

**IMPROVING RATIONAL DRUG DESIGN BY INCORPORATING NOVEL  
BIOPHYSICAL INSIGHT**

by

**Matthew Patrick Baumgartner**

B.S. in Biological Sciences, University of Pittsburgh, 2010

Submitted to the Graduate Faculty of  
the School of Medicine in partial fulfillment  
of the requirements for the degree of  
Ph.D. in Computational Biology

University of Pittsburgh

2016

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This dissertation was presented

by

Matthew Patrick Baumgartner

It was defended on

February 12, 2016

and approved by

Dr. Lansing Taylor, Ph.D., Allegheny Foundation Professor

Dr. Thomas E. Smithgall, Ph.D., William S. McEllroy Professor and Chair

Dr. Gordon S. Rule, Ph.D., Professor

Dissertation Advisor: Dr. Carlos J. Camacho, Ph.D., Associate Professor

Copyright © by Matthew Patrick Baumgartner

2016

# IMPROVING RATIONAL DRUG DESIGN BY INCORPORATING NOVEL BIOPHYSICAL INSIGHT

Matthew Patrick Baumgartner, Ph.D.

University of Pittsburgh, 2016

Computer-aided drug design is a valuable and effective complement to conventional experimental drug discovery methods. In this thesis, we will discuss our contributions to advancing a number of outstanding challenges in computational drug discovery: understanding protein flexibility and dynamics, the role of water in small molecule binding and using and understanding large amounts of data. First, we describe the molecular steps involved in the induced-fit binding mechanism of p53 and MDM2. We use molecular dynamics simulations to understand the key chemistry responsible for the dynamic transition between the *apo* and *holo* structures of MDM2. This chemistry involves not only the indole side chain of the anchor residue of p53, Trp23, but surprisingly, the beta-carbon as well. We demonstrate that this chemistry plays a key role in opening the binding site by coordinating the position and orientation of MDM2 residues, Val93 and His96, through a previously undescribed transition state. We confirm these findings by observing that this chemistry is preserved in all available inhibitor-bound MDM2 co-crystal structures. Second, we discuss our advances in understanding water molecules in ligand binding sites by data mining the structural information of water molecules found in X-ray crystal structures. We examine a large set of paired bound and unbound proteins and compare the water molecules found in the binding site of the unbound structure to the functional groups on the ligand that displace them upon binding. We identify a number of generalized functional groups that are associated with characteristic clusters of water molecules. This information has been utilized in



several successful and ongoing virtual screens. Third, we discuss software that we have developed that allows for very efficient exploration and selection of virtual screening results. Implemented as a PyMOL plugin, ClusterMols clusters compounds based on a user-defined level of chemical similarity. The software also provides advanced visualization tools and a number of controls for quickly navigating and selecting compounds of interest, as well as the ability to check online for available vendors. Finally, we present several published examples of modeling protein-lipid and protein-small molecules interactions for a number of important targets including ABL, c-Src and 5-LOX.

## TABLE OF CONTENTS

<b>1.0</b>	<b>INTRODUCTION.....</b>	<b>1</b>
1.1	OUTLINE .....	1
<b>2.0</b>	<b>CRITICAL STEPS IN THE INDUCED-FIT BINDING MECHANISM OF MDM2- P53 .....</b>	<b>3</b>
2.1	INTRODUCTION .....	3
2.2	METHODS .....	7
2.2.1	Simulation of Inhibitor-Bound Structures.....	8
2.2.2	Simulation Parameters.....	8
2.2.3	Analysis Tools.....	9
2.2.4	Preparation of Known Inhibitor Structures .....	10
2.3	RESULTS .....	10
2.3.1	RMSD Calculations.....	11
2.3.2	Binding Pocket Distance Measurements.....	11
2.3.3	Free Energy Calculations .....	12
2.3.4	Residue Overlaps.....	13
2.3.5	Val93 Rotamer Analysis .....	16
2.3.6	Val93 Position and Orientation .....	18
2.3.7	His96 Rotamer Analysis .....	19
2.3.8	His96 Position and Orientation .....	21
2.3.9	Examination of co-crystal structures of MDM2 inhibitors .....	24
2.4	DISCUSSION AND CONCLUSION .....	25

