

Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Mi-Sun Kim

4-3-12

Approval Sheet

Part I. Design of a novel class of reversible non-covalent small molecule inhibitors
for human Granzyme B (hGrB)
Part II. Curcumin and mimics as proteasome inhibitors
Part III. Design of novel coactivator binding inhibitors (CBIs) for the estrogen receptor α
: break the 1 μ M barrier

By

Mi-Sun Kim
Doctor of Philosophy

Chemistry

Dennis C. Liotta, Ph.D.
Advisor

James P. Snyder, Ph.D.
Advisor

David Lynn, Ph.D.
Committee Member

James T. Kindt, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Abstract Cover Page

Part I. Design of a novel class of reversible non-covalent small molecule inhibitors
for human Granzyme B (hGrB)

Part II. Curcumin and mimics as proteasome inhibitors

Part III. Design of novel coactivator binding inhibitors (CBIs) for the estrogen receptor α
: break the 1 μ M barrier

By

Mi-Sun Kim

B.S. of Science, Sungkyunkwan University, 1997

M.S. of Science, Sungkyunkwan University, 2000

Advisors:

Dennis C. Liotta, Ph.D.

James P. Snyder, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in chemistry
2012

Abstract

- Part I. Design of a novel class of reversible non-covalent small molecule inhibitors for human Granzyme B (hGrB)
Part II. Curcumin and mimics as proteasome inhibitors
Part III. Design of novel coactivator binding inhibitors (CBIs) for the estrogen receptor α : break the 1 μ M barrier

By Mi-Sun Kim

Part I. Design of a novel class of reversible non-covalent small molecule inhibitors for human Granzyme B (hGrB)

Granzyme B, lymphocyte serine protease, plays a critical role in controlling graft versus leukemia and graft-versus-host diseases. A key aim of this study is to design non-covalent small molecule inhibitors using a computational model and screening approach: 1) computational solvent mapping to identify “hot spots” in the active site; 2) virtual screening with three constraints based on the solvent mapping results; 3) measure enzyme activity and selectivity; 4) validate by modeling known covalent blockers. As a result, novel classes of hGrB inhibitors have been identified. To extend the pool of scaffolds, ‘scaffold hopping’ has been carried out to perform shape-based searching.

Part II. Curcumin and mimics as proteasome inhibitors

The proteasome is an important target of curcumin, and several groups have reported that inhibition of proteasome activity by curcumin and chalcone-based derivatives cause apoptosis in human cancer cell lines. In those studies, DFT calculations and *in silico* docking studies were performed to suggest binding poses in the CT-like subunit. However, these studies introduce critical and most-likely incorrect assumptions: 1) the LUMOs were calculated for the unstable diketo form; 2) the LUMO characteristics were interpreted to suggest that the carbonyl carbons should be the site of the terminal Thr1 nucleophilic attack instead of the β -carbon of the enone; 3) an incorrect cis-structure for the chalcone analogs was employed for the docking. Thus, the specific aim of this study is to identify the correct binding poses of curcumin and chalcone-based derivatives and to explore the binding pockets of the three active β subunits to explain why curcumin potentially inhibits the CT-like activity.

Part III. Design of novel coactivator binding inhibitors (CBIs) for the estrogen receptor α : break the 1 μ M barrier

Novel classes of coactivator binding inhibitors against estrogen receptor α have been developed. However, they fail to deliver IC₅₀ values below 1 μ M in the reporter gene assay. To understand these observations and break the 1 μ M barrier, we performed MD simulations to examine solvent-based entropic contributions to the free energy of ligand binding. The coactivator is far more effective at expelling water molecules from the binding site than the CBIs. These observations strongly suggest that the next generation of small molecule CBIs should span more of the peptide space, particularly on the shelf adjacent to the binding groove.

Cover Page

Part I. Design of a novel class of reversible non-covalent small molecule inhibitors
for human Granzyme B (hGrB)

Part II. Curcumin and mimics as proteasome inhibitors

Part III. Design of novel coactivator binding inhibitors (CBIs) for the estrogen receptor α
: break the 1 μ M barrier

By

Mi-Sun Kim

B.S. of Science, Sungkyunkwan University, 1997

M.S. of Science, Sungkyunkwan University, 2000

Advisors:

Dennis C. Liotta, Ph.D.

James P. Snyder, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in chemistry
2012

Acknowledgements

Advisors

I deeply appreciate mentoring and support from my advisors, Drs. James P. Snyder and Dennis Liotta. I had a hard time, especially after giving birth of my first son, however, they encouraged and trusted me at all times. So, now I can be here.

Committee members

Drs. David Lynn and James T. Kindt, for your advice, thoughtful discussions, and encouragements.

Collaborators

Drs. Philip G Ashton-Rickardt, Dean Heatcote, Anthony G M Barrett, and Chris Braddock, and Lauriane Bussion at Imperial College London for granzyme B project.

Drs. Terry William Moore and Lawrence Boise and Clint Kimzey for proteasome / curcumin project.

Drs. Aiming Sun, Terry W. Moore, Yuhong Du, Haiyan Fu at Emory university, and Drs. Jillian R. Gunther, Eric Rhoden, John A. Katzenellenbogen at University of Illinois for estrogen receptor α project.

Thank you for turning around the bio-data quickly and efficiently, and their best works to improve activity. I deeply appreciate their incredible collaborations in biology and chemistry.

Dr. Michael Groll at Technische Universität München. He has been working on functional and structural characterization of multifunctional protein complexes as the expert of the proteasome research. Thank you for your great advice and sharing good information.

Group members

Drs. Ana A. Alcaraz, Andrew Prussia, Pakh Thepchatri, Haipeng Hu, and Gordon Wells. Jin Liu and Qi Shi. You guys have been my friends as well as mentors for my graduate school experience. I have enjoyed working with all of you.

Family, church members, and friends

My mom, father-in-law, and mother-in-law. I deeply appreciate their so much love, support, and pray for me. After giving birth of my first son, mother-in-law supported me to study without any concerns. Unfortunately, she passed away after several months of returning to Korea, however, she gave me a special present that I got an email that I passed the cume at her funeral. I am missing her.

My lovely husband, Jay. He is the most important person in my life. Thank you for your so much love, trust, encouragement, support, and praying.

My precious two sons, Kon and Min. While preparing dissertation and defense, I could not be with Kon for much time, however, he was so patient and was counting the defense date. After having Min, I completed two big things, original proposal and dissertation. I thank him for finishing them without any concerns about his growing in me.

In addition, I thank my church members and friends for their love and praying.

Finally, I glorify God for leading and blessing my life. All things were from Him.

Table of Contents

Part 1. Design of a novel class of reversible non-covalent small molecule inhibitors for human Granzyme B (hGrB)

Chapter 1. Introduction	1
1.1 Graft-versus-leukemia (GVL) vs graft-versus-host disease (GVHD).....	1
1.2 Specific aims.....	2
Chapter 2. Background	3
2.1 Granzymes (lymphocyte serine proteases).....	3
2.2 Granzyme B.....	4
2.3 Substrate for granzyme B.....	4
2.4 Covalent serine protease inhibitors.....	5
2.5 Disadvantages of covalent serine protease inhibitors.....	6
2.6 Strategy for non-covalent inhibitor design.....	6
Chapter 3. Preliminary study	7
3.1 X-ray structure of apo-hGrB (PDB ID: 1FQ3).....	7
3.1.1 Overall Structure.....	7
3.1.2 Active Site.....	8
3.2 X-ray structure of hGrB complex with tetrapeptide aldehyde inhibitor (PDB ID: 1IAU).....	9
3.3 Comparison between apo- and inhibitor-bound GrB.....	10
3.3.1 S1 subsite / P1-Asp.....	10
3.3.2 Oxyanion Hole.....	11
3.3.3 S2 subsite / P2-Pro.....	12
3.3.4 S3 subsite / P3-Glu.....	13
3.3.5 S4 subsite / P4-Ile.....	14
3.3.6 S1' subsite / P1'-Trp.....	15
3.3.7 S2' subsite / P2'-Gly.....	16
Chapter 4. Computational solvent mapping	17
4.1 FTMAP algorithm.....	17

4.2 Computational solvent mapping results comparison before and after inhibitor binding.....	19
4.2.1 S1 subsite / P1-Asp.....	19
4.2.2 S2 subsite / P2-Pro.....	21
4.2.3 S3 subsite / P3-Glu.....	22
4.2.4 S4 subsite / P4-Ile.....	23
4.2.5 S1' subsite / P1'-Trp.....	24
4.2.6 S2' subsite / P2'-Gly.....	25
4.3 Summary of GrB mapping results.....	26
4.4 Distribution analysis of non-bonded interactions and hydrogen bonds between probes and hGrB residues.....	27
4.5 Three constraints.....	29
Chapter 5. Virtual screening.....	30
5.1 Libraries.....	30
5.1.1 Maybridge Screening Collection (MSC).....	30
5.1.2 TimTec ActiTarg-P (Protease targeted library).....	31
5.1.3 ChemDiv.....	31
5.2 Glide Docking.....	31
Chapter 6. Bioassay results.....	32
6.1 Hits from TimTec ActiTarg-P.....	33
6.1.1 Thiotetrazole series.....	33
6.1.2 Thiotriazole series.....	35
6.1.3 Thiazolidinedione series.....	37
6.1.4 Diazolidinedione series.....	41
6.1.5 Others.....	44
6.2 Hits from ChemDiv.....	46
6.2.1 Sulfonamide series.....	46
6.2.2 Tricyclic series.....	48
6.2.3 Other.....	50
6.3 Summary of the characteristics of the various classes of compounds.....	50
Chapter 7. Selectivity.....	51

7.1 Caspase-3 vs Caspase-8.....	51
7.2 GrB vs Caspase-3 and 8.....	52
Chapter 8. Scaffold hopping.....	54
8.1 Merck compounds modeling.....	54
8.1.1 Binding pose of Merck 15.....	55
8.1.2 Binding pose of Merck 19.....	56
8.1.3 Binding pose of Merck 20.....	57
8.2 Search Queries.....	60
8.3 Library: protease-targeted libraries.....	62
8.4 Computational procedures.....	63
8.4.1 Create a query.....	63
8.4.2 Perform a ROCS run.....	63
8.4.3 Docking using Glide.....	64
8.4.4 Induced fit docking using Glide and Prime.....	65
8.5 Results.....	67
8.5.1 Example: Enamine Casp-3 Target Library Results.....	68
8.6 Molecular property calculations using QikProp.....	70
Chapter 9 Synthesis, Purchase, and Biological test.....	72
Chapter 10. Mid-Micromolar Activities; What Could be Done?.....	72
Chapter 11. Conclusions.....	75

Part II. Curcumin and Mimics as Proteasome Inhibitors

Chapter 1. Introduction.....	77
1.1 Ubiquitin-proteasome pathway (UPP) and 26S proteasome.....	77
1.2 Proteasome inhibitors.....	79
1.2.1 Covalent reversible inhibitors.....	79
1.2.2 Covalent irreversible inhibitors.....	81
1.2.3 Non-covalent reversible inhibitors.....	84
1.3 Specific aims.....	85
Chapter 2. Background.....	86
2.1 26S proteasome and 20S proteasome.....	86

2.2 Curcumin and chalcone-based derivatives as proteasome inhibitors.....	88
2.3 Critical issues common to both papers.....	89
2.3.1 Issue 1: Misuse of the diketo form of curcumin.....	90
2.3.2 Issue 2: Implausible binding mode of RA-1 with the cis-trans configuration.....	91
2.3.3 Issue 3: N-terminal Thr1-OH nucleophilic attack on C=O.....	91
Chapter 3. Structural examination and reactivity prediction.....	92
3.1 Structural examination of curcuminoids in the Cambridge Structural Database (CSD).....	92
3.2 Optimization and single-point calculations.....	94
3.3 Quantification of the anticipated reactivity.....	95
3.4 X-ray structure of the peptide inhibitor possessing the vinyl sulfone moiety.....	102
Chapter 4. Modeling the subunit targets.....	104
4.1 Active sites analysis.....	104
4.2 Non-covalent docking studies.....	106
4.3 Docking of curcumin.....	107
4.3.1 The predicted binding pose of curcumin at β 1 and β 2.....	108
4.3.2 The predicted binding pose of curcumin at β 5.....	110
4.3.3 Summary of the predicted binding poses of curcumin at all three catalytic subunits.....	113
4.4 Docking of RA-1 analogs.....	114
4.4.1 The predicted binding pose of RA-1.....	116
4.4.2 The predicted binding pose of RA-2.....	121
4.4.3 The predicted binding pose of RA-3.....	124
4.5 Summary of the docking results for curcumin and RA-1, 2, and 3.....	126
4.6 Extending the curcumin analog SAR.....	128
Chapter 5. Conclusions.....	132

Part III. Design of Novel Coactivator Binding Inhibitors (CBIs) for the Estrogen Receptor α : *Break the 1 μ M barrier*

Chapter 1. Introduction.....	133
1.1 Estrogen receptor (ER).....	133
1.2 Blockers of ER agonist signaling.....	134
1.3 Specific aims.....	134
Chapter 2. Preliminary study.....	135
2.1 Structure of the ER α	135
2.2 High-throughput screening (HTS) and optimization.....	137
2.2.1 Modification of the A and C.....	138
2.2.2 Modification of phenyl piperazine sectors A and B.....	139
2.2.3 Modification of phenyl quinazoline region D.....	140
2.2.4 Modification of benzothiazole.....	141
Chapter 3. Molecular modeling.....	142
3.1 Molecular modeling.....	142
3.2 Induced fit docking.....	143
3.3 Molecular volume and qikprop calculations.....	144
3.4 MM-GBSA energy evaluation.....	145
3.5 Desmond molecular dynamics simulations and determination of numbers of displaced waters.....	145
Chapter 4. Results.....	146
4.1 Possible competition at the ligand binding pocket.....	146
4.2 Phenyl piperazine scaffold.....	148
4.3 Benzothiazole scaffold.....	154
Chapter 5. What the current CBIs offer and what they lack.....	156
5.1 Entropic effects originating from water displacement.....	157
Chapter 6. Conclusion.....	159
References.....	161

List of Figures

Part I. Design of a novel class of reversible non-covalent small molecule inhibitors for human Granzyme B (hGrB)

Figure 1. Recent work with granzyme B (GrB) deficient mice and humans has shown a critical role for GrB in controlling GVL and GVHD.....	2
Figure 2. GrB molecules A and B.....	7
Figure 3. Subsites (S4-S2') and related residues for hexapeptide substrate.....	8
Figure 4. Tetrapeptide aldehyde Ac-IEPD-CH.....	9
Figure 5. Movements of residues in the active site upon inhibitor binding (S1).....	10
Figure 6. Movements of residues in the oxyanion hole upon inhibitor binding (oxyanion hole).....	11
Figure 7. Movements of residues in the active site upon inhibitor binding (S2).....	12
Figure 8. Movements of residues in the active site upon inhibitor binding (S3).....	13
Figure 9. Movements of residues in the active site upon inhibitor binding (S4).....	14
Figure 10. Movements of residues in the active site upon inhibitor binding (S1').....	15
Figure 11. Movements of residues in the active site upon inhibitor binding (S2').....	16
Figure 12. Outline of computational solvent mapping algorithms used by FTMAP.....	18
Figure 13. Water molecules in the S1 pocket and alignment of consensus sites for the apo and inhibitor bound forms of GrB.....	20
Figure 14. Alignments of consensus sites resulting from computational solvent mapping (S2).....	21
Figure 15. Alignments of consensus sites resulting from computational solvent mapping (S3).....	22
Figure 16. Alignments of consensus sites resulting from computational solvent mapping (S4).....	23
Figure 17. Alignments of consensus sites resulting from computational solvent mapping (S1').....	24

Figure 18. Alignments of consensus sites resulting from computational solvent mapping (S2').....	25
Figure 19. Consensus sites in the binding site.....	27
Figure 20. Intermolecular non-bonded interactions between probes and residues for apo-GrB and Ac-IEPD-CHO-bound GrB without the ligand.....	28
Figure 21. Intermolecular hydrogen bond interactions between probes and residues for apo-GrB and Ac-IEPD-CHO-bound GrB without the ligand.....	28
Figure 22. Selected constraints based on mapping results for the docking study.....	30
Figure 23. The predicted binding mode of ST053449	34
Figure 24. The predicted binding mode of ST057416	36
Figure 25. The predicted binding mode of ST057924	40
Figure 26. The predicted binding mode of ST057833	40
Figure 27. The predicted binding mode of ST058025	43
Figure 28. The predicted binding mode of ST014575	45
Figure 29. The predicted binding mode of L581-022	47
Figure 30. The predicted binding modes of (a) G202-0149 and (b) E627-066449	49
Figure 31. The structure of Merck 15	55
Figure 32. Merck 15 at the active site of GrB.....	55
Figure 33. The structure of Merck 19	56
Figure 34. (a) Alignment with IEPD-CHO and (b) the predicted binding pose of Merck 19	57
Figure 35. The structure of Merck 20	57
Figure 36. (a) Alignment with IEPD-CHO and (b) the predicted binding pose of Merck 20	58
Figure 37. (a) Alignment with Merck 19 and Merck 20 and (b) the cation- π interaction between Arg217 and P5-phenyl and benzothiophene rings of Merck 19 and Merck 20 , respectively.....	59
Figure 38. Alignment of (a) G202-0149 and (b) E627-0664 with Merck 20	59
Figure 39. All three portions (molecule, shape and color) are generated by ROCS.....	63
Figure 40. A hit molecule is aligned to the query molecule.....	64
Figure 41. Screening procedure.....	65

Figure 42. 2 nd round screening with top 20 compounds per each calculation.....	66
Figure 43. Molecular property calculations using QikProp.....	71
Figure 44. The crystal structures of elastase complexed with (a) trifluoroacetyl (TFA)-Lys-Pro-isopropylanilide (ISO).....	73
Figure 45. The crystal structures of elastase complexed with (a) TFA-Lys-Phe-ISO and (b) TFA-Lys-Leu-ISO.....	74
Figure 46. All inhibitors are aligned showing different binding modes in spite of chemically similar structures.....	74

Part II. Curcumin and Mimics as Proteasome Inhibitors

Figure 1. Ubiquitination process.....	78
Figure 2. Covalent reversible inhibitors.....	80
Figure 3. Covalent irreversible inhibitors.....	83
Figure 4. Non-covalent reversible inhibitors (TMC-95s) represented by TMC-95A.....	84
Figure 5. The structure of the 20S proteasome that contains three active catalytic sites with PGPH-like (β 1), trypsin-like (β 2), and chymotrypsin-like (β 5) activities in the inner β rings.....	86
Figure 6. The proteasomal S1 specificity pockets.....	87
Figure 7. The structures of curcumin and RA-1.....	89
Figure 8. The structure of curcumin keto-enol form.....	90
Figure 9. The crystal structures of (a) curcumin keto-enol form and (b) an RA-1 analog.....	93
Figure 10. Conjugated enones.....	96
Figure 11-14. The atom with the highest value of Atomic Fukui Indices is marked in red.....	97, 98, 99, 100
Figure 15. The structures of β 2 inhibitors: MB1 (Ac-PRLN-VS) is selective, while MB2 (Ac-YLLN-VS) which inhibits all active sites.....	103
Figure 16. The proteasome bound MB1 crystal structure.....	104
Figure 17. The three catalytic subunits of the 20S proteasome.....	105
Figure 18. Different residue environments.....	106

Figure 19. The predicted binding pose of the curcumin keto-enol form to the $\beta 2$ subunit.....	108
Figure 20. Superposition of curcumin with crystal structures of other known inhibitors at $\beta 2$	109
Figure 21. Predicted binding pose of curcumin keto-enol form to the $\beta 5$ subunit.....	111
Figure 22. The different residues at $\beta 2$ (pink) and $\beta 5$ (light blue).....	112
Figure 23. Superposition of curcumin with crystal structures of other known inhibitors $\beta 5$	113
Figure 24. Relative inhibition of the three catalytic subunits by compounds 2-4	116
Figure 25. The predicted binding poses of RA-1 to $\beta 1$ (a), $\beta 2$ (c) and $\beta 5$ (e), and on the molecular surfaces of $\beta 1$ (b), $\beta 2$ (d) and $\beta 5$ (f).....	118
Figure 26. RA-1 docking.....	119
Figure 27. Superposition of RA-1 with the crystal structures of other known inhibitors at $\beta 1$	120
Figure 28. The predicted binding poses of RA-2 to $\beta 2$ (a), $\beta 5$ (c), and on the molecular surface of $\beta 2$ (b) and $\beta 5$ (d).....	122
Figure 29. Superposition of RA-2 with the crystal structures of other known inhibitors at $\beta 5$	123
Figure 30. The predicted binding poses of RA-3 to $\beta 5$ (a), and residence on the molecular surface of $\beta 5$ (b).....	125
Figure 31. RA-1 derivative modified at the P position.....	130
Figure 32. The Cjk1-10 derivative modified at the P position.....	131

Part III. Design of Novel Coactivator Binding Inhibitors (CBIs) for the Estrogen Receptor α : *Break the 1 μ M barrier*

Figure 1 a) Overall crystal structure of ER α including agonist (DES) and coactivator proteins (GRIP1 NR box II). b) Crystal structure of GRIP1 peptide on the coactivator recognition surface of the ER α	136
---	-----

Figure 2. Alignment of 1g and 4-hydroxytamoxifen in 3ERT (hER α -HT complex) as determined by Glide docking.....	147
--	-----

Figure 3. a) Induced fit docked pose for **1b** at the coactivator binding site of the ER α receptor (PDB X-ray structure code 3ERD). Displacements of Glu542 and Ile358 side chains resulting from docking and H-bond distances are indicated in Å; b) Alignment of **1b** (stick) and the coactivator peptide (purple ribbon; 3ERD). Hydrophobic residues Leu690 and Leu694 are matched by ligand hydrophobes.....149

Figure 4. a) Induced fit docked pose for **1q** at the coactivator binding site of the ER α receptor (PDB X-ray structure code 3ERD). Displacements of Lys362, Leu372, Glu380, Glu542 and Met543 side chains relative to 3ERD and H-bonds are indicated in Å; b) and c) Relative alignments of **1b** and **1q**.....151

Figure 5. a) Docked pose of **1e** illustrating the three ligand-protein hydrogen bonds and side chain movement upon docking in Å; b) Alignment of **1e** and coactivator peptide.....153

Figure 6. a) Docked pose of **20b** showing three hydrogen bond anchors and side chain movement upon docking in Å; b) Alignment of **20b** and coactivator peptide.155

Figure 7. Number of overlapped waters within 2 Å of coactivator peptide and the three ligands **1b**, **1e** and **4o**.....158

List of Tables

Part I. Design of a novel class of reversible non-covalent small molecule inhibitors for human Granzyme B (hGrB)

Table 1. Human granzymes (lymphocyte serine proteases).....	3
Table 2. Preferred substrate sequence based on screening by positional scanning with a combinatorial peptide library.....	5
Table 3. Summary of GrB mapping results.....	26
Table 4. Thiotetrazole series of compounds showing inhibition against GrB and selectivity results.....	33
Table 5. Thiotriazole series showing inhibition against GrB and selectivity results.....	35
Table 6. Thiazolidinedione series showing inhibition against GrB and selectivity results.....	37
Table 7. Diazolidinedione series showing inhibition against GrB and selectivity results.....	41
Table 8. Other compounds showing inhibition against GrB and selectivity results.....	44
Table 9. Sulfonamide series showing inhibition against GrB and selectivity results.....	46
Table 10. Tricyclic series showing inhibition against GrB and selectivity results.....	48
Table 11. 5782-5949 showing weak inhibition against GrB.....	50
Table 12. Caspase-3 and Caspase-8 preferences, HBs and hydrophobic interactions for each subsite.....	53
Table 13. Common and different features for GrB and Caspase3-8.....	53
Table 14. Structures used as ROCS queries. Two types of templates were used.....	61
Table 15. The protease-targeted libraries for scaffold hopping.....	62
Table 16. Summary of the calculations carried out using ROCS and subsequent docking studies.....	67
Table 17. Some results from the Enamine Casp-3 Target Library search.....	68

Part II. Curcumin and Mimics as Proteasome Inhibitors

Table 1. Total and relative DFT energies of some conformers of curcumin and RA-1.....	94
Table 2. Atomic Fukui indices, f_{NN} for the LUMO for curcumin keto-enol form.....	97
Table 3. Atomic Fukui indices, f_{NN} , for the LUMO for α,β -unsaturated carbonyl moiety.....	98
Table 4. Atomic Fukui indices, f_{NN} for the LUMO of RA-1.....	99
Table 5. Atomic Fukui indices, f_{NN} for the LUMO for Cjk1-11.....	101
Table 6. The structures of chalcone-based derivatives.....	114
Table 7. Cell killing activities of compounds 2-10 on HeLa and CaSki cervical cancer cell lines.....	115
Table 8. In vitro inhibition of the proteolytic activities of proteasome purified from lymphoblastoid cell lines (LCLs).....	115
Table 9. The residues which comprise the S4 surface of all three subunits.....	117

Part III. Design of Novel Coactivator Binding Inhibitors (CBIs) for the Estrogen Receptor α : *Break the 1 μ M barrier*

Table 1. Modification of the chlorophenyl region A and linker region C.....	138
Table 2. Modification of phenyl piperazine sectors A and B.....	139
Table 3. Modification of phenyl quinazoline region D.....	140
Table 4. Modification of benzothiazole.....	141

Part I. Design of a novel class of reversible non-covalent small molecule inhibitors for human Granzyme B (hGrB)

Chapter 1. Introduction

1.1 Graft-versus-leukemia (GVL) vs graft-versus-host disease (GVHD)

Over the past several decades, allogeneic bone marrow transplantation (BMT) has emerged as an important therapeutic option for a number of malignant diseases, especially such as acute myeloid leukemia (AML), acute lymphoid leukemia (ALL) and chronic myeloid leukemia (CML). The success of BMT relies on the graft-versus-leukemia (GVL) effect which eradicates remaining malignant cells via immunologic mechanisms, however, BMT also faces a major obstacle, namely, acute graft-versus-host disease (GVHD) in which functional immune cells in the transplanted marrow recognize the recipient as “foreign” and leads to immunologic attack¹.

Cytotoxic T cell (CTL), which belongs to a sub-group of T lymphocytes (a type of white blood cell), is capable of recognizing specific infected target cells and inducing apoptosis, programmed cell death. Recent work with granzyme B (GrB) deficient mice and humans has shown a critical role for GrB, one of the most abundant granzymes in CTL granules, in controlling GVL and GVHD: GrB results in GVHD by killing normal cells and suppresses GVL by killing useful tumor-specific CTL². Therefore, inhibition of GrB has been considered to be an attractive target to promote GVL and prevent GVHD (Figure 1).



Figure 1. Recent work with granzyme B (GrB) deficient mice and humans has shown a critical role for GrB in controlling GVL and GVHD: GrB results in GVHD by killing normal cells and suppressing GVL by destroying useful tumor-specific CTL³. Therefore, inhibition of GrB is considered to be an attractive target to promote GVL and prevent GVHD

1.2 Specific aims

Thus, the specific aim of this study is to identify at least one novel class of non-covalent and reversible small molecule inhibitors for human Granzyme B (hGrB) by employing various computational techniques. Since there are no successful noncovalent small molecule inhibitors, virtual screening has been carried out initially in the search for lead compounds using the X-ray structures of human GrB in complex with a potent covalently bound tetrapeptide aldehyde inhibitor.

Chapter 2. Background

2.1 Granzymes (lymphocyte serine proteases)

Granzymes are lymphocyte serine proteases that are released by cytoplasmic granules within cytotoxic T cells and natural killer cells which defend cells against viral infections and the growth of tumors. Exocytosis of lymphocyte granules is one mechanism exhibited by infected cells. The cytotoxic granules consists of perforin, a pore-forming protein, and a family of serine proteases called granzymes.¹ They are named because of their location, namely in cytoplasmic granules of CTL and NK cells⁴. Other names used for granzymes are cytotoxic cell proteases (CCP), cytotoxic serine proteases (CSP), fragmentins, and rat natural killer cell proteases (RNKP)⁵.

Granzymes differ in their primary substrate specificities. Four different enzymatic activities have been detected in humans: tryptase (cleaving after Arg or Lys), chymase (cleaving after Phe, Trp or Tyr), Asp-ase (cleaving after Asp), and Met-ase (cleaving after Met, Nle or Leu).⁴

Table 1. Human granzymes (lymphocyte serine proteases)¹

Names	Species	Enzyme activity	Predicted P1 cleavage site
Granzyme A	Human	Tryptase	Arg/Lys
Granzyme B	Human	Asp-ase	Asp/Glu
Granzyme H	Human	Chymase	Phe
Granzyme K	Human	Tryptase	Arg/Lys
Granzyme M	Human	Met-ase	Met/Nle/Leu

2.2 Granzyme B

Granzymes A and B are the most abundant granzymes in cytotoxic lymphocyte and are expressed earlier than other granzymes⁶. GrB is especially unique among mammalian serine proteases resulting from its strong requirement for Asp in the substrate P1 position like a number of members of the caspase family, a family of cysteine proteases involved in apoptosis. This preference is complemented by Arg226 in the enzyme's S1 pocket. Please see section 2.2 Substrate for more details.

According to crystal structure analysis, the Asp-specific S1 subsite of human granzyme B is significantly larger and less charged than the caspases, and also larger than rat granzyme B⁷.

The main player of the catalytic mechanism of GrB is the catalytic triad located in the active site of the enzyme, where catalysis occurs, and is conserved in all serine proteases. The triad consists of three essential amino acids; Ser195, His57 and Asp102 residues (**Figure 4**).

2.3 Substrate for granzyme B

Potential natural substrates for GrB include procaspases (caspases-3, 6, 7, 9 and 10) and other proteins involved in cell death (PARP, DNA-PKcs and NuMA) which have similar peptide sequences⁴.

Combinatorial substrate libraries have profiled the GrB substrate interactions furnishing a clear-cut profile: Ile or Val at P4, Glu, Met, or Gln at P3, broad preference at P2, Asp at P1, an uncharged residue at P1' and Gly or Ala at P2'⁸. **Table 2** shows the

preferred substrate sequence based on screening by positional scanning with a combinatorial peptide library.

Table 2. Preferred substrate sequence based on screening by positional scanning with a combinatorial peptide library

P4	P3	P2	P1	P1'	P2'
Ile (I)	Glu (E)	Pro (P)	Asp (D)	Xaa	Gly

2.4 Covalent serine protease inhibitors

Small molecule serine inhibitors naturally divide into two classes, covalent and non-covalent: the former binds covalently to the nucleophilic serine hydroxyl of the Asp-His-Ser catalytic triad⁹. In those covalent bound compounds, the electrophilic group is referred to as the serine trap. For example, electrophilic carbonyls can reversibly form a hemiacetal with the serine hydroxyl and acylating and alkylating agents, such as β -lactams and mono-halomethylketones. These are grouped into the irreversible covalent inhibitors class⁷. After finding the reversible inhibitor Leupeptin¹⁰, a peptide aldehyde natural product, various aldehyde analogs were reported¹¹. However, since the recognition of the inherent chemical and metabolic instability of aldehyde serine traps, a number of groups developed alternative electrophilic carbonyl groups; amides, trifluoromethyl ketones, α -ketoesters, α -diketones, α -ketoheterocycles and organoboronic acids¹².

2.5 Disadvantages of covalent serine protease inhibitors

Why is it difficult to develop covalent inhibitors of mammalian serine proteases?

First, a 'serine trap' possesses a reactive functional group which makes the molecule more vulnerable to metabolism and may also lead to chemical instability⁷.

Furthermore, it can lead to a lack of specificity because of the highly conserved Asp-His-Ser catalytic triad of all serine proteases⁷.

Secondly, compounds containing a 'serine trap' tend to have slow binding kinetics because of slow engagement of the active site of serine¹³. Since the serine must react with the keto form, the degree of hydration of an electrophilic carbonyl serine trap will affect the rate of binding.

Thirdly, covalent inhibitors can show undesirable side effects. Generally, they bind irreversibly to an enzyme and consequently cause changes to the active site, which cannot be reversed.¹⁴ Then, substrates can no longer bind to the active site, the structure of which may be totally changed. Ultimately the protein is denatured.¹⁵ As a result, if proteins in a living cell are denatured, enzymes lose their activity leading to the disruption of cell activity and possibly cell death.¹⁶

Therefore, non-covalent inhibitors which may be chemically and metabolically less reactive and selective have been developed.

2.6 Strategy for non-covalent inhibitor design

According to many intensive reports aimed at the development of thrombin and factor Xa inhibitors with good selectivity and pharmacokinetic properties, the best strategy has been to remove the electrophilic serine trap and rely solely on ionic

interactions and hydrogen bonding at P1 as well as hydrophobic contacts with other parts of the binding groove⁷.

Chapter 3. Preliminary study

3.1 X-ray structure of apo-hGrB (PDB ID: 1FQ3)¹⁷

3.1.1 Overall Structure

Human granzyme B was crystallized without inhibitor, and its structure was determined to 3.1 Å resolution. It crystallizes as a dimer, and each molecule is folded into two six-stranded β -barrels resembling an ellipse (**Figure 2**).

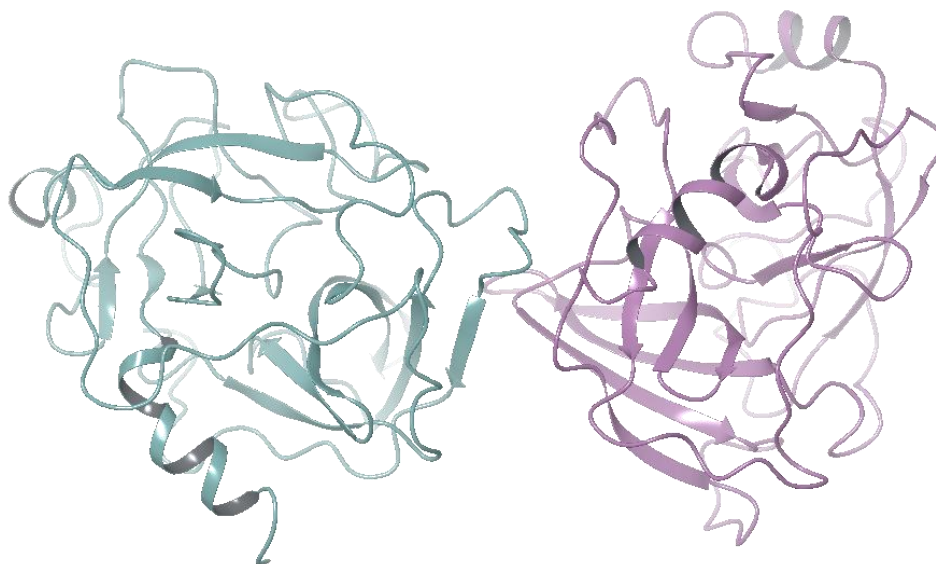


Figure 2. GrB molecules A (blue) and B (purple)

3.1.2 Active Site (Figure 3)

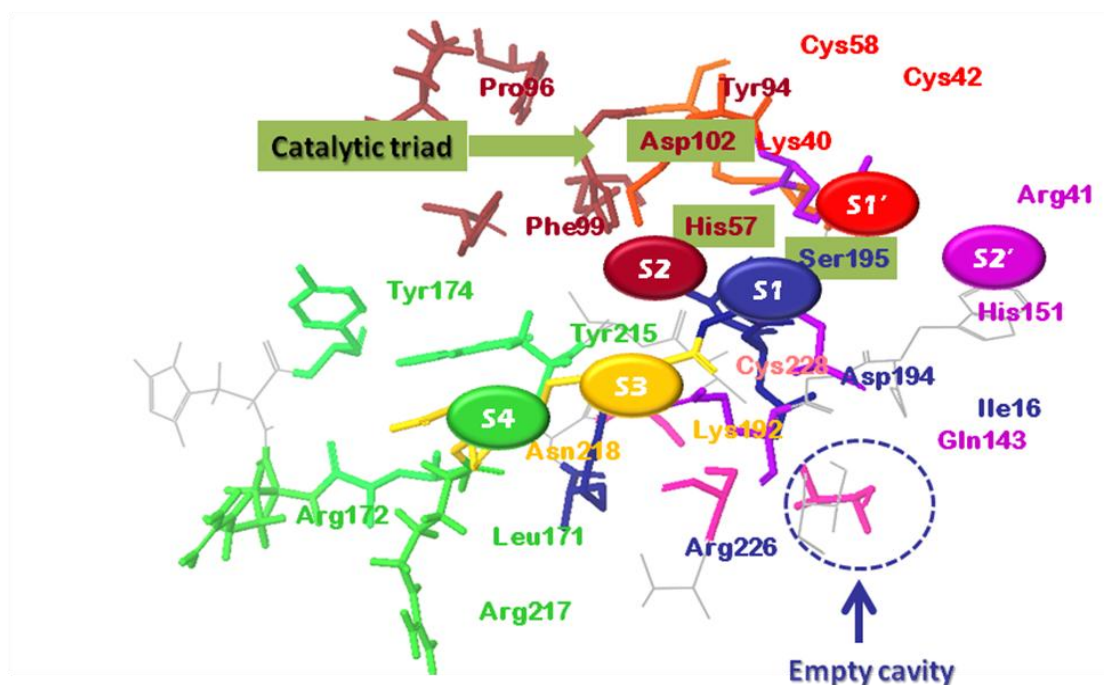


Figure 3. Subsites (S4-S2') and related residues for hexapeptide substrate

The specificity pocket (**S1**) strongly prefers a P1-Asp due to the Arg226 in the S1 binding pocket. In addition, Asp194 forms an internal salt bridge with Ile16 to generate the S1 pocket and the oxyanion hole.

The hydrophobic **S2** groove formed by Phe99 and His57 is shaped to accommodate small to medium sized hydrophobic functional groups with a wide exit toward the solvent to expose longer and more polar side chains.

The side chain of a P3 inhibitor would run along the hydrophobic molecular surface (**S3**), pointing between side chains of Lys192 and Asn218 whose terminal ammonium and carboxamide groups can form hydrogen bond interactions with polar side chains of P3. This seems to be one reason why GrB strongly prefers P3-Glu with another

factor. Namely the induction of an overall positive electrostatic surface by the underlying Arg226 side chain.

The **S4** subsite is quite narrow in the free enzyme, so first it must be opened to allow insertion of the P4 side chain. It looks as if rigid β -branched Ile or Val might be well suited to displace the Tyr174 phenolic group. In the voluminous hydrophobic cavity **S1'**, positively charged Lys40 provides an explanation for the preference of GrB towards Asp and Glu, but not towards Arg and Lys. Since the **S2'** site is flat and limited in size by the imidazole ring of His151, small P2' moieties should be acceptable in agreement with GrB's preference for P2'-Gly.

3.2 X-ray structure of hGrB complex with tetrapeptide aldehyde inhibitor (PDB ID: 1IAU)¹⁸

The 3D structure of the human granzyme B complex with a potent tetrapeptide aldehyde inhibitor (Ac-IEPD-CHO, $K_{i_{hGrB}} = 80$ nM, **Figure 4**) was determined to 2.0 Å resolution.

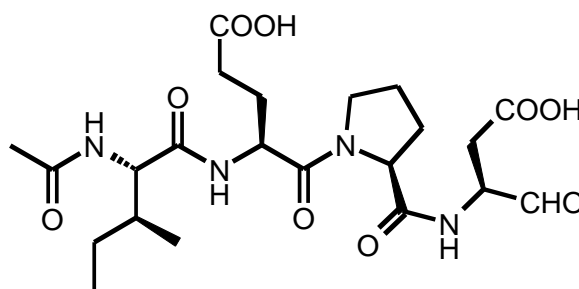


Figure 4. Tetrapeptide aldehyde Ac-IEPD-CH

The Asp-specific S1 subsite is significantly larger and less charged than caspases because of the presence of two arginines. The side chains at P1 and P4 project into the binding groove, whereas the side chains of P2 and P3 point away from the body of the protein.

3.3 Comparison between apo- and inhibitor-bound GrB

3.3.1 S1 subsite / P1-Asp

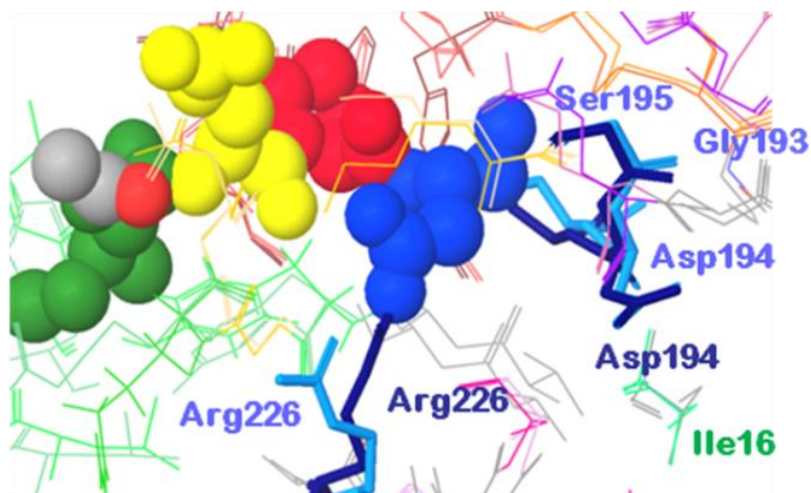


Figure 5. Movements of residues in the active site upon inhibitor binding. Protein residues and inhibitor are marked with tube and CPK, respectively. S1 subsite; dark blue (before inhibitor binding) vs light blue (after inhibitor binding).

Upon ligand binding, the positively charged GrB Arg226 side chain moves away to interact optimally with the negatively charged P1-Asp. In addition, Ile16 forms an internal salt bridge with Asp194, creating the oxyanion hole and functional S1 pocket (**Figure 5**).

3.3.2 Oxyanion Hole

To stabilize the oxyanion of the reaction intermediates formed during catalysis, an ‘oxyanion hole’ is created by the Gly193, Asp194 and Ser195 (**Figure 6**).

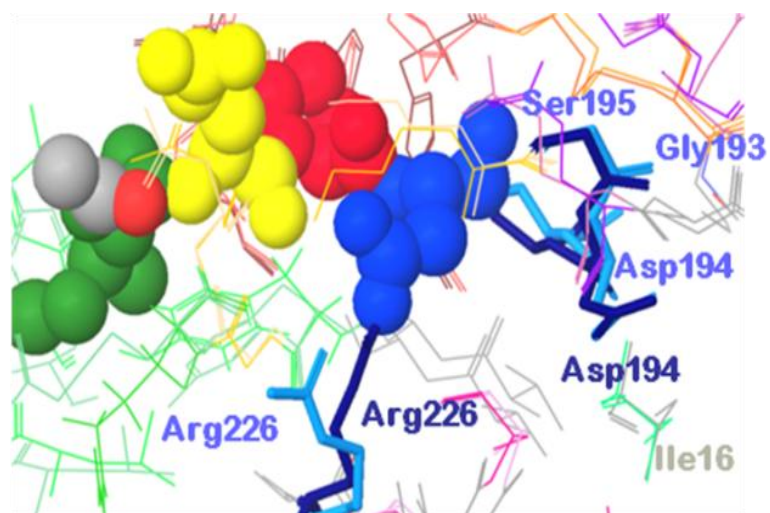


Figure 6. Movements of residues in the oxyanion hole upon inhibitor binding; dark blue (before inhibitor binding) vs light blue (after inhibitor binding).

3.3.3 S2 subsite / P2-Pro

A hydrophobic wall is generated by Phe99 and the flipping of His57. P2-proline points away from the body of the protein and makes no specific interactions with the protein. Instead, it induces a slight kink in the inhibitor allowing the P1-Asp to reach its optimal conformation in S1 (**Figure 7**).

