, K. Y.; Wong, L.; Gao, G.; Chen, S.; Chen, Z.; Ma, D.; Bartlam, M.; Rao, Z. Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biol.* **2005**, *3*, e324.
125. Fan, K.; Wei, P.; Feng, Q.; Chen, S.; Huang, C.; Ma, L.; Lai, B.; Pei, J.; Liu, Y.; Chen, J.; Lai, L. Biosynthesis, Purification, and Substrate Specificity of Severe Acute Respiratory Syndrome Coronavirus 3C-like Proteinase. *J. Biol. Chem.* **2004**, *279*, 1637-1642.
126. Goetz, D. H.; Choe, Y.; Hansell, E.; Chen, Y. T.; McDowell, M.; Jonsson, C. B.; Roush, W. R.; McKerrow, J.; Craik, C. S. Substrate specificity profiling and identification of a new class of inhibitor for the major protease of the SARS coronavirus. *Biochemistry* **2007**, *46*, 8744-8752.
127. Sirois, S.; Wei, D. Q.; Du, Q.; Chou, K. C. Virtual screening for SARS-CoV protease based on KZ7088 pharmacophore points. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1111-1122.
128. Xue, X.; Yang, H.; Shen, W.; Zhao, Q.; Li, J.; Yang, K.; Chen, C.; Jin, Y.; Bartlam, M.; Rao, Z. Production of authentic SARS-CoV M(pro) with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction. *J. Mol. Biol.* **2007**, *366*, 965-975.
129. Dragovich, P. S.; Prins, T. J.; Zhou, R.; Webber, S. E.; Marakovits, J. T.; Fuhrman, S. A.; Patick, A. K.; Matthews, D. A.; Lee, C. A.; Ford, C. E.; Burke, B. J.; Rejto, P. A.; Hendrickson, T. F.; Tuntland, T.; Brown, E. L.; Meador, J. W., 3rd; Ferre, R. A.; Harr, J. E.; Kosa, M. B.; Worland, S. T. Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 4. Incorporation of P1 lactam moieties as L-glutamine replacements. *J. Med. Chem.* **1999**, *42*, 1213-1224.
130. Matthews, D. A.; Dragovich, P. S.; Webber, S. E.; Fuhrman, S. A.; Patick, A. K.; Zalman, L. S.; Hendrickson, T. F.; Love, R. A.; Prins, T. J.; Marakovits, J. T.; Zhou, R.; Tikhe, J.; Ford, C. E.; Meador, J. W.; Ferre, R. A.; Brown, E. L.; Binford, S. L.; Brothers, M. A.; Delisle, D. M.; Worland, S. T. Structure-assisted design of mechanism-based irreversible inhibitors of human rhinovirus 3C protease with potent antiviral activity against multiple rhinovirus serotypes. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 11000-11007.
131. Patick, A. K.; Binford, S. L.; Brothers, M. A.; Jackson, R. L.; Ford, C. E.; Diem, M. D.; Maldonado, F.; Dragovich, P. S.; Zhou, R.; Prins, T. J.; Fuhrman, S. A.; Meador, J. W.; Zalman, L. S.; Matthews, D. A.; Worland, S. T. In vitro antiviral activity of AG7088, a potent inhibitor of human rhinovirus 3C protease. *Antimicrob. Agents Chemother.* **1999**,

- 43, 2444-2450.
132. Ghosh, A. K.; Xi, K.; Ratia, K.; Santarsiero, B. D.; Fu, W.; Harcourt, B. H.; Rota, P. A.; Baker, S. C.; Johnson, M. E.; Mesecar, A. D. Design and synthesis of peptidomimetic severe acute respiratory syndrome chymotrypsin-like protease inhibitors. *J. Med. Chem.* **2005**, *48*, 6767-6771.
 133. Shie, J.-J.; Fang, J.-M.; Kuo, T.-H.; Kuo, C.-J.; Liang, P.-H.; Huang, H.-J.; Wu, Y.-T.; Jan, J.-T.; Cheng, Y.-S. E.; Wong, C.-H. Inhibition of the severe acute respiratory syndrome 3CL protease by peptidomimetic alpha ,beta -unsaturated esters. *Bioorg. Med. Chem.* **2005**, *13*, 5240-5252.
 134. Wu, C.-Y.; Jan, J.-T.; Ma, S.-H.; Kuo, C.-J.; Juan, H.-F.; Cheng, Y.-S. E.; Hsu, H.-H.; Huang, H.-C.; Wu, D.; Brik, A.; Liang, F.-S.; Liu, R.-S.; Fang, J.-M.; Chen, S.-T.; Liang, P.-H.; Wong, C.-H. Small molecules targeting severe acute respiratory syndrome human coronavirus. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 10012-10017.
 135. Wu, C. J.; Jan, J. T.; Chen, C. M.; Hsieh, H. P.; Hwang, D. R.; Liu, H. W.; Liu, C. Y.; Huang, H. W.; Chen, S. C.; Hong, C. F.; Lin, R. K.; Chao, Y. S.; Hsu, J. T. Inhibition of severe acute respiratory syndrome coronavirus replication by niclosamide. *Antimicrob. Agents Chemother.* **2004**, *48*, 2693-2696.
 136. Shie, J.-J.; Fang, J.-M.; Kuo, C.-J.; Kuo, T.-H.; Liang, P.-H.; Huang, H.-J.; Yang, W.-B.; Lin, C.-H.; Chen, J.-L.; Wu, Y.-T.; Wong, C.-H. Discovery of Potent Anilide Inhibitors against the Severe Acute Respiratory Syndrome 3CL Protease. *J. Med. Chem.* **2005**, *48*, 4469-4473.
 137. Chou, K.-C.; Wei, D.-Q.; Zhong, W.-Z. Binding mechanism of coronavirus main proteinase with ligands and its implication to drug design against SARS. *Biochem. Biophys. Res. Commun.* **2003**, *308*, 148-151.
 138. Chou, K. C.; Wei, D. Q.; Du, Q. S.; Sirois, S.; Zhong, W. Z. Progress in computational approach to drug development against SARS. *Curr. Med. Chem.* **2006**, *13*, 3263-3270.
 139. Du, Q.-S.; Wang, S.-Q.; Zhu, Y.; Wei, D.-Q.; Guo, H.; Sirois, S.; Chou, K.-C. Polyprotein cleavage mechanism of SARS CoV Mpro and chemical modification of the octapeptide. *Peptides (N. Y., NY, U. S.)* **2004**, *25*, 1857-1864.
 140. Jain, R. P.; Vederas, J. C. Structural variations in keto-glutamines for improved inhibition against hepatitis A virus 3C proteinase. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3655-3658.
 141. Ramtohl, Y. K.; James, M. N.; Vederas, J. C. Synthesis and evaluation of keto-glutamine analogues as inhibitors of hepatitis A virus 3C proteinase. *J. Org. Chem.* **2002**, *67*, 3169-3178.
 142. Jain, R. P.; Pettersson, H. I.; Zhang, J.; Aull, K. D.; Fortin, P. D.; Huitema, C.; Eltis, L. D.; Parrish, J. C.; James, M. N. G.; Wishart, D. S.; Vederas, J. C. Synthesis and Evaluation of Keto-Glutamine Analogues as Potent Inhibitors of Severe Acute Respiratory Syndrome 3CLpro. *J. Med. Chem.* **2004**, *47*, 6113-6116.
 143. Bacha, U.; Barrila, J.; Velazquez-Campoy, A.; Leavitt, S. A.; Freire, E. Identification of Novel Inhibitors of the SARS Coronavirus Main Protease 3CLpro. *Biochemistry* **2004**, *43*, 4906-4912.
 144. Zhang, H. Z.; Zhang, H.; Kemnitzer, W.; Tseng, B.; Cinatl, J., Jr.; Michaelis, M.; Doerr, H. W.; Cai, S. X. Design and synthesis of dipeptidyl glutaminy fluoromethyl ketones as potent severe acute respiratory syndrome coronavirus (SARS-CoV) inhibitors. *J. Med.*