2.5	SUPPLEMENTARY FIGURES AND TABLES.....	27
<b>3.0</b>	<b>KNOWLEDGE BASED INSIGHT OF LEVERAGING CRYSTAL WATER MOLECULES IN RATIONAL DRUG DESIGN.....</b>	<b>33</b>
3.1	BACKGROUND .....	33
3.2	DATASET ASSEMBLY .....	35
3.3	WATER CLUSTER CHARACTERIZATION .....	35
3.4	CHARACTERIZING LIGAND FUNCTIONAL GROUPS.....	39
3.5	WATER CLUSTER DISPLACEMENT BY LIGAND FUNCTIONAL GROUPS .....	41
3.6	SUMMARY.....	48
<b>4.0</b>	<b>DEVELOPMENT OF TOOLS TO INCORPORATE EXPERT KNOWLEDGE IN VIRTUAL SCREENING .....</b>	<b>50</b>
4.1	BACKGROUND .....	50
4.1.1	High-Throughput Screening.....	50
4.1.2	Virtual Screening.....	52
4.2	GENERAL METHOD.....	57
4.2.1	Similarity Calculations.....	58
4.2.2	Clustering .....	60
4.3	USER INTERFACE .....	61
4.4	SUMMARY.....	64
<b>5.0</b>	<b>CONCLUSIONS AND FUTURE RESEARCH.....</b>	<b>66</b>
	<b>APPENDIX A APPLICATIONS OF MODELING PROTEIN-LIPID INTERACTIONS. 69</b>	
A.1	MODELING SELECTIVE INHIBITORS OF C-SRC.....	69

A.2	MODELING OF ALLOSTERIC ACTIVATORS OF ABL.....	82
<b>APPENDIX B APPLICATION OF MODELING PROTEIN-LIPID INTERACTIONS .</b>		<b>108</b>
B.1	MODELING LIPID INTERACTIONS WITH H-5-LOX.....	108
<b>APPENDIX C SCRIPT FOR COMPUTING RESIDUE OVERLAPS IN MDS.....</b>		<b>124</b>
<b>BIBLIOGRAPHY .....</b>		<b>129</b>

## LIST OF TABLES

Table 2.1 Free energy difference between open and closed binding pockets.....	13
Table 2.2 Percent of frames within 15 degrees of the bound state rotamer of Val93 (-51°) .....	18
Table 2.3 Percent of frames within 15 degrees of the rotated outward rotamer of His96 (62°) ...	20
Table 2.4 List of PDB IDs of inhibitor bound structures of MDM2. ....	27
Table 3.1 List of functional groups considered for ligand characterization. ....	40
Table 3.2 Most common water clusters that are displaced by functional groups with at least 100 examples .....	43
Table 4.1. ClusterMols keyboard commands. ....	63

## LIST OF FIGURES

Figure 2.1 Simulated structures of MDM2 and cross pocket distances.....	6
Figure 2.2 Fractional overlap of p53 interface residues under different simulation conditions. .	15
Figure 2.3 Rotational coordination of MDM2 residue Val93.....	17
Figure 2.4 The rotamer of His96 plays a key role in the transition state of p53 binding to MDM2 .....	20
Figure 2.5 Heat map of the distance from the helix $\alpha 2$ to the $\alpha$ -carbon of His96 and the His96 dihedral angle.....	22
Figure 2.6 Following final complexation, His96 rotamer returns to the closed state.....	23
Figure 2.7 Co-crystallized inhibitors have CB and CA analogs. ....	24
Figure 2.8 Distance restraints used in Val93 coordination simulation. ....	28
Figure 2.9 RMSD of backbone atoms of binding pocket residues to the bound state for the NMR2 and NMR3.....	28
Figure 2.10 The distances measured for the Leu, Trp and Phe pockets NMR2 and NMR3 simulations. ....	29
Figure 2.11 Distribution of dihedral angles of side chains of critical interaction residues. ....	29
Figure 2.12 Dihedral angle of Val93 side chain in the NMR2 and NMR3 simulations.....	30
Figure 2.13 Dihedral angle of His96 side chain in the NMR2 and NMR3 simulations. ....	31
Figure 2.14 Heat map of the distance from the helix $\alpha 2$ to the $\alpha$ -carbon of His96 and the His96 dihedral angle.....	32
Figure 3.1 Polar contacts between crystal waters and the protein surface.....	36