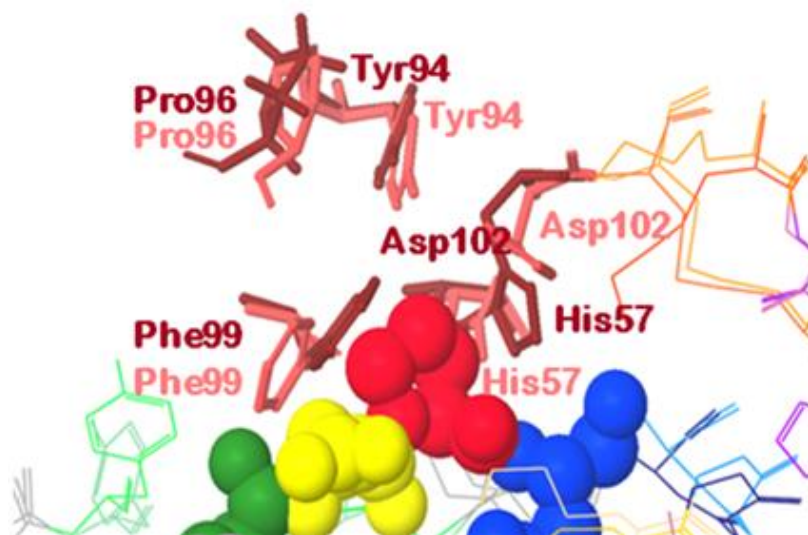


Figure 7. Movements of residues in the active site upon inhibitor binding. Protein residues and inhibitor are marked with tube and CPK, respectively. S2 subsite; dark red (before inhibitor binding) vs pink (after inhibitor binding).

3.3.4 S3 subsite / P3-Glu

Side chains of P3-Glu not only point away from the body of the protein but also form hydrogen bond interactions with Asn218, Gly216 and Lys192 (**Figure 8**).

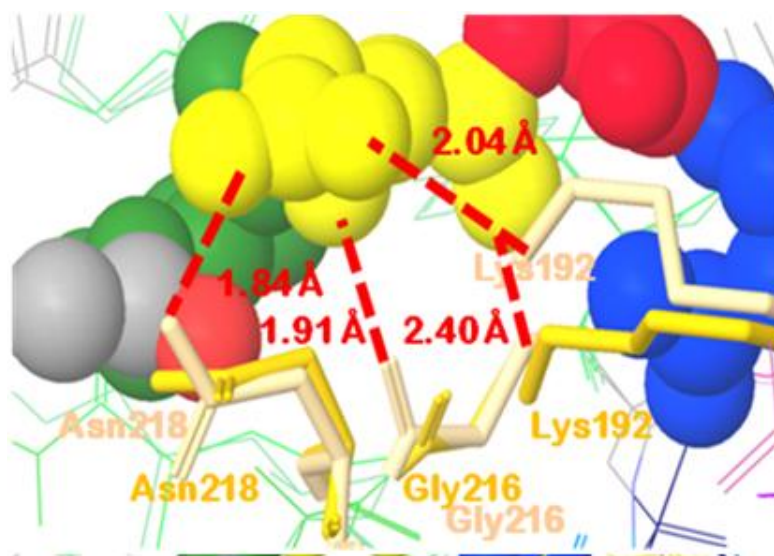


Figure 8. Movements of residues in the active site upon inhibitor binding. Protein residues and inhibitor are marked with tube and CPK, respectively. (A) S3 subsite; **dark yellow** (before inhibitor binding) vs **light yellow** (after inhibitor binding).

3.3.5 S4 subsite / P4-Ile

The shallow hydrophobic depression is formed by aromatic rings (Tyr174 and Tyr215) and the side chain of Leu171, while the surface edges are composed of two solvent exposed Arg (Arg217 and Arg172) which move a great deal upon inhibitor binding (**Figure 9**). The changed shape does not appear to be large enough for a sizeable hydrophobic group such as Phe.

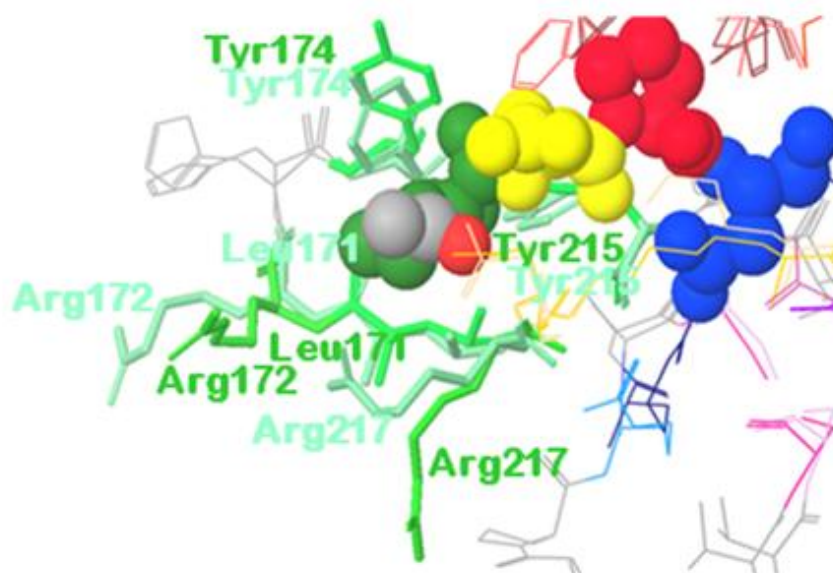


Figure 9. Movements of residues in the active site upon inhibitor binding. Protein residues and inhibitor are marked with tube and CPK, respectively. S4 subsite; dark green (before inhibitor binding) vs light green (after inhibitor binding).

3.3.6 S1' subsite / P1'-Trp (enough)

Considerable movements of the Lys40 side chain generated a large hydrophobic pocket formed by the disulfide bridge between Cys58 and Cys42, Lys40 and Ile35. This space seems generous enough to fit even a Trp (**Figure 10**).

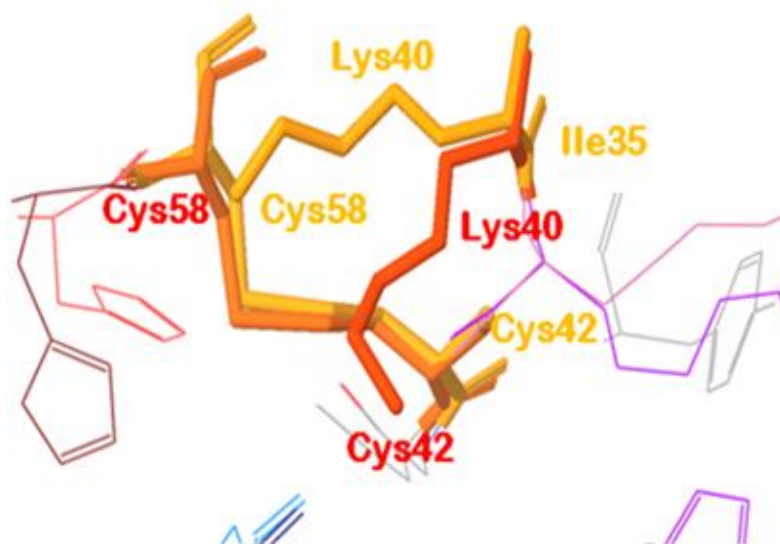


Figure 10. Movements of residues in the active site upon inhibitor binding. Protein residues and inhibitor are marked with tube and CPK, respectively. S1' subsite; dark orange (before inhibitor binding) vs light orange (after inhibitor binding).

3.3.7 S2' subsite / P2'-Gly

In the S2' subsite, a narrow corridor consistent with the specificity for Gly at P2' is generated by significant movements of side chains at His151, Gln143 and Arg41 (Figure 11).

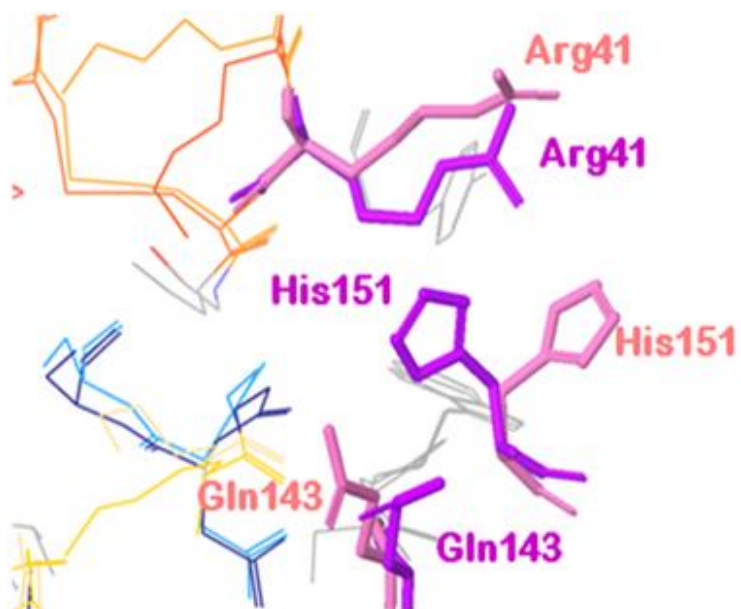


Figure 11. Movements of residues in the active site upon inhibitor binding. Protein residues and inhibitor are marked with tube and CPK, respectively. S2' subsite; dark purple (before inhibitor binding) vs light purple (after inhibitor binding).

It should be emphasized that the comments above and deductions made below depend on the X-ray structure of a single covalently bound inhibitor; see **Chapter. 10** below for discussion.

Chapter 4. Computational solvent mapping

4.1 FTMAP algorithm

To identify lead compounds from commercially available libraries, docking studies are one of the currently useful approaches. Even though various algorithms and scoring functions have been used in attempts to predict realistic binding poses and rank them, the following questions are still raised: which poses are real conformations in the receptor and which residues are important for the ligand binding.

According to various studies, it is believed that various residues in the binding region do not contribute equally to the binding free energy: smaller regions can provide major contributions to the binding free energy¹⁹. One calls those smaller regions of high or special activity ‘hot spots’. Such regions are more likely to bind small drug-like molecules with higher affinity than the rest of the binding site²⁰.

Therefore, their identification is very important for drug design. To address this problem, computational solvent mapping had been performed to explore the protein’s binding properties and understand interactions between proteins and solvent molecules by placing multiple copies of 16 different small organic molecular probes on the protein surface using the FTMAP algorithm based on the extremely efficient fast Fourier transform (FFT) correlation approach²¹. For each of the 16 probes, the mapping algorithm generates 2000 best bound positions through rigid body docking followed by energy minimization to refine the positions. The resulting conformations are clustered, and all clusters are ranked on the basis of average free energy. For each probe, six clusters with the lowest average free energy are kept and overlapping clusters of different probes are defined as consensus sites (CSs). The largest CS is generally located at the most

important subsite of the protein binding site, and the nearby smaller CSs identify other important subsites (**Figure 12**).

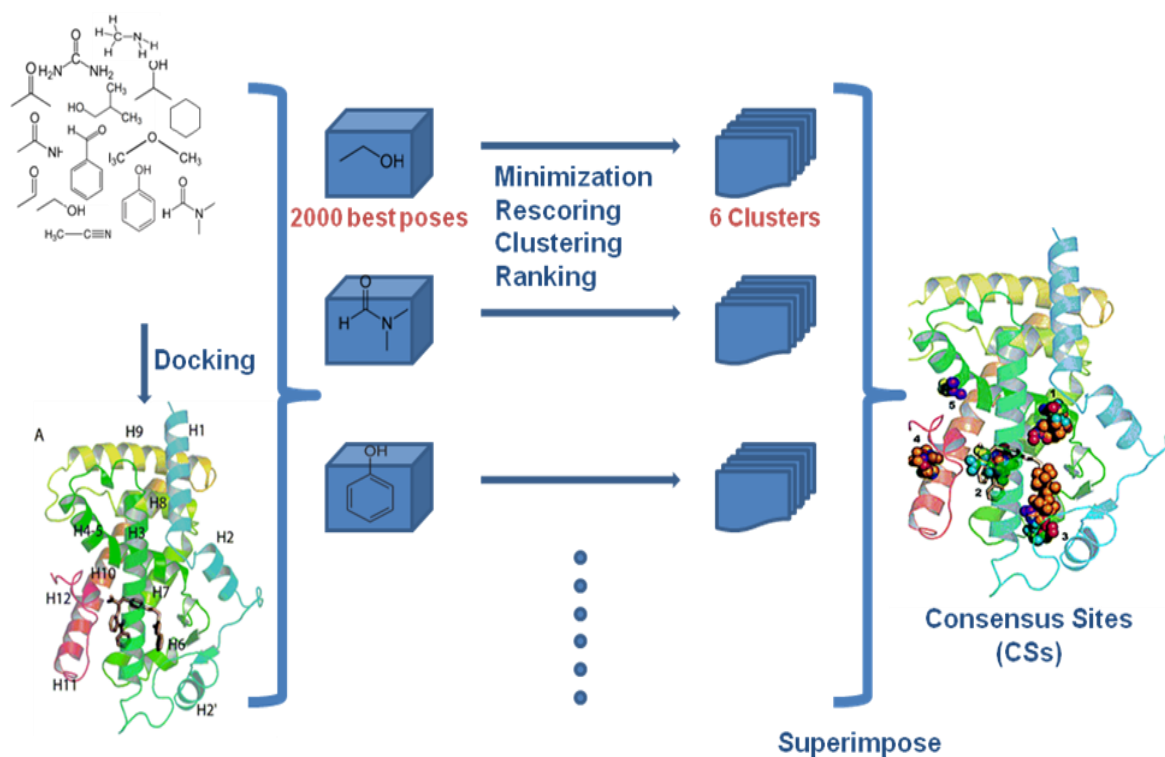


Figure 12. Outline of computational solvent mapping algorithms used by FTMAP.

Two X-ray structures (apo- (PDB ID: 1FQ3) and covalent inhibitor bound (PDB ID: 1IAU) GrB) were mapped to not only identify hot spots, but also compared the two for side chain locations in the binding sites. The comparison shows what changes occurred after inhibitor binding to the protein. Inhibitor and water molecules were removed before the mapping.

4.2 Computational solvent mapping results comparison before and after inhibitor binding

4.2.1 S1 subsite / P1-Asp

As discussed above (**Figure 5**), changes in both size and the position of S1 subsite from the apo form to the bound form increased the volume in the bound form allowing three water molecules (red spheres) to be bound along with the inhibitor (**Figure 13. (a)**).

Computational solvent mapping results also show a deeper S1 subsite than that before inhibitor binding: more probes bind in the deep S1 pocket induced by ligand binding, in good agreement with X-ray results (**Figure 13. (b)**).

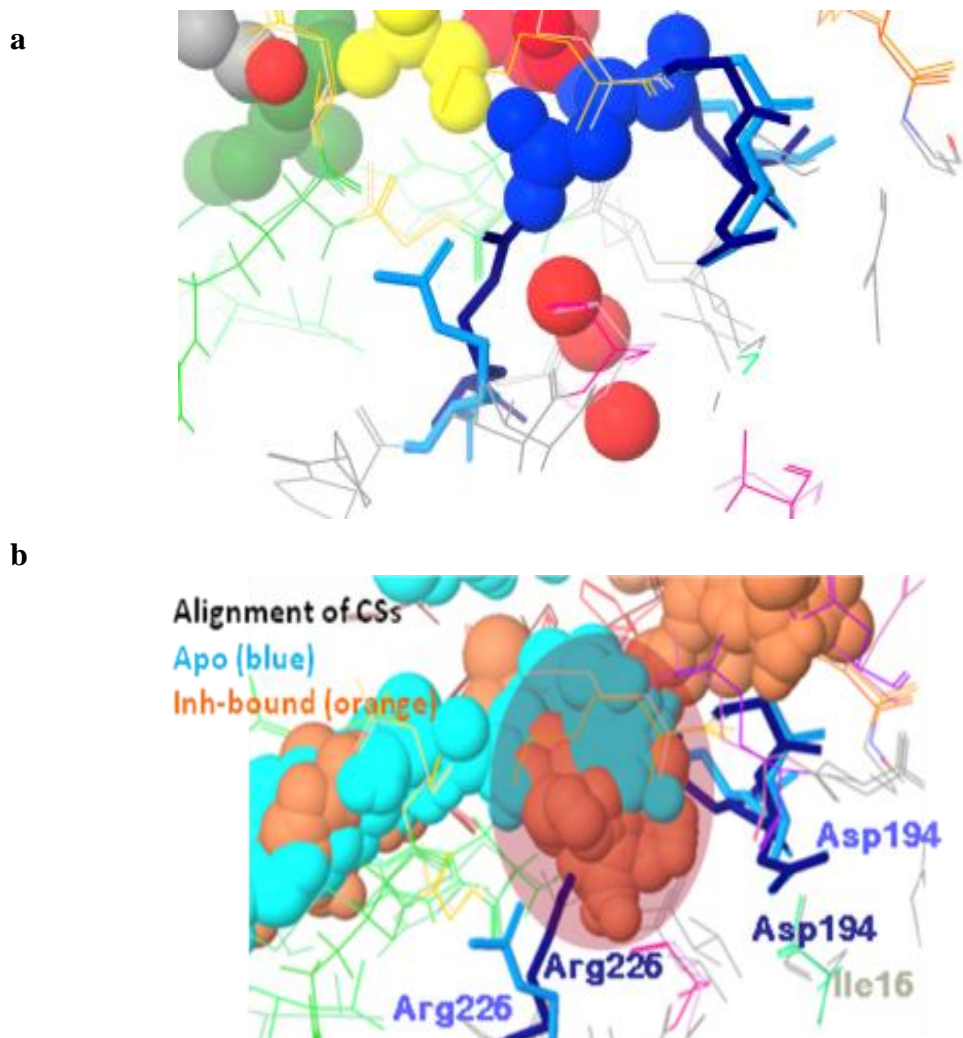


Figure 13. Water molecules in the S1 pocket and alignment of consensus sites for the apo and inhibitor bound forms of GrB. **(a)** Changes in both size and position of the S1 subsite from the apo form to the bound form (covalently bound inhibitor shown as four sets of colored CPK) is illustrated by the dark blue and light blue sticks, respectively. The volume increase in the bound form allows three water molecules (red spheres) to be bound along with the inhibitor; **(b)** Alignments of the consensus sites resulting from computational solvent mapping; blue (apo structure before inhibitor binding) vs orange (bound form after inhibitor binding).

4.2.2 S2 subsite / P2-Pro

The new generated pocket by movements of Phe99 and His57 as well as Pro96, Tyr94 and Asp102 was also found in the mapping results (**Figure 14**)

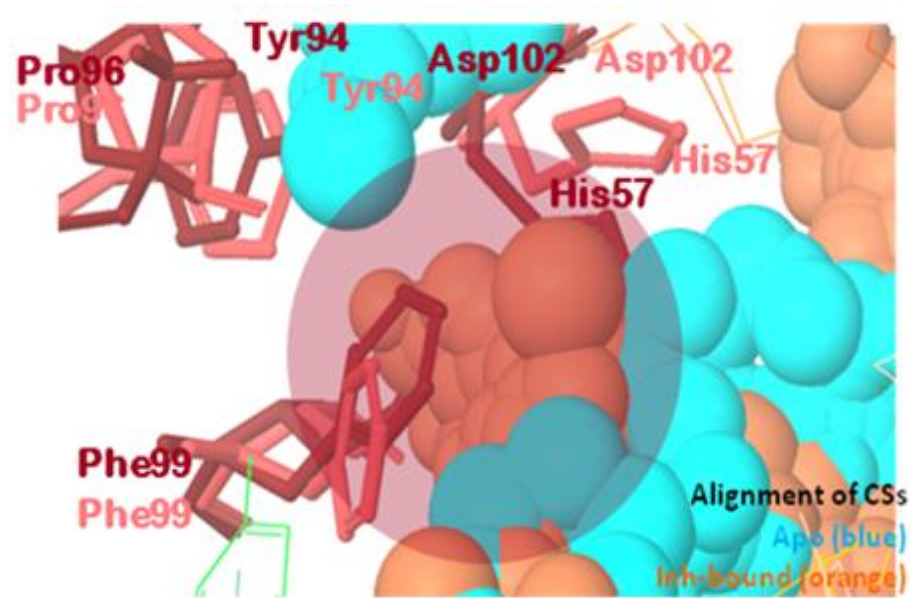


Figure 14. Alignments of consensus sites resulting from computational solvent mapping; blue (before inhibitor binding) vs orange (after inhibitor binding).

4.2.3 S3 subsite / P3-Glu

Some changes of the probes binding site from computational solvent mapping were observed due to the side chain movements of those three residues, namely Asn218, Gly216 and Lys192, to form hydrogen bond interactions with the P3-Glu (**Figure 15**).

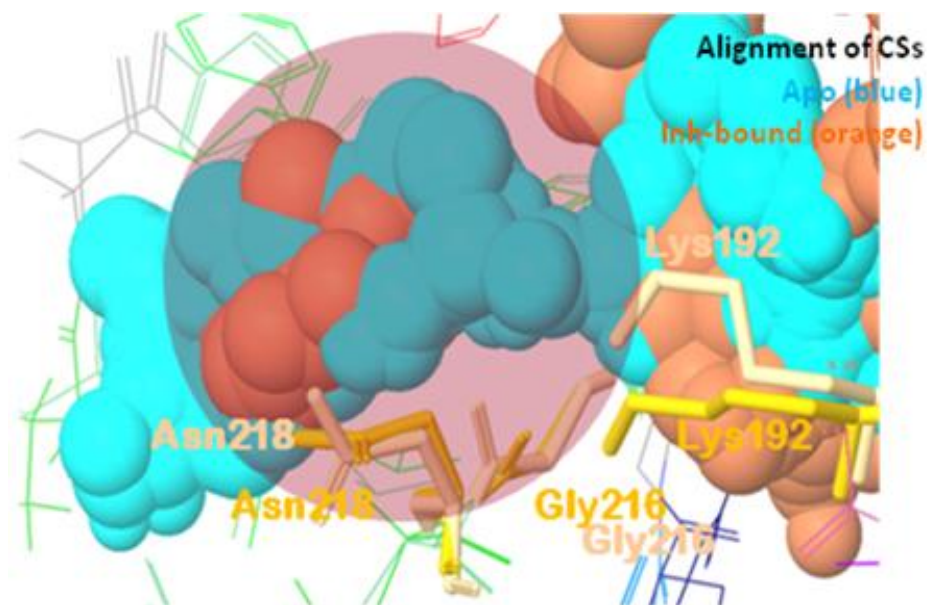


Figure 15. Alignments of consensus sites resulting from computational solvent mapping; blue (before inhibitor binding) vs orange (after inhibitor binding).

4.2.4 S4 subsite / P4-Ile

The changes in the S4 subsite were also observed by computational solvent mapping due to the side chain movements around this area, especially the two Arg (Arg217 and Arg172) (**Figure 16**).

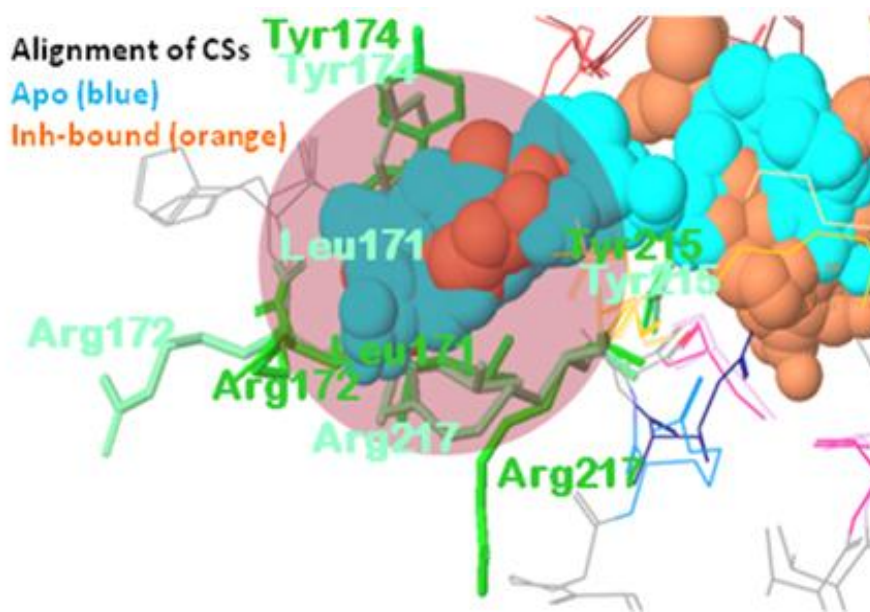


Figure 16. Alignments of consensus sites resulting from computational solvent mapping; blue (before inhibitor binding) vs orange (after inhibitor binding).

4.2.5 S1' subsite / P1'-Trp (enough)

During the computational solvent mapping, probes also bind within the large hydrophobic S1' pocket induced by the inhibitor, where no binding occurred in the apo-GrB (Figure 17)

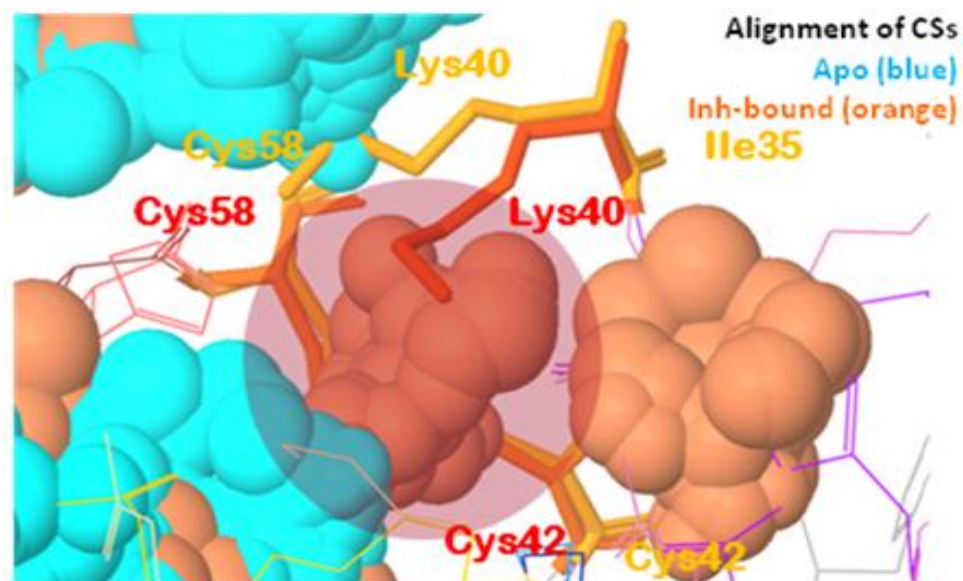


Figure 17. Alignments of consensus sites resulting from computational solvent mapping; blue (before inhibitor binding) vs orange (after inhibitor binding).

4.2.6 S2' subsite / P2'-Gly

The probes bind in this hydrophobic pocket induced by inhibitor binding, where no binding was observed in the apo-GrB (**Figure 18**).

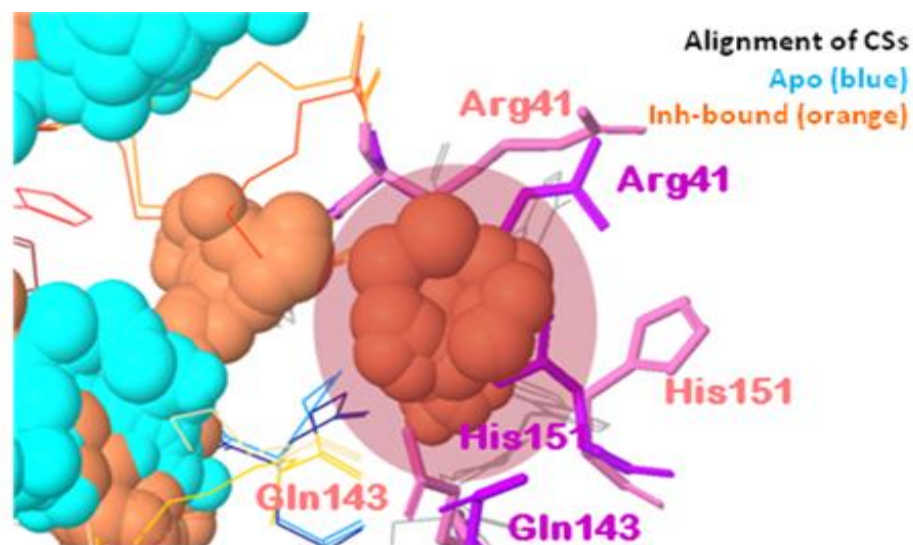


Figure 18. Alignments of consensus sites resulting from computational solvent mapping; blue (before inhibitor binding) vs orange (after inhibitor binding).

4.3 Summary of GrB mapping results

Table 3 lists the eight largest CSs located in the active site for each structure. The three largest CSs for apo-GrB are subsites S3-S4, Over S1' and S1-S2 with 27, 14 and 14 probe clusters, respectively (**Table 3** and **Figure 19**). There is also one smaller cluster in extended S2.

After inhibitor binding, the conformation of the protein is changed, especially in the binding site. Similar observations regarding plasticity in the binding site have been made in the comparison of covalent and noncovalent inhibitors of elastase.²² FTMAP also identified all subsites (S4-S2') that bind organic molecules in the overall active site (**Table 3** and **Figure 19**). Note that the prime subsites (S1'-S2') appear, and new pockets deep in S2 were generated after inhibitor binding.

Table 3. Summary of GrB mapping results

No	1FQ3		1IAU	
	Apo-GrB		Ac-IEPD-CHO-GrB	
	size	pocket	size	pocket
1	27	S3 – S4	17	S4
2	14	Over S1'	16	Not active site
3	14	S1 – S2	13	S1
4	11	Not active site	12	S2
5	9	Not active site	11	S1'
6	9	Extended S2	9	Over S2 & S1'
7	7	Not active site	8	S2'
8	2	Not active site	8	Not active site

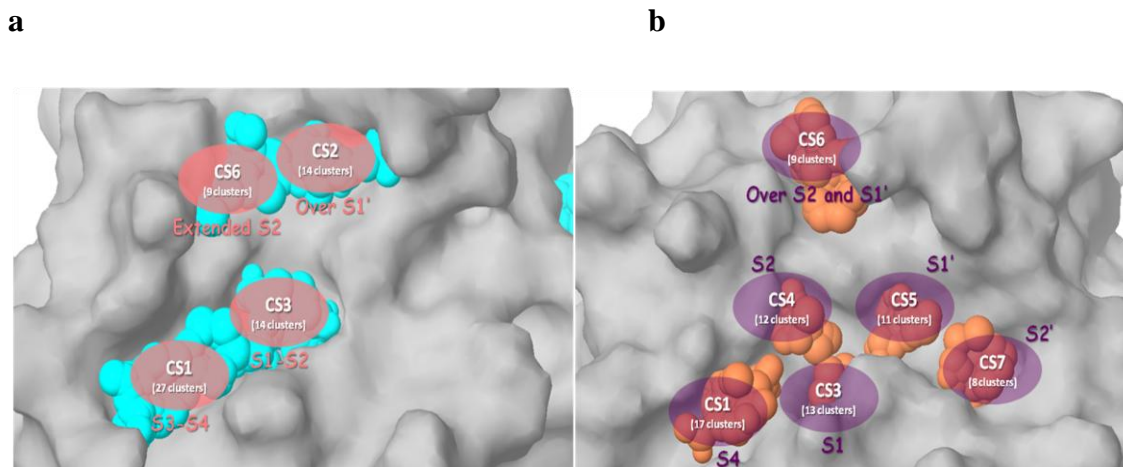


Figure 19. Consensus sites in the binding site of (A) apo GrB (PDB code 1FQ3) and (B) Ac-IEPD-CHO-GrB (PDB code 1IAU) indicating a reorganization of the binding site residues.

4.4 Distribution analysis of non-bonded interactions and hydrogen bonds between probes and hGrB residues

Figure 20 shows the distribution of non-bonded interactions between probes and GrB residues in the active site. Before inhibitor binding, major residues participating in non-bonded interactions are; Ser195 of S1; His57 and Phe99 of S2; Leu171, Tyr174 and Tyr215 of S4. After inhibitor binding, interactions with Arg41 of S2', Lys40 of S1', Phe191 of S1 and Phe99 of S2 increased. On the other hand, interactions with Leu171 of S4 decreased. In both cases, Lys192 extends from S1 to S3 and establishes a barrier between these sites and the solvent pool. This is significant for understanding how this important residue develops hydrophobic and hydrophilic interactions with inhibitors as illustrated by probe contacts shown in **Figures 20** and **21**.

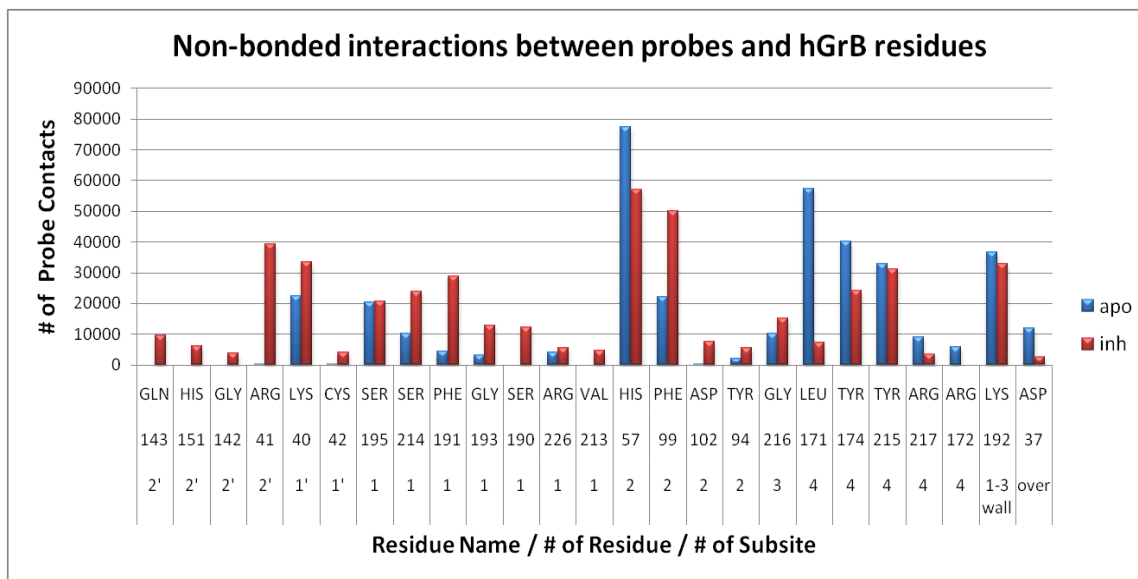


Figure 20. Intermolecular non-bonded interactions between probes and residues for apo-GrB (blue) and Ac-IEPD-CHO-bound GrB (red) without the ligand.

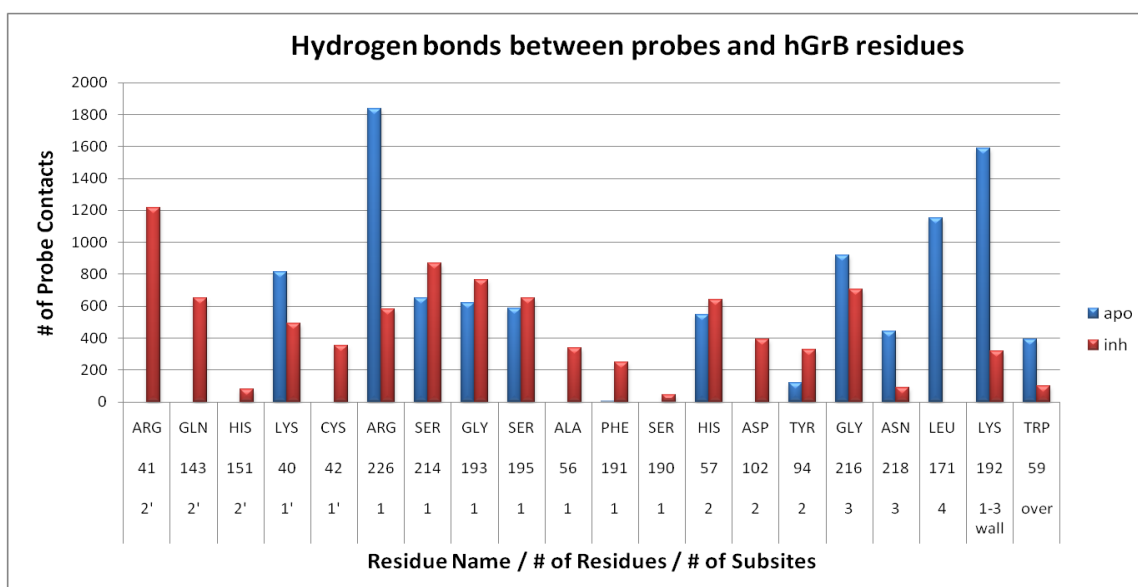


Figure 21. Intermolecular hydrogen bond interactions between probes and residues for apo-GrB (blue) and Ac-IEPD-CHO-bound GrB (red) without the ligand.

In the same way, **Figure 21** shows the distribution of hydrogen bond interactions between probes and GrB residues in the active site. Before inhibitor binding, the major residues participated in hydrogen bond interactions are; Lys40 of S1'; Arg226, Ser214, Gly193, Ser195 of S1; His57 of S2; Gly216 of S3; Leu171 of S4; and Lys192 of the wall between S1-S3. After inhibitor binding, dramatic changes occurred: interactions with residues of S2' and S1' increased. On the other hand interactions with Leu171 of S4 disappeared, possibly increasing the hydrophobicity of S4. Surprisingly, the number of probe contacts with Arg226 in the S1 decreased dramatically after inhibitor binding. By contrast, interactions with Ser214, Gly193 and Ser195 increased.

4.5 Three constraints

Based on the mapping results, three constraints were selected for docking studies utilizing the Ac-IEPD-CHO bound protein (PDB ID: 1IAU): first, hydrogen bond with Arg226 in S1 to identify molecules with P1-Asp specificity motif; second, S2 pocket that should be occupied (**Figure 16. A**); third, hydrophobic interaction in the S4 pocket (**Figure 16. B**).

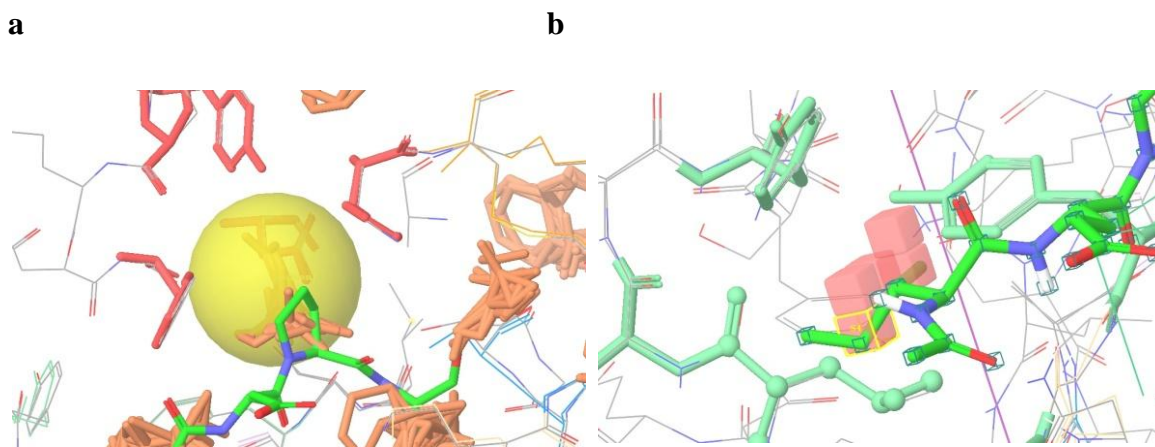


Figure 22. Selected constraints based on mapping results for the docking study: (A) deeper S2 pocket generated after inhibitor binding should be occupied and (circle in yellow) (B) hydrophobic interactions in the S4 pocket (cubes in pink).

Chapter 5. Virtual screening

5.1 Libraries

5.1.1 Maybridge Screening Collection (MSC)²³

The Maybridge Screening Collection consists of over 56,000 organic compounds available from Maybridge. The compounds are the most diverse out of 10 commercially available libraries and the ACD. Of ~400,000 pharmacophores in the world drug index, ~87 % are expressed by the MSC. Moreover, they generally obey Lipinski's 'rule of 5' and show good ADME (Absorption, Distribution, Metabolism and Excretion) profiles. The purity is greater than 90 % and reactive molecules are already excluded to reduce the number of false positives.

5.1.2 TimTec ActiTarg-P (Protease targeted library)²⁴

The ActiTarg-P collection produced by TimTec is a screening set of 1,520 molecules that possess chemical lattices present in compounds reported in the technical or patent literature to contain various protease-inhibiting properties. This targeted collection was built on the basis of structure and fragment similarity to known protease inhibitors in order to probe diverse chemical space.

5.1.3 ChemDiv (Protease targeted library and Asp recognition motif)²⁵

8,300 compounds from the protease library and 430 compounds with an aspartate recognition motif were selected for further screening. They were designed based on active ligands/inhibitors as prototypes from the patent and research literature or from a compound database, compounds of which were further substituted with various bioisosteres.

5.2 Glide Docking

All compounds from the three libraries (Maybridge Screening Collection total = over 56,000, Timtec ActiTarg-P (protease targeted library) total = 1,520, ChemDiv (protease targeted library and compounds with aspartic recognition motif) total = 8.730) were screened by Glide standard-precision (SP) docking using three constraints generated from computational solvent mapping. As mentioned above, the latter identifies binding 'hot spots' that provide major contributions to the binding free energy following placement of 16 diverse organic molecular probes on the protein surface. As a result, a

total 4,552 hits (715 from Maybridge Screening Collection, 203 from ActiTarg-P and 3,634 from ChemDiv) were selected.

Second, extra-precision (XP) docking, which was developed to weed out false positives and provide a better correlation between good poses and good scores, was performed with 715 hits of Maybridge SC and 3,634 from ChemDiv to select more reasonable poses and reduce the number of purchased compounds for the first bio-assay. Finally, a total of 511 available compounds (46 from Maybridge SC, 203 from ActiTarg-P and 262 from ChemDiv) were ordered for bio-assay.

Chapter 6. Bioassay results

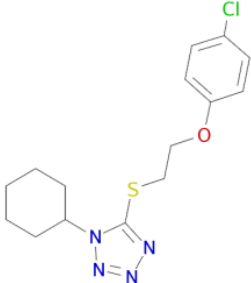
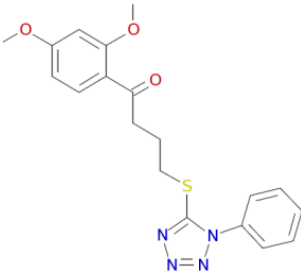
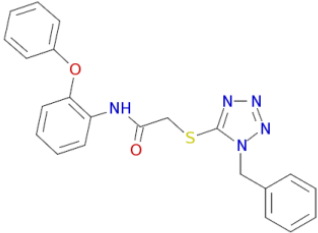
First of all, I deeply appreciate Drs. Dean Heathcote and Philip G Ashton-Rickardt of Imperial College London as collaborators who turned around the bio-data quickly and efficiently.

As a result of screening of 511 selected hits from the virtual screening of 66,250 compounds from three libraries at 100 μ M followed by serial dilution of the compounds showing greater than 40 % inhibition of enzyme activity, a total of 28 compounds (21 compounds from TimTec and 7 compounds from ChemDiv) showed inhibition and calculated IC_{50} values. In addition, any compound showing inhibition under 50 μ M was selected for selectivity testing against GrA, caspase-3 and caspase-8.

6.1 Hits from TimTec ActiTarg-P

6.1.1 Thiotetrazole series

Table 4. Thiotetrazole series of compounds showing inhibition against GrB and selectivity results.