- Chem.* **2006**, *49*, 1198-1201.
145. Kong, J. S.; Venkatraman, S.; Furness, K.; Nimkar, S.; Shepherd, T. A.; Wang, Q. M.; Aube, J.; Hanzlik, R. P. Synthesis and evaluation of peptidyl Michael acceptors that inactivate human rhinovirus 3C protease and inhibit virus replication. *J. Med. Chem.* **1998**, *41*, 2579-2587.
 146. Morris, T. S.; Frommann, S.; Shechosky, S.; Lowe, C.; Lall, M. S.; Gauss-Muller, V.; Purcell, R. H.; Emerson, S. U.; Vederas, J. C.; Malcolm, B. A. In vitro and ex vivo inhibition of hepatitis A virus 3C proteinase by a peptidyl monofluoromethyl ketone. *Bioorg. Med. Chem.* **1997**, *5*, 797-807.
 147. Gelb, M. H.; Svaren, J. P.; Abeles, R. H. Fluoro ketone inhibitors of hydrolytic enzymes. *Biochemistry* **1985**, *24*, 1813-1817.
 148. Patel, D. V.; Rielly-Gauvin, K.; Ryono, D. E.; Free, C. A.; Smith, S. A.; Petrillo, E. W., Jr. Activated ketone based inhibitors of human renin. *J. Med. Chem.* **1993**, *36*, 2431-2447.
 149. Sydnes, M. O.; Hayashi, Y.; Sharma, V. K.; Hamada, T.; Bacha, U.; Barrila, J.; Freire, E.; Kiso, Y. Synthesis of glutamic acid and glutamine peptides possessing a trifluoromethyl ketone group as SARS-CoV 3CL protease inhibitors. *Tetrahedron* **2006**, *62*, 8601-8609.
 150. Regnier, T.; Sarma, D.; Hidaka, K.; Bacha, U.; Freire, E.; Hayashi, Y.; Kiso, Y. New developments for the design, synthesis and biological evaluation of potent SARS-CoV 3CL(pro) inhibitors. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 2722-2727.
 151. Shao, Y. M.; Yang, W. B.; Kuo, T. H.; Tsai, K. C.; Lin, C. H.; Yang, A. S.; Liang, P. H.; Wong, C. H. Design, synthesis, and evaluation of trifluoromethyl ketones as inhibitors of SARS-CoV 3CL protease. *Bioorg. Med. Chem.* **2008**, *16*, 4652-4660.
 152. James, K. E.; Gotz, M. G.; Caffrey, C. R.; Hansell, E.; Carter, W.; Barrett, A. J.; McKerrow, J. H.; Powers, J. C. Aza-peptide epoxides: potent and selective inhibitors of *Schistosoma mansoni* and pig kidney legumains (asparaginyl endopeptidases). *Biol. Chem.* **2003**, *384*, 1613-1618.
 153. James, K. E.; Asgian, J. L.; Li, Z. Z.; Ekici, O. D.; Rubin, J. R.; Mikolajczyk, J.; Salvesen, G. S.; Powers, J. C. Design, synthesis, and evaluation of aza-peptide epoxides as selective and potent inhibitors of caspases-1, -3, -6, and -8. *J. Med. Chem.* **2004**, *47*, 1553-1574.
 154. Lee, T.-W.; Cherney, M. M.; Huitema, C.; Liu, J.; James, K. E.; Powers, J. C.; Eltis, L. D.; James, M. N. G. Crystal Structures of the Main Peptidase from the SARS Coronavirus Inhibited by a Substrate-like Aza-peptide Epoxide. *J. Mol. Biol.* **2005**, *353*, 1137-1151.
 155. Barabas, E.; Szell, E.; Bajusz, S. Screening for fibrinolysis inhibitory effect of synthetic thrombin inhibitors. *Blood Coagul. Fibrinolysis* **1993**, *4*, 243-248.
 156. McConnell, R. M.; York, J. L.; Frizzell, D.; Ezell, C. Inhibition studies of some serine and thiol proteinases by new leupeptin analogues. *J. Med. Chem.* **1993**, *36*, 1084-1089.
 157. Fehrentz, J. A.; Heitz, A.; Castro, B.; Cazaubon, C.; Nisato, D. Aldehydic peptides inhibiting renin. *FEBS Lett* **1984**, *167*, 273-276.
 158. Sarubbi, E.; Seneci, P. F.; Angelastro, M. R.; Peet, N. P.; Denaro, M.; Islam, K. Peptide aldehydes as inhibitors of HIV protease. *FEBS Lett.* **1993**, *319*, 253-256.