Figure 3.2 Distribution of the number of polar contacts with the protein by binding site waters.	37
Figure 3.3 Defining water clusters on the protein surface. ....	38
Figure 3.4 Distribution of the sizes of binding site water clusters .....	39
Figure 3.5 Distribution of water cluster sizes and average polar contacts.....	42
Figure 3.6 Distribution and examples of water clusters that overlap the AS(A)A functional group .....	44
Figure 3.7 Distribution and examples of water clusters that overlap the NC(C)N functional group. .....	45
Figure 3.8 Distribution and examples of water clusters that overlap the aromatic functional group. .....	46
Figure 3.9 Distribution and examples of water clusters that overlap the X (Halogen) functional group. ....	47
Figure 4.1 Overview of our structure-based virtual screening pipeline.....	54
Figure 4.2 Structure of a protein-protein interaction. ....	58
Figure 4.3 Load file dialog of ClusterMols. ....	59
Figure 4.4 Graphical User Interface (GUI) of ClusterMols.....	60
Figure 4.5 Hierarchical clustering of compounds based on their similarity. ....	61
Figure 4.6 Clustered virtual screening results showing polar contacts.....	62
Figure 4.7. Vendor information and prices of compounds of interest in ClusterMols. ....	64

## ACKNOWLEDGEMENTS

I would like to acknowledge a number of people without whom, this thesis would not exist.

To my thesis advisor, Dr. Carlos Camacho, I sincerely thank you for seeing in me what you did those years ago and for convincing others of the same. You have provided me with tremendous insight into solving complex biophysical problems and taught me to have a critical eye in evaluating my work as well as the work of others. And for that I am in your debt.

To Dr. David Koes, I have always felt tremendously lucky to have been able to work with you. You have always been incredibly supportive and helpful with any question or problem that I have brought to you. I have always thought of you as a bonus advisor who I can always count on for thoughtful discussion on nearly any topic. I am sincerely grateful for being able to learn from you.

To our frequent collaborators in the laboratory of Dr. Thomas Smithgall: Tom, Prena and Jamie, I am grateful for the years that we have been able to work together. Our discussions have always been thought-provoking and fruitful.

To the members of the Camacho Lab, past and present, who have helped me throughout the years with thoughtful discussion, advice and assistance. To Alpay, David, Weiyi, Jintao, Tim, Gustavo, Nick, Michael, Xanthe, Bentley and Zena, thank you so much for everything and I wish you all the best!

To the members of the CPCB incoming class of 2010: Shannon, Devin, Aaron, Jose, Salim and MCC. Thank you guys for making graduate school highly enjoyable! I have greatly enjoyed working closely on course work and enjoying social activities with you. More than just classmates, some of you guys are the best friends that I've ever had.



To my family, Mom and Dad, I simply cannot thank you enough for all that you have done for me throughout my entire life. You have been incredibly supportive of me no matter what I wanted to do and for that I am eternally grateful. To my four brothers: David, Eric, Alex and Todd, thanks for always being there for me. It is very reassuring to know that you guys are always there for any problems that might come along.

And finally, I am deeply grateful for the love and support of my fiancée, Maria Maggio. Thank you for being incredibly supportive and understanding of me throughout this whole endeavor. I sincerely appreciate your ongoing efforts to keep me fed and sane during the times of stress. I literally don't think that I could have done this without you.

## 1.0 INTRODUCTION

Rational drug design is the process through which drugs are designed with the intent of targeting a specific protein target which is known to be relevant for a disease of interest[1]. Structure-based drug design is one facet of rational drug design that uses computational methods to leverage information of the 3D structure of proteins to discover potential inhibitors[2-4]. The primary goal of structure-based drug design is to computationally predict whether a given compound will bind to a protein target before it is tested experimentally, which can be quite costly[5]. Virtual screening methods have successfully been used in conjunction with experimental techniques for a number of years[6-9]. Successful virtual screening methods require understanding of the protein target at a number of levels including: protein flexibility, the role of water in ligand binding and using and understanding large amounts of data. In this dissertation we will discuss our advances in these areas.

## 1.1 OUTLINE

This dissertation is organized as follows:

In Chapter 2.0, we present the atomistic induced-fit binding mechanism of MDM2 and p53, the so called “guardian of the genome”[10]. We show that the  $\alpha$  and  $\beta$  carbons of the Trp23

anchor residue of p53 are responsible for opening the binding cleft through hydrophobic interactions with MDM2 residues Val93 and His96.

In Chapter 3.0, we discuss our work in using data mining techniques to identify clusters of crystallographic water molecules that are commonly displaced by ligand functional groups. By examining pairs of *apo* and ligand-bound structures we have identified a number of water cluster/functional group associations that have useful applications in a fragment-based virtual screening approach.

In Chapter 4.0, we present a virtual screening tool call ClusterMols. Implemented as a PyMOL plugin, ClusterMols automatically clusters virtual screening results based on chemical similarity. These clusters are visualized in PyMOL in a user-friendly fashion such that the user can navigate compounds on the scaffold level using keyboard controls. Compounds of interest can be extracted and saved for later analysis or checked for commercial availability with the press of a button.

In Chapter 5.0, we give the main conclusions of the thesis and discuss possible future research directions.