Compound	Structure	MW	GrB	GrA	Casp-3	Casp-8
			IC ₅₀ (μ M)	IC ₅₀ (μ M)	IC ₅₀ (μ M)	IC ₅₀ (μ M)
ST053449		338.85	35	95	>100	82
ST054336		384.45	45	>100	75	62.5
ST025517		417.48	69			

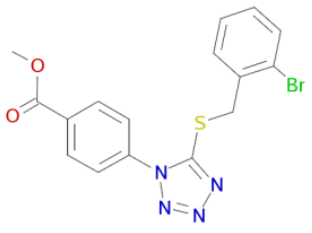
ST056810		405.27	80			
----------	---	--------	----	--	--	--

Figure 23 shows the predicted binding mode of **ST053449** ($IC_{50} = 35 \mu M$). The tetrazole ring is located in the P1-Asp recognition position, forming weak H-bond interaction with Arg226 (2.45 Å). In addition, in all these series, there is enough room to add additional functional groups for both HB and HP interactions in the deep S1 pocket. There is also room to improve activity for HP at the P2 and P4 positions.

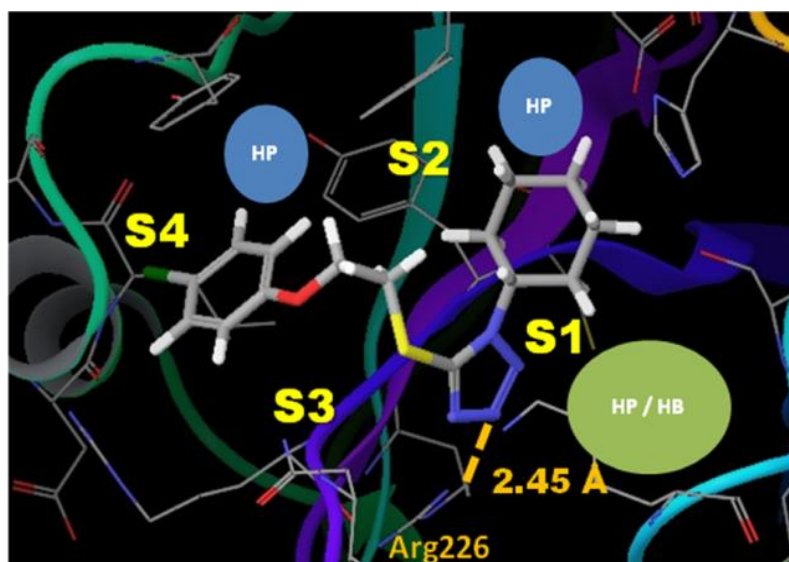


Figure 23. The predicted binding mode of ST053449. A hydrogen bond between Arg226 and N tetrazole is indicated with the yellow dotted line (2.45 Å).

6.1.2 Thiotriazole series.

Table 5. Thiotriazole series showing inhibition against GrB and selectivity results.

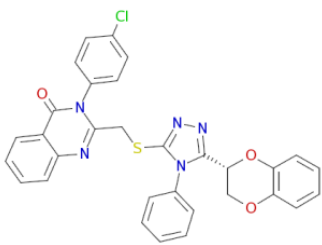
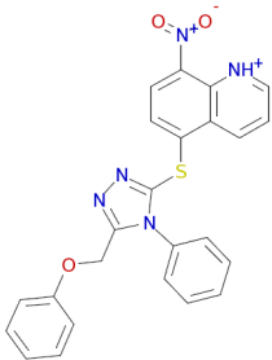
Compound	Structure	MW	GrB	GrA	Casp-3	Casp-8
			IC ₅₀ (μ M)	IC ₅₀ (μ M)	IC ₅₀ (μ M)	IC ₅₀ (μ M)
ST057416		580.06	44	>100	>100	>100
ST045738		455.49	74			

Figure 24 shows the predicted binding mode of **ST057416** ($IC_{50} = 44 \mu M$). The O of benzodioxin forms an HB with Arg226, and chlorophenyl points away from the body (2.01 Å). In addition, there is some space to add a small aliphatic group such as methyl to phenyl in the S2 to improve activity.

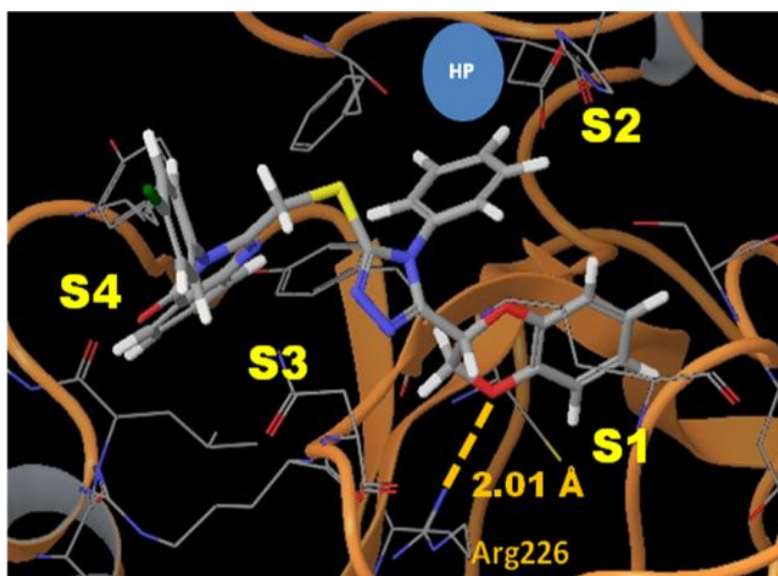
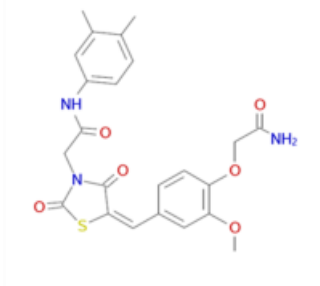
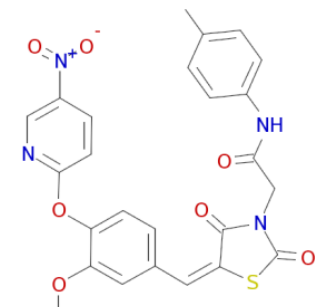
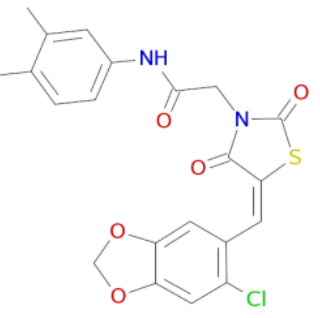
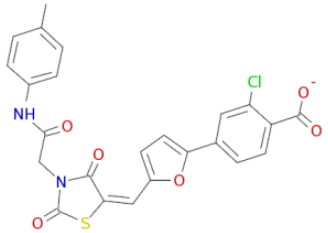
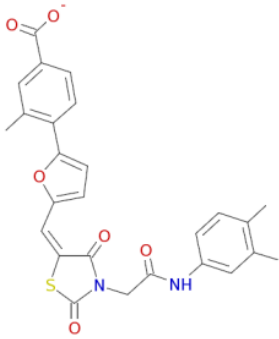
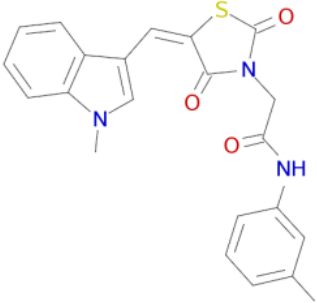
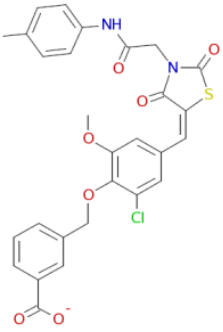


Figure 24. The predicted binding mode of **ST057416**. A hydrogen bond between Arg226 and the O of benzodioxin is indicated with yellow dotted line (2.01 Å).

6.1.3 Thiazolidinedione series

Table 6. Thiazolidinedione series showing inhibition against GrB and selectivity results.

Compound	Structure	MW	GrB	GrA	Casp-3	Casp-8
			IC ₅₀ (μ M)	IC ₅₀ (μ M)	IC ₅₀ (μ M)	IC ₅₀ (μ M)
ST057924		469.51	25	>100	>100	>100
ST057833		520.52	28	>100	>100	>100
ST057890		444.89	33	37.5	82.5	74

ST058173		496.92	33	>100	>100	44
ST057942		490.53	38			
ST057871		405.47	38	>100	47	>100
ST057842		567.01	67			

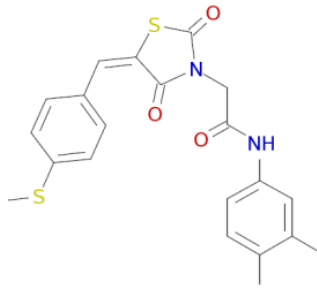
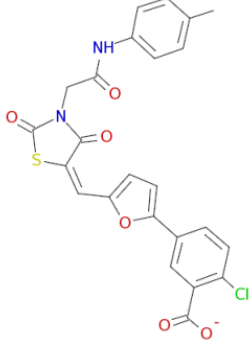
ST057992		412.52	68			
ST057824		496.92	100			

Figure 25 shows the predicted binding mode of **ST057924** ($IC_{50} = 25 \mu M$).

Acetamide is located in the P1-Asp recognition position, forming weak a HB interaction between the carbonyl O and the NH of Arg226 (2.47 Å). In addition, there is more space to add hydrophobic functional groups in both P2 and P4 positions.

In addition, **Figure 26** shows the predicted binding mode of **ST057833** ($IC_{50} = 28 \mu M$). A nitro-phenyl group is located in the P1-Asp recognition position, forming a strong HB interaction between a nitro O and NH of Arg226 (1.76 Å). In addition, there are two more HB interactions: between NH of P1-phenyl and Ser195 (1.79 Å) and between the amide O and NH of Asn218 (2.02 Å). There is also more space to add hydrophobic functional groups in both the P2 and P4 positions.

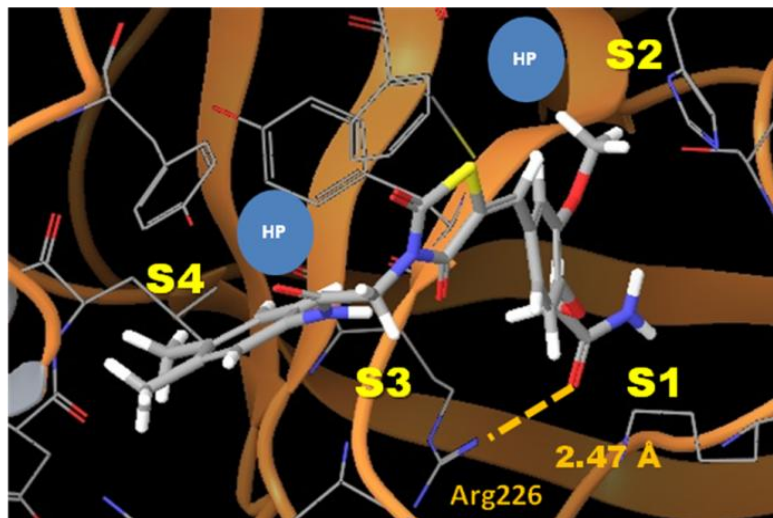


Figure 25. The predicted binding mode of ST057924. A hydrogen bond between Arg226 and the carbonyl O is indicated with yellow dotted line (2.47 Å).

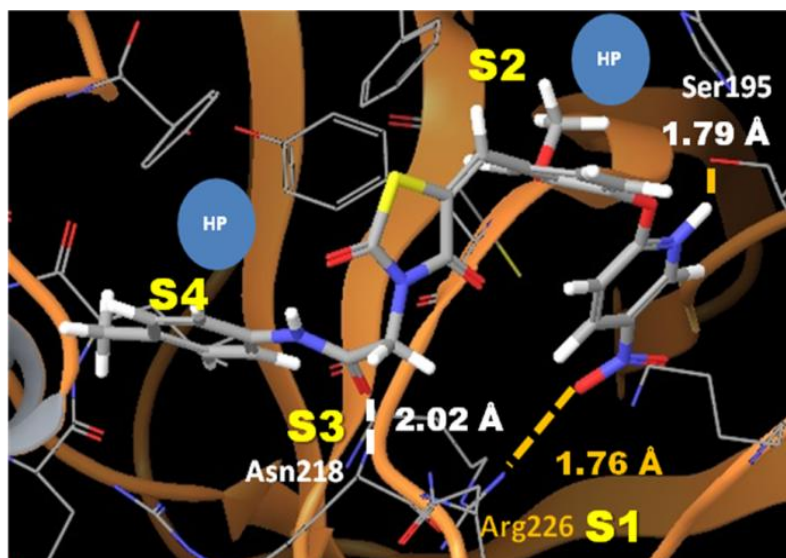
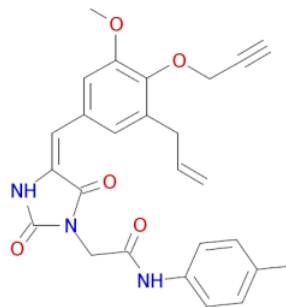
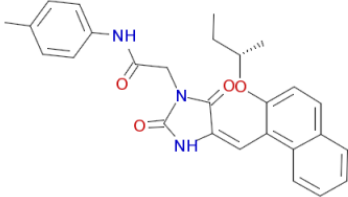


Figure 26. The predicted binding mode of ST057833. A hydrogen bond between Arg226 and the nitro O is indicated with yellow dotted line (1.76 Å).

Within this series, compounds having carboxylic acid substituents (ST058173, ST057942, ST057842, ST057824) show similar conformations in that the carboxylic acid groups and form HB interactions with Arg226 in the P1-Asp recognition position within 2.3 Å.

6.1.4 Diazolidinedione series

Table 7. Diazolidinedione series showing inhibition against GrB and selectivity results.

Compound	Structure	MW	GrB	GrA	Casp-3	Casp-8
			IC ₅₀ (μM)	IC ₅₀ (μM)	IC ₅₀ (μM)	IC ₅₀ (μM)
ST058025		459.50	44	>100	>100	>100
ST058039		457.53	53	>100	>100	67.5

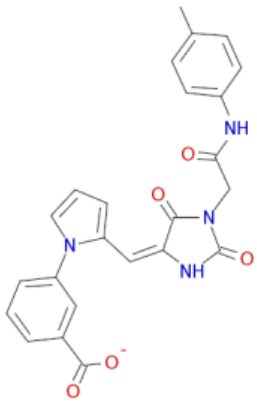
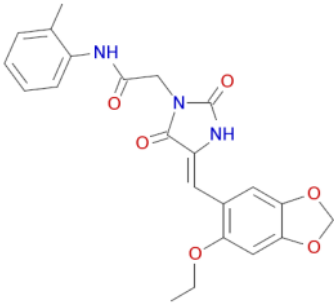
ST057786	 <p>The chemical structure of ST057786 features a central pyrazole ring. One nitrogen of the pyrazole is substituted with a 4-(4-carboxyphenyl)pyrrolidin-2-yl group. The other nitrogen is substituted with a 2-(4-aminophenyl)acetamido group. The pyrazole ring also has a carbonyl group at the 3-position and a double bond at the 5-position.</p>	444.45	87			
ST057803	 <p>The chemical structure of ST057803 features a central pyrazole ring. One nitrogen is substituted with a 2-(4-aminophenyl)acetamido group. The other nitrogen is substituted with a 2-(4-ethoxyphenoxy)phenyl group. The pyrazole ring also has a carbonyl group at the 3-position and a double bond at the 5-position.</p>	423.42	95			

Figure 27 shows the predicted binding mode of **ST058025** ($IC_{50} = 44 \mu\text{M}$). Weak H-bond interactions between the methoxy O and Arg226 (2.65 \AA) were predicted. There is also room to improve activity for HP in the P2 and P4 positions.

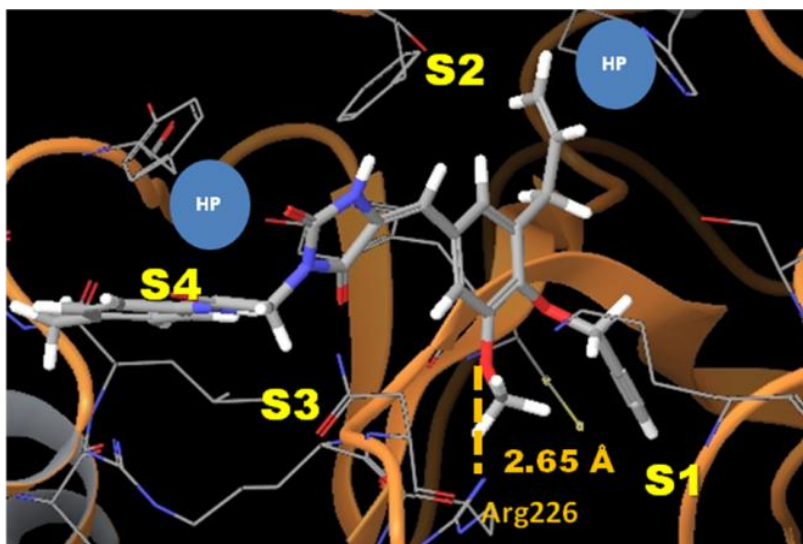


Figure 27. The predicted binding mode of **ST058025**. A hydrogen bond between Arg226 and the methoxy O is indicated with yellow dotted line (2.65 \AA).

Note that both the thiazolidinedione and diazolidinedione series possess Michael acceptors, which could make covalent bonds with Ser195. However, they don't seem to bind in the same fashion as the covalent inhibitor, Ac-IEPD-CHO, based on docking studies.

6.1.5 Others

Table 8. Other compounds showing inhibition against GrB and selectivity results.

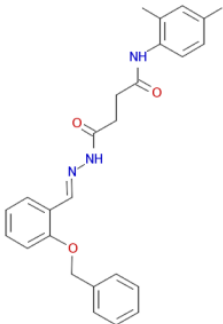
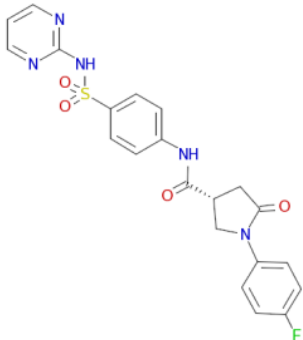
Compound	Structure	MW	GrB	GrA	Casp-3	Casp-8
			IC ₅₀ (μ M)	IC ₅₀ (μ M)	IC ₅₀ (μ M)	IC ₅₀ (μ M)
ST014575		429.52	54	>100	>100	67.5
ST066636		455.46	63			

Figure 28 shows the predicted binding mode of **ST014575** (IC₅₀ = 54 μ M). Although there are three constraints during docking including HB with Arg226, this compound doesn't have any functional group for the latter HB interaction. However, the NH and N of the aza group form strong HB interactions with O and NH of Gly216 within 1.59 Å and 2.27 Å, respectively. And to improve activity, it seems that modification of HB in P1 phenyl, HP in P2 benzyl and P4 phenyl would be effective.

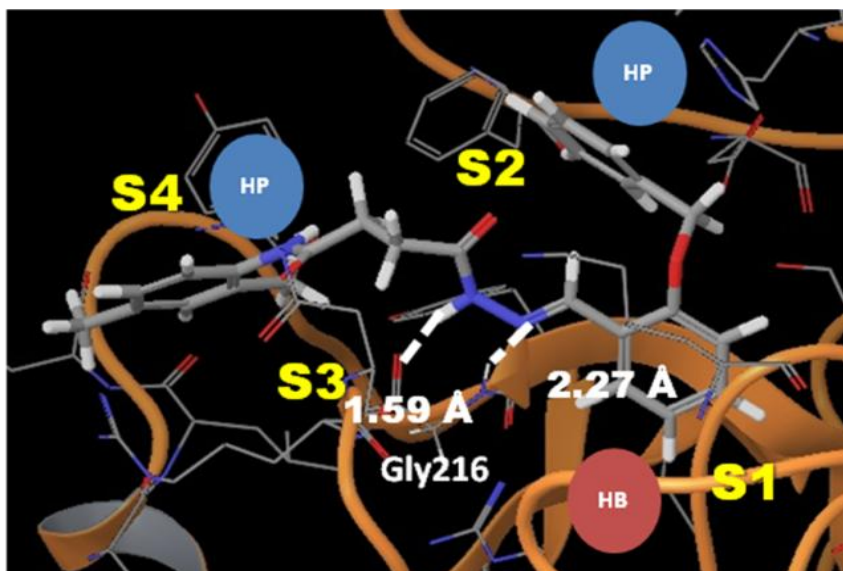


Figure 28. The predicted binding mode of **ST014575**. Hydrogen bonds between Gly216 and the NH-N(=) moiety is indicated with yellow dotted line (1.59 and 2.27 Å, respectively).

6.2 Hits from ChemDiv

6.2.1 Sulfonamide series

Table 9. Sulfonamide series showing inhibition against GrB and selectivity results.

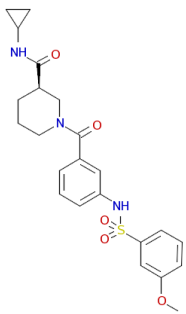
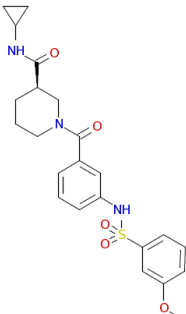
Compound	Structure	MW	GrB	GrA	Casp-3	Casp-8
			IC ₅₀ (μ M)	IC ₅₀ (μ M)	IC ₅₀ (μ M)	IC ₅₀ (μ M)
L581-0227		457.54	45	>100	>100	>100
L582-0183		457.54	50			

Figure 29 shows the predicted binding mode of **L581-0227** ($IC_{50} = 45 \mu M$). The methoxyphenyl goes into the S1 pocket, forming a HB interaction between the methoxy O and the NH of Gly216 (2.01 Å), terminal N and O of Gly216 (1.90 Å). Sulfonamide group form strong HB interactions with Gly193 and Ser195 (2.07 Å and 2.03 Å, respectively) of S1 pocket. The benzyl group in the middle of the structure occupies the S2 pocket, and the cyclopropyl group makes HP interactions in the S4 pocket.

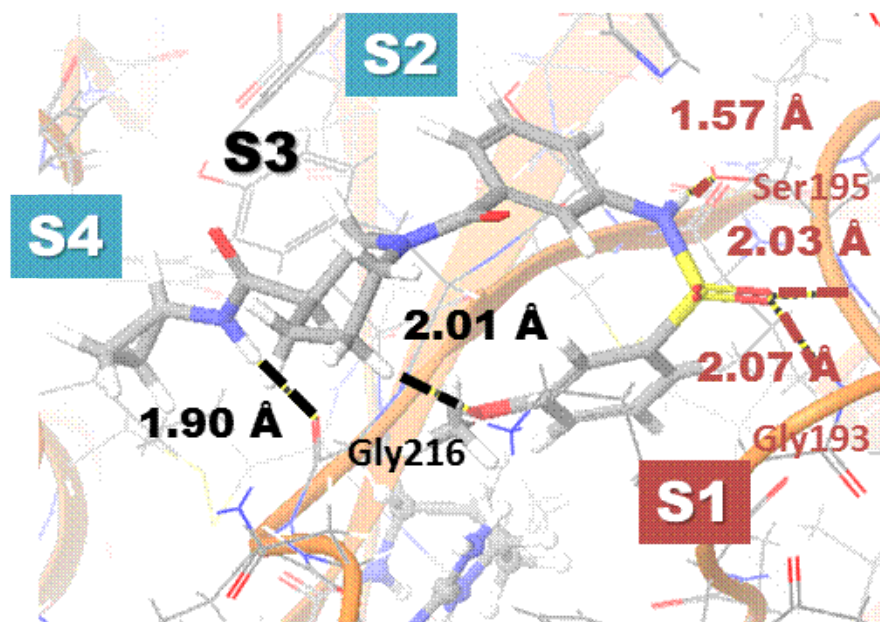
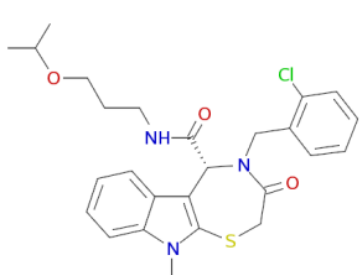
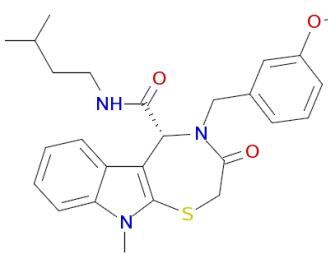
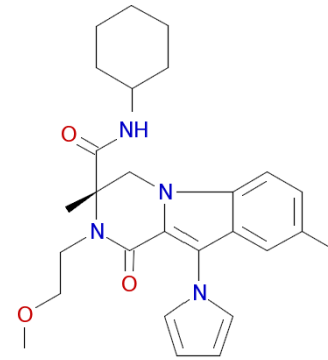


Figure 29. The predicted binding mode of **L581-0227**. Hydrogen bonds between the methoxy O and the NH of Gly216 (2.01 Å), terminal N and O of Gly216 (1.90 Å), sulfonamide group with Gly193 and Ser195 (2.07 Å and 2.03 Å, respectively) of the S1 pocket are indicated.

6.2.2 Tricyclic series

Table 10. Tricyclic series showing inhibition against GrB and selectivity results.

Compound	Structure	MW	GrB	GrA	Casp-3	Casp-8
			IC ₅₀ (μ M)	IC ₅₀ (μ M)	IC ₅₀ (μ M)	IC ₅₀ (μ M)
G202-0454		500.05	53			
G202-0149		465.61	30	50	>100	>100
E627-0402		462.59	80			

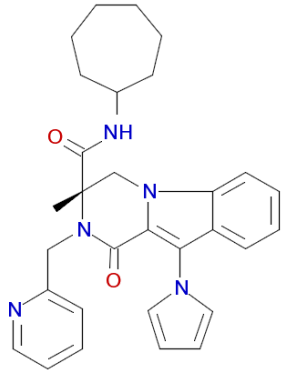
E627-0664		495.62	30	67.5	>100	>100
------------------	---	--------	----	------	------	------

Figure 30 shows binding mode of (a) **G202-0149** ($IC_{50} = 30 \mu M$) and (b) **E627-0664** ($IC_{50} = 30 \mu M$). Methoxyphenyl group of G202-0149 and pyridine ring of E627-0664 are predicted to occupy the S1 pocket; tricyclic groups go into the S2 pocket.

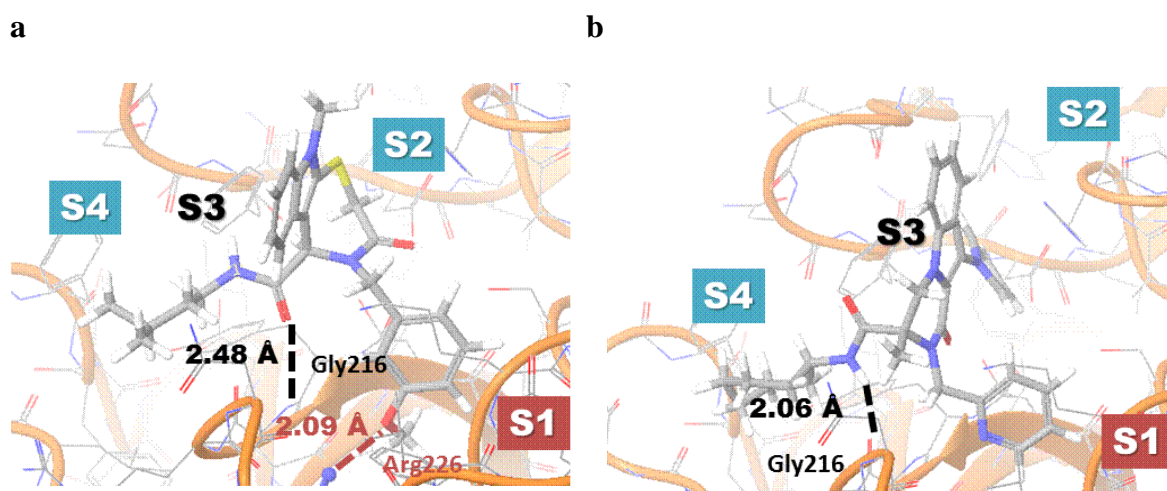
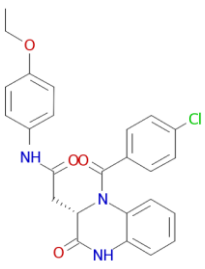


Figure 30. The predicted binding modes of (a) G202-0149 and (b) E627-0664. The methoxyphenyl of G202-0149 forms hydrogen bonds with Arg226 (2.09 Å) and Gly216 (2.48 Å). The terminal NH of amide forms a hydrogen bond with Gly216 (2.06 Å).

6.2.3 Other

Table 11. 5782-5949 showing weak inhibition against GrB

Compound	Structure	MW	GrB	GrA	Casp-3	Casp-8
			IC ₅₀ (μ M)	IC ₅₀ (μ M)	IC ₅₀ (μ M)	IC ₅₀ (μ M)
5782-5949		463.92	94			

6. 3 Summary of the characteristics of the various classes of compounds

Of 18 compounds showing IC₅₀ values below 50 μ M, triazole ST057416 (44 μ M), thiazolidinediones ST057924 (25 μ M) and ST057833 (28 μ M), diazolidinedione ST058025 (44 μ M), and sulfonamide L581-0227 (45 μ M) show good selectivity against GrA as well as Caspase-3 and 8.

Thiazolidinedione and diazolidinedione series possess Michael acceptors which could form covalent bond with Ser195, however, the possibility seems to be low based on docking studies since the conformations of the protein side chains reorganize significantly upon ligand binding.

All active compounds bind in the same way in the active site as the tetrapeptide aldehyde inhibitor, Ac-IEPD-CHO, occupying all subsites.

The smallest compound in the thiotetrazole series, ST053449 (MW 338.85), has been modified to improve activity and selectivity by chemists (Lauriane Buisson, Chris Braddock, and Anthony G M Barrett at Imperial College London).

In addition, the tricyclic series from ChemDiv has been used to supply query molecules for scaffold hopping because the location of the tricyclic groups accord with those of the Merck compounds based on docking poses.

Chapter 7. Selectivity

Several crystal structures of tetrapeptide inhibitors bound to GrB and Caspase-3 and 8 were analyzed and compared to improve selectivity between them.

GrB (serine protease) and the caspase (cysteine protease) family members are the only proteases that are involved in cell death. Even though they show distinct catalytic mechanisms from other classes of enzymes, and are unrelated in sequence, they share both substrate specificity for P1-Asp and a biological role in promoting apoptosis. Therefore, it is important for inhibitors of GrB to be selective against the caspases, especially caspase-3 (apoptosis effectors) and 8 (apoptosis signaling).

7.1 Caspase-3 vs Caspase-8

In terms of both substrate specificity and inhibitor selectivity, the **P4** position offers the most variability between all members of the caspase family²⁶. Especially, Casp-3 strongly prefers *Asp* for P4. Casp-8 can also bind a P4-*Asp* inhibitor, however, it shows a 4-fold less preference than Casp-3. In contrast, Casp-8 prefers hydrophobic residues such as *Leu* or *Val* for P4. This might explain why our compounds with hydrophobic P4

show better selectivity against Casp-3 than Casp-8: of 14 compounds in the selectivity test, 11 are inactive to Casp-3, whereas 8 are inactive to Casp-8 (**Table 4 – 10**). **Table 12** lists amino acid preferences for each subsite in addition to HB and hydrophobic interactions for Casp-3 and 8.

7.2 GrB vs Caspase-3 and 8

In addition, according to various studies,^{7, 25, 27} common features of the three enzymes (GrB and Casp-3 and 8) are: First they strongly require P1- Asp, and second they also prefer P3-Glu. Different features are: First, the S1 of GrB is larger and less charged than caspases because GrB has only one Arg whereas caspases have two such residues near S1. Second, S4 of GrB is more hydrophobic so it prefers P4-Ile. On the other hand, casp-3 strongly prefers P4-Asp and casp-8 prefers more hydrophobic residues for P4 than casp-3 such as Ile, Val, or Leu like GrB. **Table 13** summarizes more details.

In this study, from a total of 28 active compounds (21 compounds from TimTec and 7 compounds from ChemDiv), any compound showing inhibition under 50 μ M was selected for selectivity evaluation against GrA, Caspase-3 and 8 (**Table 4-11**) .

Table 12. Caspase-3 and Caspase-8 preferences, HBs and hydrophobic interactions for each subsite.

	Caspase-3	Caspase-8
P1-Asp	Covalently bound to the nucleophilic Cys / HBs between thiohemiacetal O and His and Gly / Salt bridge between P1-Asp and two Arg as well as conserved Gln	
P2-Val	Hydrophobic interactions	
P3-Glu	Salt bridge between Glu and Arg	Salt bridge between Glu and two Arg
P4-Asp	HBs with Phe and Asn / Water mediated HBs with Trp and Phe	HBs with Asn and Trp / Water mediated HBs with Asp
N-Acetyl	Two HBs with Ser / Less hydrophobic interactions	HBs with Asn / Hydrophobic interactions with Pro

Table 13. Common and different features for GrB and Caspase3-8

	Granzyme B	Caspase-3	Caspase-8
Common	Absolute requirement of P1-Asp / Preferred P3-Glu		
Different	S1 is larger and less charged because only one Arg	Two Arg	
	Less constricted and more hydrophobic S4 (Preferred P4-Ile, not enough for Phe)	Strong preference P4-Asp	Preferred P4-Ile / Leu / Val
Preferred	IEPD (Ile-Glu-Pro-Asp)	DEXD (Asp-Glu-Any-Asp)	LEXD (Leu/Val-Glu-Any-Asp)
Ki	Ac-IEPD-CHO = 80 nM	Ac-DEVD-CHO = 0.23 nM Ac-YVAD-CHO = 10 μ M	Ac-DEVD-CHO = 0.92 nM

Chapter 8. Scaffold hopping

Above, we describe various scaffolds grouped into several categories by screening. In addition to these efforts, a shape-based, *ligand*-centric approach has been carried out to identify novel scaffolds. We sought structurally dissimilar molecules with a high probability of biological activity as well as synthetic opportunities to further optimize affinity and overcome issues such as poor ADME/Tox and established patent coverage. According to various studies, this approach is often superior to the *protein*-centric approach taken by docking²⁸.

To do this, the ROCS (Rapid Overlay of Chemical Structures)²⁹ method from OpenEye Scientific Software was used, an approach designed to perform large scale 3D database searches by using shape-based superposition. Molecules are aligned by a solid-body optimization process that maximizes the overlap volume between them. In addition to a shape component, ROCS uses a “color force field” score based on defined atom types (cations, anions, hydrogen bond donors, hydrogen bond acceptors, hydrophobes, and rings) and “TanimotoCombo (shape + color) score. Therefore, one can identify compounds which are similar in shape, charge properties and chemistry.

8.1 Merck compounds modeling

To select query molecules from hit compounds discovered by previous virtual screening, a noncovalent class of Merck compounds were extensively studied³⁰. Since their crystal structures complexed with granzyme B are not known, binding poses were predicted by docking studies.

8.1.1 Binding pose of Merck 15 ($K_{iGrB} = 38 \text{ nM}$ / $IC_{50_apoptosis} = 30 \text{ }\mu\text{M}$)

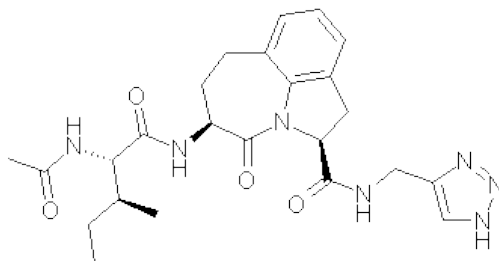


Figure 31. The structure of **Merck 15**

A rigid tricyclic group in the P2-P3 position and 1,2,3 triazole substitution at P1 (modification of the serine trap, **Figure 31**) give about 2-fold improvement ($K_{iGrB} = 38 \text{ nM}$) over the IEPD-CHO ($K_{iGrB} = 80 \text{ nM}$). According to docking evaluation in the present work, **Merck 15** shows good alignment with IEPD-CHO (**Figure 32. (a)**) and retains key HB interactions with Arg226 (2.13 Å), Ser214 (1.79 Å), Gly216 (1.88 / 2.36 Å), and Asn218 (1.79 Å) (**Figure 32. (b)**).

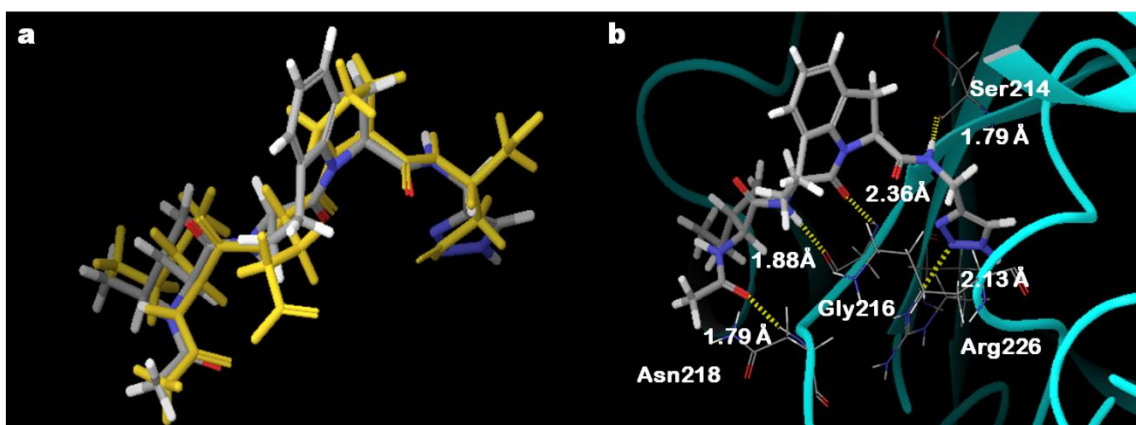


Figure 32. **Merck 15** at the active site of GrB. (a) Alignment of **15** with IEPD-CHO (yellow) in a binding pose conformation, and (b) the predicted binding pose of **Merck 15**.

In spite of loss of a HB by P3-Glu, the increased hydrophobicity of **Merck 15**, presumably as a result of its tricyclic group, seems to improve the activity.

8.1.2 Binding pose of Merck 19 ($K_{iGrB} = 13 \text{ nM}$ / $IC_{50_apoptosis} = 6.3 \text{ }\mu\text{M}$)

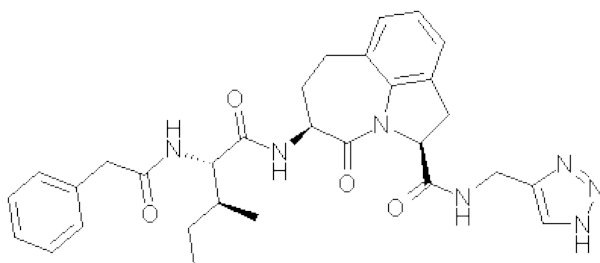


Figure 33. The structure of **Merck 19**

By adding a hydrophobic phenyl group at the P5 position (**Figure 33**), enzyme and cellular activity were improved 3- and 5-fold, respectively, over **Merck 15** presumably due to a cation- π interaction between the phenyl ring and Arg217 (**Figure 37. (b)**). A cation- π interaction is a strong, noncovalent binding interaction, in which side chains of the aromatic (Phe, Tyr, Trp) provides a surface of negative electrostatic potential that can bind to a wide range of cations through electrostatic interaction. Since, cation- π interactions contribute significantly to stabilizing protein secondary structure as well as drug-receptor interactions, these interactions along with the hydrogen bonding, ion pairing (salt bridges), and hydrophobic effect are considered as the 4th key force that contribute to macromolecular structure and molecular recognition in biology.³¹

According to the docking studies employing final scoring with MM-GBSA, **Merck 19** shows good alignment with IEPD-CHO (**Figure 34. (a)**) and keeps key HB

interactions with Arg226 (2.13 Å), Ser214 (1.79 Å), Gly216 (1.86 / 2.34 Å), and Asn218 (1.84 Å) (**Figure 34. (b)**).

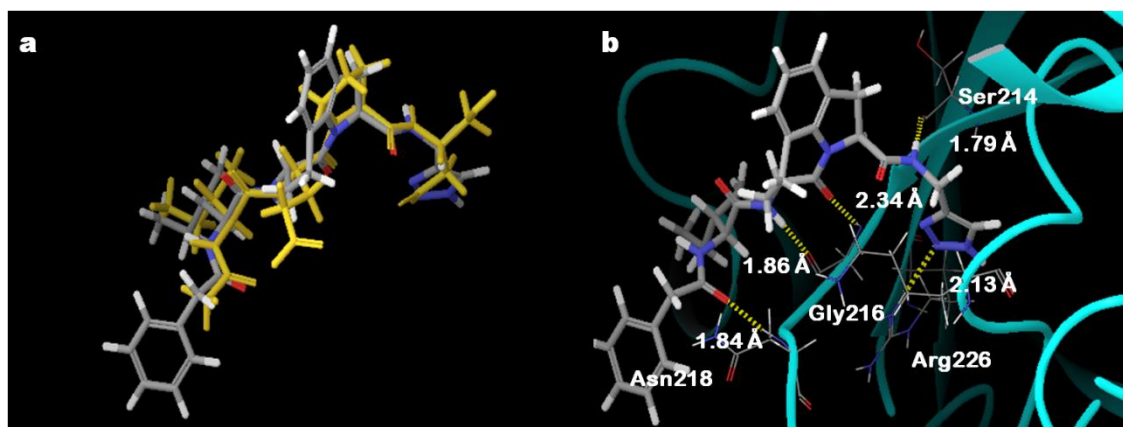


Figure 34. (a) Alignment with IEPD-CHO (yellow) and (b) the predicted binding pose of **Merck 19**.

8.1.3 Binding pose of **Merck 20** ($K_{iGrB} = 7 \text{ nM}$ / $IC_{50_apoptosis} = 3.1 \text{ }\mu\text{M}$)

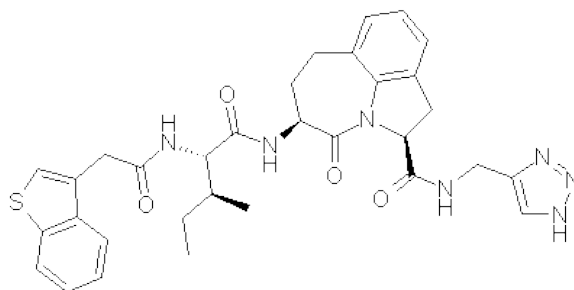


Figure 35. The structure of **Merck 20**

By adding a hydrophobic benzothiophene group in the P5 position, enzyme and cellular activity were further improved 5- and 10-fold, respectively, over **Merck 15** due

to cation- π interaction between benzothiophene ring and Arg217 (**Figure 37. (b)**).

According to the docking studies, **Merck 20** shows good alignment with IEPD-CHO (**Figure 36. (a)**) and keeps key HB interactions with Arg226 (2.14 Å), Ser214 (1.79 Å), and Gly216 (1.89 / 2.36 Å (**Figure 36. (b)**).

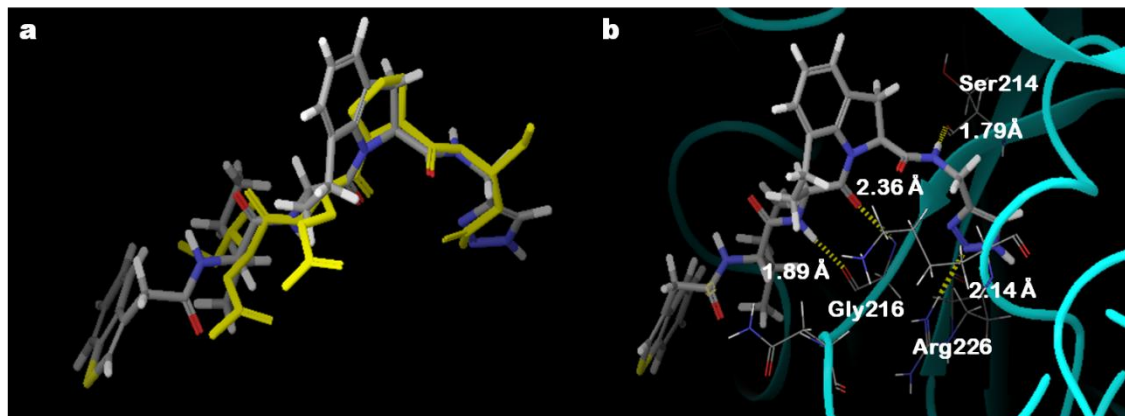


Figure 36. (a) Alignment with IEPD-CHO (yellow) and (b) the predicted binding pose of **Merck 20**.