159. Thornberry, N. A.; Bull, H. G.; Calaycay, J. R.; Chapman, K. T.; Howard, A. D.; Kostura, M. J.; Miller, D. K.; Molineaux, S. M.; Weidner, J. R.; Aunins, J.; et al. A novel heterodimeric cysteine protease is required for interleukin-1 beta processing in monocytes. *Nature* **1992**, *356*, 768-774.
160. Graybill, T. L.; Dolle, R. E.; Helaszek, C. T.; Miller, R. E.; Ator, M. A. Preparation and evaluation of peptidic aspartyl hemiacetals as reversible inhibitors of interleukin-1 beta converting enzyme (ICE). *Int. J. Pept. Protein Res.* **1994**, *44*, 173-182.
161. Al-Gharabli, S. I.; Shah, S. T. A.; Weik, S.; Schmidt, M. F.; Mesters, J. R.; Kuhn, D.; Klebe, G.; Hilgenfeld, R.; Rademann, J. An efficient method for the synthesis of peptide aldehyde libraries employed in the discovery of reversible SARS coronavirus main protease (SARS-CoV Mpro) inhibitors. *ChemBioChem* **2006**, *7*, 1048-1055.
162. Liu, Z.; Huang, C.; Fan, K.; Wei, P.; Chen, H.; Liu, S.; Pei, J.; Shi, L.; Li, B.; Yang, K.; Liu, Y.; Lai, L. Virtual Screening of Novel Noncovalent Inhibitors for SARS-CoV 3C-like Proteinase. *J. Chem. Inf. Comput. Sci.* **2005**, *45*, 10-17.
163. Chen, L.; Gui, C.; Luo, X.; Yang, Q.; Guenther, S.; Scandella, E.; Drosten, C.; Bai, D.; He, X.; Ludewig, B.; Chen, J.; Luo, H.; Yang, Y.; Yang, Y.; Zou, J.; Thiel, V.; Chen, K.; Shen, J.; Shen, X.; Jiang, H. Cinanserin is an inhibitor of the 3C-like proteinase of severe acute respiratory syndrome coronavirus and strongly reduces virus replication in vitro. *J. Virol.* **2005**, *79*, 7095-7103.
164. Clark, R. D.; Strizhev, A.; Leonard, J. M.; Blake, J. F.; Matthew, J. B. Consensus scoring for ligand/protein interactions. *J. Mol. Graph. Model.* **2002**, *20*, 281-295.
165. Lu, I. L.; Mahindroo, N.; Liang, P.-H.; Peng, Y.-H.; Kuo, C.-J.; Tsai, K.-C.; Hsieh, H.-P.; Chao, Y.-S.; Wu, S.-Y. Structure-Based Drug Design and Structural Biology Study of Novel Nonpeptide Inhibitors of Severe Acute Respiratory Syndrome Coronavirus Main Protease. *J. Med. Chem.* **2006**, *49*, 5154-5161.
166. Dooley, A. J.; Shindo, N.; Taggart, B.; Park, J.-G.; Pang, Y.-P. From genome to drug lead: Identification of a small-molecule inhibitor of the SARS virus. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 830-833.
167. Tsai, K. C.; Chen, S. Y.; Liang, P. H.; Lu, I. L.; Mahindroo, N.; Hsieh, H. P.; Chao, Y. S.; Liu, L.; Liu, D.; Lien, W.; Lin, T. H.; Wu, S. Y. Discovery of a novel family of SARS-CoV protease inhibitors by virtual screening and 3D-QSAR studies. *J. Med. Chem.* **2006**, *49*, 3485-3495.
168. Yang, H.; Bartlam, M.; Rao, Z. Drug design targeting the main protease, the achilles' heel of coronaviruses. *Curr. Pharm. Des.* **2006**, *12*, 4573-4590.
169. Kao, R. Y.; Tsui, W. H. W.; Lee, T. S. W.; Tanner, J. A.; Watt, R. M.; Huang, J.-D.; Hu, L.; Chen, G.; Chen, Z.; Zhang, L.; He, T.; Chan, K.-H.; Tse, H.; To, A. P. C.; Ng, L. W. Y.; Wong, B. C. W.; Tsoi, H.-W.; Yang, D.; Ho, D. D.; Yuen, K.-Y. Identification of Novel Small-Molecule Inhibitors of Severe Acute Respiratory Syndrome-Associated Coronavirus by Chemical Genetics. *Chem. Biol.* **2004**, *11*, 1293-1299.
170. Blanchard, J. E.; Elowe, N. H.; Huitema, C.; Fortin, P. D.; Cechetto, J. D.; Eltis, L. D.; Brown, E. D. High-Throughput Screening Identifies Inhibitors of the SARS Coronavirus Main Proteinase. *Chem. Biol.* **2004**, *11*, 1445-1453.
171. Kuo, C. J.; Liu, H. G.; Lo, Y. K.; Seong, C. M.; Lee, K. I.; Jung, Y. S.; Liang, P. H. Individual and common inhibitors of coronavirus and picornavirus main proteases. *FEBS*

- Lett.* **2009**, 583, 549-555.
172. Hsu, J. T. A.; Kuo, C.-J.; Hsieh, H.-P.; Wang, Y.-C.; Huang, K.-K.; Lin, C. P. C.; Huang, P.-F.; Chen, X.; Liang, P.-H. Evaluation of metal-conjugated compounds as inhibitors of 3CL protease of SARS-CoV. *FEBS Lett.* **2004**, 574, 116-120.
 173. Liu, Y.-C.; Huang, V.; Chao, T.-C.; Hsiao, C.-D.; Lin, A.; Chang, M.-F.; Chow, L.-P. Screening of drugs by FRET analysis identifies inhibitors of SARS-CoV 3CL protease. *Biochem. Biophys. Res. Commun.* **2005**, 333, 194-199.
 174. Chen, C. N.; Lin, C. P.; Huang, K. K.; Chen, W. C.; Hsieh, H. P.; Liang, P. H.; Hsu, J. T. Inhibition of SARS-CoV 3C-like Protease Activity by Theaflavin-3,3'-digallate (TF3). *Evid Based Complement Alternat Med* **2005**, 2, 209-215.
 175. Kaeppler, U.; Stiefl, N.; Schiller, M.; Vicik, R.; Breuning, A.; Schmitz, W.; Rupprecht, D.; Schmuck, C.; Baumann, K.; Ziebuhr, J.; Schirmeister, T. A new lead for nonpeptidic active-site-directed inhibitors of the severe acute respiratory syndrome coronavirus main protease discovered by a combination of screening and docking methods. *J. Med. Chem.* **2005**, 48, 6832-6842.
 176. Lin, C.-W.; Tsai, F.-J.; Tsai, C.-H.; Lai, C.-C.; Wan, L.; Ho, T.-Y.; Hsieh, C.-C.; Chao, P.-D. L. Anti-SARS coronavirus 3C-like protease effects of *Isatis indigotica* root and plant-derived phenolic compounds. *Antiviral Res.* **2005**, 68, 36-42.
 177. Jenwitheesuk, E.; Samudrala, R. Identifying inhibitors of the SARS coronavirus proteinase. *Bioorg. Med. Chem. Lett.* **2003**, 13, 3989-3992.
 178. Plewczynski, D.; Hoffmann, M.; von Grothuss, M.; Ginalski, K.; Rychewski, L. In silico prediction of SARS protease inhibitors by virtual high throughput screening. *Chem. Biol. Drug Des.* **2007**, 69, 269-279.
 179. Zhang, X. W.; Yap, Y. L. Old drugs as lead compounds for a new disease? Binding analysis of SARS coronavirus main proteinase with HIV, psychotic and parasite drugs. *Bioorg. Med. Chem.* **2004**, 12, 2517-2521.
 180. Savarino, A. Expanding the frontiers of existing antiviral drugs: possible effects of HIV-1 protease inhibitors against SARS and avian influenza. *J. Clin. Virol.* **2005**, 34, 170-178.
 181. Zhang, X. W.; Yap, Y. L.; Altmeyer, R. M. Generation of predictive pharmacophore model for SARS-coronavirus main proteinase. *Eur. J. Med. Chem.* **2005**, 40, 57-62.
 182. Schmidt, M. F.; Isidro-Llobet, A.; Lisurek, M.; El-Dahshan, A.; Tan, J.; Hilgenfeld, R.; Rademann, J. Sensitized detection of inhibitory fragments and iterative development of non-peptidic protease inhibitors by dynamic ligation screening. *Angew. Chem. Int. Ed. Engl.* **2008**, 47, 3275-3278.
 183. Coteron, J. M.; Catterick, D.; Castro, J.; Chaparro, M. J.; Diaz, B.; Fernandez, E.; Ferrer, S.; Gamo, F. J.; Gordo, M.; Gut, J.; de las Heras, L.; Legac, J.; Marco, M.; Miguel, J.; Munoz, V.; Porras, E.; de la Rosa, J. C.; Ruiz, J. R.; Sandoval, E.; Ventosa, P.; Rosenthal, P. J.; Fiandor, J. M. Falcipain inhibitors: optimization studies of the 2-pyrimidinecarbonitrile lead series. *J. Med. Chem.* **2010**, 53, 6129-6152.
 184. Gamboa de Dominguez, N. D.; Rosenthal, P. J. Cysteine proteinase inhibitors block early steps in hemoglobin degradation by cultured malaria parasites. *Blood* **1996**, 87, 4448-4454.
 185. Ramjee, M. K.; Flinn, N. S.; Pemberton, T. P.; Quibell, M.; Wang, Y.; Watts, J. P. Substrate mapping and inhibitor profiling of falcipain-2, falcipain-3 and berghepain-2:

- Implications for peptidase anti-malarial drug discovery. *Biochem. J* **2006**, *399*, 47-57.
186. Li, H.; Huang, J.; Chen, L.; Liu, X.; Chen, T.; Zhu, J.; Lu, W.; Shen, X.; Li, J.; Hilgenfeld, R.; Jiang, H. Identification of Novel Falcipain-2 Inhibitors as Potential Antimalarial Agents through Structure-Based Virtual Screening. *J. Med. Chem.* **2009**, *52*, 4936-4940.
 187. Oprea, T. I.; Matter, H. Integrating virtual screening in lead discovery. *Curr. Opin. Chem. Biol.* **2004**, *8*, 349-358.
 188. Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J Med Chem* **2000**, *43*, 4759-4767.
 189. Shah, F.; Mukherjee, P.; Desai, P.; Avery, M. Computational approaches for the discovery of cysteine protease inhibitors against malaria and SARS. *Curr. Comput. Aided Drug Des.* **2010**, *6*, 1-23.
 190. Mukherjee, P.; Desai, P.; Zhou, Y. D.; Avery, M. Targeting the BH3 Domain Mediated Protein-Protein Interaction of Bcl-xL through Virtual Screening. *J. Chem. Inf. Model.* **2010**, *50*, 906-923.
 191. Vijayan, R. S. K.; Prabu, M.; Mascarenhas, N. M.; Ghoshal, N. Hybrid Structure-Based Virtual Screening Protocol for the Identification of Novel BACE1 Inhibitors. *Journal of chemical information and modeling* **2009**, *49*, 647-657.
 192. Gamo, F. J.; Sanz, L. M.; Vidal, J.; de Cozar, C.; Alvarez, E.; Lavandera, J. L.; Vanderwall, D. E.; Green, D. V.; Kumar, V.; Hasan, S.; Brown, J. R.; Peishoff, C. E.; Cardon, L. R.; Garcia-Bustos, J. F. Thousands of chemical starting points for antimalarial lead identification. *Nature* **2010**, *465*, 305-310.
 193. Guiguemde, W. A.; Shelat, A. A.; Bouck, D.; Duffy, S.; Crowther, G. J.; Davis, P. H.; Smithson, D. C.; Connelly, M.; Clark, J.; Zhu, F.; Jimenez-Diaz, M. B.; Martinez, M. S.; Wilson, E. B.; Tripathi, A. K.; Gut, J.; Sharlow, E. R.; Bathurst, I.; El Mazouni, F.; Fowble, J. W.; Forquer, I.; McGinley, P. L.; Castro, S.; Angulo-Barturen, I.; Ferrer, S.; Rosenthal, P. J.; Derisi, J. L.; Sullivan, D. J.; Lazo, J. S.; Roos, D. S.; Riscoe, M. K.; Phillips, M. A.; Rathod, P. K.; Van Voorhis, W. C.; Avery, V. M.; Guy, R. K. Chemical genetics of Plasmodium falciparum. *Nature* **2010**, *465*, 311-315.
 194. *Glide*, version 5.0; Schrodinger, LLC, New York, NY, 2008.
 195. *QikProp*, version 3.1; Schrodinger, LLC, New York, NY, 2008.
 196. *Phase*, version 3.0; Schrodinger, LLC, New York, NY, 2008.
 197. *Sybyl*, version 8.1; Tripos Inc., St. Louis, MO, 2008.
 198. *Jaguar* version 7.6; Schrödinger, LLC, New York, NY, 2009.
 199. *Pymol*, Schrodinger, LLC, New York, NY, 2008.
 200. Brinen, L. S.; Hansell, E.; Cheng, J.; Roush, W. R.; McKerrow, J. H.; Fletterick, R. J. A target within the target: probing cruzain's P1' site to define structural determinants for the Chagas' disease protease. *Structure* **2000**, *8*, 831-840.
 201. Holm, L.; Park, J. DaliLite workbench for protein structure comparison. *Bioinformatics* **2000**, *16*, 566.
 202. Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49*, 6789-6801.
 203. Subramanian, S.; Hardt, M.; Choe, Y.; Niles, R. K.; Johansen, E. B.; Legac, J.; Gut, J.;

- Kerr, I. D.; Craik, C. S.; Rosenthal, P. J. Hemoglobin cleavage site-specificity of the Plasmodium falciparum cysteine proteases falcipain-2 and falcipain-3. *PLoS One* **2009**, *4*, e5156.
204. (Accessed)
205. Lambros, C.; Vanderberg, J. P. Synchronization of Plasmodium falciparum erythrocytic stages in culture. *J. Parasitol.* **1979**, *65*, 418-420.
206. Babich, H.; Borenfreund, E. Cytotoxicity of T-2 toxin and its metabolites determined with the neutral red cell viability assay. *Appl. Environ. Microbiol.* **1991**, *57*, 2101-2103.
207. Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B* **1988**, *37*, 785-789.
208. Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *Chem. Phys* **1993**, *98*, 5648-5652.
209. Bernhard Schlegel, H. Estimating the Hessian for gradient-type geometry optimizations. *Theoretica. Chimica. Acta.* **1984**, *66*, 333-340.
210. Contreras, R. R.; Fuentealba, P.; Galván, M.; Pérez, P. A direct evaluation of regional Fukui functions in molecules. *Chem. Phys. Lett.* **1999**, *304*, 405-413.
211. Bulat, F. A.; Chamorro, E.; Fuentealba, P.; Toro-Labbé, A. Condensation of frontier molecular orbital Fukui functions. *J. Phys. Chem. A* **2004**, *108*, 342-349.
212. Jansen, J. M.; Martin, E. J. Target-biased scoring approaches and expert systems in structure-based virtual screening. *Curr. Opin. Chem. Biol.* **2004**, *8*, 359-364.
213. Lee, H. J.; Shieh, C. K.; Gorbalenya, A. E.; Koonin, E. V.; La Monica, N.; Tuler, J.; Bagdzhadzhyan, A.; Lai, M. M. C. The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative proteases and RNA polymerase. *Virology* **1991**, *180*, 567-582.
214. Nägler, D. K.; Ménard, R. Human cathepsin X: A novel cysteine protease of the papain family with a very short proregion and unique insertions1. *FEBS Lett.* **1998**, *434*, 135-139.
215. Setti, E. L.; Davis, D.; Chung, T.; McCarter, J. 3,4-disubstituted azetidiones as selective inhibitors of the cysteine protease cathepsin K. Exploring P2 elements for selectivity. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2051-2053.
216. Kerr, I. D.; Lee, J. H.; Farady, C. J.; Marion, R.; Rickert, M.; Sajid, M.; Pandey, K. C.; Caffrey, C. R.; Legac, J.; Hansell, E. Vinyl sulfones as antiparasitic agents and a structural basis for drug design. *J. Biol. Chem.* **2009**, *284*, 25697.
217. Brak, K.; Kerr, I. D.; Barrett, K. T.; Fuchi, N.; Debnath, M.; Ang, K.; Engel, J. C.; McKerrow, J. H.; Doyle, P. S.; Brinen, L. S. Nonpeptidic tetrafluorophenoxymethyl ketone cruzain inhibitors as promising new leads for chagas disease chemotherapy. *J. Med. Chem.* *53*, 1763-1773.
218. Johnson, D. S.; Weerapana, E.; Cravatt, B. F. Strategies for discovering and derisking covalent, irreversible enzyme inhibitors. *Future Med Chem* **2010**, *2*, 949-964.
219. Potashman, M. H.; Duggan, M. E. Covalent modifiers: an orthogonal approach to drug design. *J. Med. Chem.* **2009**, *52*, 1231-1246.
220. Marquis, R. W.; Ru, Y.; Yamashita, D. S.; Oh, H. J.; Yen, J.; Thompson, S. K.; Carr, T. J.; Levy, M. A.; Tomaszek, T. A.; Ijames, C. F.; Smith, W. W.; Zhao, B.; Janson, C. A.; Abdel-Meguid, S. S.; D'Alessio, K. J.; McQueney, M. S.; Veber, D. F. Potent