In Appendix A, a number of applications of theory to real-world problems in the realm of modeling protein-small molecule interactions will be presented.

In Appendix B, we will present an application of modeling protein-lipid interactions.

## **2.0 CRITICAL STEPS IN THE INDUCED-FIT BINDING MECHANISM OF MDM2-P53**

### **2.1 INTRODUCTION**

The tumor suppressor p53, which is transcribed by the TP53 gene, has been shown to play a key role in most human cancers[11-19]. Often called the “guardian of the genome,” p53 is a transcription factor that plays a central role in the response to cellular stress[10, 20]. It is responsible for a number of important cellular functions including cell cycle regulation, DNA repair and the initiation of apoptosis[21]. In healthy cells, the transcription factor p53 is activated upon cellular stress, leading to cellular arrest or apoptosis[22]. During low stress, it is downregulated by mouse double minute 2 (MDM2)[10, 20-22]. MDM2 regulates cellular p53 activity through two pathways: transcriptional inhibition of the TP53 gene and increased ubiquitination and degradation of the p53 protein via direct binding of its N-terminal domain[20]. In some cancer cells, p53 activity is often lost either through mutation or inactivation of the TP53 gene[13]. In other cancer cells with wild-type p53, overexpression of MDM2 can block p53 function and promote tumor growth, making MDM2 a promising therapeutic target[22, 23].

Prior to complexation, the MDM2 binding domain on the N-terminus of p53 is unstructured[11]. The MDM2 binding motif of p53 primarily consists of three hydrophobic residues (Phe19, Trp23 and Leu26; Figure 2.1B) that become deeply buried upon binding. When

MDM2 is in the unbound (*apo*) state, the binding sites for these p53 residues are mostly occluded and direct complexation is not possible. Due to this occlusion, some amount of structural rearrangement of MDM2 is required for binding of p53[16, 17, 24-27]. Previous work has described the induced-fit binding process of MDM2, but none have determined the critical interactions for opening the binding cleft[16, 17].

Previous research has suggested that many protein-protein interactions (PPIs) may involve the burying of an anchor residue: an exterior residue of the ligand that serves as a structural marker and facilitates recognition by sampling its bound state configuration even before encountering the binding partner[28]. Docking of this anchor residue marks the formation of the encounter complex and provides time for additional triggering interactions to take place that induce structural transitions of the receptor and ligand towards the high-affinity complex. Previous work has identified Trp23 of p53 to be an anchor residue when binding to MDM2[29].

In this work, we use molecular dynamics (MD) simulations to identify the specific molecular interactions required to open the p53 binding pocket of MDM2. This approach provides direct insight into molecular recognition and how structural flexibility and the induced-fit binding process are exploited to regulate critical cellular processes. To determine the role of the anchor residue in modifying the structure of the receptor interface, we simulated MDM2 with different small fragments of the Trp23 anchor residue of p53. Trajectories were then analyzed to identify the receptor interface deformations that occur upon complexation, particularly those that lead to the formation of the binding cavity. These findings were verified via MD simulations of the PPI and inhibitor-bound crystal structures.

We were able to map distinct MDM2 interface deformations to specific ligand motifs by iteratively modifying peptide ligand mimics and analyzing receptor interface dynamics in the

resulting trajectories. We calculated equilibrium free energy differences between the open and closed receptor interface states for the *apo* and fragment-bound receptors, allowing us to quantify the energetic contribution of each atom to opening the binding pocket.

We demonstrate that the binding of the two-ringed indole of Trp23 alone is insufficient for opening the binding cavity, specifically the binding pockets for Phe19 and Leu26. However, upon the addition of the beta carbon (CB) and, to a greater extent, the alpha carbon (CA) of Trp23, the binding cleft is stabilized in the open state, allowing for binding of p53 residues Phe19 and Leu26, which initiate the formation of the high-affinity bound-state complex. In particular, we demonstrate that the CB and CA atoms are responsible for the induced-fit binding process by coordinating the position and orientation of MDM2 residue Val93 (and thereby sheets  $\beta 3$  and  $\beta 4$  and helix  $\alpha 4$ ; Figure 2.1B), which opens the Phe19 and Leu26 binding sites. Additionally, we show that the CB and CA atoms shift the side chain of His96 on helix  $\alpha 4$  to an open state observed in the *apo* NMR structures allowing for opening of the binding pocket prior to the formation of the final high-affinity bound state. A *post hoc* examination of publicly available inhibitor-bound MDM2 crystal structures indicate that almost all of the known inhibitors share the same features as the CB and CA atoms of Trp23.