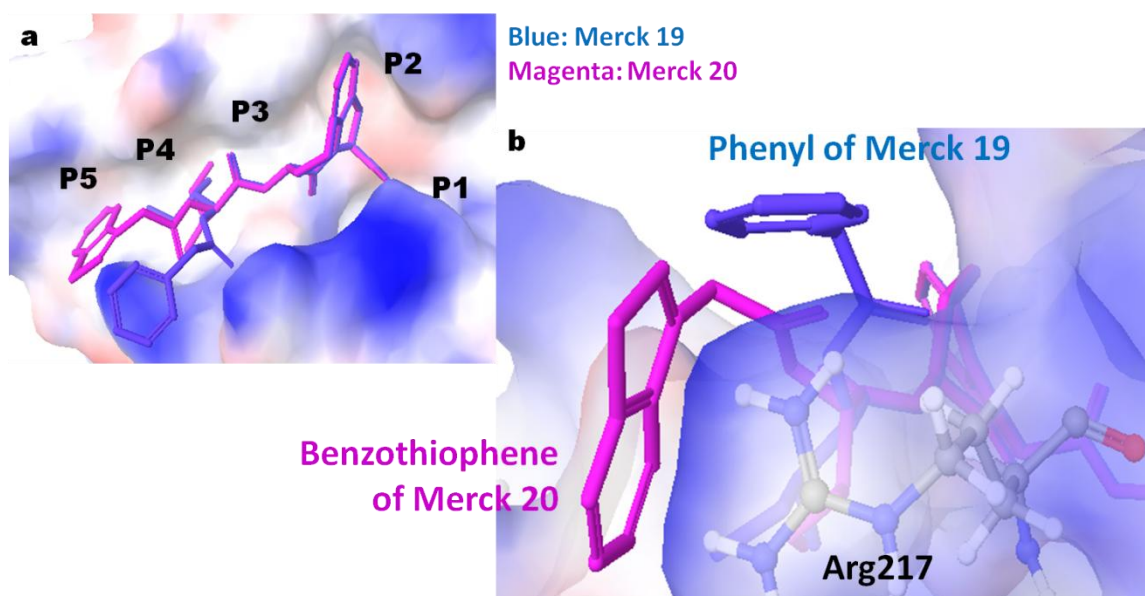


Figure 37. (a) Alignment with **Merck 19** (blue) and **Merck 20** (magenta) and (b) the cation- π interaction between Arg217 and P5-phenyl and benzothiophene rings of **Merck 19** and **Merck 20**, respectively.

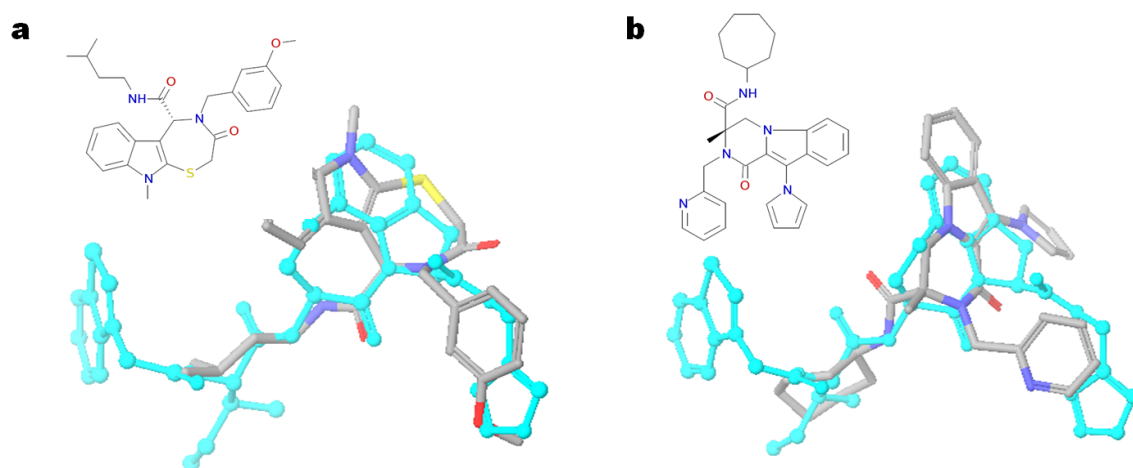


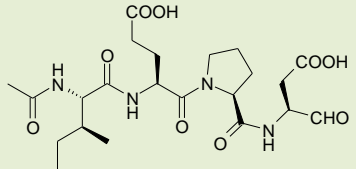
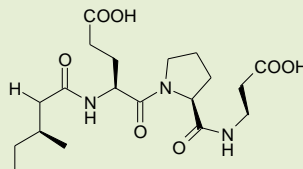
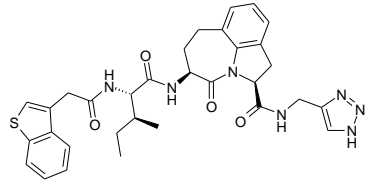
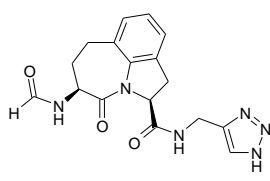
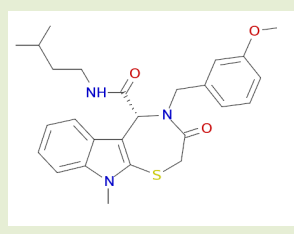
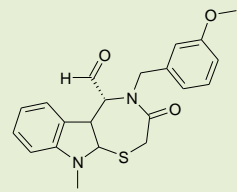
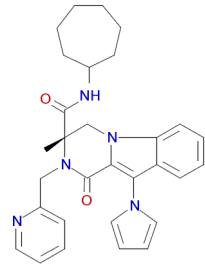
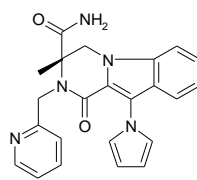
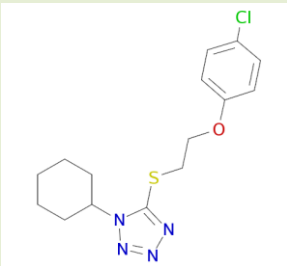
Figure 38. Alignment of (a) **G202-0149** and (b) **E627-0664** with **Merck 20** (light blue).

Of the hit compounds from virtual screening, there are several tricyclic compounds from ChemDiv libraries (**Table 10**). The predicted binding poses of **G202-0149** ($IC_{50_{GrB}} = 30 \mu M$) and **E627-0664** ($IC_{50_{GrB}} = 30 \mu M$) showing IC_{50} under $50 \mu M$ were superimposed on **Merck 20** to compare the position of tricyclic rings (**Figure 38**). Since the position of tricyclic rings of **G202-0149** and **E627-0664** were aligned well with one of **Merck 20**, the two compounds were added to the query molecules.

8.2 Search Queries

As inhibitors of hGrB, a tetrapeptide inhibitor, IEPD-CHO, and several Merck compounds have been published. And, from our current virtual screening based on computational solvent mapping, various scaffolds which show moderate activity were identified. Of them, the tricyclic compounds **G202-0149** and **E627-0664**, which align well with the Merck compounds and the tetrazole compound **ST053449** (current starting compound) were selected as search molecules for the ROCS procedure along with **IEPD-CHO** and **Merck-20**. Two types of templates (“**Full**” which used a whole structure, and “**Part**” which is a substructure) were used as ROCS queries (**Table 14**).

Table 14. Structures used as ROCS queries. Two types of templates were used: “Full” query where all heavy atoms were used, and “Part” query where substructures were used. Since **ST053449** is small, the entire molecular structure was employed.

	“Full” query	“Part” query
IEPD-CHO ($K_i = 80 \text{ nM}$)	 The full structure of IEPD-CHO is a complex molecule consisting of a central pyrrolidine ring. It is substituted with a propanoic acid chain, a propanoic acid chain, and a propanoic acid chain with an aldehyde group. The stereochemistry is indicated with wedges and dashes.	 The part structure of IEPD-CHO is a simplified version of the full structure, showing only the heavy atoms of the pyrrolidine ring and the three propanoic acid chains, with the aldehyde group.
Merck 20 ($K_i = 7 \text{ nM}$)	 The full structure of Merck 20 is a complex molecule featuring a central pyrrolidine ring. It is substituted with a benzothiazole ring, a propanoic acid chain, and a propanoic acid chain with a benzothiazole ring. The stereochemistry is indicated with wedges and dashes.	 The part structure of Merck 20 is a simplified version of the full structure, showing only the heavy atoms of the pyrrolidine ring and the two propanoic acid chains.
G202-0149 ($IC_{50} = 30 \text{ }\mu\text{M}$)	 The full structure of G202-0149 is a complex molecule featuring a central pyrrolidine ring. It is substituted with a propanoic acid chain, a propanoic acid chain, and a propanoic acid chain with a methoxy group. The stereochemistry is indicated with wedges and dashes.	 The part structure of G202-0149 is a simplified version of the full structure, showing only the heavy atoms of the pyrrolidine ring and the three propanoic acid chains.
E627-0664 ($IC_{50} = 30 \text{ }\mu\text{M}$)	 The full structure of E627-0664 is a complex molecule featuring a central pyrrolidine ring. It is substituted with a propanoic acid chain, a propanoic acid chain, and a propanoic acid chain with a methoxy group. The stereochemistry is indicated with wedges and dashes.	 The part structure of E627-0664 is a simplified version of the full structure, showing only the heavy atoms of the pyrrolidine ring and the three propanoic acid chains.
ST053449 ($IC_{50} = 35 \text{ }\mu\text{M}$)	 The full structure of ST053449 is a complex molecule featuring a central pyrrolidine ring. It is substituted with a propanoic acid chain, a propanoic acid chain, and a propanoic acid chain with a methoxy group. The stereochemistry is indicated with wedges and dashes.	

8.3 Library: protease-targeted libraries

The **Table 15** shows the screening libraries.

Table 15. The protease-targeted libraries for scaffold hopping.

Company	Library	Total # of Molecules
TimTec	ActiTarg-P	1,520
ChemDiv	Aspartic Recognition Motif	430
	Serine / Cysteine Proteases Inhibitors Lib	8,300
Enamine	Protease Targeted Lib	3,927
	Caspase-3 Targeted Lib	226

8.4 Computational procedures

8.4.1 Create a query

Each molecule generated a corresponding query (**Figure 39**). **Figure 39** is a query from **G202-0149** generated by ROCS where the molecular structure of the hit is displayed as green sticks with associated shape and color atoms.

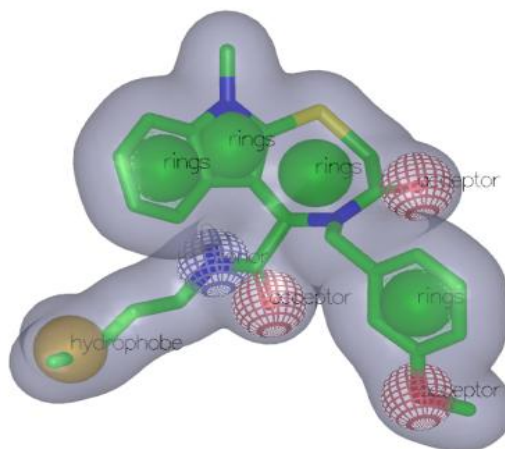


Figure 39. All three portions (molecule, shape and color) are generated by ROCS. The aligned hit molecule is shown as green sticks colored by atom type. In addition to shape, color atoms are displayed: rings (green), hydrophobe (yellow), and acceptor / donor (red / blue).

8.4.2 Perform a ROCS run

Pre-computed conformers for *each* compound in *each* library were aligned to *each* query to find and quantify the maximal overlap of the volume and color atoms. The most common score, TanimotoCombo (shape + color) is generally used when the query and library molecules are of *similar* size.

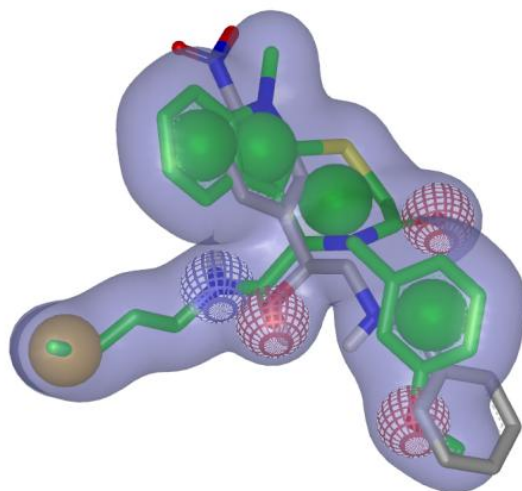


Figure 40. A hit molecule (gray stick by atom type) is aligned to the query molecule (green stick by atom type).

In the **Figure 40**, an aligned hit molecule is shown as a gray stick by atom type. As a result, the top 500 hits from each search were saved: $500 \text{ hits} \times 9 \text{ queries} \times 5 \text{ libraries} = \text{total } 22,500 \text{ compounds (including duplicates)}$.

8.4.3 Docking using Glide

The high score (high similarity) from ROCS searching doesn't *definitely* guarantee a *real* activity. Following removal of the covalent aldehyde inhibitor, all 22,500 compounds were docked into the active site of GrB using Glide to introduce structural information of the enzyme with same three constraints based on computational solvent mapping to explore more promising hits (**Figure 41**).

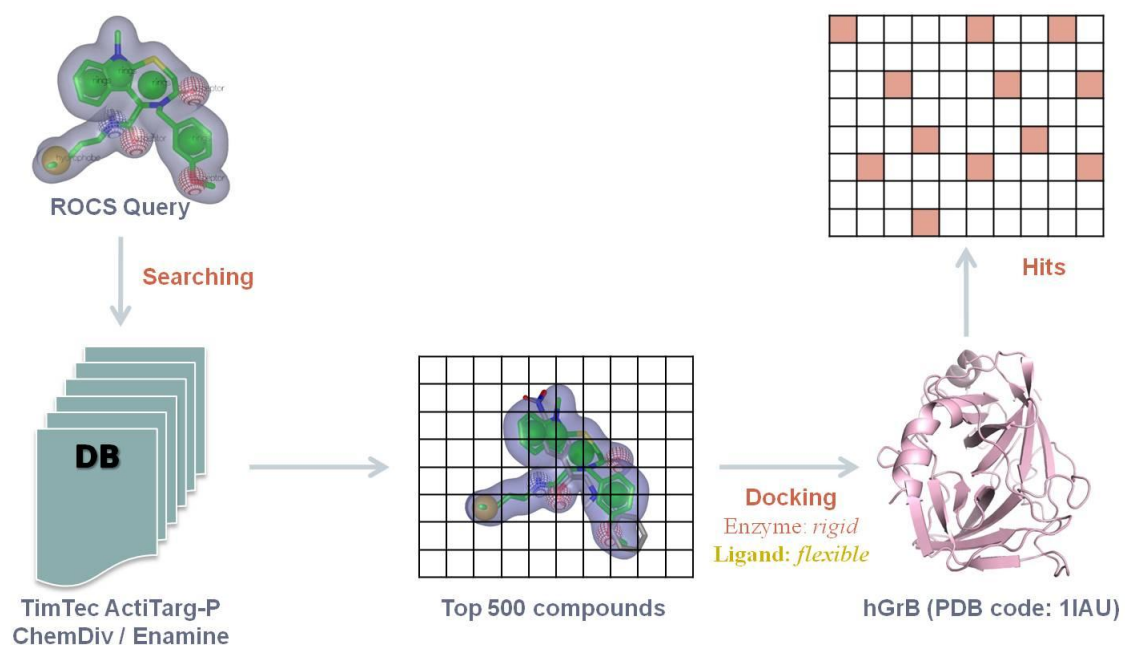


Figure 41. Screening procedure: (1) ROCS search of the prepared libraries with each query, and (2) top 500 hits from each library were docked into the active site of GrB using Glide where enzyme was kept *rigid* and ligand was allowed to be *flexible*.

8.4.4 Induced fit docking using Glide and Prime

As mentioned above, ligand-based approaches often outperform docking algorithms at averaged multiple targets. However, in this study, of the top 500 hits derived by ROCS similarity searching from each query and each library, about 10 % survived through the subsequent receptor based *rigid* docking study. Interestingly, from the top 20 compounds of the 500 hits, only a few pass both ROCS and docking evaluations. Nonetheless, many structures from the 500 not in the top 20 survive both filters. In order to reevaluate rejected structures from the top 20, induced fit docking was introduced as described below.

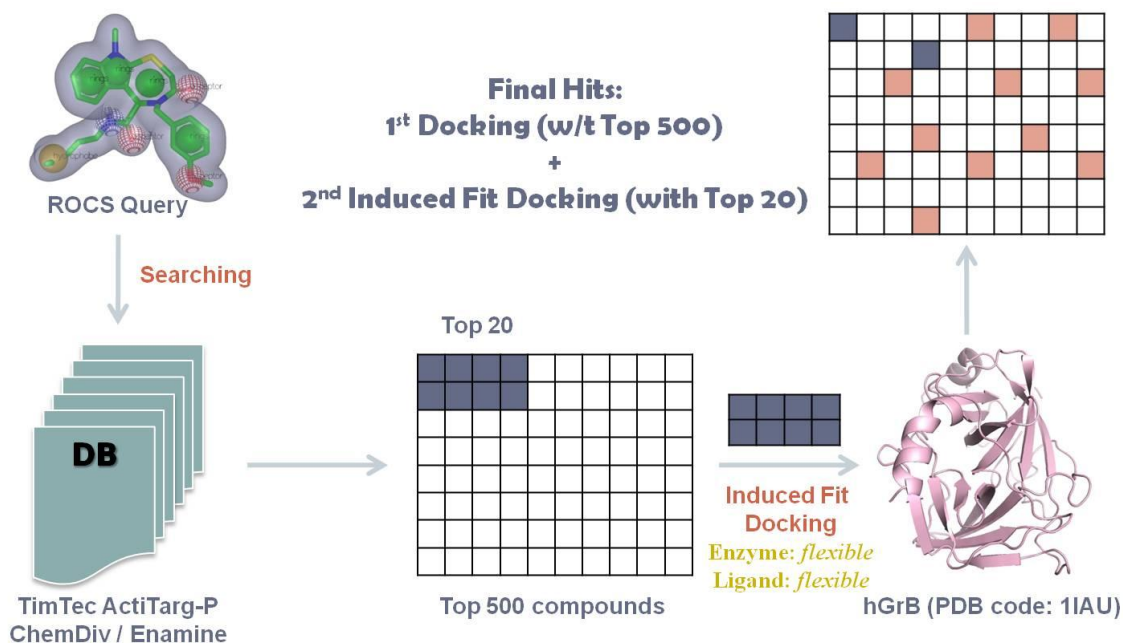


Figure 42. 2nd round screening with top 20 compounds per each calculation was carried out using Induced Fit Docking (IFD), and final hits (1st docking with ROCS top 500 hits + 2nd Induced Fit docking with ROCS top 20 hits per each calculation with each query and library) were selected as assay candidates.

So, to rescue false negatives, Induced Fit Docking (IFD) was performed, in which the enzyme is allowed to alter its binding site so that it more closely conforms to the shape and binding mode of the ligand. The assumption of a rigid receptor can give misleading results, since in reality many proteins undergo side-chain or back-bond movements, or both, upon ligand binding. Therefore, IFD is often applied to generate an accurate binding mode.

The above table summarizes the calculations carried out using ROCS and docking methods. Five protease targeted libraries were searched with two types of queries (“q_Full” and “q_Part”) and five molecular templates (Merck 20, IEPD-CHO, G202-0149, E627-0664, and ST053449). Three rounds of calculations were performed with each query against each library: (1) ROCS search, (2) Glide (G) SP docking with top 500 molecules from ROCS search, and (3) Induced Fit Docking (I) with top 20 from ROCS search to rescue false negatives due to limitations of rigid docking.

8.5.1 Example: Enamine Casp-3 Target Library Results

Table 17. Some results from the Enamine Casp-3 Target Library search

Library		Query										VS					
Enamine Casp-3		Merck 20			IEPD-CHO			G202-0149			E627-0664		ST053449				
ID		q_Full			q_Part			q_Full			q_Part		q_Full				
		G	R20	I	G	R20	I	G	R20	I	G		R20	I	G	R20	I
1	T0511-8093																
2	T0511-6405																
3	T0511-6384																
4	T0519-9186																
5	T5268703																
6	T5314444																
7	T0518-0051																
8	T0504-8930																
9	T0504-8931																
10	T0501-9588																

The above table shows some results from the Enamine Casp-3 Target Library search. On the left side, IDs of hit molecules which survived docking are shown, while the right columns show from which calculations they were selected as hits. For example, initially **T0511-8093** proved to be a hit from all 9 queries and bound to the GrB well. Moreover, it was one of the ROCS top 20 hits when “q_Part” of IEPD-CHO and E627-0664 and “q_Full” of G202-0149 and ST053449 were used as queries.

However, in another related case, even though **T0504-8930** was one of ROCS top 20 hits several times, it was discarded (**Table 9**, “d”) by both rigid and flexible docking and, thus, was not purchased for examination in the GrB inhibition assay.

In addition, this library was added as a protease targeted library after our previous VS with TimTec and ChemDiv libraries. So, the last column for VS which was carried out by the same procedures used for previous VS was added. As a result, both the 2nd and 5th molecules were also selected by VS for further assay.

Finally, the yellow-marked molecules on IDs are finally selected as assay candidates. Of the three libraries examined (TimTec, ChemDiv, and Enamine), only the 28 Enamine compounds were tested at 100 uM so far. Further tests identified one compound ($IC_{50} = 72.5 \text{ uM}$) showing significant inhibition at ~81%.

8.6 Molecular property calculations using QikProp

Of a total of 14,403 compounds from 5 libraries from 3 companies screened using ROCS and Glide, 574 compounds were selected as assay candidates, with which molecular properties were further calculated by QikProp of Schrödinger, which is an ADME prediction program designed by William L. Jorgensen³².

QikProp not only predicts physically significant descriptors and pharmaceutically relevant properties of organic molecules, but also provides the range for comparing particular molecules' properties with those of 95 % of known drugs.

Of 44 possible descriptors, six important descriptors (solubility, blockage of HERG K⁺ channels, Caco-2 cell permeability, brain/blood partition coefficient, human serum albumin, Lipinski's rule of five) were utilized in the selection of final hits to be ordered. Compounds that fall outside of the 95% range for known drugs or recommended values were discarded.

The remaining compounds were inspected by chemists (Professor Barrett's group at Imperial College London). Finally, a total 57 compounds (28 from Enamine, 27 from ChemDiv, and 3 from TimTec) were selected to be ordered for bioassay (**Figure 43**).

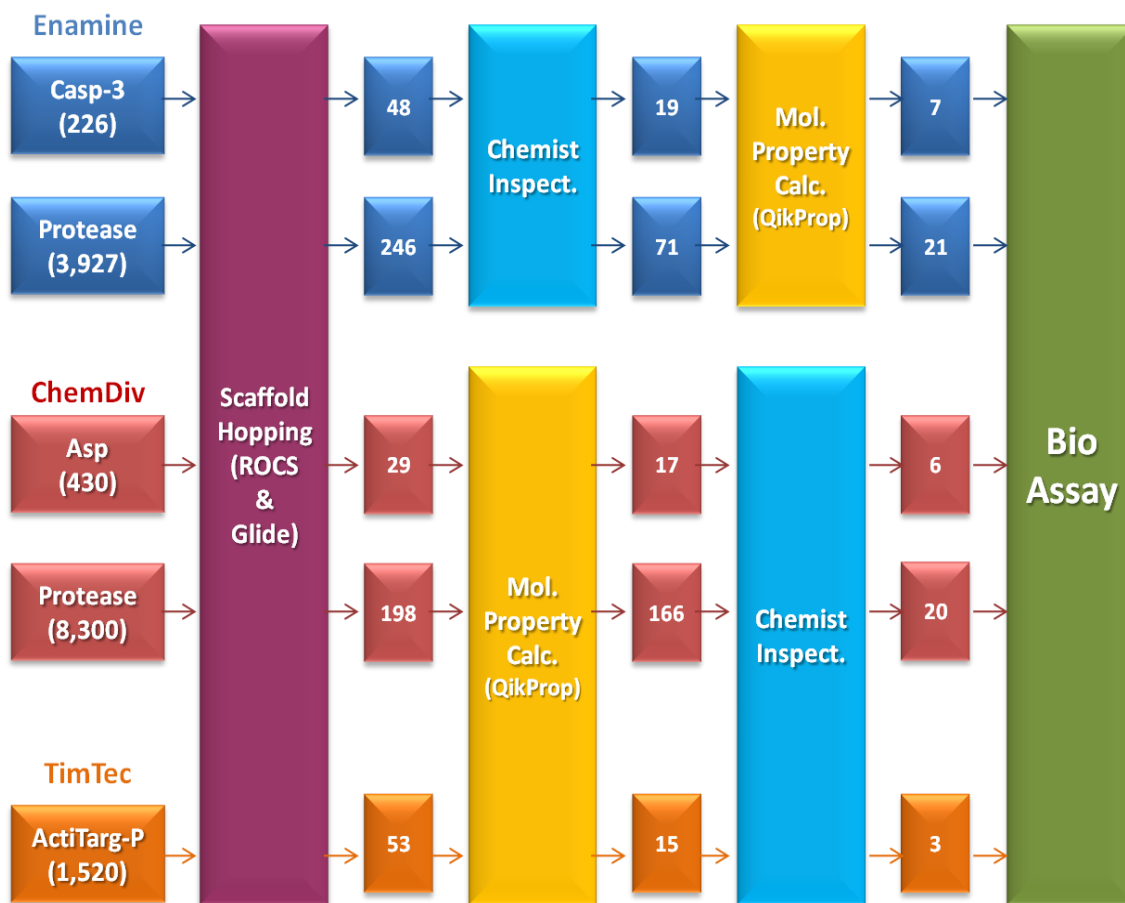


Figure 43. Molecular property calculations using QikProp. 574 compounds from 14,403 compounds from 5 libraries of 3 companies screened using ROCS and Glide were further calculated, and passed compounds within range for 95 % of known drugs or recommended values inspected by chemists. Finally, a total 57 compounds (28 from Enamine, 27 from ChemDiv, and 3 from TimTec) were selected to be ordered for bioassay.

Chapter.9 Synthesis, Purchase, and Biological test

Optimized hit compounds from the computer models has led to suggestions for synthesis that are being pursued with modification by medicinal chemists at Imperial College London. In addition, selected hits from scaffold hopping will be purchased and biological tests will be done by Dr. Dean at Imperial College London.

Chapter 10. Mid-Micromolar Activities; What Could be Done?

Even though intensive computational searches have been carried out to identify promising non-covalent candidates for GrB, the activities of the hit molecules never dropped below 25 μ M.

This issue might have resulted from the use of the X-ray structure of the covalently-bound inhibitor, Ac-IEPD-CHO, to identify non-covalent small molecule inhibitors. Doing so assumes that the enzyme has a unique binding site, thus, the structure of a single inhibitor complex might be sufficient to define the available subsites of an enzyme. However, analysis with the complex structures of elastase and covalent / non-covalent inhibitors demonstrates they can bind to the enzyme in different poses, even though they have similar structures and binding affinities. Thus, Carla Mattos et al.²² revealed the existence of multiple binding sites and different binding modes for non-covalent inhibitors with similar chemical structures: trifluoroacetyl-dipeptide-anilides (TFA-dipeptide-anilides).

Trifluoroacetyl (TFA)-Lys-Pro-isopropylanilide (ISO) binds to elastase in the same fashion as covalent inhibitors which must be oriented at the active site in the same manner as the substrate forming acyl enzyme intermediate for reaction to take place: TFA

in the S1; Lys in the S2; Pro in the S3: anilide group in the S4 (**Figure 44. (a)**). The compound also shows a similar binding mode to the covalent peptide inhibitor, MSACK (methoxysuccinyl-Ala-Ala-Pro-Ala chloromethyl ketone).

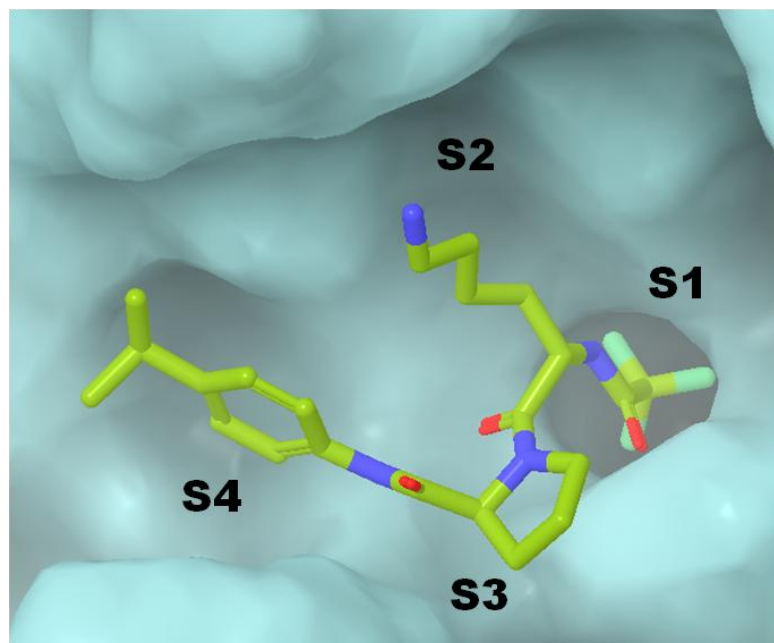


Figure 44. The crystal structures of elastase complexed with (a) trifluoroacetyl (TFA)-Lys-Pro-isopropylanilide (ISO).

However, even exchanging Pro with Phe or Leu, they (TFA-Lys-Phe-ISO and TFA-Lys-Leu-ISO) shows different binding modes. The TFA is in the oxyanion hole, the Lys is in the S2, the exchanged Phe and Leu are in the S1 instead, and anilide groups are located in the new pockets (**Figure 45**).

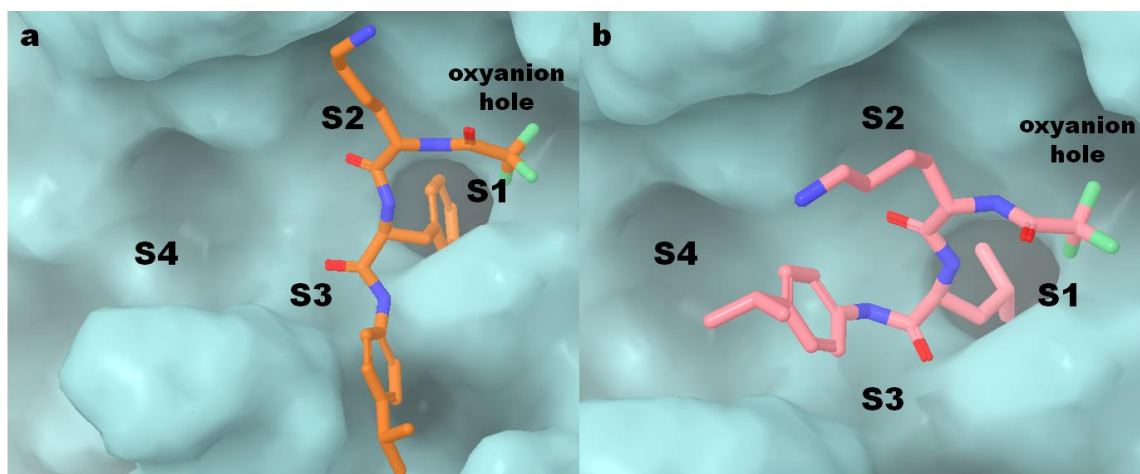


Figure 45. The crystal structures of elastase complexed with (a) TFA-Lys-Phe-ISO and (b) TFA-Lys-Leu-ISO.

The structures of elastase in the complexes are unchanged except for the conformation of Arg226. (**Figure 46**).

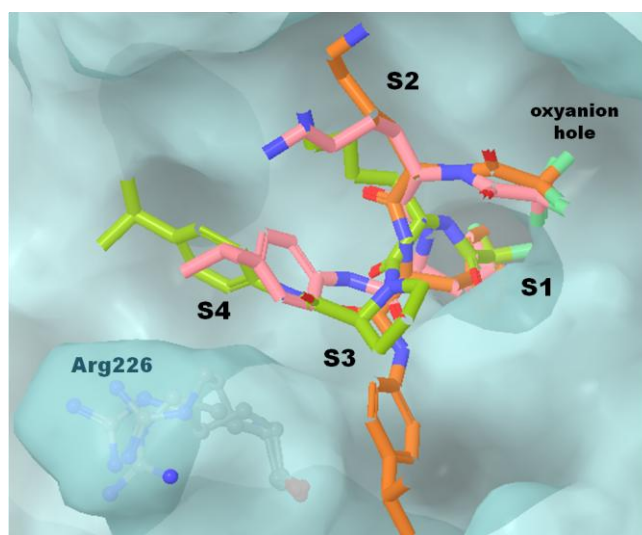


Figure 46. All inhibitors are aligned showing different binding modes in spite of chemically similar structures. They are represented with different carbon colors: TFA-Lys-Pro-ISO (green C); TFA-Lys-Phe-ISO (orange C); TFA-Lys-Leu-ISO (pink C).

Therefore, the superposition of covalent and non-covalent inhibitors illustrates that sometimes the structures complexed with covalent inhibitors might not be a good template for non-covalent design. Some hit molecules from ROCS as a ligand-based approach were also discarded after docking by introducing protein information.

In addition, the consistent use of the 3 constraints generated by computational solvent mapping might lead false-negatives, even though hit rate of virtual screening using those 3 constraints is relatively higher than that of HTS.

So, the new strategy with new protein information generated for example, by molecular dynamics without any constraints could give more chance to have novel compounds.

Chapter 11. Conclusions

Novel classes of small molecule inhibitors against human GrB have been developed by docking studies with three constraints (HBs with Arg226, hydrophobic interactions for S2 and S4 subsites) based on computational solvent mapping using FTMAP.

The most distinctive compounds identified are; ST057924 ($IC_{50} = 25 \mu M$) and ST057833 ($IC_{50} = 28 \mu M$) with a dioxo-thiazolidine group; ST058025 ($IC_{50} = 44 \mu M$) with a dioxo-diazolidinyl group; ST057416 ($IC_{50} = 44 \mu M$) with a triazole group. All show great selectivity against GrA as well as Caspase-3 and 8. In order to form HBs with Arg226 at S1, acetamide (ST057924), nitro (ST057833), methoxy (ST058025) and benzodioxin (ST057416) substituents are located in the P1-Asp recognition position.

They could also be modified by adding more hydrophobic groups at the P2 and P4 positions to improve activity.

In addition to optimization of current identified hits, similar compounds in each series of hits and new scaffolds with dissimilar structures, but having probability of biological activity against GrB, have been subsequently explored using ROCS, shape-based superposition methods and selected hits have been purchased and tested for bioassay.

Through this study, the hypothesis that more focused screening for GrB inhibitors by identifying hot spots and using protease targeted libraries could produce novel inhibitors and increase the hit rate was shown to be true. However, true drug leads can be achieved only by synthetic modification of the hits. Efforts along these lines have been initiated by Lauriane Buisson in Professor Tony Barrett's group at Imperial College.

Part II. Curcumin and mimics as proteasome inhibitors

Chapter 1. Introduction

1.1 Ubiquitin-proteasome pathway (UPP) and 26S proteasome

The ubiquitin-proteasome pathway (UPP) functions as a primary mechanism in intracellular protein degradation, in which the target proteins are tagged with a small regulatory protein, ubiquitin.³³ Ubiquitin directs unneeded proteins to the proteasome for degradation and recycling.

In the process of ubiquitination (**Figure 1**), ubiquitin is first activated by E1 (a ubiquitin-activating enzyme) in combination with ATP as an energy source to form a ubiquitin-adenylate intermediate.³⁴ The activated ubiquitin is transferred to the cysteine residue of the E1 active site, thus forming a thioester linkage between the C-terminal carboxyl group of ubiquitin and the E1 cysteine sulfhydryl group. As a consequence, AMP is released. Again, ubiquitin is transferred from E1 to the cysteine residue in the active site of E2 (a ubiquitin-conjugating enzyme). This is followed by the further transfer of ubiquitin to the target protein catalyzed by E3 (a ubiquitin ligase) to form an isopeptide bond between a lysine of the target protein and the C-terminal glycine of ubiquitin. That is, E3 not only recognizes the target protein but also interacts with E2.

In the eukaryotic cell, the 26S proteasome is the leading enzyme player in the degradation of ubiquitinated proteins designated for removal.

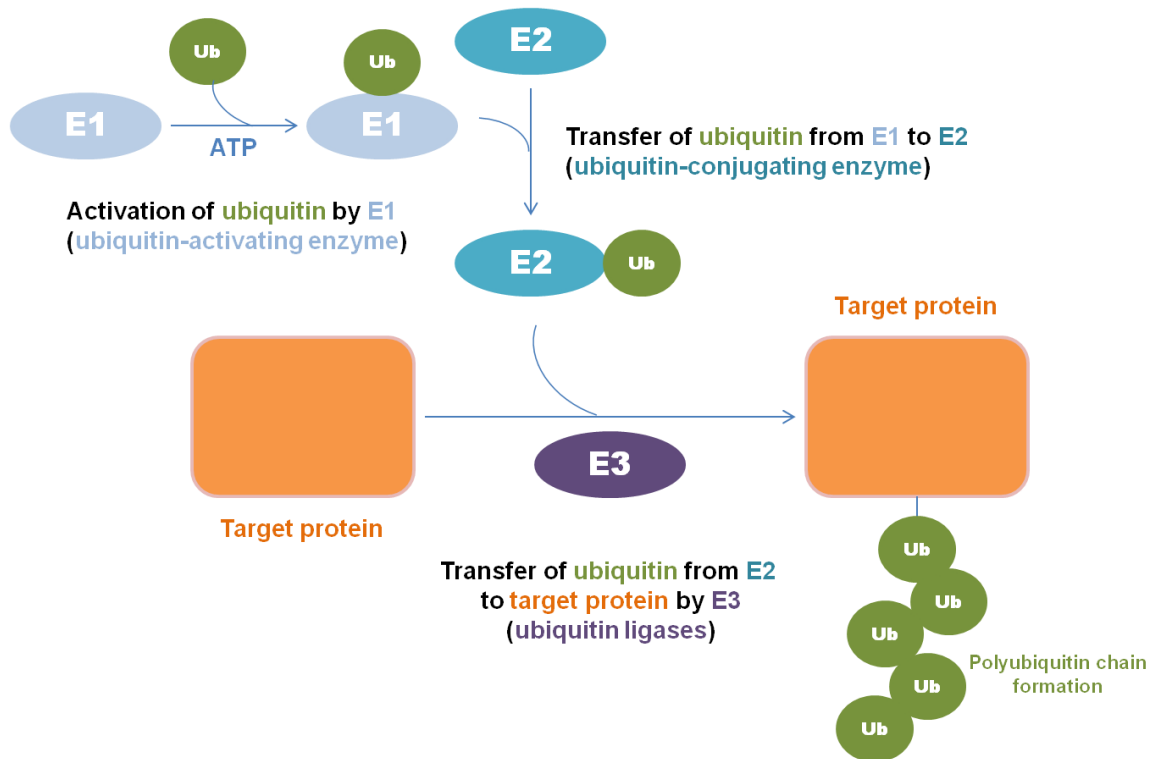


Figure 1. Ubiquitination process. Ubiquitin (small regulatory protein which bind to proteins and direct proteins to the proteasome to degrade) is first activated by E1 (ubiquitin-activating enzyme), and is transferred from E1 to E2 (ubiquitin-conjugating enzyme). Again, it is transferred from E2 to the target protein by E3 (ubiquitin ligases) which not only functions to recognize substrate but also interacts with both E2 and the target protein.

Since the ubiquitin-proteasome pathway (UPP) plays an essential role in many cellular events including signal transduction, cell cycle, cellular proliferation and apoptosis, dysregulation of this pathway causes various inherited or acquired diseases such as Alzheimer's disease, asthma, and type I diabetes, and several types of cancer.³⁵ In addition, since many factors involved in angiogenesis are regulated by the proteasome, the ability of proteasome inhibitors to inhibit angiogenesis, together with their ability to inhibit cell proliferation and induce apoptosis in human cancer cells, makes them attractive candidates as anti-cancer drugs.

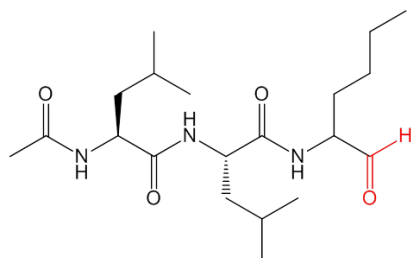
1.2 Proteasome inhibitors^{35,36}

Since the proteasome was discovered, a variety of synthetic and natural compounds have been tested for their ability to inhibit proteasomal activities. One method of classifying proteasomal inhibitors is based on their reversibility of binding to the proteasome.

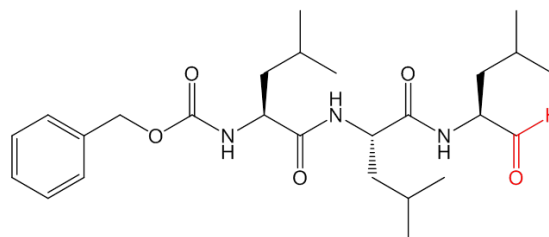
1.2.1 Covalent reversible inhibitors: peptide aldehydes and boronates (Figure 2)

Peptide aldehydes^{35,36} were the first discovered inhibitors and are still actively investigated. They are also well-known inhibitors of cysteine and serine proteases. For example, calpain inhibitor I (Ac-Leu-Leu-nLeu-al) was first identified as calpain inhibitor I and found to be 25-fold more potent against calpain and cathepsin B than the proteasome. On the other hand, MG132 (Z-Leu-Leu-Leu-al) is both more selective and potent than calpain inhibitor I against the proteasome with a K_i value of 4 nM.

Peptide Aldehydes

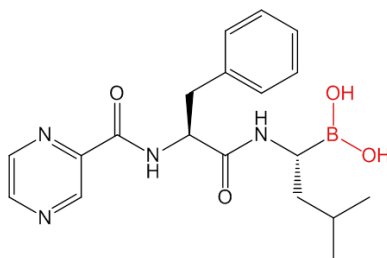


Calpain inhibitor I
(Ac-Leu-Leu-nLeu-al)



MG132
(Z-Leu-Leu-Leu-al)

Peptide Boronates



Bortezomib
(Velcade®)

Figure 2. Covalent reversible inhibitors; peptide aldehydes (calpain inhibitor I and MG132); peptide boronates (bortezomib)

Bortezomib (PS-341 or VELCADE[®], Millennium Pharmaceuticals, Cambridge, MA), a dipeptide boronic acid, was the first therapeutic proteasome inhibitor to be approved by the U.S. FDA for treatment of relapsed multiple myeloma in 2003 and mantle cell lymphoma in 2006.

The peptide boronates have become one of the most popular families of proteasome inhibitors because the boronic acid moiety increases specificity for the

proteasome due to its high affinity for hard oxygen nucleophilicities in contrast to soft cysteine sulfur nucleophilicities.³⁷ The boron compounds improved on earlier generated synthetic inhibitors, such as peptide aldehydes, which show cross-reactivity toward cysteine proteases and low metabolic stability. In addition, under physiological conditions, bortezomib preferentially inhibits the proteasome CT-like activity at low nanomolar concentrations.