- dipeptidylketone inhibitors of the cysteine protease cathepsin K. *Bioorg. Med. Chem.* **1999**, *7*, 581-588.
221. LaLonde, J. M.; Zhao, B.; Smith, W. W.; Janson, C. A.; DesJarlais, R. L.; Tomaszek, T. A.; Carr, T. J.; Thompson, S. K.; Oh, H. J.; Yamashita, D. S.; Veber, D. F.; Abdel-Meguid, S. S. Use of papain as a model for the structure-based design of cathepsin K inhibitors: crystal structures of two papain-inhibitor complexes demonstrate binding to S'-subsites. *J. Med. Chem.* **1998**, *41*, 4567-4576.
222. Steert, K.; Berg, M.; Mottram, J. C.; Westrop, G. D.; Coombs, G. H.; Cos, P.; Maes, L.; Joossens, J.; Van der Veken, P.; Haemers, A.; Augustyns, K. alpha-ketoheterocycles as inhibitors of *Leishmania mexicana* cysteine protease CPB. *ChemMedChem* *5*, 1734-1748.
223. Lee, B. J.; Singh, A.; Chiang, P.; Kemp, S. J.; Goldman, E. A.; Weinhouse, M. I.; Vlasuk, G. P.; Rosenthal, P. J. Antimalarial activities of novel synthetic cysteine protease inhibitors. *Antimicrob. Agents Chemother.* **2003**, *47*, 3810-3814.
224. *Jaguar user manual*, version 7.6; Schrödinger, LLC, New York, NY, 2009.
225. Rietjens, I.; Soffers, A.; Hooiveld, G.; Veeger, C.; Vervoort, J. Quantitative structure-activity relationships based on computer calculated parameters for the overall rate of glutathione S-transferase catalyzed conjugation of a series of fluoronitrobenzenes. *Chem. Res. Toxicol.* **1995**, *8*, 481-488.
226. Dominguez, J. N.; Lopez, S.; Charris, J.; Iarruso, L.; Lobo, G.; Semenov, A.; Olson, J. E.; Rosenthal, P. J. Synthesis and antimalarial effects of phenothiazine inhibitors of a *Plasmodium falciparum* cysteine protease. *J. Med. Chem.* **1997**, *40*, 2726-2732.
227. Pinto, A. V.; de Castro, S. L. The trypanocidal activity of naphthoquinones: a review. *Molecules* **2009**, *14*, 4570-4590.
228. Sabnis, Y. A.; Desai, P. V.; Rosenthal, P. J.; Avery, M. A. Probing the structure of falcipain 3, a cysteine protease from *Plasmodium falciparum*: Comparative protein modeling and docking studies. *Protein Sci.* **2003**, *12*, 501-509.
229. Shah, F.; Mukherjee, P.; Gut, J.; Legac, J.; Rosenthal, P. J.; Tekwani, B. L.; Avery, M. A. Identification of Novel Malarial Cysteine Protease Inhibitors Using Structure-Based Virtual Screening of a Focused Cysteine Protease Inhibitor Library. *J. Chem. Inf. Mod.* **2011**, *51*, 852-864.
230. Ciapetti, P.; Giethlen, B., *Molecular Variation Based on Isosteric Replacements*. Third ed.; Academic Press: 2008.
231. Sijwali, P. S.; Koo, J.; Singh, N.; Rosenthal, P. J. Gene disruptions demonstrate independent roles for the four falcipain cysteine proteases of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **2006**, *150*, 96-106.
232. Huang, L.; Lee, A.; Ellman, J. A. Identification of potent and selective mechanism-based inhibitors of the cysteine protease cruzain using solid-phase parallel synthesis. *J. Med. Chem.* **2002**, *45*, 676-684.
233. Brak, K.; Doyle, P. S.; McKerrow, J. H.; Ellman, J. A. Identification of a new class of nonpeptidic inhibitors of cruzain. *J. Am. Chem. Soc.* **2008**, *130*, 6404-6410.
234. Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J. Am. Chem. Soc.* **2008**, *130*, 2817-2831.
235. Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. Motifs for molecular

- recognition exploiting hydrophobic enclosure in protein–ligand binding. *Proceedings of the National Academy of Sciences* **2007**, *104*, 808.
236. Pellicciari, R.; Camaioni, E.; Gilbert, A. M.; Macchiarulo, A.; Bikker, J. A.; Shah, F.; Bard, J.; Costantino, G.; Gioiello, A.; Robertson, G. M. Discovery and characterization of novel potent PARP-1 inhibitors endowed with neuroprotective properties: From TIQ-A to HYDAMTIQ. *Med. Chem. Commun.* **2011**, *2*, 559-565.
237. Chrencik, J. E.; Patny, A.; Leung, I. K.; Korniski, B.; Emmons, T. L.; Hall, T.; Weinberg, R. A.; Gormley, J. A.; Williams, J. M.; Day, J. E. Structural and Thermodynamic Characterization of the TYK2 and JAK3 Kinase Domains in Complex with CP-690550 and CMP-6. *J. Mol. Biol.* **400**, 413-433.
238. Abel, R.; Salam, N. K.; Shelley, J.; Farid, R.; Friesner, R. A.; Sherman, W. Contribution of Explicit Solvent Effects to the Binding Affinity of Small Molecule Inhibitors in Blood Coagulation Factor Serine Proteases. *ChemMedChem* **6**, 1049-1066.
239. Higgs, C.; Beuming, T.; Sherman, W. Hydration Site Thermodynamics Explain SARs for Triazolylpurines Analogues Binding to the A2A Receptor. *ACS Medicinal Chemistry Letters* **1**, 160-164.
240. Robinson, D. D.; Sherman, W.; Farid, R. Understanding kinase selectivity through energetic analysis of binding site waters. *ChemMedChem* **5**, 618-627.
241. Snyder, P. W.; MecinoviÄ‡, J.; Moustakas, D. T.; Thomas Iii, S. W.; Harder, M.; Mack, E. T.; Lockett, M. R.; HÃ©roux, A.; Sherman, W.; Whitesides, G. M. Mechanism of the hydrophobic effect in the biomolecular recognition of arylsulfonamides by carbonic anhydrase. *Proceedings of the National Academy of Sciences*.
242. Oballa, R. M.; Truchon, J. F.; Bayly, C. I.; Chauret, N.; Day, S.; Crane, S.; Berthelette, C. A generally applicable method for assessing the electrophilicity and reactivity of diverse nitrile-containing compounds. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 998-1002.
243. Lazaridis, T. Inhomogeneous fluid approach to solvation thermodynamics. 1. Theory. *The Journal of Physical Chemistry B* **1998**, *102*, 3531-3541.
244. Lazaridis, T. Inhomogeneous fluid approach to solvation thermodynamics. 2. Applications to simple fluids. *The Journal of Physical Chemistry B* **1998**, *102*, 3542-3550.
245. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society* **1996**, *118*, 11225-11236.
246. Shah, F.; Wu, Y.; Gut, J.; Pedduri, Y.; Legac, J.; Rosenthal, P. J.; Avery, M. A. Design, synthesis and biological evaluation of novel benzothiazole and triazole analogs as falcipain inhibitors. *Med. Chem. Commun.* **2011**, DOI: 10.1039/C1031MD00129A.
247. Anand, K.; Ziebuhr, J.; Wadhvani, P.; Mesters, J. R.; Hilgenfeld, R. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* **2003**, *300*, 1763.
248. Massova, I.; Kollman, P. A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discovery Des.* **2000**, *18*, 113-135.
249. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127-

- 6129.
250. Ramjee, M. K.; Flinn, N. S.; Pemberton, T. P.; Quibell, M.; Wang, Y.; Watts, J. P. Substrate mapping and inhibitor profiling of falcipain-2, falcipain-3 and berghepain-2: implications for peptidase anti-malarial drug discovery. *Biochem. J* **2006**, *399*, 47.
 251. Beuming, T.; Farid, R.; Sherman, W. High energy water sites determine peptide binding affinity and specificity of PDZ domains. *Protein Science* **2009**, *18*, 1609-1619.
 252. Sabnis, Y.; Rosenthal, P. J.; Desai, P.; Avery, M. A. Homology modeling of falcipain-2: validation, de novo ligand design and synthesis of novel inhibitors. *J Biomol Struct Dyn* **2002**, *19*, 765-774.
 253. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Daniel, T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739-1749.
 254. Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006**, *49*, 6177-6196.
 255. Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J. F.; Honig, B.; Shaw, D. E.; Friesner, R. A. A hierarchical approach to all-atom protein loop prediction. *PROTEINS-NEW YORK-* **2004**, *55*, 351-367.
 256. Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B. On the role of the crystal environment in determining protein side-chain conformations. *Journal of molecular biology* **2002**, *320*, 597-608.
 257. Pearlstein, R. A.; Hu, Q. Y.; Zhou, J.; Yowe, D.; Levell, J.; Dale, B.; Kaushik, V. K.; Daniels, D.; Hanrahan, S.; Sherman, W. New hypotheses about the structure–function of proprotein convertase subtilisin/kexin type 9: Analysis of the epidermal growth factor like repeat A docking site using WaterMap. *Proteins: Structure, Function, and Bioinformatics* *78*, 2571-2586.
 258. Pellicciari, R.; Camaioni, E.; Gilbert, A. M.; Macchiarulo, A.; Bikker, J. A.; Shah, F.; Bard, J.; Costantino, G.; Gioiello, A.; Robertson, G. M. Discovery and characterization of novel potent PARP-1 inhibitors endowed with neuroprotective properties: From TIQ-A to HYDAMTIQ. *Med. Chem. Commun.* **2011**.
 259. Shah, F.; Mukherjee, P.; Gut, J.; Legac, J.; Rosenthal, P. J.; Tekwani, B. L.; Avery, M. A. Identification of Novel Malarial Cysteine Protease Inhibitors Using Structure-Based Virtual Screening of a Focused Cysteine Protease Inhibitor Library. *Journal of chemical information and modeling*.
 260. Fouchier, R. A.; Kuiken, T.; Schutten, M.; van Amerongen, G.; van Doornum, G. J.; van den Hoogen, B. G.; Peiris, M.; Lim, W.; Stohr, K.; Osterhaus, A. D. Aetiology: Koch's postulates fulfilled for SARS virus. *Nature* **2003**, *423*, 240.
 261. Tsang Kenneth, W.; Ho Pak, L.; Ooi Gaik, C.; Yee Wilson, K.; Wang, T.; Chan-Yeung, M.; Lam Wah, K.; Seto Wing, H.; Yam Loretta, Y.; Cheung Thomas, M.; Wong Poon, C.; Lam, B.; Ip Mary, S.; Chan, J.; Yuen Kwok, Y.; Lai Kar, N. A cluster of cases of severe acute respiratory syndrome in Hong Kong. *The New England journal of medicine* **2003**, *348*, 1977-1985.

262. Drosten, C.; Preiser, W.; Gunther, S.; Schmitz, H.; Doerr Hans, W. Severe acute respiratory syndrome: identification of the etiological agent. *Trends in molecular medicine* **2003**, *9*, 325-327.
263. Li, W.; Shi, Z.; Yu, M.; Ren, W.; Smith, C.; Epstein, J. H.; Wang, H.; Crameri, G.; Hu, Z.; Zhang, H.; Zhang, J.; McEachern, J.; Field, H.; Daszak, P.; Eaton, B. T.; Zhang, S.; Wang, L.-F. Bats are natural reservoirs of SARS-like coronaviruses. *Science* **2005**, *310*, 676-679.
264. Lau, S. K. P.; Woo, P. C. Y.; Li, K. S. M.; Huang, Y.; Tsoi, H.-W.; Wong, B. H. L.; Wong, S. S. Y.; Leung, S.-Y.; Chan, K.-H.; Yuen, K.-Y. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 14040-14045.
265. Guan, Y.; Zheng, B. J.; He, Y. Q.; Liu, X. L.; Zhuang, Z. X.; Cheung, C. L.; Luo, S. W.; Li, P. H.; Zhang, L. J.; Guan, Y. J.; Butt, K. M.; Wong, K. L.; Chan, K. W.; Lim, W.; Shortridge, K. F.; Yuen, K. Y.; Peiris, J. S. M.; Poon, L. L. M. Isolation and Characterization of Viruses Related to the SARS Coronavirus from Animals in Southern China. *Science* **2003**, *302*, 276-279.
266. Tong, T. R. Drug targets in severe acute respiratory syndrome (SARS) virus and other coronavirus infections. *Infectious Disorders: Drug Targets* **2009**, *9*, 223-245.
267. Mukherjee, P.; Shah, F.; Desai, P.; Avery, M. Inhibitors of SARS-3CLpro: Virtual Screening, Biological Evaluation, and Molecular Dynamics Simulation Studies. *Journal of chemical information and modeling* **2011**, *51*, 1376-1392.
268. Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684-3690.
269. Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant pressure molecular dynamics algorithms. *FIELD Full Journal Title:Journal of Chemical Physics* **1994**, *101*, 4177-4189.
270. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **1977**, *23*, 327-341.
271. Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible multiple time scale molecular dynamics. *Journal of Chemical Physics* **1992**, *97*, 1990-2001.
272. 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*, 10089-10092.
273. Liu, H.-L.; Lin, J.-C.; Ho, Y.; Chen, C.-W. Homology models of main proteinase from coronavirus associated with SARS. *Chemical Physics Letters* **2005**, *401*, 24-29.