1.2.2 Covalent irreversible inhibitors: lactones, vinyl sulfones, peptide epoxyketones (Figure 3)

Omuralide³⁸ is a clasto-lactacystin- β -lactone spontaneously hydrolyzed from lactacystin, a natural compound produced by *Streptomyces* at neutral pH. The crystal structure³⁹ of the yeast proteasome-omuralide complex shows that omuralide is covalently bound only to the $\beta 5$ subunit of the proteasome due to the apolar nature of the S1 specificity pocket of this subunit. It also provides strong evidence that an acyl enzyme conjugate is indeed an intermediate in catalysis by the proteasome.

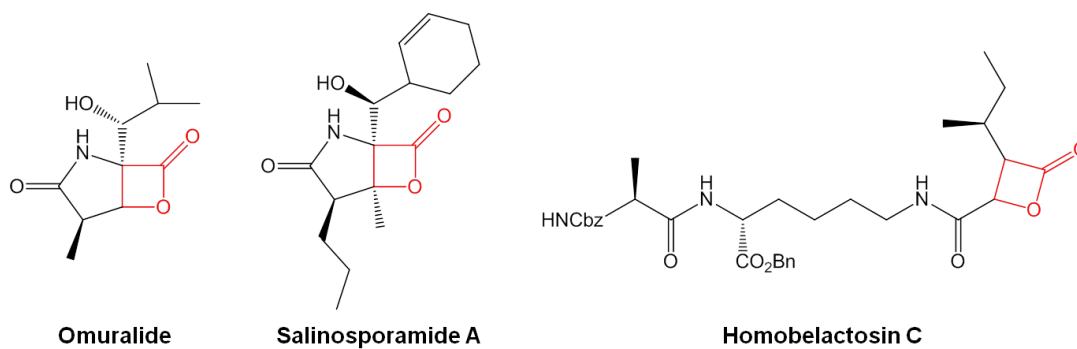
Salinosporamide A is a secondary metabolite of the marine actinonycete *Salinispora tropica*, which is a highly potent and selective inhibitor of the proteasome. It is structurally related to omuralide, however several unique substituents, including a cyclohexene ring and a chloroethyl group in place of isopropyl and methyl groups, respectively, improve its potency both *in vitro* and *in vivo*.⁴⁰ The covalent acyl ester binding to the proteasomal active site was confirmed by X-ray structure analysis. In contrast to omuralide, salinosporamide A is bound to all six catalytic subunits.

Homobelactosin C is also a natural product from *Sterptomyces*, which shows low nM IC₅₀ value against human colon cancer and pancreoma cells resulting from inhibition of the proteasomal activity. The X-ray structure of the yeast proteasome-homobelactosin C complex reveals acylation of the hydroxyl group of the Thr1 O^γ in the β5 subunit.⁴¹

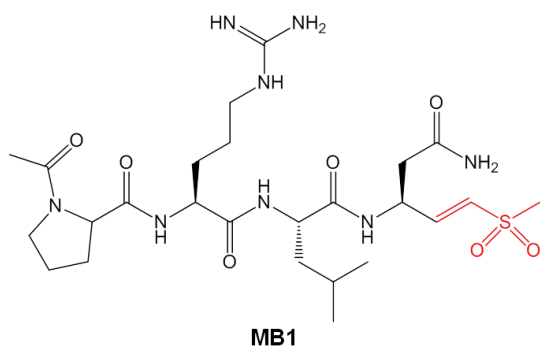
Another class of covalent inhibitors contains peptides possessing a vinyl sulfone moiety, which irreversibly binds to the proteasome but is less reactive than aldehydes. They were first discovered as inhibitors of cysteine proteases, such as cathepsins, because vinyl sulfone acts as a Michael acceptor for soft nucleophilicities such as thiols, leading to covalent bond formation with Thr1 O^γ. Their selectivity can be modified by alternating the peptide parts, so MB1 (Ac-Pro-Arg-Leu-Asn-VS) was developed as a β2 selective inhibitor. Interestingly, an analog, MB2 (Ac-Tyr-Leu-Leu-Asn-VS), is able to inhibit all three activities. These two inhibitors, which differ only in two residues, P3 and P4, demonstrate that binding interactions distal to Thr1 control substrate specificity against the proteasome.

Epoxomicin, α',β'-epoxyketone peptide natural product, is specific for the proteasome and forms morpholino ring in the reaction between the epoxyketone and Thr1 O^γ. Therefore, the compound does not inhibit other proteases since it does not possess an amino acid terminal nucleophilic residue to promote formation of a morpholino adduct with the proteases.⁴² It shows a high degree of selectivity against CT-like activity at lower concentrations, however, at higher concentrations, it covalently binds to all three active subunits.

Lactones



Vinyl Sulfones



Peptide Epoxyketones

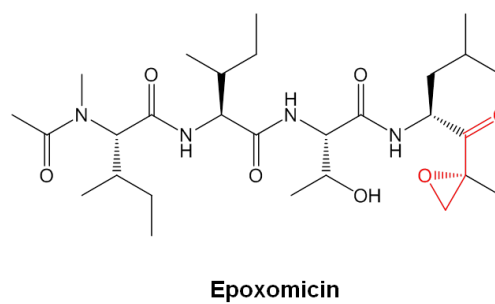


Figure 3. Covalent irreversible inhibitors; lactones (omuralide, salinosporamide A, and homobelactosin C); peptides possessing a vinyl sulfone moiety (MB1); peptide epoxyketone (epoxomicin).

1.2.3 Non-covalent reversible inhibitors: TMC-95s

A new class of specific proteasome inhibitors (TMC-95s) discovered by isolation from *Apiospora montagnei* is the first known category of non-covalent reversible inhibitors. The class binds non-covalently and blocks all three catalytic subunits, but with different activities depending on the characteristics of the P1 and P3 residues. This result arises because the S1 and S3 pockets in the proteasome are major determinants for differential binding.⁴³ Since covalent inhibitors generally cause cell death by induction of apoptosis, non-covalent and time-limited inactivation of the different proteasomal activities might reduce the cytotoxic effects.

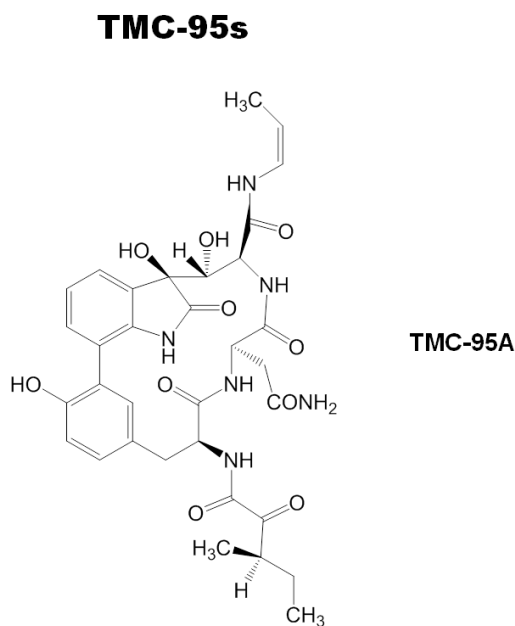


Figure 4. Non-covalent reversible inhibitors (TMC-95s) represented by TMC-95A.

1.3 Specific aims

As mentioned above, bortezomib has shown significant clinical benefit, however, some toxicity was also observed such as nausea, fatigue and diarrhea. More adverse side effects include peripheral neuropathy, neutropenia, lymphopenia, and hyponatremia, etc.⁴⁴ Therefore, various studies to search for new agents with less or no toxicity have been carried out, e. g. by examining naturally occurring or nutritional compounds which are generally more tolerated by the human body.

Curcumin³⁴ and some of its chalcone-based mimics⁴⁵ have also been reported to serve the role as proteasome inhibitors. In attempts to establish a mechanism for this action, these same studies have utilized molecular docking approaches to generate protein-ligand complexes and quantum chemical analysis to suggest that the critical proteasome catalytic residue, Thr1, engages in nucleophilic attack at a carbonyl group of curcumin and its analogs.

However, the molecular structures employed and the mechanistic arguments presented are at odds with the known structures of these molecules and modern reactivity principles, respectively.

The present work addresses these issues by using structurally validated geometries, describing biological data for a series of curcumin analogs acting on the proteasome and offering a testable 3D model for proteasome-curcumin analog interaction. The latter includes a computer-directed survey of inhibitor binding at each of the N-terminal sites within the three catalytically active β subunits (peptidylgultamyl peptide hydrolytic (PGPH)-like, trypsin (T)-like, and chymotrypsin (CT)-like) in order to explore relative selectivity.

Chapter 2. Background

2.1 26S proteasome and 20S proteasome

The 26S proteasome, the executioner of the ubiquitin-proteasome pathway in proteolysis (more than 90% of cell protein degradation), is an extraordinarily large protein aggregate found in all eukaryotes and archaea and some bacteria. It is composed of a 700,000 Da proteolytic core particle (20S proteasome) and two 900,000 Da 19S regulatory particles. The 20S proteasome is cylindrical in shape and consists of 28 subunits that comprise four stacked seven-membered protein rings: two outer α -rings and two inner β -rings (**Figure. 5**). The outer α -rings interact with the 19S regulatory particles,

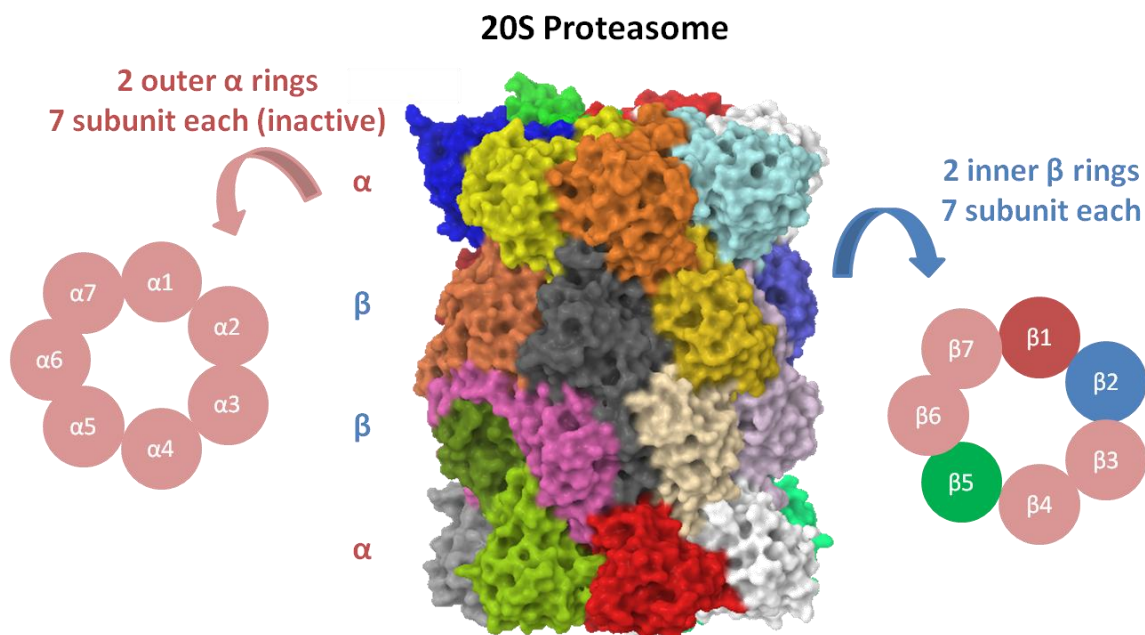


Figure 5. The structure of the 20S proteasome that contains three active catalytic sites with PGPH-like ($\beta 1$), trypsin-like ($\beta 2$), and chymotrypsin-like ($\beta 5$) activities in the inner β rings.

whereas each β -ring hosts the three proteolytically active sites: peptidylglutamyl peptide hydrolytic (PGPH)-like (β 1, cleavage after acidic residues or branched chain amino acid residues), trypsin (T)-like (β 2, cleavage after basic residues) and chymotrypsin (CT)-like (β 5, cleavage after hydrophobic residues) subunits named for the character of the P1 cleavage sites.

The depth and polarity of the proteasomal S1 specificity pockets are governed mainly by residue 45: (1) In the S1 pocket of the β 1 subunit, Arg45 preferentially interacts with acidic residues and favors limited branched chain amino acid residues (PGPH-like activity); (2) The β 2 subunit presents Gly45 in a cleft confined by Glu53, so the spacious S1 pocket accepts very large basic residues (trypsin-like activity); (3) Within β 5, Met45 contributes chymotrypsin-like activity accommodating hydrophobic residues.

(Figure. 6).³⁶

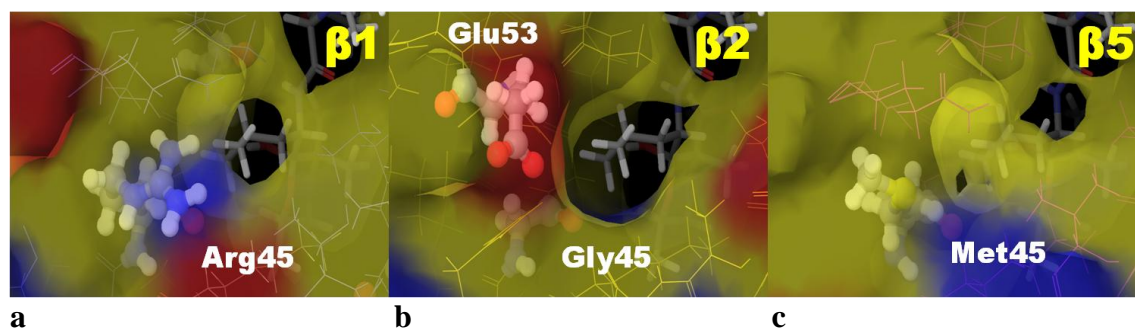


Figure 6. The proteasomal S1 specificity pockets: **a**) β 1 subunit (PGPH-like activity), **b**) β 2 subunit (T-like activity), **c**) β 5 subunit (CT-like activity).

Since catalytic activities of these three active subunits are associated with their individual N-terminal threonine (Thr1) residues as the nucleophile, the 20S proteasome is classified as an N-terminal nucleophilic (Ntn) hydrolases.

2.2 Curcumin and chalcone-based derivatives as proteasome inhibitors

Curcumin, the active ingredient of the natural spice turmeric, has been shown to inhibit the growth of transformed cells and is currently in clinical trials for treatment of various cancers, including multiple myeloma, pancreatic cancer and colon cancer. Moreover, there is no toxicity associated with curcumin even at very high doses.

In this connection, in 2008, Vesna Milacic et al.⁴⁴ reported that curcumin inhibits the proteasome activity in human colon cancer HCT-116 and SW480 cell lines. In order to provide evidence for curcumin's direct inhibition of the proteasome, they incubated purified rabbit 20S proteasome with different concentration of curcumin. As a result, curcumin inhibited all three proteasomal activities with the highest potency for CT-like activity, showing similar potency to MG132. Western blot analysis was also carried out with extract cells to further confirm, resulting in a dose-dependent accumulation of ubiquitinated protein in both cell lines. In addition, curcumin not only induced production of 90.7 % apoptotic SW480 cells at the highest concentration, but also inhibited tumor growth in ICR-SCID mice bearing HCT-116 xenografts after 21 day treatment.

More recently, another group, Martina Bazzaro et al., published that chalcone-based compounds blocks proteasomal activity and simultaneously exhibit dose-dependent antiproliferative and proapoptotic properties in HeLa cervical cancer cells containing human papillomavirus.⁴⁵ In that study, they explored whether the carbonyl group of an

α,β -unsaturated carbonyl system functions as a substrate for the Thr1 O^y by functionalizing the compounds with various amino acid substitutions on the amino group of the 4-piperidone (cf. **Table 6**). Some of them also showed inhibition against proteasomal activities in vitro for proteasome purified from lymphobalstoid cell lines (LCLs) and exhibited a selective pattern toward the three catalytic activities. In addition, combination treatment with bortezomib (FDA approved proteasome inhibitor) and lead compound RA-1 led to greater apoptosis-associated morphological changes in HeLa cervical cells compared to single treatment of the compounds.

2.3 Critical issues common to both papers

Rationalizations of the biochemistry in the three cell types rest on generation of curcumin and RA-1 proteasome complexes followed by assumed Thr1 attack on the carbon of C=O groups.

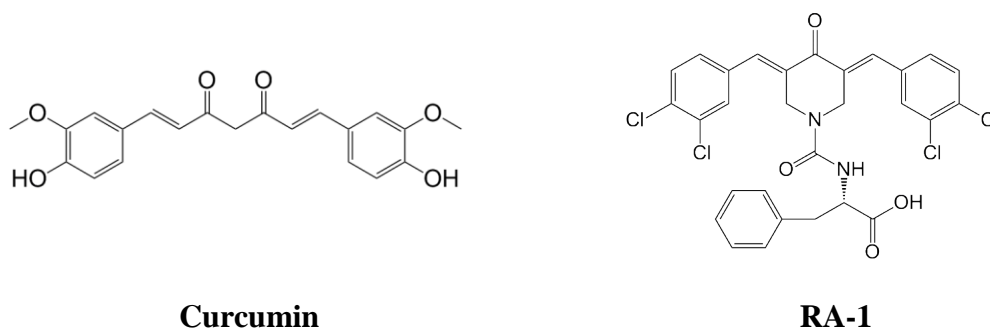


Figure 7. The structures of curcumin and RA-1.

However, two critical issues common to both papers are 1) analysis of proteasome complexes with chemically unreasonable structures of curcumin and RA-1 (**Figure 7**); 2) assignment of lowest unoccupied MO (LUMO) density as the basis for C=O attack by

Thr1 in spite of a rich literature which teaches that conjugate (or Michael) addition is favored.

2.3.1 Issue 1: Misuse of the diketo form of curcumin

The above-mentioned papers^{44,45} analyzed the predicted binding mode of curcumin in the proteasome with chemically unreasonable structures. The crystal structure of curcumin was first determined by single crystal X-ray diffraction in 1982^{46a} and by NMR in 2007.^{46b} In both states, the curcumin molecule exists in the keto-enol form with no detectable presence of the diketo tautomer (**Figure 8**).

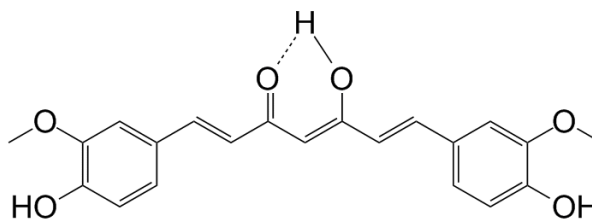


Figure 8. The structure of curcumin keto-enol form

The measured IR and Raman spectra in the solid state and in different solvents also show that curcumin exists only in the keto-enol form: even in nonpolar solvents CCl₄, the keto-enol form predominates.⁴⁷

2.3.2 Issue 2: Implausible binding mode of RA-1 with the cis-trans configuration

In the case of RA-1, docking to the CT-like site of the proteasome was performed with cis-trans configurations around the C=C bonds emanating from the central piperidone.

2.3.3 Issue 3: N-terminal Thr1-OH nucleophilic attack on C=O

The previous authors presumed the carbon of C=O to be the electrophilic partner of the Thr1 OH nucleophile during binding of the curcuminoids to the proteasome.^{44, 45, 46} Neglected, however, was the general observation that Michael acceptors (i.e. C*=C-C=O) uniformly exhibit a reactivity profile that directs the nucleophile to C* in a 1,4-addition reaction that generates a resonance-stabilized enolate anion as a first intermediate. In fact, a number of such acceptors have served as an electrophilic trap and were subsequently captured in X-ray structures of blocked proteasomes. They uniformly reveal a Thr-O-C* covalent attachment.⁴⁸

Chapter 3. Structural examination and reactivity prediction

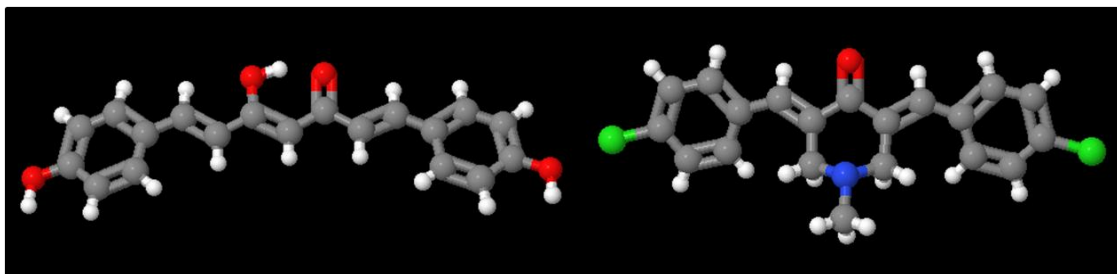
3.1 Structural examination of curcuminoids in the Cambridge Structural Database (CSD)

A search of the CSD⁴⁹ reveals numerous curcumin analogs, none of which adopt the diketo form with two exceptions; namely, when the central carbon of the 1,3-diketone involves a double bond that prevents enolization (**Figure 9 (c) – (d)**).⁵⁰ Replacement of the central CH₂ with alkyl groups in FLLL31 and FLLL32, STAT3 blockers, achieves the same purpose (**Figure 9 (e) – (f)**).⁵¹

In addition, examination of similar structures to RA-1 with a central 5, 6-membered ring and terminal aromatic rings reveals that no case sustains the cis-trans geometry. The trans-trans isomer is uniformly presented in the solid state. **Figure 9** below shows the CSD crystal structures of curcumin and an RA-1 analogs.

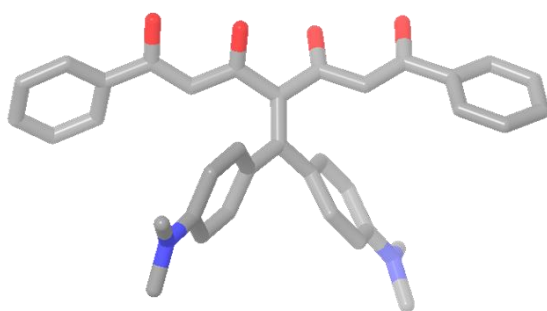
a

b

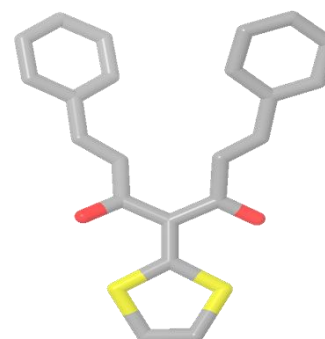


c

d



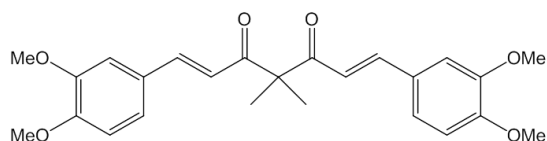
MUKHEF



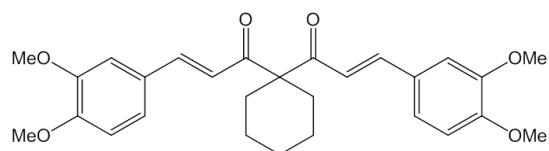
MINRIK

e

f



FLLL31



FLLL32

Figure 9. The crystal structures of (a) curcumin keto-enol form and (b) an RA-1 analog.

They are results of searching CSD. Diketone exceptions; (c) MUKHEF and (d) MINRIK, in which the central carbon of the 1,3-diketone involves a double bond that prevents enolization; STAT3 inhibitors (d) FLLL31 and (f) FLLL32.

3. 2 Optimization and single-point calculations

Density functional (DFT) single-point calculations at the B3LYP/6-311G** level were performed with optimized structures of curcumin (keto-enol / diketo forms) and 4 configurations of RA-1 with B3LYP/6-31G** to determine relative energies between them using Jaguar of Schrödinger.⁵²

Table 1. Total and relative DFT energies of some conformers of curcumin and RA-1.

	B3LYP/6-311G**	
	Total energy (hartrees)	E_{rel} (kJ/mol)
Curcumin		
keto-enol	-1263.897991	0.0
diketo	-1263.887334	28.0
RA-1		
trans-trans	-3370.021042	0.0
trans-cis	-3370.016010	13.2
cis-trans	-3370.014425	17.4
cis-cis	-3370.011144	26.0

The planar keto-enol structure of curcumin is predicted to be more stable than the nonplanar diketo isomer by 28.0 kJ/mol (**Table 1**). This is consistent with the DFT results of Kolev et al.⁴⁷ In the case of RA-1, trans-trans configuration is the most stable form by

13-26 kJ/mol (**Table 1**). It confirms that the trans-trans isomer is a fundamental property of the molecular structure and not a crystal packing effect.

In the subsequent docking studies at the three proteasome 20S catalytic sites, the most stable isomers, keto-enol form of curcumin and trans-trans RA-1, were employed.

3. 3 Quantification of the anticipated reactivity

Atomic Fukui indices are frequently calculated to predict and quantify the relative reactivities of atoms in a molecule in various types of reactions. They are based on Fukui functions which are molecular properties and partial derivatives of the electron or spin density with respect to a change in either the electron count or the unpaired spin count: a change in the electron count would be induced by reaction with another molecule or by any external charge transfer mechanism, while a change in the unpaired spin count would be induced by electromagnetic radiation. We have used Jaguar from Schrödinger⁵³ to carry out such calculations. In this implementation, atomic Fukui indices are calculated only for an N-electron system, using only the MO coefficients from the frontier orbitals. Fukui indices are characterized by two subscripts, N or S, which represent the electron density and the spin density, respectively. The first index indicates the property that responds to a change in the property of the second index. Thus, for example, f_{NN} indicates the changes in the electron density when the molecule experiences a reaction in which its electron density is perturbed. While the atoms in an electrophile that are most reactive to nucleophilic attack are indicated by high positive values of f_{NN} for the LUMO, the most nucleophilic atoms of the attacking agent are indicated by the highest positive values of f_{NN} for the HOMO.

Since curcumin and RA-1 analogs act as Michael acceptors, atomic Fukui indices were calculated with the structure optimized at the B3LYP/6-311G** level to determine the atoms that are most reactive towards Thr1 nucleophilic attack.

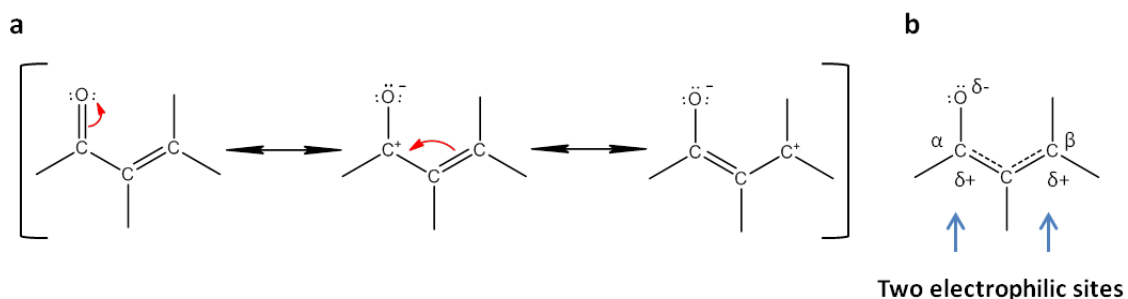


Figure 10. Conjugated enones. (a) Three resonance structures and (b) two electrophilic sites for an α,β -unsaturated carbonyl compound

Generally, the Michael reaction involves two carbonyl components, the enolate of one carbonyl compound and an α,β -unsaturated carbonyl compound. The latter can be described by three resonance structures (**Figure 10. (a)**). Therefore, it has two electrophilic sites; the carbon of carbonyl group and the β carbon (**Figure 10. (b)**). The calculated atomic Fukui indices are consistent with this general fact. As mentioned above, atoms that are most reactive towards nucleophilic attack are indicated by high positive values of f_{NN} for the LUMO.

In the case of the curcumin keto-enol form, the highest value is assigned to the β carbon of the enol, not the carbonyl (C18, red marked in **Figure 11**) with 0.14 (**Table 2**).

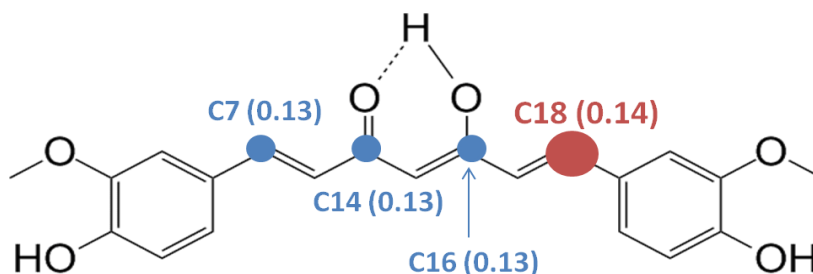


Figure 11. The atom with the highest value of Atomic Fukui Indices is marked in red.

Table 2. Atomic Fukui indices, f_{NN} for the LUMO for curcumin keto-enol form.

Atom	F _{NN} (LUMO)	Atom	F _{NN} (LUMO)	Atom	F _{NN} (LUMO)	Atom	F _{NN} (LUMO)	Atom	F _{NN} (LUMO)
C1	0.01	H11	0.00	C21	0.03	O31	0.04	H41	0.00
C2	0.03	H12	0.00	C22	0.01	H32	0.00	H42	0.00
C3	0.00	C13	0.04	C23	0.04	C33	0.00	H43	-0.00
C4	0.04	C14	0.13	C24	0.00	H34	-0.00	H44	0.00
C5	0.00	C15	0.02	O25	0.01	H35	0.00	H45	0.00
C6	0.04	C16	0.13	O26	0.00	H36	0.00	H46	0.00
C7	0.13	C17	0.04	H27	-0.00	H37	0.00	H47	-0.00
H8	0.00	C18	0.14	H28	0.00	H38	0.00		
O9	0.00	C19	0.04	H29	0.00	C39	0.00		
O10	0.01	C20	0.00	O30	0.09	H40	-0.00		

However, this is not consistent with the general feature that a nucleophile attacks the β carbon of the carbonyl group. As a control, only the α,β -unsaturated carbonyl moiety as incorporated in the following structure was calculated. It is evident that the β carbon of the carbonyl group (**Figure 12**) has the highest value (0.21, **Table 3**).

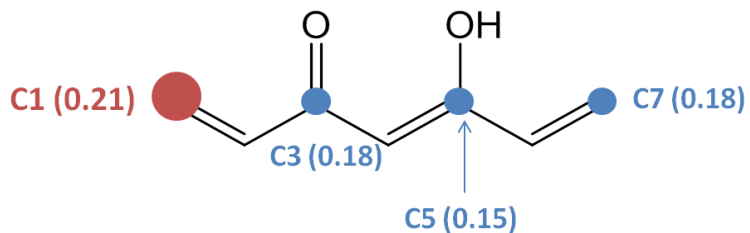


Figure 12. The atom with the highest value of Atomic Fukui Indices is marked in red.

Table 3. Atomic Fukui indices, f_{NN} , for the LUMO for α,β -unsaturated carbonyl moiety.

Atom	F _{NN} (LUMO)	Atom	F _{NN} (LUMO)	Atom	F _{NN} (LUMO)	Atom	F _{NN} (LUMO)
C1	0.21	C5	0.15	O9	0.04	H13	0.00
C2	0.04	C6	0.03	H10	0.00	H14	0.00
C3	0.18	C7	0.18	H11	0.00	H15	0.00
C4	0.03	O8	0.13	H12	0.00	H16	-0.00

For RA-1, two carbons (C3 and C49, red marked in **Figure 13**) at the carbonyl and β positions have the highest value (0.16, **Table 4**).

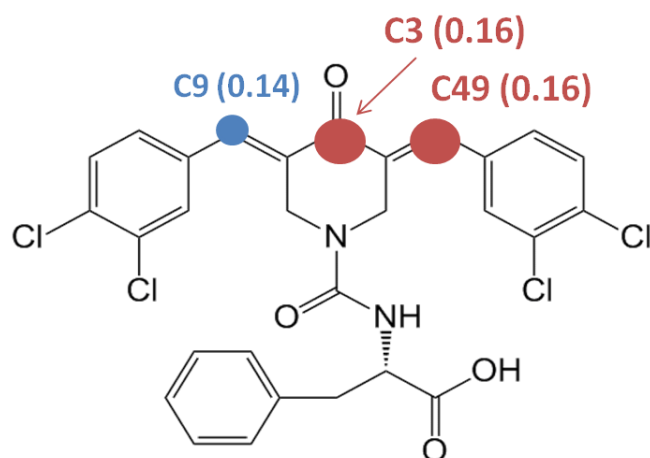


Figure 13. The atom with the highest value of Atomic Fukui Indices is marked in red.

Table 4. Atomic Fukui indices, f_{NN} for the LUMO of RA-1.

Atom	F _{NN} (LUMO)	Atom	F _{NN} (LUMO)	Atom	F _{NN} (LUMO)	Atom	F _{NN} (LUMO)	Atom	F _{NN} (LUMO)
C1	-0.00	H14	0.00	C27	-0.00	C40	0.02	C53	0.02
C2	0.04	N15	0.00	C28	0.00	C41	0.01	C54	0.04
C3	0.16	O16	-0.00	C29	-0.00	C42	0.03	C55	0.00
C4	0.06	C17	0.00	C30	0.00	C143	0.01	C156	0.01
C5	0.00	H18	0.00	C31	-0.00	C144	0.01	H57	0.00
N6	0.01	C19	-0.00	C32	0.00	H45	0.00	H58	0.00
H7	0.00	C20	-0.00	H33	0.00	H46	0.00	H59	-0.00
H8	0.00	H21	0.00	H34	0.00	H47	-0.00	C160	0.00
C9	0.14	H22	0.00	H35	0.00	C148	0.00	H61	0.00
O10	0.14	H23	-0.00	H36	0.00	C49	0.16		
H11	0.00	O24	0.00	H37	0.00	C50	0.05		
H12	0.00	O25	-0.00	C38	0.04	C51	0.01		
C13	0.00	H26	0.00	C39	0.00	C52	0.03		

This calculated indices are consistent with the general feature that the α,β -unsaturated carbonyl compound has two electrophilic sites; the carbon of the carbonyl group and the β carbon (**Figure 10**). While these two sites would seem equally electrophilic, the calculation does not take into account the formation of a delocalized enolate in the one case and a simple alkoxide anion in the other. Thus, in practice the system that forms the most stable intermediate would seem to be favored.

In addition, our lab has also synthesized RA-1 derivatives with various mono- and di-peptide substituents at the nitrogen atom of the piperidone ring to validate the activity of the series. Atomic Fukui indices were calculated for them, and they show results similar to RA-1. The outcome for one of them, Cjk1-11, is shown in the **Figure 14**. In this case, the β carbon (C7) has the highest value, 0.16 (**Table 5**).

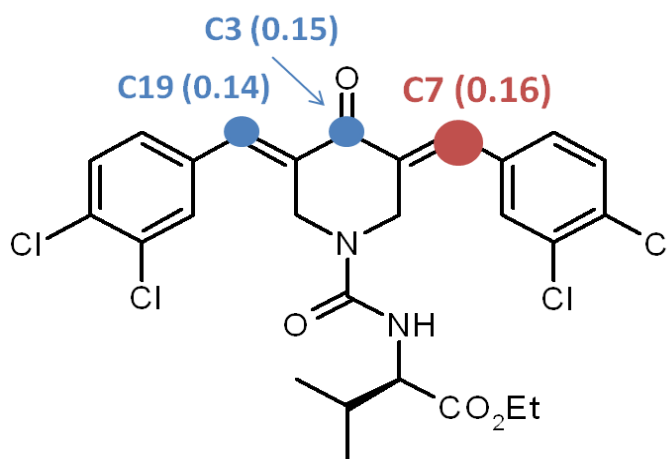


Figure 14. The atom with the highest value of Atomic Fukui Indices is marked in red.

Table 5. Atomic Fukui indices, f_{NN} for the LUMO for Cjk1-11.

Atom	F _{NN} (LUMO)	Atom	F _{NN} (LUMO)	Atom	F _{NN} (LUMO)	Atom	F _{NN} (LUMO)	Atom	F _{NN} (LUMO)
C1	0.00	C14	0.04	H27	0.00	N40	-0.00	H53	0.00
C2	0.06	C15	0.00	H28	0.00	C41	-0.00	C54	0.00
C3	0.15	H16	0.00	H29	0.00	H42	-0.00	H55	0.00
C4	0.04	H17	0.00	C130	0.01	C43	-0.00	H56	0.00
C5	0.00	H18	-0.00	C131	0.00	H44	0.00	H57	0.00
N6	0.01	C19	0.14	C132	0.01	C45	-0.00	H58	0.00
C7	0.16	C20	0.04	C133	-0.00	O46	0.00	H59	0.00
O8	0.13	C21	0.01	H34	0.00	O47	-0.00	H60	0.00
H9	0.00	C22	0.03	H35	0.00	H48	0.00	H61	-0.00
C10	0.05	C23	0.01	H36	0.00	C49	-0.00	H62	0.00
C11	0.01	C24	0.04	H37	0.00	C50	-0.00	H63	0.00
C12	0.03	C25	0.00	C38	0.00	C51	0.00		
C13	0.02	H26	0.00	O39	0.00	H52	0.00		

The Fukui indices predict approximately equal chance of nucleophilic attack at the b-carbon and the adjacent C=O for curcumin. Removal of the phenyl rings (**Figure 12**) tends to favor the b-carbon in agreement with the literature which makes it clear that Michael addition is the preferred form of reactivity. In this reaction in the laboratory, it would appear that the formation of a stabilized enolate anion drives the reaction by comparison with formation of an unstabilized C-O⁻. Thus, Fukui analysis for a broad range of enones appears to be ambiguous and was not very helpful in the present study.

This not inconsistent with previous reports that Fukui indices are not predictive in certain reactions.⁵⁴ Interestingly, this approach appears to suffer limitations similar to molecular orbital analysis reported by Milacic and Bazzaro et al. which makes a similar prediction. In the following, I accept the experimental outcomes and focus on classical Michael reactivity.

3. 4 X-ray structure of the peptide inhibitor possessing the vinyl sulfone moiety

Even though α , β -unsaturated carbonyl compounds have two electrophilic sites, the carbon of the carbonyl group and the β carbon as discussed above (**Figure 10**), it is generally observed that Michael acceptors uniformly exhibit a reactivity profile that directs the nucleophile to β carbon in a 1,4-addition reaction that generates a resonance-stabilized enolate anion as a first intermediate.

In fact, a number of such acceptors have served as an electrophilic trap and were subsequently captured in X-ray structures of blocked proteasomes. They uniformly reveal a Thr-O-C β covalent attachment.⁴⁸

An example is illustrated by compounds possessing a vinyl sulfone moiety as proteasome inhibitors.⁵⁵ It is interesting that two structures which only differ in the P3 and P4 positions, show dramatic differences with respect to their subunit selectivity: MB1 (Ac-PRLN-VS) acts as β 2 selective inhibitor, whereas MB2 (Ac-YLLN-VS) (**Figure 15**) binds to the all three catalytic subunits with similar binding modes.

P3 Arg of MB1 forms hydrogen bond interactions with Ser20 of β 2 (Thr20 of β 1 / Ala20 of β 5, instead) and Cys118 of adjacent β 3 (Ser118 of both β 2 and β 6, instead).

However, $\beta 1/ \beta 2$ and $\beta 5/ \beta 6$ lack the residues for hydrogen bond interactions with the P3 Arg due to different locations and character of these residue side chains (Figure 17, 18, and 22). In addition, P4 Pro forces a slight kink in the inhibitor that causes the tail portion of the inhibitor to bend away from the cavity wall of S4.

Therefore, the binding poses of MB1 and MB2 explain the favorable interactions between P3 and S3, generated at the interface of β_i and β_{i+1} . The poses not only affect the selectivity against different catalytic subunits, but also demonstrate that these interactions distal to the catalytic site Thr1 control the absolute substrate specificity of the multiple active sites of the proteasome.

According to the crystal structures, these two peptide vinyl sulfones bind to the yeast 20S proteasome, forming covalent bonds between C* of C*-C-SO₂ and O of Thr1 with a distance 1.54 Å (**Figure 16**).

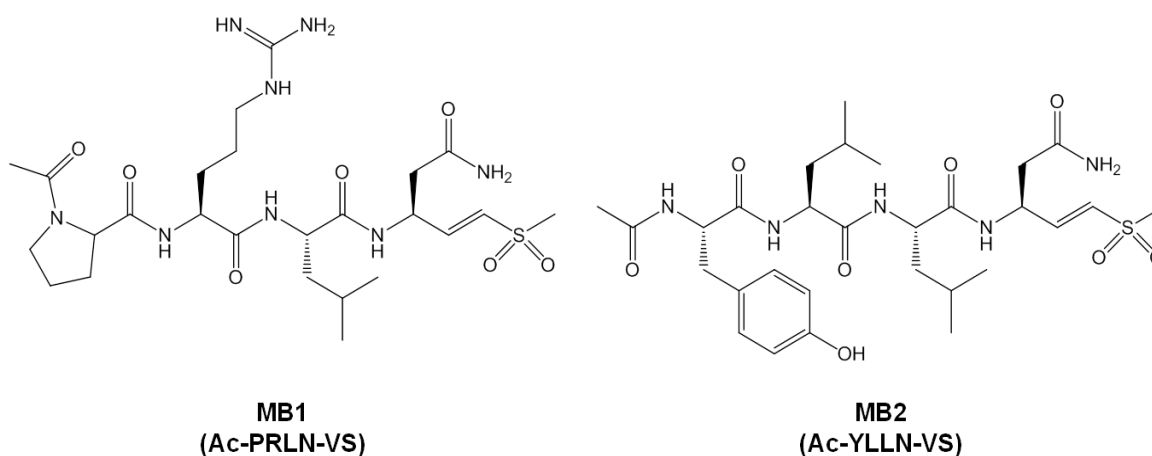


Figure 15. The structures of $\beta 2$ inhibitors: **MB1** (Ac-PRLN-VS) is selective, while **MB2** (Ac-YLLN-VS) which inhibits all active sites.

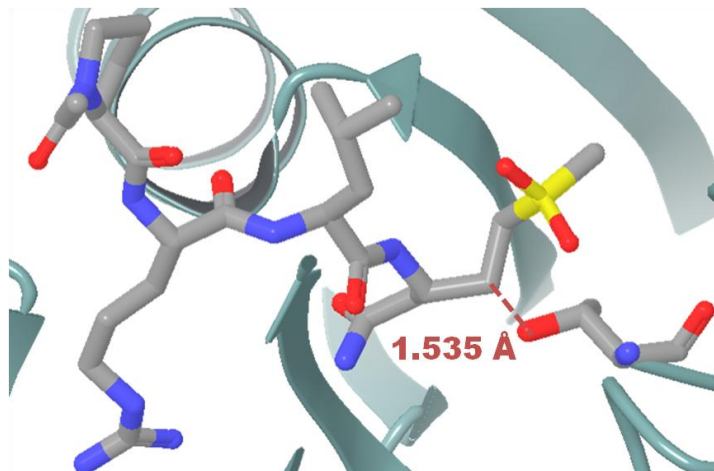


Figure 16. The proteasome bound MB1 crystal structure.

Chapter 4. Modeling the subunit targets

4.1 Active sites analysis

As discussed above (**Figure 6**), the proteasomal S1 specificity pockets are mainly formed by residue 45 of the corresponding β -subunit. The positively charged S1 of β_1 due to Arg45 is the smallest pocket, whereas the negatively charged S1 of β_2 due to Gly45 and Glu53 is spacious.

In contrast to S1, the S2 pocket is not in contact with P2 which points into empty space. Therefore, bulky and spatially demanding groups might be introduced at P2 of the ligands to enhance the specificity of the inhibitors for the proteasome against other proteases which might have space-limiting S2 pockets. That is, while the proteasome might accommodate large P2 substituents, other proteases might not.