Vitae

FALGUN SHAH

417, Faser hall, Department of Medicinal Chemistry, School of Pharmacy, University of Mississippi, University, MS-38677
Email: falgunhshah@gmail.com, Office: 662-915-1286

EDUCATION

2006-Present, **Ph.D.** in Pharmaceutical Sciences with a major in Medicinal Chemistry from the University of Mississippi, Oxford, MS, Dissertation title: Utilization of structure and ligand-based drug design tools in the discovery of small molecule inhibitors of cysteine proteases for the treatment of Malaria and SARS Infection. Mentor: - Prof. Mitchell A. Avery

2003-2005, **Master of Pharmacy** in Pharmaceutical Chemistry from University of Mumbai, Mumbai, India, Dissertation title: Design, synthesis and structure activity relationships of newer antidepressants. Mentor: - Dr. Urmila J. Joshi

2003-2004, **Post-graduate Diploma in Contract Clinical Research** from K. M. Kundnani College of Pharmacy, Mumbai, India.

1999-2003, **Bachelor of Pharmaceutical Science** from University of South Gujarat, Gujarat, India. GATE-Graduate Aptitude Test in Engineering, Rank:-330/~20000

RESEARCH EXPERIENCE

Post-doctoral fellowship, University of California, San Diego, December 1st, 2011 to present

Visiting Scholar, University of California, San Diego, September 1st to November 4th, 2011. Mentor: Prof. Ruben Abagyan

- Developing ligand-based models for hERG toxicity prediction

Summer Intern, Tissue Repair Division, Pfizer Global Research & Development, Groton, CT, June 14th to Sep 24th, 2010. Mentor: Dr. Jack A Bikker

- Development of (pipeline pilot-based) **protocol** to automatically pre-populate **compound attrition** or **progression decisions** based on the knowledge of a project screening tree generated from proprietary database.
- Application of **decision capture** protocol to generate easily interpretable graphical views to visualize the screening tree, most relevant assays and highly profiled compounds for the project of interest
- **Understanding** the role of **waters** in binding affinity of PARP inhibitors.

Summer Trainee, University of North Carolina, Chapel Hill, NC, June 1st – July 31st, 2009

- Hands on various **QSAR approaches** such as *K*-Nearest Neighborhood method, Support Vector Machine. Mentor: Dr. Alexander Tropsha

Research Assistant, University of Mississippi, Oxford, MS, August 2006-present

- Employed **virtual screening** to identify drug-like small molecule inhibitors of parasitic cysteine proteases (malaria, leishmania) using targeted cysteine protease compound library generated based on soft-electrophiles of interest.
- Carried out **substructure/similarity searching** to identify the analogs of virtual screening hits. **Structure-based optimization** of obtained hits and development of combinatorial library of selected compounds.
- **Solvent thermodynamic** analysis Falcipain Ligand binding domain, to gain insight in to structure-activity relationships of FP-2 hits
- Performed *de novo design* and docking studies of plasmodial cysteine protease inhibitors of falcipain-II (malaria), and viral cysteine protease (SARS 3CL^{pro}).
- Performed **scoring-based enrichment studies** for falcipain-2 (Malarial cysteine protease) inhibitors.
- Calculation of **LUMO density** for predicting the most electrophilic centers in the cysteine protease inhibitors
- Investigate the binding mode of Bcl-xL (anticancer target) inhibitors through **steered molecular dynamics** study and conventional molecular dynamics approach.
- Conducted **QSAR** study of artemisinin analogs (malaria) and arylpiperazines class of 5-HT_{1A} receptor antagonists (depression). Developed Multiple Linear Regression (MLR), Support Vector machine (**SVM**) and *K*-Nearest Neighbor (**KNN**) classification models for artemisinin and its derivatives.
- Performed **pharmacophore modeling** of angiotensin-II type-I (AT₁) and dual AT₁-PPAR_γ ligands.
- Developed **homology model** of 5-HT_{1A} and TXA₂ receptor (**GPCRs**).
- **Synthesized** (4 step) small molecule library containing benzthizaole scaffold as novel malarial cysteine protease inhibitors and 2-naphthyloxyethylamines derivatives as 5-HT_{1A} receptor antagonists.
- Design of hypothesis and protocol for the validation of proposed artemisinin mechanism through inhibition of *Pf*ATP6-SERCA.
- Defended **Original Research Proposal** (NIH, R01) entitled “Targeting Influenza A virus (H1N1, H3N2, H5N1): Development of Cap-Dependent Endonuclease inhibitors as new Anti-Influenza therapeutics” as a part of Ph.D. curriculum.

TECHNICAL SKILLS

- **Structure-based modeling:** Homology modeling and refinement, pose-validation, enrichment, structure-based virtual screening of commercial database, receptor-based pharmacophore and structure-based drug design,
- **Ligand-based modeling:** 2D/3D QSAR, pharmacophore modeling and screening, clustering, diversity analysis, similarity/substructure searching, property-based filtration, database handling, ADME/Toxicity modeling

- **Simulation techniques:** WaterMap analysis, Conformational analysis, Molecular Dynamics simulation and Steered Molecular Dynamics simulation
- **Synthetics:** Synthesis of library of small molecules bearing benzthiazole and substituted 2-naphthyloxyethylamines scaffold, olefination reactions, peptide coupling etc.
- **Programming skills:** shell scripting, scripting languages such as python, ICM reference language, pilot scripting, SQL

MODELING TOOLS AND SOFTWARE

- **Pipeline Pilot**, Discovery studio, **Spotfire**
- **Schrodinger**:- Glide, WaterMap, Phase, Canvas, Desmond, Macromodel, Impact, Prime, Strike, QikProp, Jaguar
- **ICM**:- 4D docking, APF, QSAR modeling
- **Sybyl**, **MOE**, **GOLD**, **Catalyst**
- **Miscellaneous:** NAMD, WEKA, LIBSVM, ISIDA, PYMOL, VMD, ChemOffice Ultra, E-notebook

AWARDS/AFFILIATION

- **CCG Excellence Student Travel Award**, Fall 2011, ACS meeting at Denver.
- **Travel award for Practical Computing Biologist** workshop organized by NESCent Academy at University of North Carolina, Raleigh, June 5th to June 15th, 2011.
- **American Chemical Society's** Ole Miss Local Section's **Graduate Student Research Award** in the Pharmaceutical Sciences for year 2011.
- **NIH Pre-Doctoral Fellowship**, NIH Center of Biomedical Research Excellence (**COBRE**), 2009-2010, "Potential of artemisinin and its derivatives in the treatment of brain cancer and neurological manifestations of Malaria: Design of novel artemisinin derivatives using QSAR models, prediction of pharmacokinetic properties and insight into the mechanism of the action of this class of drugs".
- **Research Grant Award**, Graduate Student Council, University of Mississippi, 2008-09, "Design of small molecule reversible inhibitors of cathepsin K for the treatment of osteoporosis using computer aided drug designing"
- **Scholarship for Academic Excellence**, Sir Ratan Tata and J. R. D. Tata Trust, Mumbai, India, 2004-05.
- **Junior Research Fellowship**, All India Council For Technical Education, 2003-2004
- **Professional affiliation:** American Chemical Society (COMP, MEDI), Royal Society of Chemistry (AMRSC), Rho chi chapter at Olemiss, University of Mississippi.