The largest pocket of the catalytic subunit is S3 which is located at the interface between the primary catalytic subunit and the backside of the adjacent β subunit. Of the

three subunits, the S3 pocket of $\beta 1$ is the smallest, while the S3 cleft of $\beta 5$ is deeper than that of $\beta 2$ (**Figure 17**).

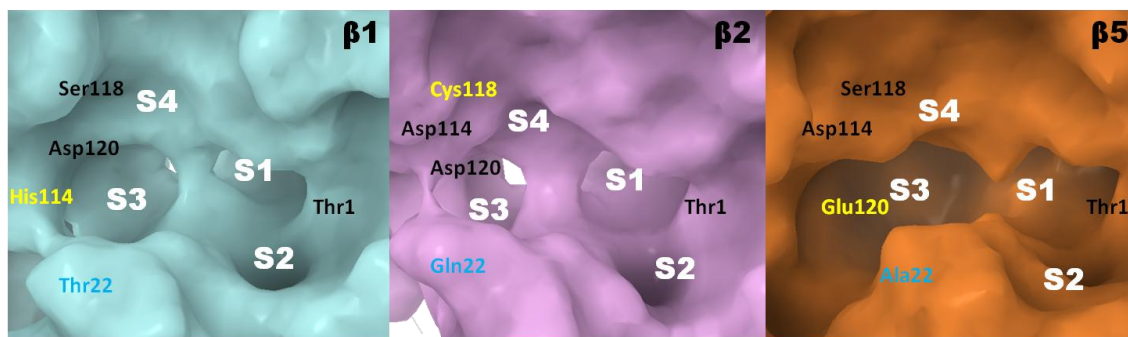


Figure 17. The three catalytic subunits of the 20S proteasome.

In addition, there are Asp114 residues in $\beta 3$ adjacent to $\beta 2$ and $\beta 6$ adjacent to $\beta 5$ which can make HB interactions with inhibitors. On the other hand, the S3 pocket of $\beta 1$ consists of His114 from the adjacent $\beta 2$. Moreover, the S3 pockets of $\beta 1$ and $\beta 5$ include Ser118, whereas $\beta 2$ incorporates Cys118 instead of Ser118. Thus, S3 can also serve as another specificity pocket to distinguish the three proteasome subunits by variation of P3.

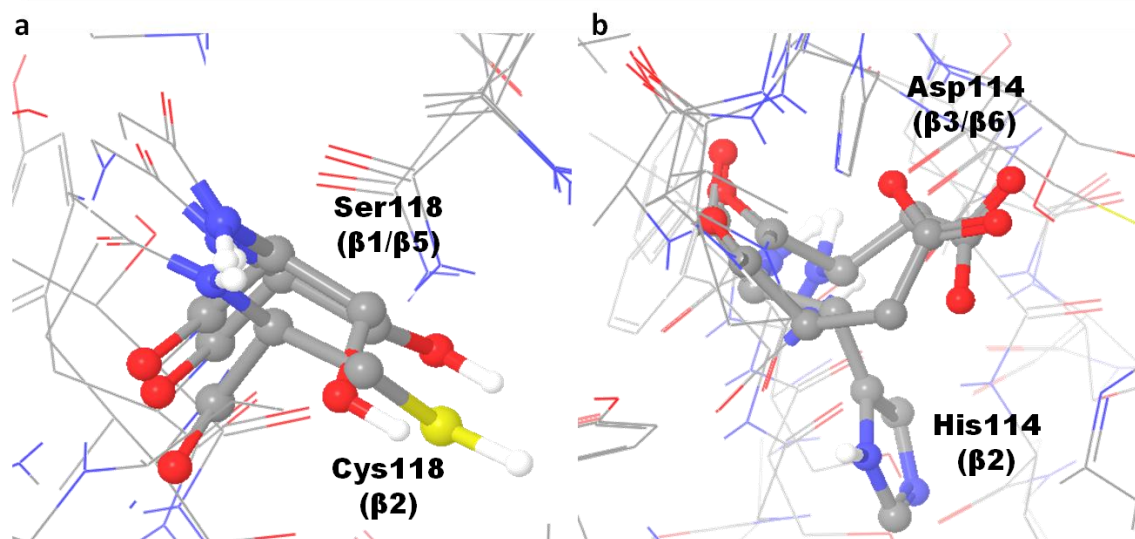


Figure 18. Different residue environments. (a) While Asp114 in $\beta 3$ and $\beta 6$ is adjacent to $\beta 2$ and $\beta 5$ forming the S3 pocket, $\beta 2$ adjacent to $\beta 1$ incorporates His114 instead. (b) Ser118 resides in $\beta 1$ and $\beta 5$, whereas $\beta 2$ incorporates Cys118 instead of Ser118.

The S4 pocket of $\beta 2$ is quite small, so bulky hydrophobic P4 substituents can be excluded from binding to the $\beta 2$ subunit (**Figure 17**).

As a result, alterations to the P1 and P3 may facilitate the further improvement of binding properties of proteasomal inhibitors with tunable selectivity for each of the catalytic subunits.

4.2 Non-covalent docking studies

To investigate curcumin's and its analogs' mechanistic profiles as proteasome inhibitors, non-covalent docking studies have been carried out with Schrödinger's Glide for the curcumin keto-enol form, RA-1 and its derivatives, and several Emory compounds at all three proteasome catalytic subunits. While the literature regards Thr1 OH as the

primary nucleophile, the X-ray structures reveal several nucleophiles in the binding pockets with potential to complex with or attack diverse ligands.

In non-covalent docking of the small molecules separately to the three binding centers as a preliminary to covalent attachment, we pay special attention to the geometric relationship of these nucleophiles to the β carbon of Michael acceptor center. RA-1, a special case incorporating a phenylalanine tail on the central piperidone ring, is the single curcumin analog that shows sub-micromolar activity on HeLa cells, while all other amino acid substitutions are 30-fold less active illustrating a flat SAR.³⁵ By contrast, the Leu and Tyr analogs exhibit in vitro μM IC_{50} values for proteolytic activities of isolated proteasome purified from lymphoblastoid cell lines (LCLs) as does the Phe analog. Even though the authors introduced diverse amino acid tails at the P position presuming selectivity toward the three proteasome subunits, they didn't mention their role in the paper.

Therefore, the undetermined role of the amino acid tail in RA-1 was also investigated. Based on modeling studies, I propose that the tails do not populate the S1 specificity pocket, but themselves on the S3 or S4 surface instead. This might be the reason why the compounds inhibit all three proteasomal activities even though the amino acid tails of the active compounds are hydrophobic. Section 4.4 elaborates this point in detail.

4.3 Docking of curcumin

According to the study carried out by Milacic et al., curcumin was shown to be able to inhibit all three proteasomal activities with the highest potency at the CT-like

catalytic center.⁴⁴ In addition, Michael Groll and colleagues also demonstrated that curcumin is quite active in-vitro: it inhibits all active sites of the 20S yeast proteasome in the mid to low μM range ($\beta 1$, 72 μM ; $\beta 2$, 20 μM and $\beta 5$, 10 μM).

Above, it is proposed that the β carbon of curcumin is susceptible to nucleophilic attack by Thr1 O based on Atomic Fukui indices and the crystal structure of vinyl sulfone inhibitors MB1 and MB2. Therefore, binding poses that place the β carbon of curcumin within the 4.0 Å of Thr1 O were examined.

4.3.1 The predicted binding pose of curcumin at $\beta 1$ and $\beta 2$

First of all, curcumin doesn't seem to be able to bind at $\beta 1$ with good affinity because of the small size of the binding pockets.

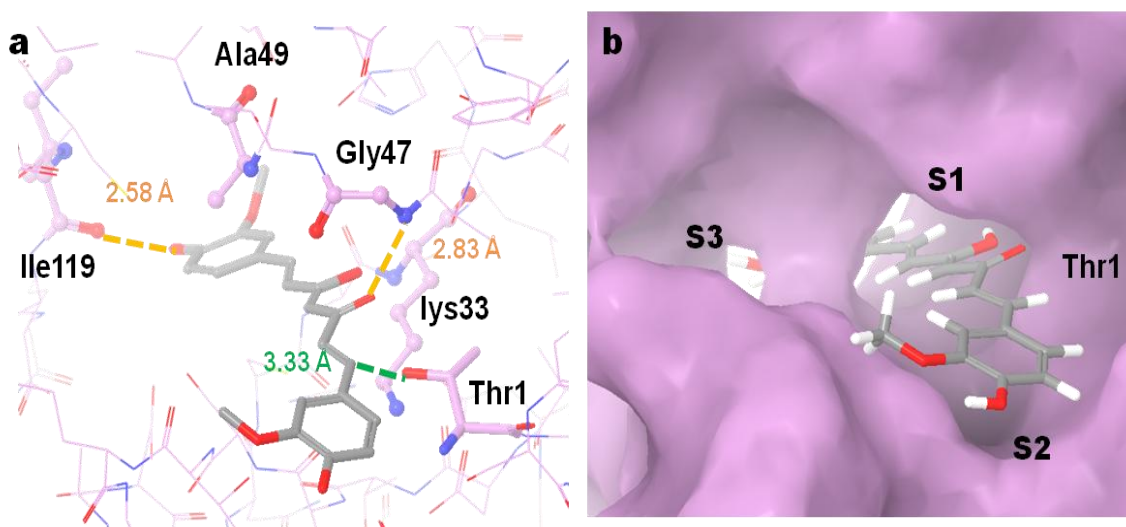


Figure 19. The predicted binding pose of the curcumin keto-enol form to the $\beta 2$ subunit: (a) Distance between the β carbon of carbonyl group and the O of Thr1 is 3.33 Å, while

two hydrogen bonds (2.58 and 2.83 Å, respectively) with backbone atoms O of Ile119 and N of Gly47 anchor the ligand in the pocket, and (b) on the molecular surface of $\beta 2$.

On the other hand, at $\beta 2$ curcumin fills the spacious S1 pocket, and one of the terminal aromatic rings is directed onto the S2 surface (**Figure 19. (b)**). The β carbon conjugated to C=O is close to O of Thr1 at a distance of 3.33 Å, while curcumin forms hydrogen bonds with the backbone atoms O of Ile119 and N of Gly47 with distances of 2.58 Å and 2.83 Å, respectively. In addition, the natural product seems to make good hydrophobic contacts with the side chains of Ala49 and Lys33 (**Figure 19. (a)**)

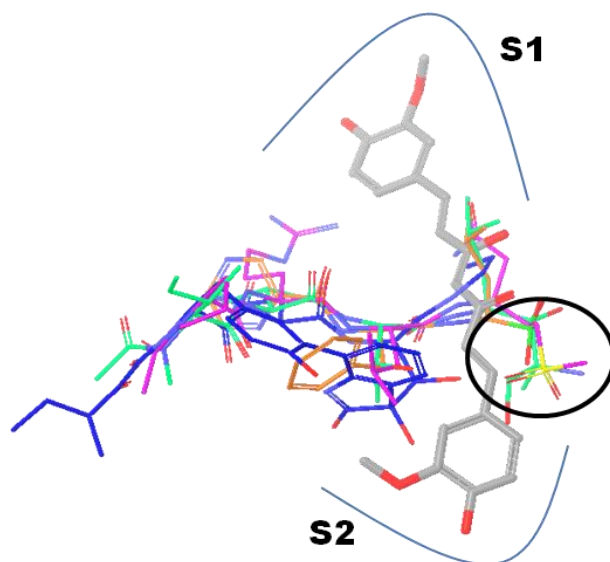


Figure 20. Superposition of curcumin with crystal structures of other known inhibitors at $\beta 2$: curcumin (gray carbon), epoxomicin (green carbon), bortezomib (orange carbon), vinyl sulfone inhibitor MB1 (pink carbon) and non-covalent inhibitor TMC-95A (blue carbon). The position for covalent-bond formation is marked with a black circle.

Figure 20 shows the superposition of the crystal structures of other known inhibitors at $\beta 2$ with different carbon colors: curcumin (gray carbon), epoxomicin (green carbon), bortezomib (orange carbon), vinyl sulfone inhibitor MB1 (pink carbon) and the non-covalent inhibitor TMC-95A (blue carbon). As shown in the figure, the position of the Michael acceptor β carbon of curcumin is very close to the site where covalent bonds are formed between various inhibitors and Thr1 (black circle): the distance between the β carbons and the electrophilic atoms of known inhibitors is 3.03 Å, 2.47 Å, and 2.63 Å for epoxomicin, bortezomib, and MB1, respectively. Importantly, the carbon chain of curcumin is aligned with the hydrophobic side chains of other inhibitors, such as Leu or n-propylene which make hydrophobic contacts with Lys33.

4.3.2 The predicted binding pose of curcumin at $\beta 5$

At $\beta 5$, curcumin shows a binding mode similar to that at $\beta 2$, filling the S1 pocket and pointing onto the S2 surface. In addition, the β carbon is close to the O of Thr1 with a separation of 3.66 Å. Since S3 of $\beta 5$ is deeper and more spacious than $\beta 2$ (**Figure 21. (b)**) as a result of a conformational reorganization of Glu120 compared to Asp120 and the presence of Ala20 instead of Ser20 (**Figure 22**), curcumin also fills the deep S3 pocket forming hydrogen bonds with the backbone O of Ser112 (3.12 Å) (**Figure 21. (a)**).

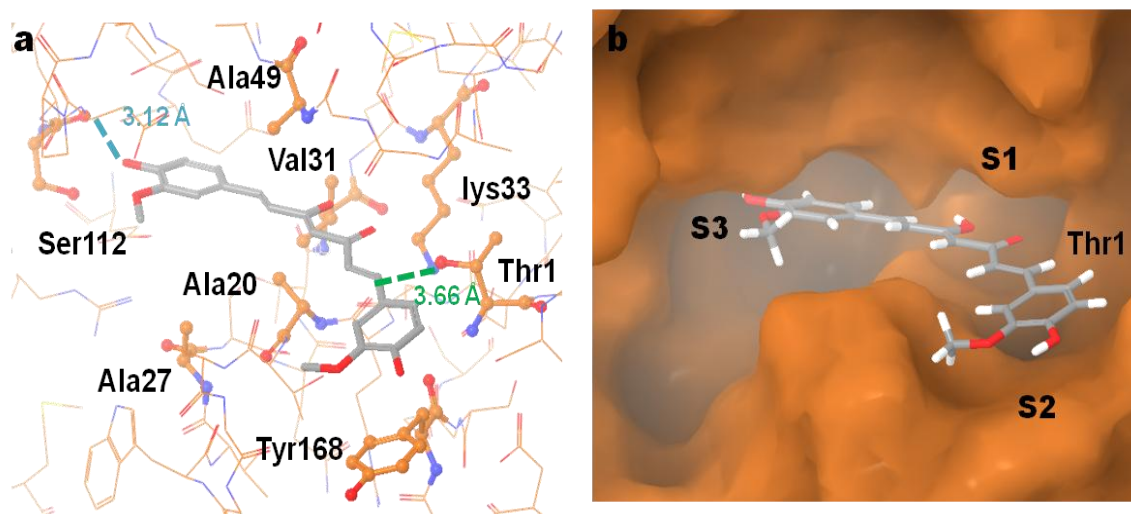


Figure 21. Predicted binding pose of curcumin keto-enol form to the $\beta 5$ subunit: (a) The β carbon of carbonyl group and the O of Thr1 are separated by 3.66 Å. Ligand anchoring results from an H-bond between the backbone O of Ser112 and the hydroxyl O of curcumin (3.12 Å) and a π - π interaction to Tyr168 on the S2 surface with a close distance (4.41 Å) between the carbons of the phenyl rings, and (b) on the molecular surface of $\beta 5$.

In addition, curcumin seems to make good hydrophobic contacts with the side chains of Ala20, Ala27, Ala49, Val31 and Lys33. Moreover, one of the terminal aromatic rings makes a π - π interaction with Tyr168 on the S2 surface with a close distance of 4.41 Å between the phenyl ring carbons (**Figure 21 (a)**).

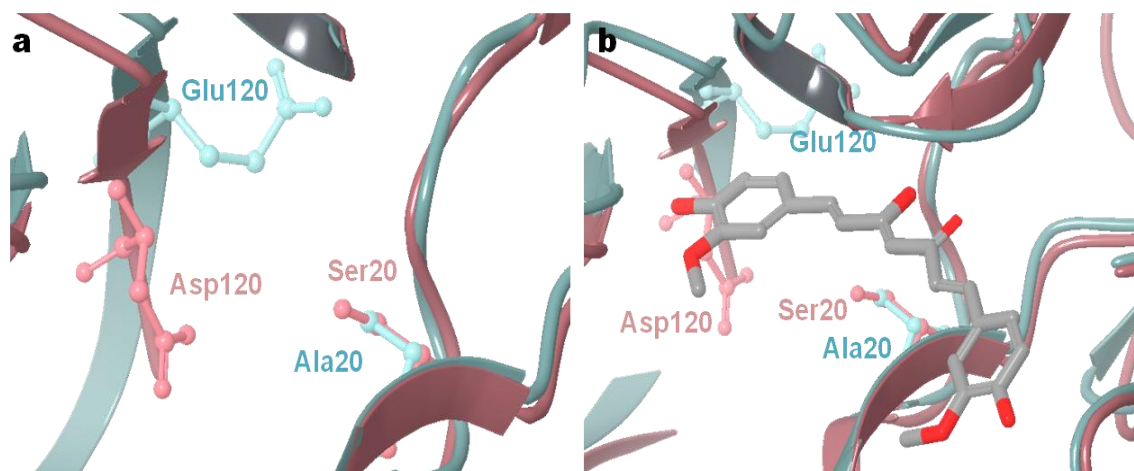


Figure 22. The different residues at $\beta 2$ (pink) and $\beta 5$ (light blue). The pocket of $\beta 5$ is more spacious than $\beta 2$ because of a refolded Glu120 instead of Asp120 and Ala20 instead of Ser20.

Figure 23 illustrates the superposition of curcumin with the crystal structures of other known inhibitors at $\beta 5$ in different carbon colors: curcumin (gray carbon), epoxomicin (green carbon), bortezomib (orange carbon). The position of the Michael acceptor of curcumin is close to the location where covalent bonds are formed between the inhibitor and Thr1 (black circle) as indicated by the distances between β carbon and electrophilic atoms of known inhibitors is 2.71 Å, and 2.55 Å with epoxomicin, and bortezomib, respectively. One 4-hydroxy-3-methoxyphenyl unit of curcumin resides on the S2 surface, whereas the other aromatic ring inserts deeply into the S3 pocket similar to the P3 Ile and Asn side chains of epoxomicin and TMC-95A, respectively, and the P3-pyrazine of bortezomib.

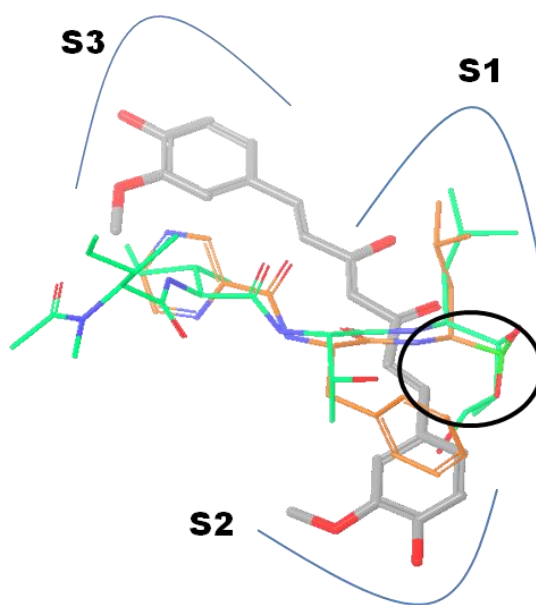


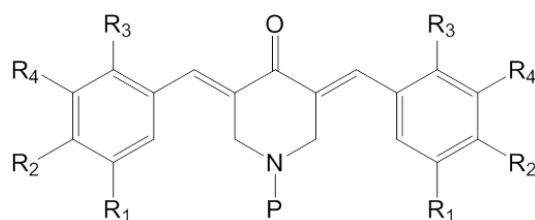
Figure 23. Superposition of curcumin with crystal structures of other known inhibitors $\beta 5$: curcumin (gray carbon), epoxomicin (green carbon), bortezomib (orange carbon). The position for covalent-bond formation is marked with a black circle.

4.3.3 Summary of the predicted binding poses of curcumin at all three catalytic subunits

In summary, the docking studies of curcumin just described imply a better binding mode at $\beta 5$ by filling the two specific pockets, S1 and S3, as well as interacting with Tyr168 at the S2 surface. These results are in accord with the studies carried out by the Milacic and Groll groups showing that curcumin has the highest potency at $\beta 5$, strongly demonstrating that the docking studies follow experiments.

4.4 Docking of RA-1 analogs

Table 6. The structures of chalcone-based derivatives⁴⁵



Compound	R1	R2	R3	R4	P	
2 RA-1	Cl	Cl	H	H	-CO-phenylalanine	
3 RA-2	Cl	Cl	H	H	-CO-tyrosine	CT-like
4 RA-3	Cl	Cl	H	H	-CO-leucine	
5 RA-4	Cl	Cl	H	H	-CO-glutamic acid	PGPH-like
6 RA-5	Cl	Cl	H	H	-CO-lysine	T-like
7 RA-6	Cl	Cl	H	H	-CO-phenylalanine-tyrosine	
8 RA-7	Cl	Cl	H	H	-CO-glutamic acid-aspartic acid	
9 RA-8	Cl	Cl	H	H	-CO-lysine-arginine	
10 RA-9	Cl	Cl	H	H	-CO-arginine-lysine	

Martina Bazzaro et. al functionalized chalcone-based structures with either aromatic (**2** and **3**), hydrophobic (**4**), acidic (**5**), or basic (**6**) amino acidic substitutions at position P, predicting the selectivity toward chymotrypsin-like (**2-4**, hydrophobic functional group), PGPH-like (**5**, acidic functional group), and trypsin-like (**6**, basic functional group), respectively (**Table 6**). In addition, the length of the amino-acid portion at position P was extended by dipeptides in compounds **7-10** to explore the potential affinity.

Compounds **2-4** showed antiproliferative activity, whereas the action of compounds **5-10** dramatically decreased with $IC_{50} > 50 \mu M$ (**Table 7**).⁴⁵

Table 7. Cell killing activities of compounds **2-10** on HeLa and CaSki cervical cancer cell lines⁴⁵

Compound		IC_{50} (μM) (HeLa cells)	IC_{50} (μM) (CaSki cells)
2	RA-1	0.32	1.5
3	RA-2	10	10
4	RA-3	10	20
5	RA-4	> 50	> 50
6	RA-5	> 50	> 50
7	RA-6	> 50	> 50
8	RA-7	> 50	> 50
9	RA-8	> 50	> 50
10	RA-9	> 50	> 50

Table 8. In vitro inhibition of the proteolytic activities of proteasome purified from lymphoblastoid cell lines (LCLs)⁴⁵

Compound		Isolated enzyme IC_{50} (μM)		
		PGPH-like ($\beta 1$)	T-like ($\beta 2$)	CT-like ($\beta 5$)
2	RA-1	6.2	6.6	4.8
3	RA-2	> 10	8.4	4.2
4	RA-3	> 10	> 10	9.3

Moreover, 10 μM of compound **2** nearly fully inhibited PGPH-like, T-like and CT-like activities of the purified proteasome.

On the other hand, compounds **3** and **4** only induced partial inhibition with a selective inhibition pattern toward T / CT-like activities and CT-like activity, respectively (**Table 8, Figure 24**).⁴⁵ In the Bazzaro study,¹³ the role of the amino acid tail at the P position of RA-1 analogs was neither determined nor rationalized. However, the present work suggests that the tail resides in the hydrophobic extension at P, i.e. Phe, Tyr, Ile.

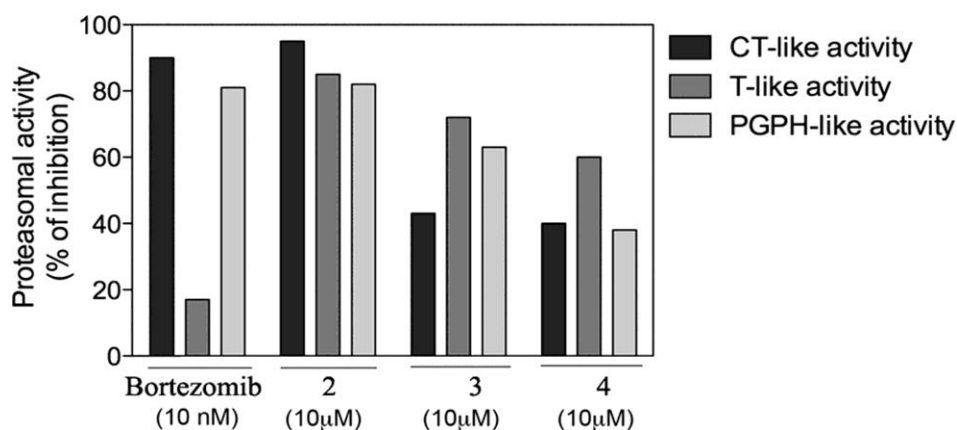


Figure 24. Relative inhibition of the three catalytic subunits by compounds **2-4**.

Bortezomib was used as a positive control.⁴⁵

4.4.1 The predicted binding pose of RA-1 ($\text{IC}_{50} = 6.2$ ($\beta 1$), 6.6 ($\beta 2$), and 4.8 ($\beta 5$)) at all three catalytic subunits

The docking results generate good binding poses at all three subunits (**Figure 25**).

First of all, one β carbon is close to the Thr1 oxygen atoms (2.87 Å ($\beta 1$), 3.80 Å ($\beta 2$), and 3.17 Å ($\beta 5$)). The corresponding O---C=C angles are -98.4 ($\beta 1$), 66.9 ($\beta 2$),

and 90.9 ($\beta 5$) and all lie near a common plane, implying rather poor geometries for nucleophilic attack. Although $\beta 1$ reveals the closest contact between the β carbon and Thr-O, all three might in principle serve as Michael nucleophiles in view of the dynamic behavior of proteins at ambient temperature which can possibly lead to productive reaction trajectories.

One of the di-chloro phenyl rings sits in the S1 pocket, while the other lays on the S2 surface. The former penetrates deeply into the S1 pockets of $\beta 2$ and $\beta 5$ because of pocket size (**Figure 26. (a)**). Furthermore, the short π - π interactions at $\beta 5$ between the latter chlorinated ring and Tyr168 with the short 4.18 Å separation seems to be one of the reasons for the higher activity at $\beta 5$ relative to $\beta 1$ and $\beta 2$ (**Figure 25. (e)**).

One telling structural feature is the locus of the Phe phenyl ring at P4 which is directed at the S4 hydrophobic surface in $\beta 2$ rather than the S3 pockets, presumably because of stronger hydrophobic contacts with the side chains of Thr48, Leu115 and Ile116 (**Table 9**).

Table 9. The residues which comprise the S4 surface of all three subunits.

$\beta 1$	$\beta 2$	$\beta 5$
Ser48	Thr48	Gly48
Ala115	Leu115	Pro115
His116	Ile116	Val116

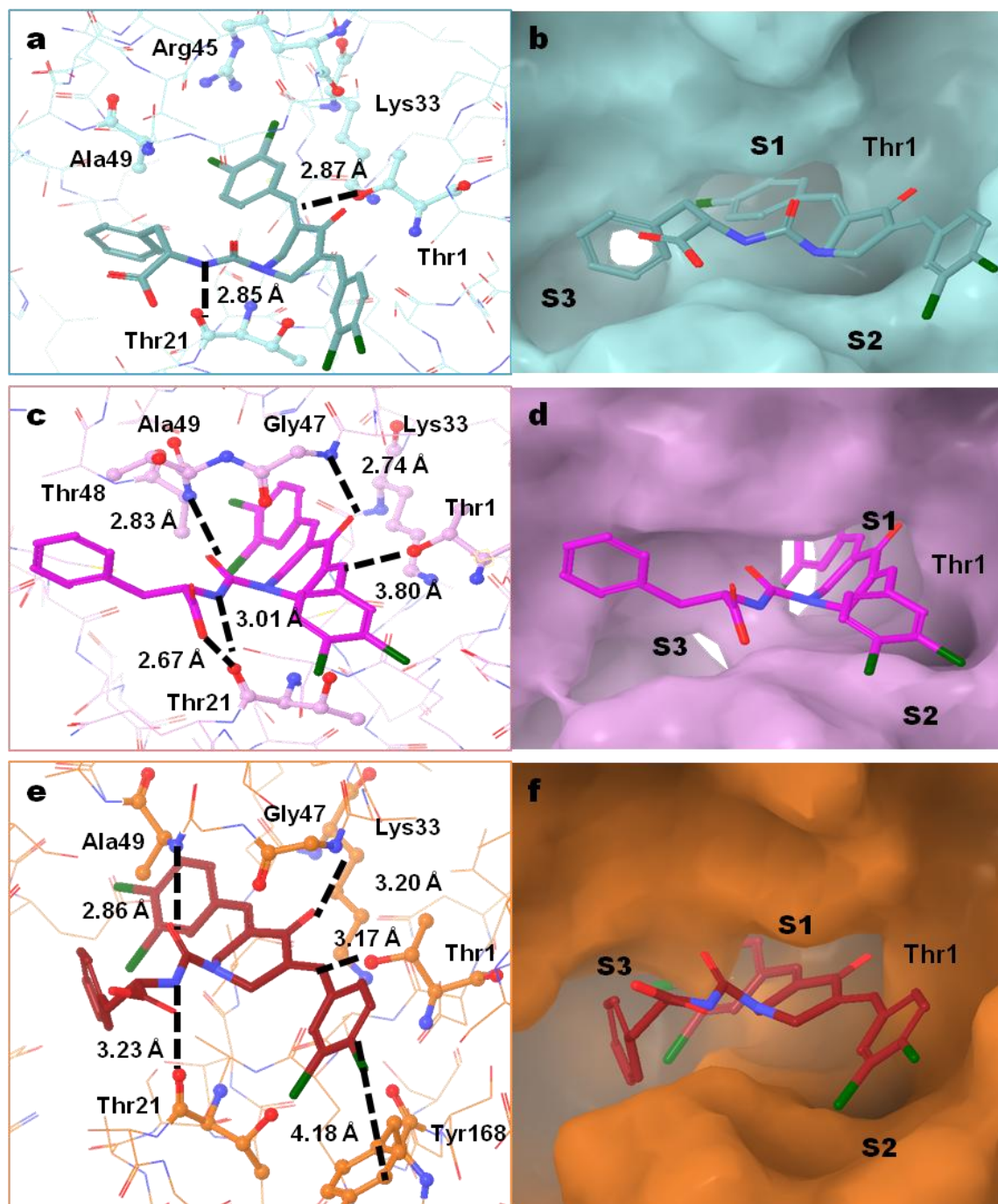


Figure 25. The predicted binding poses of RA-1 to $\beta 1$ (a), $\beta 2$ (c) and $\beta 5$ (e), and on the molecular surfaces of $\beta 1$ (b), $\beta 2$ (d) and $\beta 5$ (f). The residues which form hydrogen bonds and hydrophobic interactions with RA-1 are shown by the interatomic distances (dotted lines, Å) between heavy atoms.

In addition, RA-1 forms hydrogen bonding interactions with the backbone O of Thr21 (2.85 Å) at β 1, the backbone N of Gly47 (2.74 Å) and Ala49 (2.83 Å) and the backbone O of Thr21 (2.67 Å) at β 2, and the backbone N of Gly47 (3.20 Å) and Ala49 (2.86 Å) and the backbone O of Thr21 (3.23 Å) at β 5 (**Figure 25, (a, c, e)**). These interactions play a key role in orienting the RA-1 within the binding site.

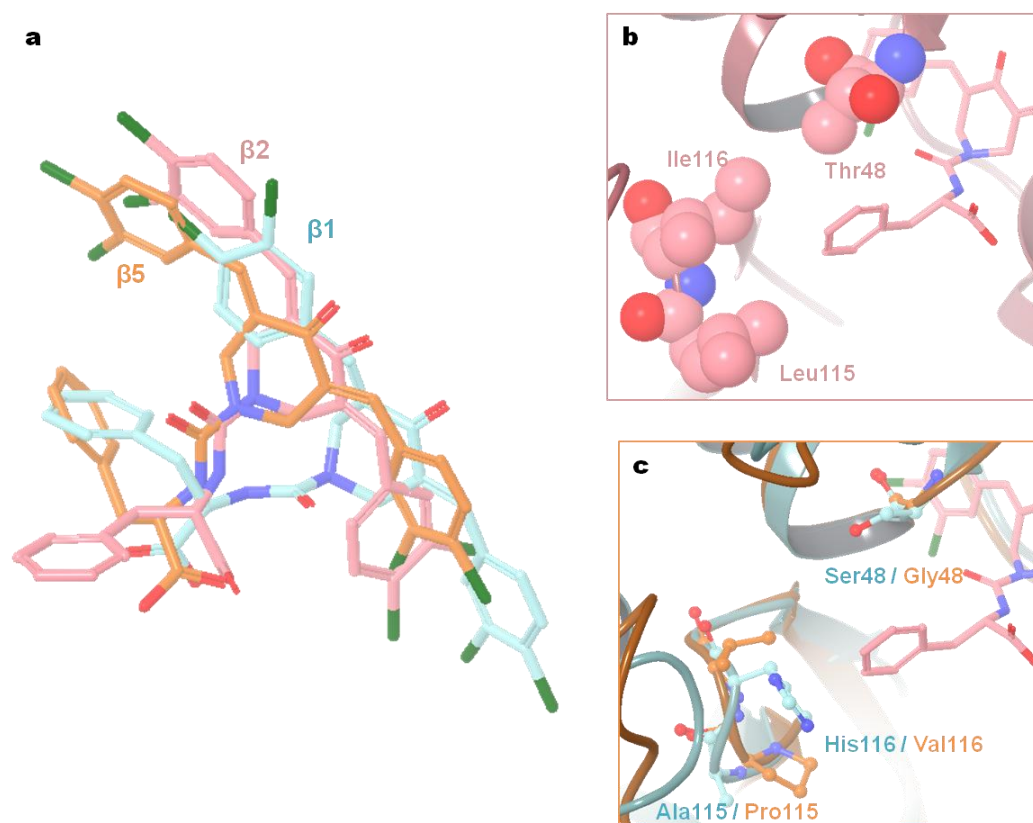


Figure 26. RA-1 docking: (a) The superimposed poses of RA-1 at β 1 (turquoise), β 2 (pink), and β 5 (orange). (b-c) The phenyl ring at β 2 sits on the S4 hydrophobic surface rather than in the S3 pocket because of stronger hydrophobic contacts with side chains of Thr48 (Ser48 at β 1 / Gly48 at β 5), Leu115 (Ala115 at β 1 / Pro115 at β 5) and Ile116 (His116 at β 1 / Val116 at β 5).

Figure 27 shows the superposition of RA-1 with the crystal structures of other known inhibitors with different carbon color at β 1s: RA-1 (light blue), epoxomicin (green carbon), bortezomib (orange carbon), TMC-95A (blue). The position of the Michael acceptor atom of RA-1 is close to the place where covalent bonds are formed between other inhibitors and Thr1 (yellow circle). The distances between the β carbon and the electrophilic atoms of known inhibitors are 1.70 Å and 1.55 Å with epoxomicin, and bortezomib, respectively. The phenyl ring of Phe at P4 position slides deeply into the S3 pocket as do the P3 Ile and Asn side chains of epoxomicin and TMC-95A, and the P3-pyrazine of bortezomib.

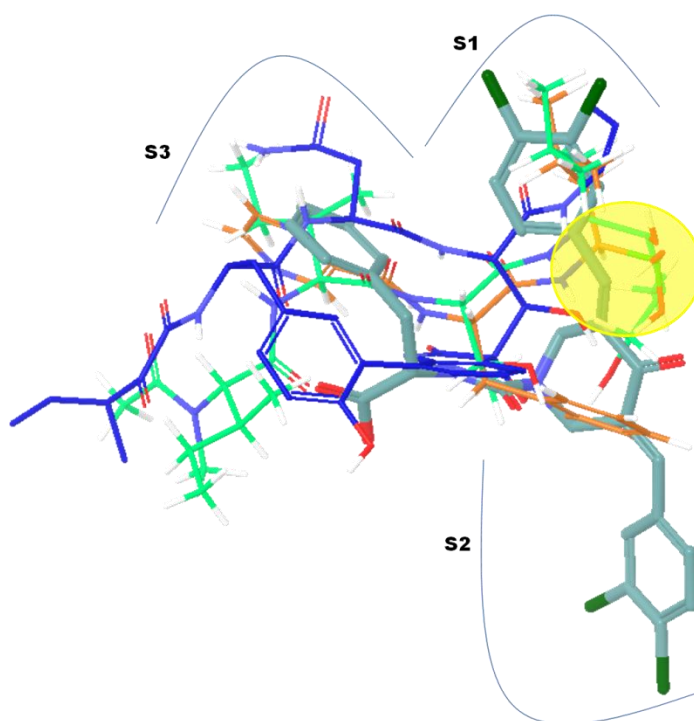


Figure 27. Superposition of RA-1 with the crystal structures of other known inhibitors at β 1: RA-1 (light blue), epoxomicin (green carbon), bortezomib (orange carbon), TMC-95A (blue). The position for covalent-bond formation is marked with a yellow circle.

4.4.2 The predicted binding pose of RA-2 ($IC_{50} > 10$ ($\beta 1$), = 8.4 ($\beta 2$), and 4.2 ($\beta 5$)) at all three catalytic subunits

RA-2 with a CO-tyrosine moiety at the P position shows antiproliferative activity against both HeLa and CaSki cervical cancer cell lines at 10 μ M, and induced partial inhibition of T and CT-like activities.

In this case, RA-2 shows a different binding mode at $\beta 1$ relative to $\beta 2$ and $\beta 5$ in which the chalcone moiety is placed in a solvent exposed area. This is because the Tyr OH experiences a steric clash with the S3 surface due to the smaller size of the overall active sites of the other two subunits. Therefore, it is likely to occupy the S1 or S3 pockets, pushing out other parts of the compound into the solvent exposed area as predicted. Thus, this compound is unlikely to engage in a Michael reaction.

On the other hand, at $\beta 2$ and $\beta 5$, RA-2 generates poses similar to RA-1, in which one β carbon comes close to the Thr1 O atoms (3.68 Å ($\beta 2$), and 3.17 Å ($\beta 5$), respectively). Positioning of the β carbon in this way suggests the possibility of a nucleophilic attack by Thr1 (**Figure 28**). One of the di-chloro phenyl rings occupies the S1 pocket, while the other points to the S2 surface. Furthermore, the short π - π interactions at $\beta 5$ between the latter and Tyr168 (shortest distance between the carbons of two phenyl rings, 3.85 Å) seems to be one of the reasons for the higher predicted activity relative to $\beta 2$.

RA-2 forms hydrogen bonding interactions with the backbone N of Gly47 (2.80 Å) and Ala49 (2.85 Å) and the backbone O of Thr21 (3.03 Å) at $\beta 2$; the backbone N of Gly47 (3.23 Å) and Ala49 (2.73 Å) at $\beta 5$ (**Figure 28, (a, c)**). In addition, the side chain of Tyr establishes hydrophobic interactions with side chains of Ala49 and Gln22.

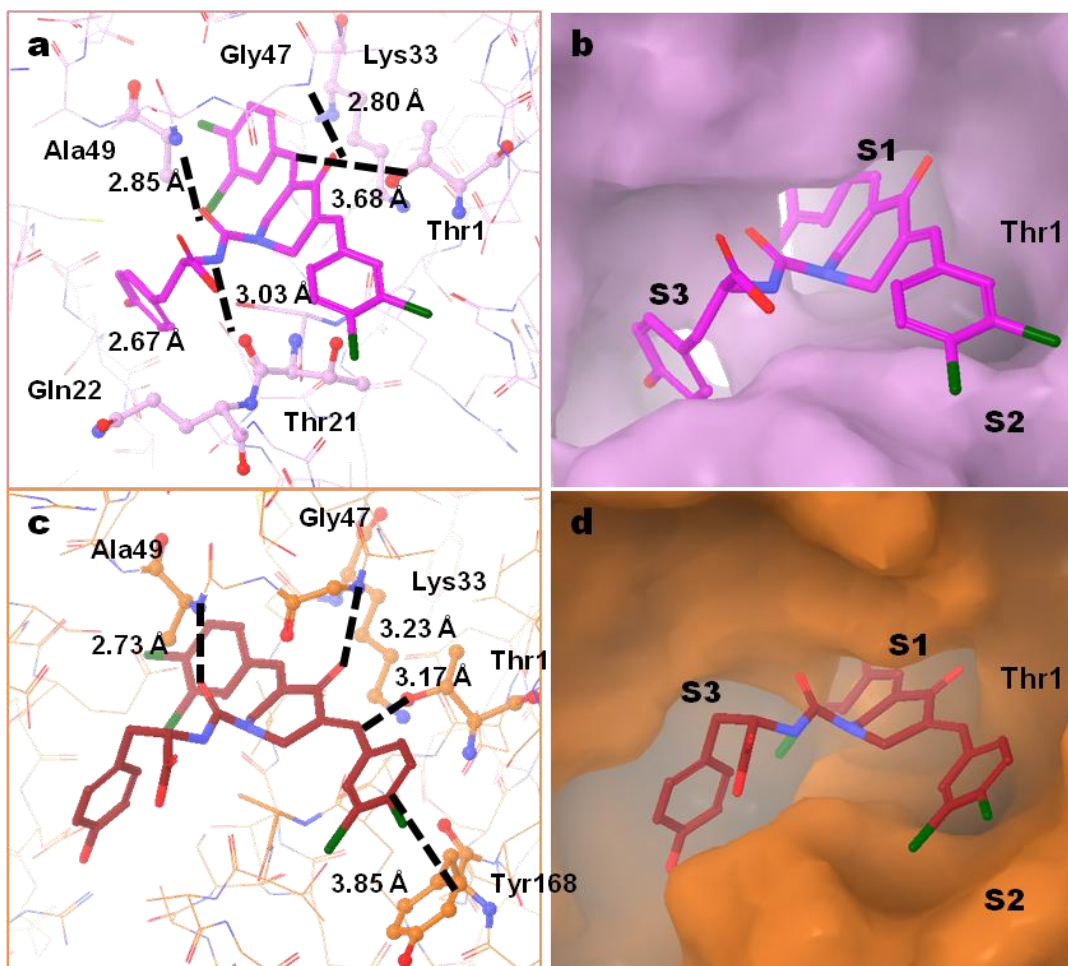


Figure 28. The predicted binding poses of RA-2 to $\beta 2$ (a), $\beta 5$ (c), and on the molecular surface of $\beta 2$ (b) and $\beta 5$ (d). The residues which form hydrogen bonds and hydrophobic interactions with RA-1 are shown with the distance (\AA) between heavy atoms.

Figure 29 displays the superposition of the crystal structures of other known inhibitors with different carbon color at $\beta 5$: RA-2 (pink), epoxomicin (green carbon), bortezomib (orange carbon). The position of the Michael acceptor of RA-2 is near the site where covalent bonds are formed between inhibitors and Thr1 (yellow circle). The distances between the β carbon and the electrophilic atoms of known inhibitors are 2.45

Å and 2.21 Å with epoxomicin, and bortezomib, respectively. The phenyl ring of Phe at the P4 position inserts deeply into the S3 pocket as does the P3 Ile and Asn side chains of epoxomicin and the P3-pyrazine of bortezomib.

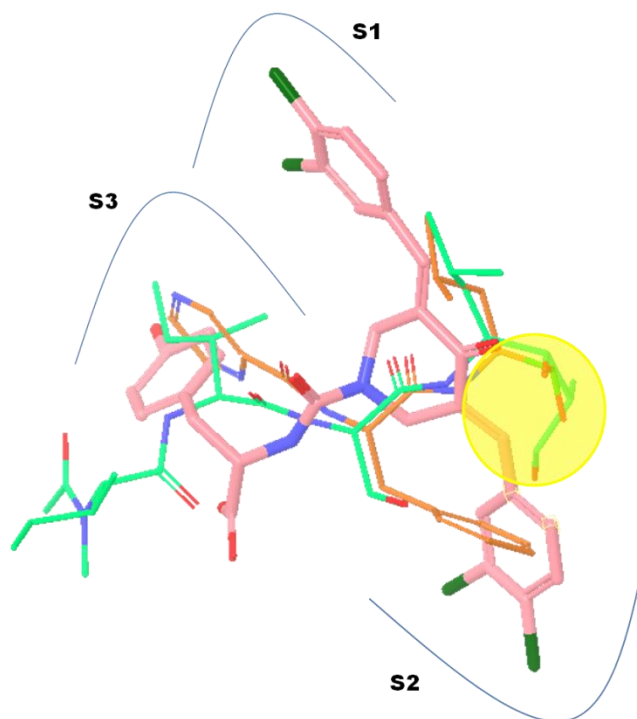


Figure 29. Superposition of RA-2 with the crystal structures of other known inhibitors at $\beta 5$: RA-1 (pink), epoxomicin (green carbon), bortezomib (orange carbon). The position for covalent-bond formation is marked with a yellow circle.

4.4.3 The predicted binding pose of RA-3 ($IC_{50} > 10$ ($\beta 1$ and $\beta 2$), and = 9.3 ($\beta 5$)) at all three catalytic subunits

Even though RA-3 induced partial inhibition with a selective inhibition toward CT-like activity, the binding modes at $\beta 2$ and $\beta 5$ are similar to those predicted for RA-1 and RA-2. However, at $\beta 1$ Leu at the P position resides in the S1 pocket instead of S3 pocket. The representative pose at $\beta 5$ is shown in **Figure 30**.

One β carbon is close to the O of Thr1 (3.28 Å), again setting the molecule up for possible nucleophilic attack by Thr1. One of di-chloro phenyl rings fits into the S1 pocket, and the other one points onto the S2 surface. Furthermore, the short π - π interactions at $\beta 5$ between the latter and Tyr168 (shortest distance between aromatic carbons is 3.85 Å) seems to contribute to the higher activity relative to $\beta 1$ and $\beta 2$.

This is consistent with a review of π - π stacking interactions taken from the protein database.⁵⁶ Georgia B. McGaughey et al. examined a representative set of high resolution X-ray crystal structures of nonhomologous proteins to determine the preferred positions and orientations of noncovalent interactions between the aromatic side chains of the amino acids Phe, Tyr, His, Trp. They found that pairs (dimers) of aromatic side chains preferentially align in an off-centered parallel orientation, which is 0.5-0.75 kcal/mol more stable than a T-shaped structure for Phe interactions and 1 kcal/mol more stable than a T-shaped structure for the full set of aromatic side chains.

In addition, closest contact distances (R_{clo}) between the respective C and N atoms were calculated. The numerical distribution of R_{clo} has a prominent minimum between 4.5 - 5 Å. Distances beyond this range are unproductive contributors to binding. The π - π

interactions at $\beta 5$ between Tyr168 and a ring of inhibitors resulting from docking should exist and their interactions could result in higher activity.

RA-3 forms hydrogen bonding interactions with the backbone N of Ala50 (2.91 Å) and Gly47 (3.28 Å) and the side chain of Asp114 (2.58 Å). However, it doesn't seem to occupy the S3 specificity pocket.

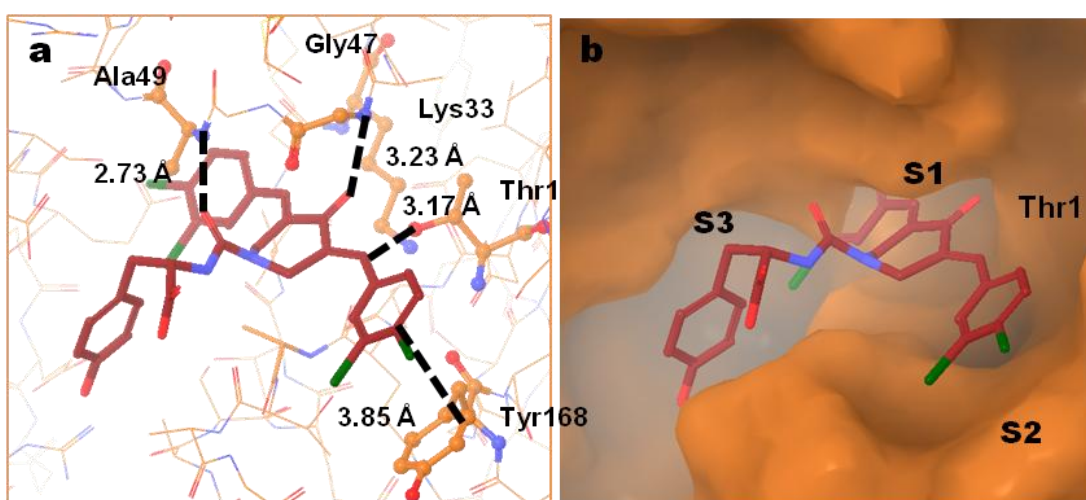


Figure 30. The predicted binding poses of RA-3 to $\beta 5$ (a), and residence on the molecular surface of $\beta 5$ (b). The residues which form hydrogen bonds and hydrophobic interactions with RA-3 are shown with distances (Å) between heavy atoms.

4.5 Summary of the docking results for curcumin and RA-1, 2, and 3

Non-covalent docking studies have been carried out to investigate curcumin and its analogs' mechanistic profiles as proteasome inhibitors at all three catalytic subunits where Thr1 O is regarded as the primary nucleophile. Since the X-ray structure of vinyl sulfone inhibitor, MB1, revealed a Thr-O-C β covalent attachment, we pay attention to the geometric relationship of Thr1 O with the β carbon of a Michael acceptor center for several promising nucleophiles in the proteasome active sites. Most binding poses place the β carbon of Michael acceptor within the 4.0 Å of those nucleophilicities, however, reasonable binding poses were generated only for nearby Thr1 O.

According to superposition of the predicted binding poses of various curcumin analogs with the X-ray structures of known inhibitors, the distances between β carbon and electrophilic atoms of known inhibitors are within 1.55 to 3.23 Å. In those conformations, several hydrogen bonds with the backbone O or N of Thr21, Gly47, Ala49, Ala50, Ile119, or Ser112 play an essential role in anchoring the inhibitor within the catalytic subunits. In addition, favorable hydrophobic contacts with the side chains of Gln22, Lys33, Ala49, Thr48, Leu115, or Ile116 assist the stabilization of the inhibitors. Especially interesting, one of the terminal aromatic rings makes a π - π interaction with Tyr168 on the S2 surface of β 5.

Curcumin and its analogs don't seem to be able to bind at β 1 with good affinity because of the small size of binding pockets. Moreover, the binding site analysis and the predicted binding poses imply a better binding mode at β 5 by filling the two specific pockets, S1 and S3, as well as interacting with Tyr168 at the S2 surface. Importantly, these results are inconsistent with the study carried out Milacic and Michael Groll groups

that curcumin and its analogs have the highest potency to CT-like activity, strongly demonstrating that the present docking studies follow experiments.

According to in vitro experiments with the proteasome purified from lymphoblastoid cell lines, while RA-1 shows activity at all three subunits, RA-2 does so at $\beta 2$ and $\beta 5$. In the case of RA-3, it operates at $\beta 5$, but is considerably weaker. This is because the Tyr OH of RA-2 experiences a steric clash with the S3 surface at $\beta 1$ due to the smaller size, pushing out other parts of the compound into the solvent exposed area as predicted. For RA-3 at $\beta 1$, Leu at the P position slides into the S1 pocket instead of the S3 cleft. Therefore, these compounds are unlikely to engage in a Michael reaction. In addition, the predicted binding pose of RA-3 doesn't seem to occupy S3 pockets in $\beta 2$ making a hydrogen bond with the backbone O of Trh21 instead, which might result in the inactivity at $\beta 2$.

Since the authors of the Bazzaro et al. paper did not report in vitro activity for RA-4-9, the inactivity against in vivo experiments might result from toxicity, solubility or other problems. But, if we presume there is no activity against the purified proteasome, the intensive active site analysis and docking studies above can explain their inactivity in the three subunits.

First of all, di-peptide analogs like RA6-9 are too large and long to bind any of the catalytic subunits; their molecular weights are more than 700. Thus, the chalcone moiety is placed in the solvent exposed area, and consequently, one cannot expect Michael reaction between them and the Thr1 O of each subunit.

In the case of RA-4, the side chain of Glu at the P position inserted into S1 of $\beta 1$ due to Arg45 could as predicted direct the selectivity towards PGPH like activity.

However, this event is unlikely to cause a Michael reaction as well. For RA-5, the side chain of Lys interacts with His114 and Asp120 in S3 of β 1 because this side chain is too long to go into the S1. In addition, no reasonable binding poses for a Michael reaction were generated. These two molecules generate several promising binding poses at β 2 and β 5 similar to other active compounds. Therefore, their inactivities against in vivo experiments might result from other molecular properties not binding issues. Now, we are repeating these experiments in a collaboration with Lawrence Boise of Winship Cancer Institute of Emory university to confirm the reported activities.

4.6 Extending the curcumin analog SAR

In addition to validation of activities for RA-1, 2, and 3, we have synthesized curcumin analogs with various side chains at the P position in the hope that they might show proteasome activity similar or better to that of RA-1. Docking studies with them have been carried out based on results of the work with curcumin and RA series described above to design better compounds for cervical cancer and blockade of the papilloma virus, i.e. for better than the low μ M results described.

One important feature of the catalytic subunits is that the S3 pocket between the three subunits differs significantly in terms of pocket size and residue composition. Especially, the interactions with residue 114 and 118 provide subunit selectivity. The P3 pyrazine of bortezomib, backbone NH of epoxomicin, and Asn of TMC-95A interact with the O of Asp114 at β 5 adjacent to β 6. However, curcumin and RA-1, 2, and 3 fail to deeply fill the S3 pockets and interact with either residue 114 or 118.

In order to improve ligand binding in the S3 pocket, fragment replacement has been carried out using BROOD of OpenEye.⁵⁷ BROOD uses not only the shape and attachment geometry of a query fragment but also imposes constraints for required interactions to identify bioisosteric replacements by searching a default database of over 6 million fragments covering 1-15 heavy atoms and 1-3 attachment points.

If the protein is also given as input, it will check close-contacts between protein and ligand and discard fragments which clash with the protein. In addition, this program calculates not only various physical properties including molecular weight, logP, Lipinski Rule of 5, etc. but also synthetic accessibility for every analog molecule. Therefore, BROOD can be very helpful to search interesting fragment replacements for compound optimization to improve activity as well as selectivity.

Figure 31-32 shows two examples (**compound A** and **B**) generated by BROOD followed by Glide docking at $\beta 5$. One β carbon in both compounds A and B is close to the O of Thr1 (3.05 Å and 3.14 Å, respectively), setting the molecule up for possible nucleophilic attack by Thr1. One of the di-chloro phenyl rings fits into the S1 pocket, and the other one points onto the S2 surface, making short π - π interactions at $\beta 5$ between the latter and Tyr168 (shortest distance between aromatic carbons is 4.16 Å for **compound A** and 4.18 Å for **compound B**).

These compounds form hydrogen bonding interactions with the backbone N of Ala49 (2.99 Å) and Gly47 (3.17 Å) and O of Thr21 (3.09 Å) for **compound A** and the backbone N of Ala49 (3.11 Å) and O of Thr21 (2.98 Å) for **compound B**. In addition, H of the positively charged N can establish a strong electrostatic interaction with Asp114 filling the S3 pocket. The structures superimposed onto the X-ray structures of

epoxomicin and bortezomib (**Figure 31-32. (c)**) show that the replaced fragment overlaps well with them.

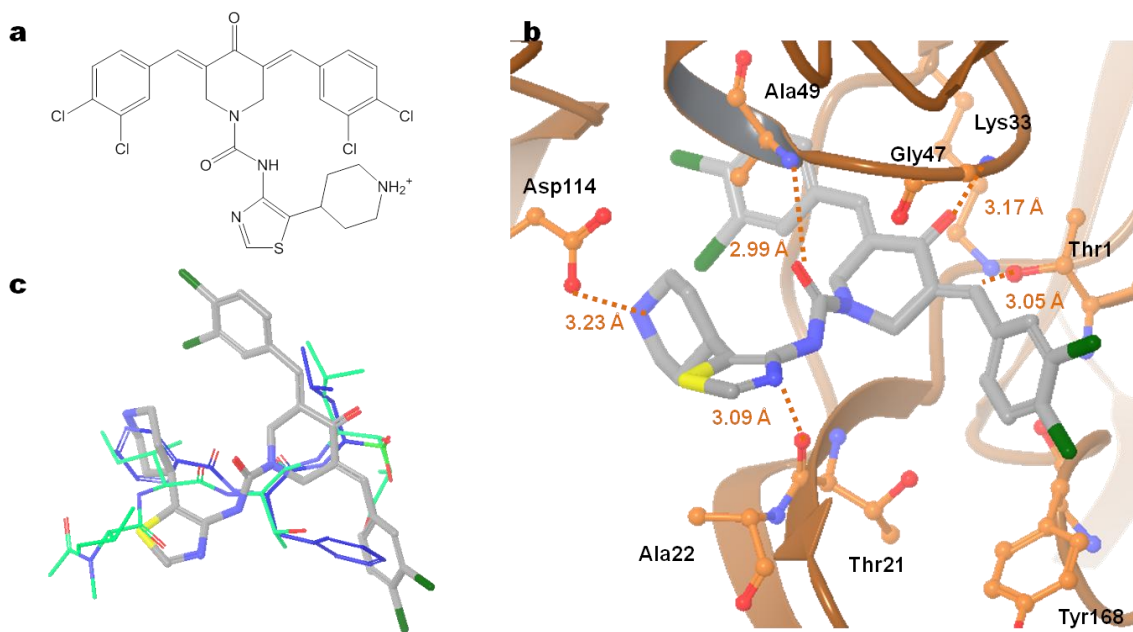


Figure 31. RA-1 derivative modified at the P position. (a) The 2D structure and (b) the predicted binding pose of **compound A** in $\beta 5$. The residues which form hydrogen bonds and hydrophobic interactions with **compound A** are shown with distances (\AA) between heavy atoms. (c) Superposition of **compound A** with the crystal structures of other known inhibitors at $\beta 5$: **compound A** (gray carbon), epoxomicin (green carbon), bortezomib (blue carbon).

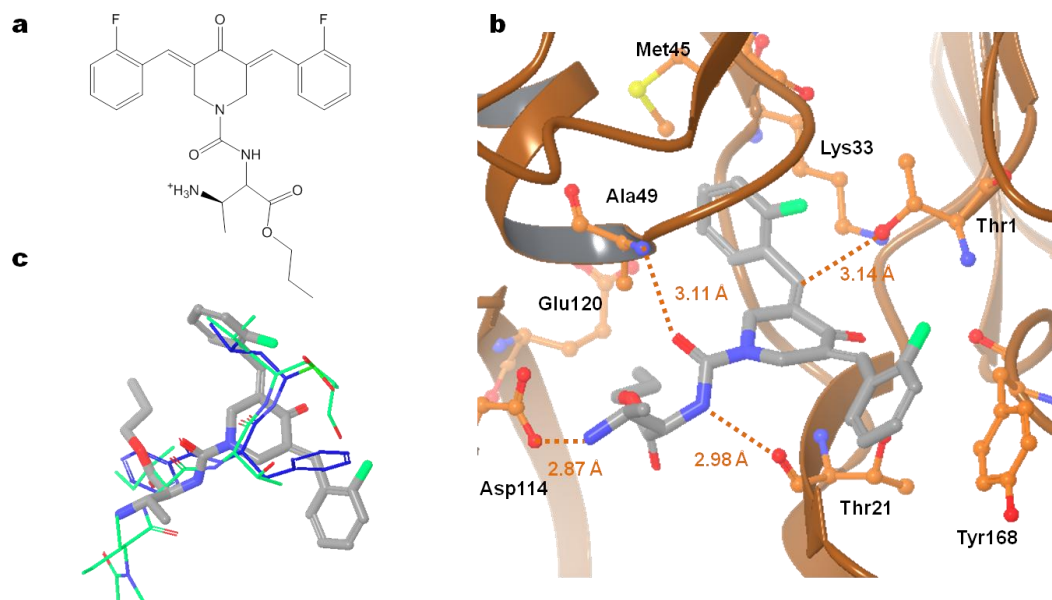


Figure 32. The Cjk1-10 derivative modified at the P position. (a) The 2D structure and (b) the predicted binding pose of **compound B** in $\beta 5$. The residues which form hydrogen bonds and hydrophobic interactions with **compound B** are shown with distances (\AA) between heavy atoms. (c) Superposition of **compound B** with the crystal structures of other known inhibitors at $\beta 5$: **compound B** (gray carbon), epoxomicin (green carbon), bortezomib (blue carbon).

Chapter 5. Conclusions

Even though both curcumin and chalcone-based derivatives (RA-1, 2, and 3) inhibit proteasomal activities *in vitro* as well as *in vivo*, their *in vitro* IC₅₀ values fall in the low μ M range. According to the docking results, one of the critical reasons for the latter is low binding affinities in the S3 pockets of each subunit leading to reduced occupancy, thus increasing the IC₅₀ values. Of particular note, the interactions with the side chain carbonyl of Asp114 of β 6 adjacent to β 5 play an important role for the high affinity of epoxomicin, bortezomib or the non-covalent inhibitor, TMC-95A, based on analysis of their crystal structures. However, curcumin and chalcone-based derivatives (RA-1, 2, and 3) lack those interactions, occupying deep S3 pockets instead. In section 4.5, a strategy for overcoming this problem with isosteric replacements in order to form strong interactions with Asp14 and simultaneously fill the S3 pocket is described.

In addition, as the assay results appear to demonstrate, the functionalized P position doesn't affect activities. This is because the various amino acid substitutions seem not to occupy the S1 pocket. Instead, it is assumed that the Michael reaction occurs during binding as in the case for the vinyl sulfone inhibitor, MB1. Moreover, various studies demonstrated that even with unfavorable P1 interactions, it may be possible for inhibitors to bind to catalytic subunits. Indeed, the P1 of many inhibitors are hydrophobic groups, such as the side chains of Leu, Ile or Phe.³⁵

Therefore, in order to improve the activity and specificity of curcuminoids, strong favorable interactions at S3 may facilitate the further improvement of highly selective proteasomal inhibitors. Section 4.5 elaborates on this point in detail.

Part III. Design of novel coactivator binding inhibitors (CBIs) for the estrogen receptor α : *Break the 1 μ M barrier?*

Chapter 1. Introduction¹

1.1 Estrogen receptor (ER)

The Estrogen Receptor (ER) is a member of the nuclear receptor (NR) superfamily of ligand-activated transcription factors associated with a diverse family of diseases including cardiovascular conditions, osteoporosis, and breast cancer. Transcriptional activation by ER α is mediated by two separate action functions (AFs), such as AF-1 in the N terminus and AF-2 in the C terminal ligand binding domain (LBD). AF-1 is regulated by growth factors acting on the MAP kinase pathway⁵⁸, whereas AF-2 is responsive to ligand binding⁵⁹. The ligand-dependent activation of transcription is mediated by interactions with coactivators: receptor agonists promote coactivator binding, whereas antagonists, block coactivator binding.

All biology, chemistry and modeling were carried out as a part of collaboration between Emory University and University of Illinois (Aiming Sun, Terry W. Moore, Jillian R. Gunther, Mi-Sun Kim, Eric Rhoden, Yuhong Du, Haiyan Fu, James P. Snyder, and John A. Katzenellenbogen).⁶⁰ My part involves all the modeling to support this work and is described below.

¹ A number of the sections of the following ER α discussion have been taken from the manuscript submitted to ChemMedChem that was ultimately published.⁶⁰

1.2 Blockers of ER agonist signaling

The traditional strategy of blocking ER agonist signaling has been achieved by using competing antagonists that bind to the ligand binding domain (LBD) of the receptor. However, most patients with hormone-responsive breast cancer eventually develop resistance to traditional anti-estrogens such as Tamoxifen. According to the various studies, the biological mechanisms of the intrinsic and acquired resistance to Tamoxifen may include loss of ER expression, mutations of the ER genes, reduced intra-tumoral concentrations of Tamoxifen, or increased agonistic metabolites rather than anti-estrogenic metabolites for ER activation⁶¹. Therefore, an alternative approach, the design of nonpeptidic small molecules that directly inhibit the interaction between agonist bound ER and coactivator proteins, has been undertaken.

1.3 Specific aims

The Emory high-throughput screening center as part of the MLSCN (Molecular Libraries Screening Center Network) identified two hits (**1** (Phenyl piperazine scaffold) and **2** (Benzothiazole) scaffold) by high-throughput screening (HTS). These novel coactivator binding inhibitors are shown to be active at low μM concentration by using a cell-based, luciferase reporter gene assay (developed by the John Katzenallenbogen group, our collaborators at the University of Illinois), which measures ER α transcriptional activity. The compounds were also demonstrated to directly block the interaction between ER α and the coactivator proteins by means of a time-resolved fluorescence resonance energy transfer (TR-FRET) assay. However, they fail to deliver IC₅₀ values below 1 μM in the reporter gene assay.

Thus, the specific aim of this study is to understand these observations and break the 1 μ M barrier. Since there are X-ray structures of ER α in complex with agonist, antagonist, and coactivators, docking was carried out to develop a description of ligand binding and understand structure-activity relationships as a preliminary study. In addition, MD simulations were performed to examine solvent-based entropic contributions to the free energy of ligand binding. Based on the results from the docking and MD simulations, several further molecular modeling approaches have been proposed to design improved coactivator binding inhibitors in order to break the 1 μ M barrier.

Chapter 2. Preliminary study

2.1 Structure of the ER α

Figure 1 shows the overall structure of the ER α ligand binding domain (**LBD**)-agonist (DES, diethylstilbestrol)-coactivator peptide (GRIP1 NR box II) complex. This protein monomer is a wedge-shaped molecule in which the agonist (yellow) and its binding cavity are within the interior of the LBD, while the peptide binds as a short α helix (orange) to a hydrophobic groove on the surface of the LBD (**Fig. 1 a**).

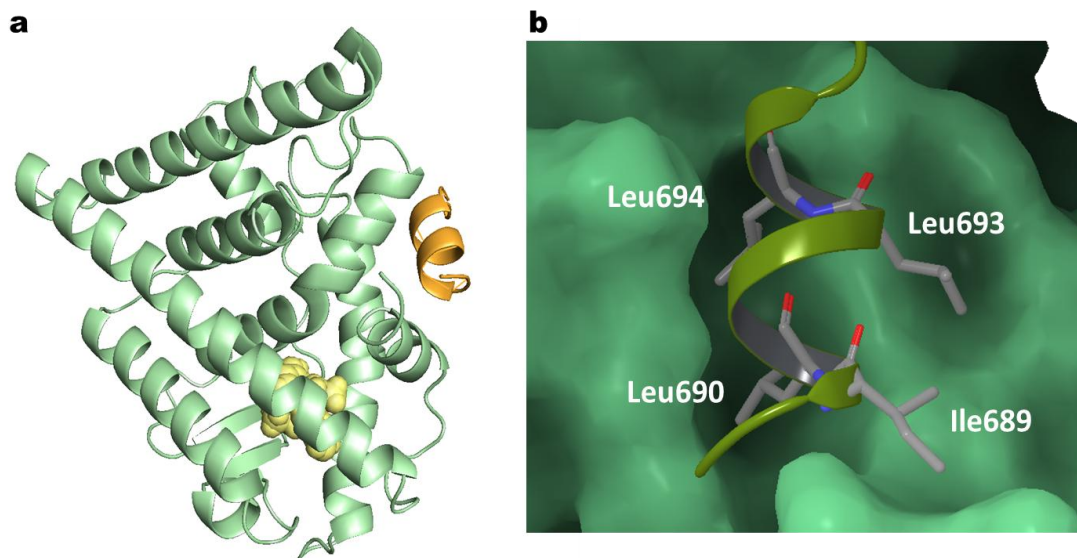


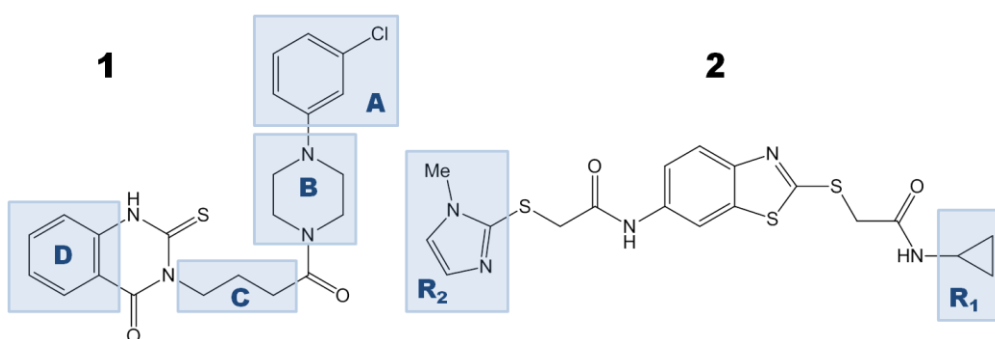
Figure 1 a) Overall crystal structure of ER α including agonist (DES, yellow) and coactivator proteins (GRIP1 NR box II, orange). b) Crystal structure of GRIP1 peptide on the coactivator recognition surface of the ER α

The coactivator protein is an α -helical structure with the conserved sequence of LXXLL, also known as an NR box, which forms the core of the short α helix that is recognized by a hydrophobic groove on the surface of the LBD (**Fig. 1 b**). This coactivator recognition surface is composed of a shallow hydrophobic groove which is highly nonpolar and capped on either end by two oppositely charged residues (charge clamp; Lys362 and Glu542). These two residues are positioned at opposite ends of the LBD surface to stabilize the conformation of the coactivator protein. Shown here is just a small part of the coactivator protein, the so called NR box containing the LXXLL amino acid sequence motif. Leu690 and Leu694 are entirely embedded in the hydrophobic groove, forming deep and strong hydrophobic interactions. Ile689 and Leu693 interact

with the receptor on one half of their surface, forming weaker but apparently significant surface-associated hydrophobic interactions.

2.2 High-throughput screening (HTS) and optimization

Two hits (**1** (Phenyl piperazine scaffold) and **2** (Benzothiazole scaffold)) was identified by high-throughput screening (HTS).



1 was modified by synthesis in four sectors: chlorophenyl **A** and piperazine **B** with a variety substituents including aromatic rings; linker **C** length altered; phenyl quinazoline **D** decorated with -OMe, -COOMe and -COOH groups. **2** was also modified in the two side rings, namely **R₁** and **R₂**. The following tables is part of the published paper mentioned above.⁶⁰

2.2.1 Modification of the chlorophenyl region A and linker region C.

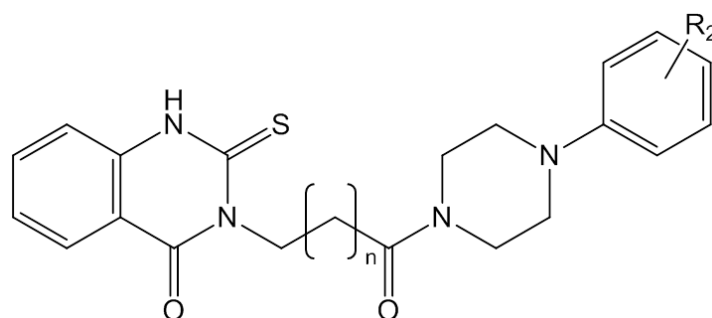


Table 1. Modification of the chlorophenyl region A and linker region C

Compd	R ₂	n	Luciferase activity (IC ₅₀ , μM)	Compd	R ₂	n	Luciferase activity (IC ₅₀ , μM)
Peptide	-	-		1n	2-CN	2	7.5 ^b
1	3-Cl	2	14.8 ^a	1o	3, 4-dichloro	4	4.2
1a	4-Cl	2	7.1	1p	4-OH	4	10.8 ^a
1b	3,4-dichloro	2	5.5	1q	4-CF ₃	4	4.4
1c	4-OMe	2	7.4	1r	4-CN	4	8.0
1d	4-OH	2	inactive	1s	2-CN	4	3.5
1e	2-OH	2	4.8	1t	3-CN	4	9.9
1f	3-OH	2	10.4	1u	3-NO ₂	4	6.6
1g	4-CF ₃	2	2.3	1v	3-Cl	4	5.2 ^a
1h	3-CF ₃	2	35 ^a	1w	4-Cl	4	9.3 ^a
1i	2-CF ₃	2	11.9	1x	3-CF ₃	4	25 ^a
1j	3-CF ₃ , 4-Cl	2	inactive	1y	3-Cl	6	4.2
1k	4- ⁱ Pr	2	26 ^a	1z	2-CN	6	2.7 ^c
1l	3-NO ₂	2	8.2 ^b	1aa	3, 4-di-Cl	6	inactive
1m	4-CN	2	inactive	1bb	4-CF ₃	0	inactive

^aCompound showed partial inhibition in reporter gene assay; ^bSome toxicity in β-galactosidase assay; ^cSignificant toxicity in β-galactosidase assay.

2.2.2 Modification of phenyl piperazine sectors A and B.

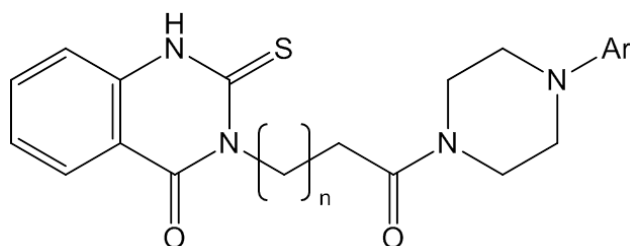


Table 2. Modification of phenyl piperazine sectors A and B

Compd	n	Ar	reporter gene (IC ₅₀ , μM)
9a	2		8.6
9b	2		inactive
9c	2		10.9
9d	4		6.6
9e	4		inactive
9f	4		7.0 ^a

2.2.3 Modification of phenyl quinazoline region D.

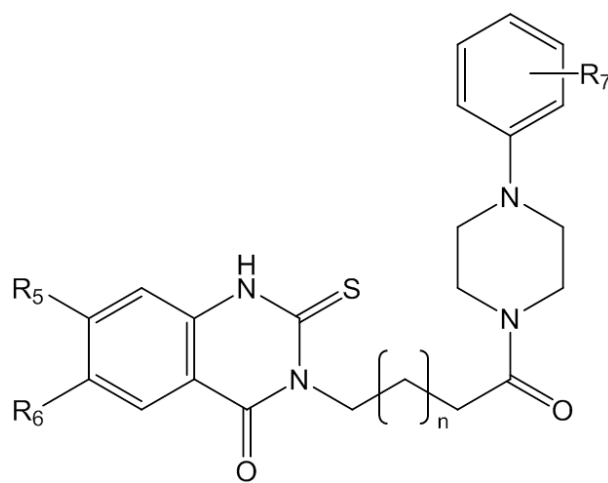


Table 3. Modification of phenyl quinazoline region D

Compd	R ₅	R ₆	R ₇	n	reporter gene (IC ₅₀ , μM)
11a	OMe	OMe	3-Cl	1	3.7 ^a
11b	COOMe	H	3-Cl	1	7.3
11c	OMe	OMe	4-CF ₃	1	21 ^a
11d	COOH	H	3-Cl	1	inactive
11e	OMe	OMe	4-Cl	1	inactive
11f	OMe	OMe	3, 4-di-Cl	1	22 ^a
11g	OMe	OMe	4-CF ₃	3	inactive

2.2.4 Modification of benzothiazole

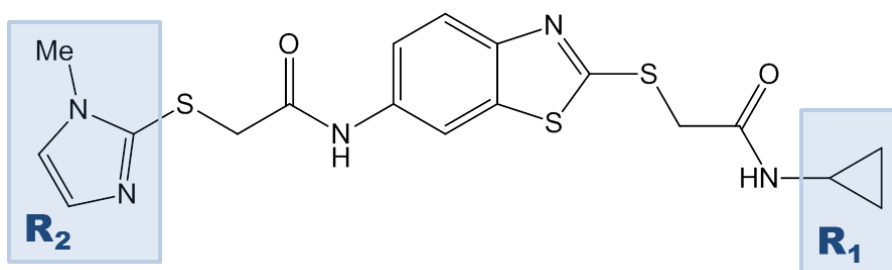


Table 4. Modification of benzothiazole

Compd	R ₁	R ₂	Luciferase activity (IC ₅₀ , μM)	Compd	R ₁	R ₂	Luciferase activity (IC ₅₀ , μM)
4	<i>c</i> -Pr	<i>N</i> -MeIm	20-50	4y	<i>i</i> -Bu	<i>p</i> - <i>t</i> -BuPh	>50
4a	<i>c</i> -Pr	Im	>50*	4z	<i>c</i> -Pr	<i>p</i> -OMePh	20-50
4b	<i>c</i> -Pr	<i>o</i> -CF ₃ Ph	6.3	4aa	<i>i</i> -Pr	<i>p</i> -OMePh	>50
4c	<i>c</i> -Pr	<i>m</i> -CF ₃ Ph	20-50	4bb	<i>i</i> -Bu	<i>p</i> -OMePh	>50
4d	<i>i</i> -Bu	2,5-diClPh	9.1	4cc	<i>i</i> -Pr	<i>p</i> -ClPh	20-50
4e	<i>c</i> -Pr	<i>o</i> -MePh	>50	4dd	<i>i</i> -Bu	<i>p</i> -ClPh	>50
4f	<i>i</i> -Bu	<i>o</i> -MePh	>50	4ee	<i>c</i> -Pr	<i>p</i> -CF ₃ Ph	6.0
4g	<i>i</i> -Bu	<i>o</i> - <i>i</i> -PrPh	>50	4ff	<i>i</i> -Pr	<i>p</i> -CF ₃ Ph	15
4h	<i>i</i> -Bu	<i>o</i> - <i>t</i> -BuPh	5.8	4gg	<i>i</i> -Bu	<i>p</i> -CF ₃ Ph	17
4i	<i>i</i> -Bu	<i>o</i> -ClPh	20-50	4hh	<i>c</i> -Pr	<i>p</i> -BrPh	20-50
4j	<i>i</i> -Bu	<i>m</i> -ClPh	20-50	4ii	<i>i</i> -Pr	<i>p</i> -BrPh	20-50
4k	<i>i</i> -Bu	<i>o</i> -CF ₃ Ph	20-50	4jj	<i>i</i> -Bu	<i>p</i> -BrPh	>50
4l	<i>i</i> -Bu	<i>m</i> -CF ₃ Ph	>50	4kk	<i>c</i> -Pr	2-Naph	8.7
4m	<i>t</i> -Bu	<i>o</i> -EtPh	20-50	4ll	<i>i</i> -Pr	2-Naph	10
4n	<i>i</i> -Pr	<i>o</i> -ClPh	20-50	4mm	<i>i</i> -Bu	2-Naph	12

4o	<i>i</i> -Pr	<i>m</i> -ClPh	6.3	4nn	<i>i</i> -Pr	3,4-diClPh	>50
4p	<i>i</i> -Pr	<i>o</i> -CF ₃ Ph	20-50	4oo	<i>i</i> -Bu	3,4-diClPh	20-50
4q	<i>i</i> -Pr	<i>m</i> -CF ₃ Ph	>50	4pp	<i>i</i> -Bu	2,4-diMePh	20-50
4r	<i>c</i> -Pr	<i>p</i> -MePh	20-50	4qq	<i>i</i> -Pr	2,4-diMePh	20-50
4s	<i>i</i> -Pr	<i>p</i> -MePh	>50	4rr	<i>i</i> -Pr	<i>p</i> -EtPh	20-50
4t	<i>i</i> -Bu	<i>p</i> -MePh	>50	4ss	<i>i</i> -Bu	<i>p</i> -EtPh	>50
4u	<i>c</i> -Pr	<i>p</i> -EtPh	20-50	4tt	<i>c</i> -Pr	<i>p</i> - <i>i</i> -PrPh	20-50
4v	<i>i</i> -Bu	<i>p</i> - <i>i</i> -PrPh	>50	4uu	<i>i</i> -Pr	<i>p</i> - <i>i</i> -PrPh	>50
4w	<i>c</i> -Pr	<i>p</i> - <i>t</i> -BuPh	20-50	4vv	<i>i</i> -Pr	2,5-diClPh	6.1
4x	<i>i</i> -Pr	<i>p</i> - <i>t</i> -BuPh	>50				

Chapter 3. Molecular modeling

3.1 Molecular modeling

To develop binding models of the ER α coactivator inhibitors described herein and elucidate initial structure activity relationships, we employed the crystal structure of the human estrogen receptor α (hER α) ligand binding domain (LBD) bound to the agonist diethylstilbestrol (DES) and a peptide derived from the NR box II region of the coactivator GRIP1 (2 Å resolution; pdb code 3ERD). The peptide was deleted and flexible docking of the ligands to the coactivator site was performed with the Induced Fit Docking module of the Schrödinger Suite (2008)⁶².

Since the quality of pose prediction depends strongly on reasonable starting structures, prior to ligand docking the protein LBD was first prepared in a form suitable for docking followed by subsequent MM-GBSA calculations and MD simulation with the

“Protein Preparation Wizard” in Maestro (v. 9.0.211, 2009). Thus, coactivator peptide, and tightly bound water molecules were deleted; bond orders, assigned; hydrogens, added; protein termini, capped with ACE (N-acetyl) and NMA (N-methyl amide); structures, fixed; and labeling, systematized.

In addition, since several residues with missing side chains were detected within and near loops, these were automatically added and conformations for them were predicted using Schrödinger’s “Prime side-chain prediction” module and the OPLS-2000 protein optimized all-atom force field.⁶³ The procedure follows.

For the prediction of the first residue, the side-chain rotamer library is searched to find the rotamer with the lowest predicted energy, while keeping all other side chains fixed. Then the next residue is considered, and so forth, until all residues have been treated.

Once the process is complete, the steps are repeated until no residues change rotamer states; i.e. rotamer convergence. Once this was achieved, structure optimization was run on all of the side-chain atoms while keeping backbone atoms fixed. Finally, restrained minimization of the protein was carried out by the IMPREF utility in “Protein Preparation Wizard” which performs protein refinement with OPLS-2000 until heavy atom positions deviate from the crystal structure with an RMSD of no more than 0.3 Å.

3.2 Induced fit docking.⁶²

Frequently, ligands are docked flexibly into the binding site of a rigid protein receptor. However, in many situations this yields misleading results as many proteins

experience side chain or backbone movements upon ligand binding. Small-molecule blockade of protein-protein interactions, as in the present case, is particularly susceptible to such effects. Therefore, although more time-consuming, we decided to incorporate both ligand and protein flexibility in all docking exercises in an effort to maximize accuracy.

As a result, binding site residues including Ile358, Lys362, Leu372, Glu380, Glu542, and Met543 experienced movements to accommodate docked blockers. In particular, Lys362 was displaced up to 4 Å towards solvent during the course of induced-fitting of various inhibitors.

3.3 Molecular volume and qikprop calculations.

In order to investigate the possibility that ERα coactivator inhibitors might bind to the agonist binding site or compete with antagonists, molecular volumes of a classical agonist, diethylstilbestrol (DES), and a traditional antiestrogen, 4-hydroxytamoxifen (HT), as well as **1g** were calculated with Schrödinger's QikProp facility⁶⁴ as applied to either the bound X-ray crystal structure or the most favored docked pose.

In addition, the volume of the antagonist binding pocket was calculated with CASTp (Computed Atlas of Surface Topography of proteins). In order to explore why CBIs show low IC₅₀ values in the reporter gene assay, the Caco-2 cell permeabilities of three key ligands, agonist DES and antagonist HT were predicted with QikProp: **1b**, 745; **1e**, 488; **4o**, 572; DES, 207 907; HT, 626 624 (< 25 poor, > 500 great).

3.4 MM-GBSA energy evaluation.

In order to identify the optimal binding poses generated by docking, MM-GBSA calculations^{65, 66} were performed within Prime following induced-fit docking. The method is known to be able to identify experimental protein-ligand complexes (i.e. the X-ray structure), over alternative poses by computing an estimate of the free energy of binding. The procedure is superior to GlideScore, the scoring function employed by Prime to rank poses during their generation. Consequently, it is prudently used to re-score the list of ligand/receptor docking geometries to select the energetically most favorable complex.⁶⁶

3.5 Desmond molecular dynamics simulations and determination of numbers of displaced waters.

MD simulations were performed using the Schrodinger Desmond module developed by D. E. Shaw Research.⁶⁷ The above-described crystal structure (3ERD) prepared for induced-fit docking by the “Protein Preparation Wizard” was chosen for a Desmond simulation as well.

The coactivator peptide was removed and the protein was solvated in an orthorhombic SPC water box with a outer boundary of 10 Å from the protein. The system was initially subjected to OPLS-2000 optimization to relax the system to the nearest local energy minimum. Trajectory data were recorded every 1.2 ps for a total of 1ns simulation at 300K and a pressure of 1.0 bar using an NPT ensemble class. In addition, seven sodium cations were added to neutralize the system, and the Ewald method was used to affect efficient and accurate long-range electrostatics.

The last structure of the 1 ns simulation was taken for analysis of the waters in the binding site since its energy is very similar to that of the average energy value over the nearly isoenergetic 1 ns time course. To guarantee that the ER α protein has equilibrated faithfully, the 1 ns structure was superposed with the starting complex to show that both the protein backbone and the diethylstilbestrol agonist are in essentially identical spatial locations.

After the MD treatment, the crystal structure of the coactivator peptide and the docking poses of **1b**, **1e** and **4o** were individually placed in the binding site water pool by superposing the respective protein complex backbones with that of the solvated and simulated apo-protein. Only waters 2 Å or less from any atoms of coactivator peptide, **1b**, **1e** and **4o** were considered to be ligand-overlapped and, therefore, candidates for extrusion from the binding site.

Chapter 4. Results

4.1 Possible competition at the ligand binding pocket.

The molecules identified by HTS and analogues in this study were pursued with the understanding that they compete with the co-activator proteins and peptides and do not bind at the agonist (estradiol)/antagonist (tamoxifen) binding pocket. The lead modification principle flows largely from characterization of the coactivator peptide as a pharmacophore anchored by a “charge clamp” superimposed on a family of hydrophobic residues residing between and below the cationic and anionic centers.

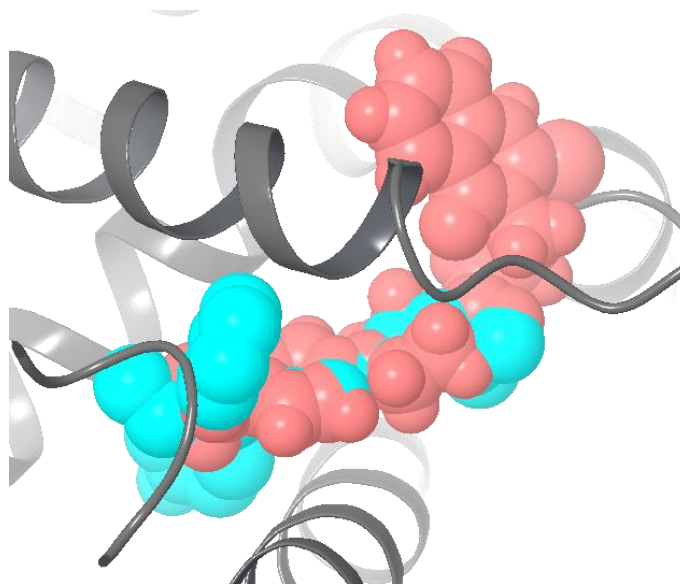


Figure 2. Alignment of **1g** (pink) and **4-hydroxytamoxifen** (blue) in 3ERT (hER α -HT complex) as determined by Glide docking. Much of **1g** is predicted to bind well outside the ligand binding site.