LIST OF PUBLICATIONS AND PRESENTATIONS

PUBLICATIONS

1. Weldon, D., Shah F., Chittiboyina, A., Jiri, G., Rosenthal, P., Shivakumar, D. Sherman W., Desai, P., Jung J. C., Avery M. A. Design, Synthesis, and Biological Results of α -Keto Substituted Peptidomimetics for the Inhibition of *Plasmodium falciparum*. (**Under review**)
2. Shah F., Jiri G., Legac, J., Shivakumar, D., Sherman W., Rosenthal, P., Avery M. A. Computer-Aided Drug Design of Falcipain Inhibitors: Virtual Screening, Structure-Activity Relationships, Solvent Thermodynamics and Reactivity Analysis (**Under review**).
3. Shah F., Wu, Y., Jiri G., Pedurri, Y., Rosenthal, P., Avery M. A. Design, synthesis and biological evaluation of novel benzthiazole and triazole analogs as inhibitors of falcipain, cysteine protease of malaria parasite Plasmodium falciparum, **MedChemComm**, **2011**_DOI: 10.1039/C1MD00129A
4. Shah F., Mukherjee, P., Jiri G., Legac, J., Rosenthal, P., Tekwani B., Avery M. A. Identification of Novel Malarial Cysteine Protease Inhibitors Using Structure-Based Virtual Screening of a Focused Cysteine Protease Inhibitor Library, **Journal of Chemical Information and Modeling**, **2011**, 51 (4), 852.
5. Shah, F., Mukherjee, P., Desai, P., Avery M. A., Computational approaches for the discovery of cysteine protease inhibitors against Malaria and SARS. **Current Computer Aided Drug Designing**, **2010**, 6, 1.
6. Shah, F., Zhang, S., Kandhari, S. P., Mukherjee, P., Chittiboyina, A., Avery, M. A., Avery, B. A., *In vitro* erythrocytic uptake studies of artemisinin and selected derivatives using LC-MS and 2D-QSAR analysis of uptake in parasitized erythrocytes. **Bioorganic & Medicinal Chemistry**, **2009**, 17(14), 5325.
7. Mukherjee, P., Shah, F., Desai, P., Avery M. A. SARS-3CLpro: Virtual Screening, Biological Evaluation and Molecular Dynamics Simulation Studies, **Journal of Chemical Information and Modeling**, **2011**, 51 (6), 1376.
8. Sun, L., Shah F., Helal, M., Wu, Y., Pedduri, Y, Carvalho, P. D., Jiri G., Rosenthal, P., Avery M. A. Design, synthesis and development of novel Guaianolide-endoperoxide as a potential anti-malarial agent, **Journal of Medicinal Chemistry**, **2010**, 53(21), 7864.
9. Pellicciari, R., Camaioni, E., Gilbert, A. M., Macchiarulo, A., Bikker J., Shah, F., Bard, J., Costantino, G., Gioello, A., Robertson, M. et al., Discovery and characterization of novel potent PARP-1 inhibitors endowed with neuroprotective properties: From TIQ-A to HYDAMTIQ, **Med Chem Comm.**, **2011**, 2, 559.
10. Mizuno, C., S., Chittiboyina, A. G., Shah, F., Patny, A., Kurtz, T. W., Pershadsingh, H. A., Speth, R. C., Karamyan, V., Carvalho, P. B., Avery. M. A., Design, Synthesis and Docking Studies of Novel Benzimidazoles for the Treatment of Metabolic Syndrome, **Journal of Medicinal Chemistry**, **2010**, 53 (3), 1076.

11. Mukherjee, P.; Pradhan A.; Shah, F., Tekwani B., Avery, M. A., Structural insights into the plasmodium falciparum histone deacetylase 1 (PfHDAC-1): A novel target for the development of antimalarial therapy. *Bioorganic & Medicinal Chemistry*, **2008**, *16*, 5254.
12. Joshi, U. J., Shah, F., Tikhele, S., Homology model of 5-hydroxytryptamine_{1A} receptor based on crystal structure of bovine rhodopsin. *Internet Electronic Journal of Molecular Deigning*, **2006**, *5*, 403.
13. Joshi U. J., Shah F., 5-HT_{1A} receptor: structure and implications for drug discovery. *Indian Drugs*, **2007**, *44*, 169.
14. Joshi, U. J., Tikhele, S., Shah, F., 2D QSAR of Arylpiperazines as 5-HT_{1A} Receptor Agonists, *Indian Journal of Pharmaceutical Science*, **2007**, *69*, 800.
15. Joshi, U. J., Dube, R. K., Shah, F., Naik, S. R., Design and synthesis of substituted 2-naphthoxyethylamines as potential 5-HT_{1A} antagonists. *Indian Journal of Pharmaceutical Science*, **2007**, *69*, 814.

SELECTED PRESENTATIONS

1. Shah, F., Mukherjee, P., Jiri G., Rosenthal, P., Legac, J., Avery M. A. Designing reversible, non-peptidic and selective inhibitors of falcipain, cysteine protease of malaria parasite Plasmodium falciparum, **accepted** at *Gordon Research Conference on Computer Aided Drug Designing* at West Dover, VT, 2011 (accepted for open mic session).
2. Shah, F., Bikker J. A., Decision capture: improve decision-making and alleviate attrition rate; *Pfizer internal conference*, Groton, CT, August 03, 2010.
3. Shah, F., Wu, Yunshan, Golbraikh, A., Wu, Tony, Wang, X. S., Fourches, D., Tropsha, A., Avery, M. A. Development of binary QSAR models to predict the biological activity of semi-synthetic artemisinin derivatives by kNN and SVM approach, at 239th *ACS National Meeting* at San Francisco, CA, Spring 2010.
4. Shah, F., Mukherjee, P., Jiri G., Rosenthal, P., Tekwani, B., Avery M. A. Identification of novel malarial cysteine protease inhibitors: Virtual screening study and optimization of hits, at 239th *ACS National Meeting* at San Francisco, CA, Spring 2010.
5. Shah, F., Mukherjee, P., Avery, M. A. Targeting Falcipain-II, a cysteine protease from *P. Falciparum*: a structure based virtual screening study at *Gordon Research Conference on Computer Aided Drug Designing* at Tilton, NH, 2009
6. Shah, F., Mukherjee, P., Avery, M. A. Investigating the binding mode of ligands of Bcl-xL by steered molecular dynamics simulation, at 237th *ACS National Meeting* at Salt Lake City, UT, Spring 2009.

7. Shah, F., Mukherjee, P., Avery, M. A. Docking based pharmacophore modeling of combined AT₁-PPAR gamma ligands at 235th *ACS National Meeting* at New Orleans, LA, Spring 2008.
8. Shah, F., Tikhele, S., Joshi, U. J., The 5-HT₇ Receptor at the 4th *International Symposia on Biomolecular Chemistry organized by Royal Society of Chemistry* at the University of Sheffield, U.K., 2004.
9. Shah, F., Tikhele, S., Joshi, U. J., Homology model of 5-HT_{1A} using crystal structure of Bovine rhodopsin at 1st *Indo- Japanese International Conference on Advances in Pharmaceutical Research and Technology* at Mumbai, India, November 25, 2005.