To initiate the ligand-protein docking studies, we were curious to know if the inhibitors under investigation could conceivably compete with the classical ligands. First, we examined the diethylstilbestrol agonist complex (PDB code 3ERD) in which helix-12 occludes the agonist binding pocket. The molecular volume of diethylstilbestrol (DES) is 947 Å³, while that of inhibitor **1g** is 1376 Å³. The volume differential is consistent with the inability to find a single pose for **1g** in the highly constrained pocket. The possibility that CBI's might bind competitively with ER α antagonists was also examined in the context of the X-ray structure in which 4-hydroxytamoxifen (HT, molecular volume 1335 Å³) occupies the ligand site (volume 1476 Å³) and H-12 has folded into the coactivator binding site (PDB code 3ERT).

Since HT and **1g** have very similar molecular volumes, in principle both can occupy the expansive pocket. Docking **1g** into the emptied HT site shows that the phenyl piperazine-quinazoline scaffold is able to penetrate the site, but the best binding pose is quite different from that of HT. Consequently, the molecule occupies the pocket only partially, while much of the structure resides in solvent (**Figure 2**).

In addition, MM-GBSA calculations were carried out to compare estimated binding free energies ($\Delta G_{\text{binding}} = E_{\text{complex}}(\text{minimized}) - E_{\text{ligand}}(\text{minimized}) - E_{\text{receptor}}$) for HT and **1g**. The latter complex as pictured in **Figure 2** is posited to be much less stable than that for HT ($\Delta\Delta G_{\text{binding}} \sim 17$ kcal/mol). The calculations are in complete accord with conclusions from the mechanistic studies described above, namely that the present classes of blockers are genuine CBIs that prevent coactivator peptides from complexing with the ER α receptor. The correlation between biology and modeling offers a measure of confidence that the CBI docking results reported below might be useful for guiding future design strategies.

4. 2 Phenyl piperazine scaffold.

All active compounds in **Table 1** with the exception of **1e** and **1f** bearing an OH hydrogen donor, exhibit similar docked poses. These show good alignment with the X-ray structure of the bound ER α -coactivator peptide and simultaneously form productive hydrophobic contacts and H-bonds with Gln375, Lys362, or Val368. In addition, as a result of the induced fit docking, each is accompanied by Glu380, Glu542, Met 543, Ile358 and/or Leu372 side chain movements in order to better match dipoles of the functional groups on the phenyl associated with the piperazine ring or hydrophobic

contacts with the benzene center of the quinazoline ring. It is noteworthy that the center of mass of **1b** does not fall along the axis of the peptide helix, but below it and deeper into the groove (**Figure 3b**). The docking has provided an intuitively reasonable spatial overlap by superposing the dichlorobenzene ring onto Leu690 and the piperazine ring onto Leu694; hydrophobic onto hydrophobic.

To test the possibility that docking might insert **1b** arbitrarily deep within the groove, the coactivator peptide from 3ERD was rigidly docked into the binding site. In none of the resulting poses do the Leu and Ile side chains penetrate deeper into the pocket than observed in the X-ray structure, indicating that the hydrophobic basin of the protein is the arbiter of hydrophobe location.

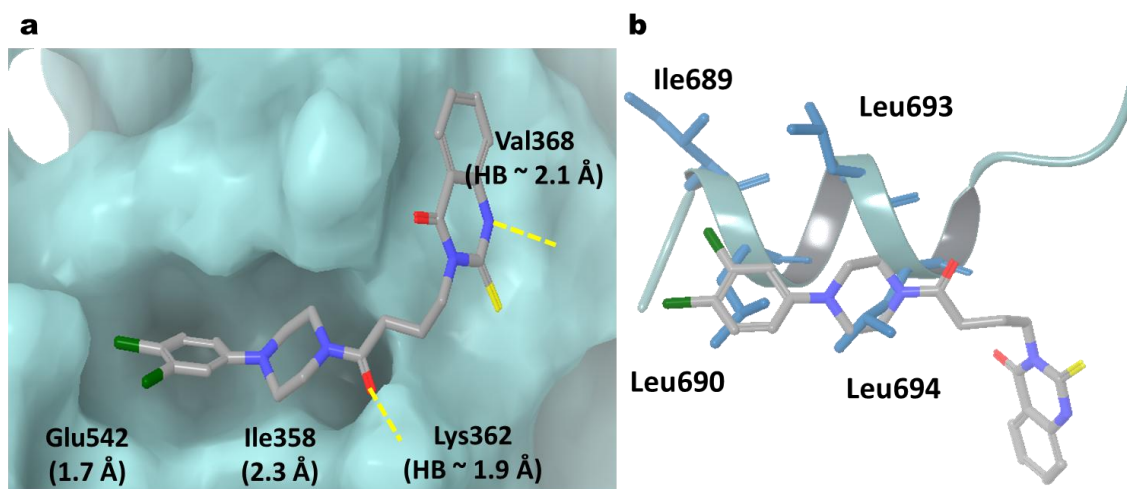


Figure 3. a) Induced fit docked pose for **1b** at the coactivator binding site of the ER α receptor (PDB X-ray structure code 3ERD). Displacements of Glu542 and Ile358 side chains resulting from docking and H-bond distances are indicated in Å); b) Alignment of **1b** (stick) and the coactivator peptide (blue ribbon; 3ERD). Hydrophobic residues Leu690 and Leu694 are matched by ligand hydrophobes.

Pronounced side chain relocation can also occur for Lys362, a cationic charge clamp residue. For example, H-bonding to Lys362 by the piperazine-associated carbonyl group in **1b** (3-carbon linker, $n = 2$) can contribute to modifying the locus of the side chain. **Figure 3** depicts the docked pose of **1b** in which the A-sector 3,4-dichloro phenyl piperazine moiety slips deeply into hydrophobic groove, while the D-sector phenyl quinazoline resides outside of it. During ligand fitting, Glu542 and Ile358 were relocated by 1.7 and 2.3 Å, respectively, to accommodate the two chlorines and the piperazine ring. In addition to the Lys362---O=C H-bond (1.9 Å), the quinazoline NH and the Val368 C=O interact similarly at 2.1 Å.

For certain active CBI analogs, such as **1q**, induced fit docking leads to displacements of Lys362 up to 4 Å into regions of solvent exposure in order to accommodate the longer five-carbon linker (**Table 1**, $n = 4$) as shown in **Figure 4a**. Moreover, not only have Glu380, Glu542, and Met543 been translated in order to provide space for the 4-CF₃, but Leu372 has also shifted 2.1 Å for compatibility with the quinazoline ring. When comparing **1q** with **1b** and its shorter three-carbon linker, there are three significant differences as depicted by **Figure 4b-c**. The latter illustrates that the longer linker of **1q** is embedded much more deeply in the hydrophobic groove than the shorter linker of **1b** (**Figure 4b**). This affects the positions of the phenyl and quinazoline rings. Thus, the 4-CF₃ substituted phenyl ring of **1q** shifts left (**Figure 4c**), while the quinazoline ring rotates so that the ligand's carbonyl forms an H-bond with Gln375 rather than with Val368 as for **1b** (**Figure 4a / Figure 4c** right). All in all, the compounds in this series appear to exhibit quite similar binding poses by adopting compensating local adjustments resulting from linker length variation.

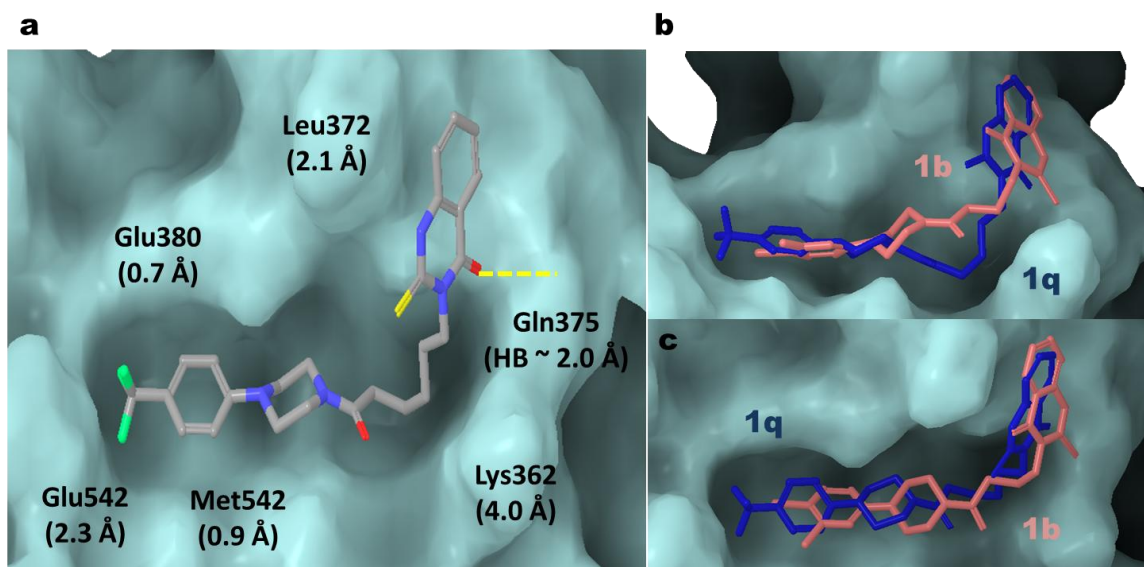


Figure 4. a) Induced fit docked pose for **1q** at the coactivator binding site of the ER α receptor (PDB X-ray structure code 3ERD). Displacements of Lys362, Leu372, Glu380, Glu542 and Met543 side chains relative to 3ERD and H-bonds are indicated in Å; b) and c) Relative alignments of **1b** and **1q**.

Other short/long linker analog pairs with similar activities presumably experience the same phenomenon (**1a** (7.1)/**1w** (9.3^a), **1b** (5.5)/**1o** (4.2), **1h** (35^a)/**1x** (25^a), **1l** (8.2^b)/**1u** (6.6), **1n** (7.5^b)/**1s** (3.5)) (**Table 1** IC₅₀ values and footnotes in parentheses). Some derivatives with 5- and 6-carbon linkers exhibiting superior activity relative to their 3-carbon linker counterparts (e.g. **1** (14.8^a)/**1v** (5.2^a), **1** (14.8^a)/**1y** (4.2), **1m** (inactive)/**1r** (8.0), **1n** (7.5^b)/**1z** (2.7^c)) appear to experience a particular advantage in this respect. The last two comments need to be tempered with the knowledge that some members of the pairs either show only partial inhibition in the reporter gene assay or significant toxicity against the target cells (cf. footnotes to **Table 1**)

Surprisingly, introduction of an appropriately placed hydrogen donor group (2-OH, **1e**, Table 1) resulted in a rather different U-shaped binding mode (**Figure 5**). In this case, the orientation of the ligand has been reversed, the terminal rings effectively switching places relative to the pose pictured in **Figure 3**. Thus, the quinazoline benzene unit mimics the Leu690 coactivator peptide side chain, while the piperazine ring and the head of its attached phenol ring occupy the corresponding Leu694 subsite (**Figure 5b**). The hydrophobic edge of the phenol is posited to reside rather deep in the groove in a hydrophobic subpocket formed by Ile358, Ala361, Phe367, Val368, Leu379 and the methylene chain of Lys362. Relative to 3ERD, the pocket has been reshaped by residue movements of 1-2.4 Å and is accompanied by (0.6-0.9 Å) shifts of Glu380 (0.9), Glu542 (0.8) and Met543 (0.6) as well to accommodate the quinazoline ring. A significant outcome is that the OH of the ligand's phenol ring is predicted to engage in hydrogen bonds with both Gln375 (1.7 Å) and Lys362 (2.1 Å), the latter having been translated by 4.2 Å during docking (**Figure 5a**). Moreover, the NH of the quinazoline ring also forms an H-bond with Glu542 (1.8 Å). These non-covalent interactions with the charge clamp residues Lys362 and Gln542 serve to anchor **1e** in the CBI pocket and to provide excellent alignment with the locus of the coactivator peptide (**Figure 5b**).

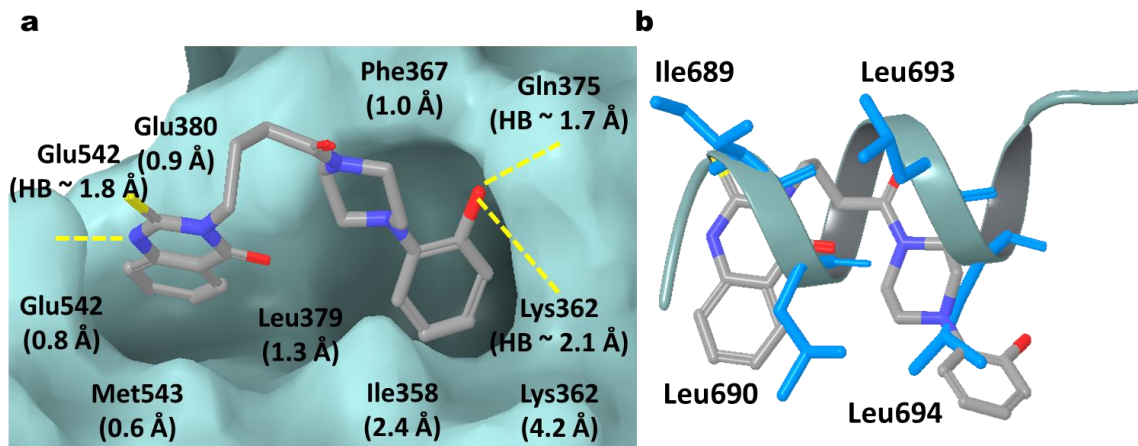


Figure 5. (a) Docked pose of **1e** illustrating the three ligand-protein hydrogen bonds and side chain movement upon docking in Å; (b) Alignment of **1e** (stick) and coactivator peptide (purple ribbon).

When the sector-A phenyl ring in series **1** was replaced with other aromatic moieties, compounds **9a**, **9c** and **9d** delivered similar activities. Consistently, the compounds revealed similar poses as well as good hydrophobic contacts and hydrogen bonding interactions with Gln375, Val 368 and Lys 362. However, due to linker length variation, small differences are evident. That is, the quinazoline ring rotated in **9a** and **9c**, while it shifted somewhat to the right in **Figure 3** and away from the hydrophobic groove in the case of **9d**.

By contrast, when the D-sector quinazoline ring was decorated with various functional groups, only the methyl ester **11b** was active (**Figure 5, Table 3**). As depicted in **Figure 3**, the D-sector rings are suggested to reside outside the coactivator peptide groove on an open surface area of the protein exposed to solvent. This is consistent with the observation that increased hydrophobicity on the quinazoline ring arising from

multiple methoxy substitution (**11a**, **11c**, **11e-g**) confers no activity advantage. Surprisingly, the carboxylic acid **11d** also proves to be inactive. In the context of ligand-protein interactions, however, docking suggests that the COOH resides well into solvent pulling the hydrophobically stabilized quinazoline ring of **Figure 3** into the solvent pool and away from the binding pocket. All in all, the docked inactive compounds in this series are suggested to enjoy fewer hydrophobic contacts and hydrogen bonding interactions. Moreover, many of them were also predicted to complex with the ER α receptor by entirely different binding modes upon induced-fit docking. If the pose of **Figure 3** represents the best that can be achieved in this series, the alternative complexes imply a loss in potency consistent with observation.

4.3 Benzothiazole scaffold.

Compound **4o**, one of the more active members of this series, provides a satisfying binding CBI model as illustrated by **Figure 6**. The structure not only fits well in the hydrophobic groove, but also forms good hydrogen bonding interactions with the charge clamp residues Lys362, Glu380 and Gln375 (2.0 Å). Like inhibitors **1b** and **1e** (**Figures 3** and **5**, respectively), this structure likewise offers hydrophobic features that appear to serve as Leu690 and Leu694 side chain surrogates; i.e. CH₂S, one edge of the thiazole ring and the centrally placed benzene ring.

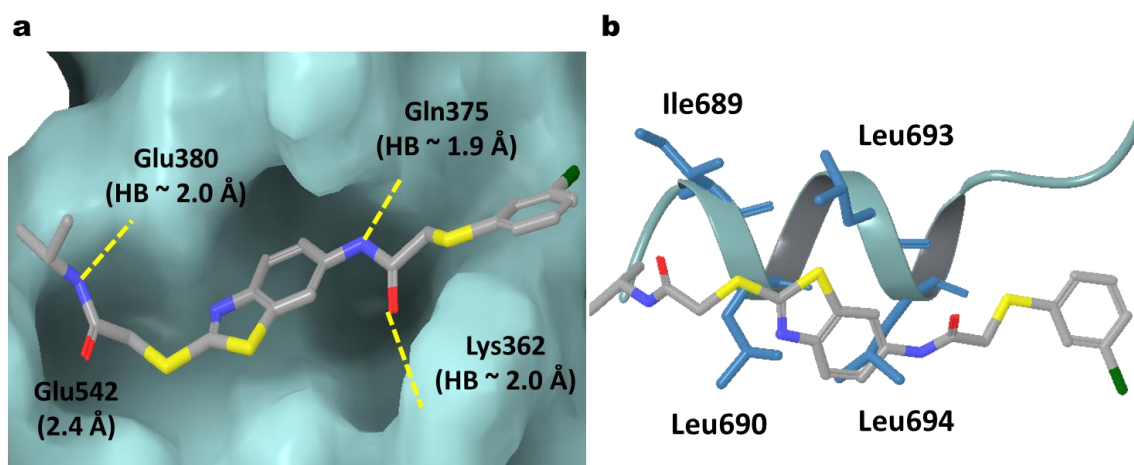


Figure 6. a) Docked pose of **4o** showing three hydrogen bond anchors and side chain movement upon docking in Å; b) Alignment of **4o** (stick) and coactivator peptide (purple ribbon).

The model accommodates a number of the SAR trends discussed above. For example, activities appear to be relatively unperturbed by replacement of the terminal *i*-Pr group in **4o** with other hydrophobes. **Figure 6** illustrates the deep apolar pocket in which the NHR₁ is sited. At the other end of the molecule, the SR₂ moiety is exposed to both a rather hydrophobic protein surface and aqueous solvent, suggesting that amphiphilic substitution might best serve binding. Indeed, the CF₃ group appears to be a superior substituent for SAR, the CF₃ group tolerant of both polar and nonpolar environments.⁶⁸ Replacement of fluorinated methyl with CH₃ or OMe depletes activity seriously in accord with water layering. One puzzling observation is the complete loss of activity when methyl is removed from the imidazole in **4** to give **4a**. In the context of the model of **Figure 6**, this might have its origin in the observed loss of efficacy for carboxylic acid **17** relative to its methyl ester. Both the quinazoline COOH and imidazole NH are exposed to

solvent (**Figures 3b** and **6b**, respectively). Just as we speculated that **17** may be extracted into solvent and out of the binding cleft, such a phenomenon may operate for **4a** as well.

Chapter 5. What the current CBIs offer and what they lack.

Compounds **1b**, **1e** and **4o** populate the coactivator peptide binding groove with differential but productive hydrogen bonding and favorable hydrophobic contacts. According to induced-fit docking, these structural similarities are achieved by a reasonable spatial match of the organic structures and the part of the coactivator peptide that binds deepest in the groove. In spite of these common and favorable features, all three compounds derived by different strategies fail to deliver IC₅₀ values that fall to 1 μM or below in the reporter gene assay. This could be the result of poor penetration of the cellular membrane present in the latter, although an estimate of Caco-2 membrane permeabilities suggests that the values for **1b**, **1e** and **4o** do not differ significantly from those for ERα ligands DES and HT (See Methods, Molecular Modeling).

More likely, the blockers are simply too small to compete effectively with the α-helical coactivator peptide. The docked poses of the synthetic inhibitors reveal an important characteristic which speaks to this point. Assuming that compounds **1b** (**Figure 3b**), **1e** (**Figure 5b**) and **4o** (**Figure 6b**) are representative of their classes, they fill important parts of the groove particularly those occupied by the Leu690 and Leu694 side chains of the coactivator peptide. However, none of the present analogs present structure to the protein shelf on which the Ile689 and Leu693 hydrophobes reside. Given the importance of entropic contributions to the free energy of binding, the focus on groove-

only inhibitors may be misplaced. The next section takes up this issue in terms of water displacement resulting from inhibitor complexation at the CBI site.

5.1 Entropic effects originating from water displacement.

Many studies have shown that ligand-protein binding at membrane bound receptors is frequently entropy-driven,⁶⁹ and that the source of the large $-T\Delta S$ contribution to ΔG arises by the release of water molecules from the binding pocket during the ligand complexation event.⁷⁰ For very tightly bound waters, the contribution can be as much as 2 kcal/mol⁷¹ with lesser values for bound waters with a greater degree of freedom. With the goal of developing a relative semi-quantitative estimate of the importance of this phenomenon in the present case, the coactivator-bound X-ray structure (PDB code 3ERD) was relieved of the coactivator peptide, surrounded by a box containing 8742 SPC waters 10 Å from the ER α protein and subjected to 1 ns molecular dynamics at 300 °C with the Desmond protocol (see Experimental Section). The X-ray structure of the coactivator peptide and the docked structures of **1b**, **1e**, and **4o** were separately superposed on the solvated binding site and, for each structure, all waters were deleted except those that overlapped the ligand severely or displayed oxygen atom-ligand atom distances below the sum of the van der Waals radii.⁷² The superpositions are depicted in **Figure 7**.

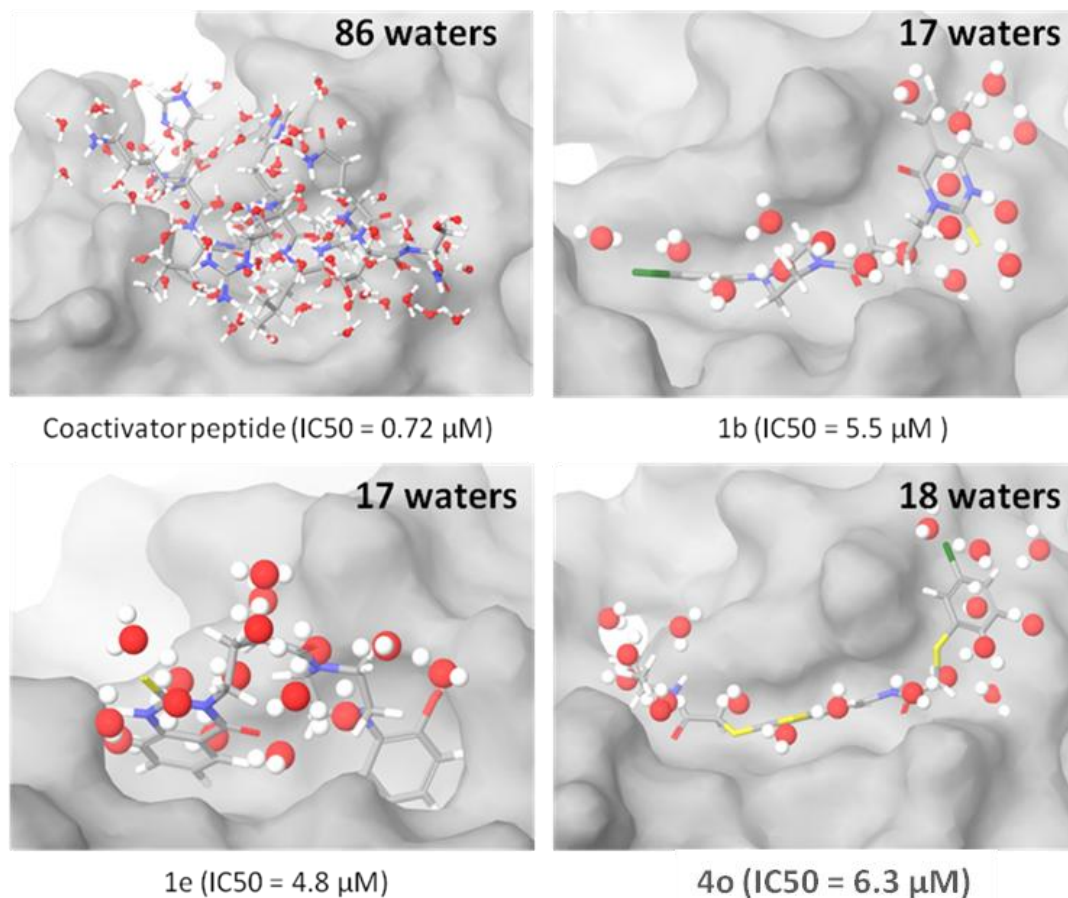


Figure 7. Number of overlapped waters within 2 Å of coactivator peptide and the three ligands **1b**, **1e** and **4o**. Reporter gene assay IC₅₀ values in parentheses.

Very clearly the peptide displaces **5** fold more water molecules than the non-peptidic system, **13** of which line the shelf occupied by the side chains of Ile689 and Leu693. If a conservative but average 0.2 kcal/mol were assigned to each displaced water at ambient temperatures, the peptide then contributes about 14 kcal/mol more binding free energy from $-T\Delta S$ at the ER α receptor than **1b**, **1e** and **4o**. It is equally significant that Ile689 and Leu693, which line the shelf adjacent to the coactivator binding groove,

displace as many water molecules (total of **13**) as the same three micromolar active inhibitors.

Chapter 6. Conclusion

The development of compounds that can block the interaction between the estrogen-activated ER and important coactivator peptides could provide unique pharmacological tools for interrupting the signal transduction cascade by which this transcription factor regulates gene activity and might provide a lead for novel therapeutic agents. Nevertheless, inhibition of protein-protein interaction with small molecules remains a significant challenge.^{73,74} Here, we described the synthetic and cell-based assay follow up of a selection of hits that came from a high-throughput screening effort to search for small molecule coactivator binding inhibitors (CBI) in a large compound library. The screening protocol employed a TR-FRET assay for inhibition of the interaction of ER α with the important coactivator SRC1 that we have described previously.

Compound **1** was identified as the most promising hit, and optimization through analog synthesis and further biological evaluation in a cell-based reporter gene assay yielded several compounds that were active in the low micromolar range. Other active analogs prepared were derived from a second hit, compound **4**. The mechanism of action of the potential CBIs were further examined to verify the site of action: (a) The inhibitory activity of representative compounds in the reporter gene assay was shown to be non-surmountable because the IC₅₀ values were the same in the presence of either 10⁻⁷ M or 10⁻⁹ M estradiol. This result indirectly confirms that activity results from direct displacement of coactivator, not by competition with estradiol at the ligand binding site. (b) By contrast,

in the TR-FRET assay, the IC_{50} value of representative compounds was affected by increasing the concentration of the SRC component. This outcome further confirms that the compounds are competing directly with SRC for binding to the ER. Thus, the new ER blockers fulfill the mechanistic criteria for coactivator binding inhibitors: direct competition with SRC binding and no competition with estradiol binding.

The most distinctive compounds identified are **1b** and **1g** with 3, 4-dichloro- and a *para*-CF₃ groups on the phenyl ring in sector A, respectively. Both compounds exhibit low micromolar potency in the cell-based reporter gene assay (IC_{50} around 6 μ M). The structures of these two substances and that of **4o** were subjected to extensive induced fit docking experiments. Depicted in **Figures 3-6**, the resulting binding models are characterized by protein side chain movements tailored to each ligand, a situation is parallel to that suggested for side chain rearrangements involving different ligands that perturb GPCRs.⁷⁵ The models not only provide insights into a variety of aspects of the evolving CBI SAR, but illuminate that a key aspect of SRC blockade is mimicry of two of the key hydrophobic leucine side chains (Leu690 and Leu694). To the extent that water release is a contributor to the free energy of binding, it is clear from **Figure 7** that the coactivator peptide is far more effective, and that the remaining two shelf-oriented apolar side chains displacing 13 water molecules (Ile 689 and Leu693) have no counterparts in the weakly active ligands. These results suggest that the next generation of small molecule CBIs should span more of the peptide space, particularly on the shelf adjacent to the deep binding groove bordered by helices H-3, H-4, H-5 and H-12. This might well come at the cost of inhibitor molecular weights beyond the Lipinski ideal of 500,⁷⁶ but if other molecular properties are satisfactory, the potency gain could certainly compensate.

References

- ¹ Kenneth R. Cooke et al. (2001). LPS antagonism reduces graft-versus-host disease and preserves graft-versus-leukemia activity after experimental bone marrow transplantation. *The Journal of Clinical Investigation*. 107:1581-1589
- ² Cao, X et al. (2007). Granzyme B and perforin are important for regulatory T cell-mediated suppression of tumor clearance. *Immunity*. 27: 635-646
- ³ Cao, X et al. (2007). Granzyme B and perforin are important for regulatory T cell-mediated suppression of tumor clearance. *Immunity*. 27: 635-646
- ⁴ D. Masson et al. (1987). *Cell*. 49: 679-685
- ⁵ Chih-Min Kam et al. (2000). Granzymes (lymphocyte serine proteases): characterization with natural and synthetic substrates and inhibitors. *Biochimica et Biophysica Acta*. 1477: 307-323
- ⁶ J.A. Prendergast et al. (1992). *The Journal of Biological Chemistry*. 267: 5090-5095
- ⁷ Jennifer Rotonda et al. (2001). The 3D structure of human granzyme B compared to caspase-3, key mediators of cell death with cleavage specificity for aspartic acid in P1. *Chemistry & Biology*. 8:357-368
- ⁸ Sandra M. Waugh et al. (2000). The structure of the pro-apoptotic protease granzyme B reveals the molecular determinants of its specificity. *Natural Structural Biology*. 7: 762-765
- ⁹ Philip E. J. Sanderson. (1999). Small, noncovalent serine protease inhibitors. *Medicinal Research Reviews*. 19: 179-197

-
- ¹⁰ Anyagi T. et al. (1969). *The Journal of Antibiotics*. 22:558
- ¹¹ Thompson RC. (1973). *Biochemistry*. 12:47; Westerik JO. et al. (1972). *The Journal of Biological Chemistry*. 247: 8195; Bajusz S. et al. (1978). *International Journal of Peptide & Protein Research*. 12:217; Bajusz S. et al. (1990). *Journal of Medicinal Chemistry*. 33:1729; Jackson CV et al. (1996). *Clinical and Applied Thrombosis/Haemostasis*. 2:258
- ¹² Das J. et al. (1995). *Bioorganic and Medicinal Chemistry*. 3:999; Edwards PD et al. (1994). *Medicinal Research Reviews*. 14:127; Kettner CA et al. (1984). *The Journal of Biological Chemistry*. 259:15106; Kettner CA et al. (1990). *The Journal of Biological Chemistry*. 265:18289
- ¹³ Edwards PD et al. (1994). Synthetic inhibitors of Elastase. *Medicinal Research Reviews*. 14:127
- ¹⁴ http://en.wikibooks.org/wiki/Structural_Biochemistry/Enzyme/Reversible_Inhibitors
- ¹⁵ <http://www.elmhurst.edu/~chm/vchembook/573inhibit.html>
- ¹⁶ [http://en.wikipedia.org/wiki/Denaturation_\(biochemistry\)](http://en.wikipedia.org/wiki/Denaturation_(biochemistry))
- ¹⁷ Eva Estebanez-Perpina et al. (2000). Crystal structure of the caspase activator human Granzyme B, a proteinase highly specific for an Asp-P1 residue. *Biological chemistry*. 381: 1203-1214
- ¹⁸ Jennifer Rotonda et al. (2001). The 3D structure of human Granzyme B compared to caspase-3, key mediators of cell death with cleavage specificity for Asp in P1. *Chemistry & Biology*. 8: 357-368
- ¹⁹ Hajduk, P. J. et al. (2005). Druggability indices for protein targets derived from NMR-based screening data. *Journal of Medicinal Chemistry*. 48: 2518-2525

-
- ²⁰ DeLano, W. L. et al (2002). Unraveling hot spots in binding interfaces: progress and challenges. *Current Opinion in Structural Biology*. 12: 14-20
- ²¹ Ryan Brenke et al. (2009). Fragment-based identification of druggable ‘hot spots’ of proteins using Fourier domain correlation techniques. *Bioinformatics*. 25: 621-627
- ²² Carla Mattos et al. (1994). Analogous inhibitors of elastase do not always bind analogously. *Nature*. 1: 55-58
- ²³ http://www.maybridge.com/portal/alias__Rainbow/lang__en/tabID__146/DesktopDefault.aspx
- ²⁴ <http://www.actimol.com/protease-inhibitors-actitarg-p-library.html>
- ²⁵ <http://www.actimol.com/protease-inhibitors-actitarg-p-library.html>
- ²⁶ Robert V. Talanian et al. (2000). Caspases as targets for anti-inflammatory and anti-apoptotic drug discovery. *Journal of Medicinal Chemistry*. 43: 3351-3371
- ²⁷ Yunyi Wei et al. (2000). The structure of caspases-1, -3, -7 and -8 reveal the basis for substrate and inhibitor selectivity. *Chem. Biol*. 7: 423-432
- ²⁸ Violeta I/ Pérez-Nueno et al. (2008). Comparison of ligand-based and receptor-based virtual screening of HIV entry inhibitors for the CXCR4 and CCR5 receptors using 3D ligand shape matching and ligand-receptor docking. *J. Chem. Inf. Model*. 48: 509-533
- ²⁹ <http://www.eyesopen.com/rocs>
- ³⁰ Christopher A. Willoughby et al. (2002). Discovery of potent, selective human granzyme B inhibitors that inhibit CTL mediated apoptosis. *Bioorg. Med. Chem. Lett*. 12: 2197-2200
- ³¹ Dennis A. Dougherty. (2007). Cation- π interactions involving aromatic amino acids. *J. Nutr*. 137: 1504S – 1508S

-
- ³² QikProp, version 3.4, Schrödinger, LLC, New York, NY, 2011
- ³³ Sachiko Tsukamoto and Hideyoshi Yokosawa. (2009). Targeting the proteasome pathway. *Expert Opinion*. 13, 5:605-621
- ³⁴ <http://en.wikipedia.org/wiki/Ubiquitin>
- ³⁵ Alexei F. Kisselev and Alfred L. Goldberg. (2001). Proteasome inhibitors: from research tools to drug candidates. *Chemistry & Biology*. 8: 739-758
- ³⁶ Ljudmila Borissenko and Michael Groll. (2007). 20S proteasome and its inhibitors: crystallographic knowledge for drug development. *Chem. Rev.* 107: 687-717
- ³⁷ E. Genin et al. (2010). Proteasome inhibitors: recent advances and new perspectives in medicinal chemistry. *Curr. Top. Med. Chem.* 10: 232-256
- ³⁸ E. J. Corey and Wei-Dong Z. Li. (1999). Total synthesis and biological activity of lactacystin, Omuralide and analogs. *Chem. Pharm. Bull. (Tokyo)*. 47, 1:1-10
- ³⁹ Michael Groll et al. (1997). Structure of 20S proteasome from yeast 2.4 Å resolution. *Nature*. 386: 463
- ⁴⁰ Venkat R. Macherla et al. (2005). Structure-activity relationship studies of salinosporamide A (NPI-0052), a novel marine derived proteasome inhibitor. *J. Med. Chem.* 48: 3684-3687
- ⁴¹ Michael Groll et al. (2006). Inhibitor-binding mode of homobelactosin C to proteasomes: New insights into class I MHC ligand generation. *PNAS*. 103: 4576-4579
- ⁴² Michael Groll et al. (2000). Crystal structure of epoxomicin: 20S proteasome reveals a molecular basis for selectivity of α' , β' -epoxyketone proteasome inhibitors. *J. Am. Chem. Soc.* 122: 1237-1238

-
- ⁴³ Michael Groll et al. (2001). Crystal structure of the 20S proteasome: TMC-95A complex: a non-covalent proteasome inhibitor. *J. Mol. Biol.* 31: 543-548
- ⁴⁴ Vesna Milacic et al. (2008). Curcumin inhibits the proteasome activity in human colon cancer cells in vitro and in vivo. *Cancer Res.* 68, 18: 7283-7292
- ⁴⁵ Martina Bazzaro, Ravi K. (2011). Anchoori, Mohana Krishna R. Mudiam, et al. α,β -Unsaturated carbonyl system of chalcone-based derivatives is responsible for broad inhibition of proteasomal activity and preferential killing of human papilloma virus (HPV) positive cervical cancer cells. *J. Med. Chem.* 54: 449-456
- ⁴⁶ a) Hanne Hjorth Tønnesen, Jan Karlsen and Arvid Mostad. (1982). Structural studies of curcuminoids. I. The crystal structure of curcumin. *Acta Chem. Scand. B.* 36: 475-479; b) Florastina Payton, Peter Sandusky, and William L. Alworth. (2007). NMR Study of the Solution Structure of Curcumin. *J. Nat. Prod.* 70: 143-146.
- ⁴⁷ Tsonko M. Kolev, Evelina A. Velcheva et al. (2005). DFT and experimental studies of the structure and vibrational spectra of curcumin. *Int. J. Quant. Chem.* 102: 1069-1079
- ⁴⁸ a) Michael Groll, Robert Huber and Luis Moroder. (2009). The persisting challenge of selective and specific proteasome inhibition. *J. Pept. Sci.* 15: 58-66; b) Ljudmila Borissenko and Michael Groll. (2007). 20S Proteasome and Its Inhibitors: Crystallographic Knowledge for Drug Development. *Chem. Rev.* 107: 687-717 ; c) Wouter A. van der Linden, Paul P. Geurink, Chris Oskam et al. (2010). Proteasome selectivity towards Michael acceptor containing oligopeptide-based inhibitors. *Org. Biomol. Chem.* 8:1885-1893.
- ⁴⁹ <http://webcsd.ccdc.cam.ac.uk/index.php>

-
- ⁵⁰ a) R. GOMPPER, K. POLBORN, **2002**, Ccd PRIVATE COMMUNICATION; b) B.K.Joseph et al. *Chem Commun*, **2002** (736)
- ⁵¹ Li Lin et al. (**2010**). Novel STAT3 phosphorylation inhibitors exhibit potent growth-suppressive activity in pancreatic and breast cancer cells. *Cancer Research*. 10: 2445-2454
- ⁵² Jaguar, version 7.8, Schrödinger, LLC, New York, NY, 2011.
- ⁵³ a) Renato R. Contreras et al. (**1999**). A direct evaluation of regional Fukui functions in molecules. *Chem. Phys. Lett.* 304: 405-413; b) Eduardo Chamorro et al. *J. Chem. Phys.* (**2005**). 123: 114107
- ⁵⁴ a) Junia Melin et al. (**2007**). Removing electrons can increase the electron density: a computational study of negative Fukui functions. *J. Phys. Chem. A*. 111: 10017-10019; b) Junia Melin et al. (**2004**). Is the Fukui function a right descriptor of Hard-Hard interactions? *J. Phys. Chem. A*. 108: 2487-2491
- ⁵⁵ Michael Groll et al. (**2002**). Probing structural determinants distal to the site of hydrolysis that control substrate specificity of the 20S proteasome. *Chem. & Biol.* 9: 655-662
- ⁵⁶ Georgia B. McGaughey et al. (**1998**). π -stacking interactions. *J. Biol. Chem.* 273: 15458-15463
- ⁵⁷ <http://www.eyesopen.com/brood>
- ⁵⁸ Kato, S., Endoh, H., Masuhiro, Y., Kitamoto, T., Uchiyama, S., Sasaki, H., Masushige, S., Gotoh, Y., Nishida, E., Kawashima, H., et al. (**1995**). Activation of the estrogen

receptor through phosphorylation by mitogen-activated protein kinase. *Science*. 270: 1491–1494.

⁵⁹ Kumar, V., Green, S., Stack, G., Berry, M., Jin, J.-R., and Chambon, P. (1987).

Functional domains of the human estrogen receptor. *Cell*. 51: 941–951

⁶⁰ Aiming Sun et al. (2011). Discovering Small-Molecule Estrogen Receptor

a/Coactivator Binding Inhibitors: High-Throughput Screening, Ligand Development, and Models for Enhanced Potency, *ChemMedChem*, 6: 654-666

⁶¹ Ring, A.; Dowsett, M. (2004). Mechanisms of Tamoxifen Resistance, *Endocrine-Related Cancer*. 11: 643-658

⁶² W. Sherman et al. *J. Med. Chem.* (2006) 49: 534-553

⁶³ K. Zhu et al. *J. Chem. Theory. Comput.* (2007) 3: 2108-2119

⁶⁴ W. L. Jorgensen, *Science* (2004) 303: 1813-1818

⁶⁵ Niu, H.; Kalyanaraman, C.; Irwin, J. J.; Jacobson, M. P. (2006). Physics-based Scoring of Protein-Ligand Complexes: Enrichment of Known Inhibitors in Large-Scale Virtual Screening. *J. Chem. Inf. Model.* 46: 243-253.

⁶⁶ Guimaraes, C.R.; Cardozo, M. (2008). MM-GB/SA Rescoring of Docking Poses in Structure-Based Lead Optimization. *J. Chem. Inf. Model.* 48: 958-970.

⁶⁷ a) M. O. Jensen et al. *PNAS*, (2008) 105, 14430-14435; b) P. Maragakis et al. *J. Phys. Chem. B*, (2008) 112: 6155-6158

⁶⁸ Sun, A.; Yoon, J.-J.; Yin, Y.; Prussia, A.; Yang, Y.; Min, J.; Plemper, R. K.; Snyder, J. P. (2008). Potent Non-Nucleoside Inhibitors of the Measles Virus RNA-Dependent RNA Polymerase Complex. *J. Med. Chem.* 51: 3731-3741.

⁶⁹ a) Gilli, P.; Gilli, G.; Borea, P. A.; Varani, K.; Scatturin, A.; Dalpiaz A. (2005)

Binding thermodynamics as a tool to investigate the mechanisms of drug-receptor interactions: thermodynamics of cytoplasmic steroid/nuclear receptors in comparison with membrane receptors. *J. Med. Chem.* 48: 2026-2035; b) Borea, P. A.; Dalpiaz A.; Varani, K.; Gilli, P.; Gilli, G. (2000) Can thermodynamic measurements of receptor binding yield information on drug affinity and efficacy? *Biochem Pharmacol.* 60: 1549-1556.

⁷⁰ a) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. (2008). Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding. *J. Am. Chem. Soc.* 130: 2817-2831; b) Freire, E. (2008) Do enthalpy and entropy distinguish first in class from best in class? *Drug Discov. Today.* 13: 869-874; c) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. (2007) Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. *Proc. Nat. Acad. Sci. U.S.A.* 104: 808-813; d) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. (2006) Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* 49: 6177-6196.

⁷¹ a) Jack D. Dunitz, J. D. (1994). The entropic cost of bound water in crystals and biomolecules. *Science* 264: 670; b) Zhang, L.; Hermans, J. (1996) Hydrophilicity of Cavities in Proteins. *Proteins Struc. Func. Genetics* 24: 433-438.

⁷² Bondi, A. van der Waals Volumes and Radii. *J. Phys Chem.* (1964) 68: 441-451.

⁷³ Toogood, P. L. (2002) Inhibition of protein-protein association by small molecules: Approaches and progress. *Journal of Medicinal Chemistry* 45: 1543-1558.

-
- ⁷⁴ Pecuh, M. W.; Hamilton, A. D. (2000). Peptide and protein recognition by designed molecules. *Chemical Reviews* 100: 2479-2493.
- ⁷⁵ Rosenbaum, D. M.; Cherezov, V.; Hanson, M. A.; Rasmussen, S. G. F.; Thian, F. S.; Kobilka, T. S.; Choi, H.-J.; Yao, X.-J.; Weis, W. I.; Stevens, R. C.; Kobilka, B. K. (2007) GPCR Engineering Yields High-Resolution Structural Insights into b2-Adrenergic Receptor Function. *Science* 318: 1266-1273.
- ⁷⁶ Lipinski, C.A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 23: 3-26.