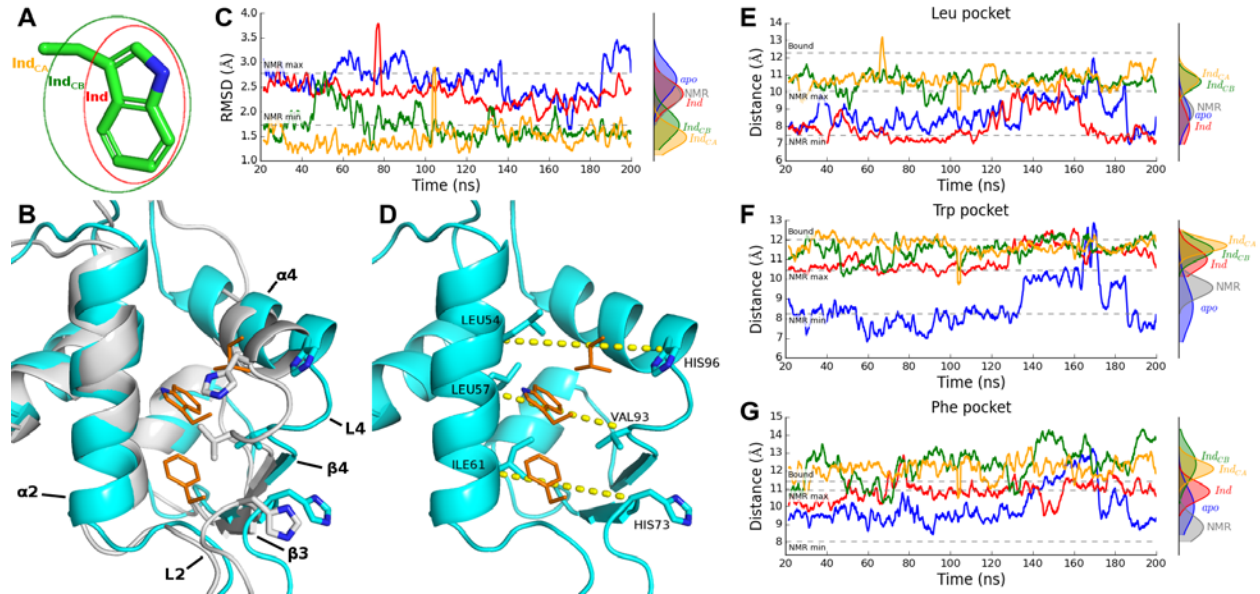


Figure 2.1 Simulated structures of MDM2 and cross pocket distances. Sections of Trp23 from p53 used in the Ind, IndCB and IndCA simulations for MDM2. B) PPI crystal structure (PDB ID: 1YCR) of MDM2 (blue cartoon) and the interface residues of p53 (orange lines) overlaid on apo NMR structure (PDB ID: 1Z1M) of MDM2 (white cartoon). Three residues on the interface are shown as sticks to highlight the closed and open states of the pocket. C) RMSD of backbone atoms of binding pocket residues to the bound state for the simulations starting from the first NMR structure and fit distributions. The lines are colored (here and throughout) as follows: apo: blue, Ind: red, IndCB: green, IndCA: yellow, NMR: grey. The horizontal dashed lines indicate the minimum and maximum RMSDs of the 24 NMR structures. Plots of the simulations starting from the second and third NMR structures are shown in Figure 2.9. D) Illustration of the distances used to measure pocket opening and closing. The PPI bound structure (blue) is shown with the interface residues of p53 (orange). The distances (dashed yellow lines) calculated are between of the alpha carbons of His96 and Leu54 (Leu pocket; top), His73 and Ile61 (Trp pocket; middle) and Val93

and Leu57 (Phe pocket; bottom). The distances measured for the E) Leu, F) Trp and G) Phe pockets calculated from the simulations of the first NMR structure. The horizontal dashed lines indicate the distances in the bound structure and the minimum and maximum distances observed in the 24 unbound NMR structures. Plots of the simulations starting from the second and third NMR structures are shown in Figure 2.10.

## 2.2 METHODS

To determine the atomistic mechanism of opening the binding site, we ran MD simulations using the first three NMR structures of the *apo* form of human MDM2 (PDB ID: 1Z1M[30]). The disordered N- and C-termini were discarded and only residues 25 to 109 were simulated[21, 31], which are the residues that were crystallized in the bound PPI structure that we used as a reference (PDB 1YCR[32]). We conducted simulations under four different conditions: one *apo* receptor simulation with no ligand and three simulations (*Ind*, *Ind<sub>CB</sub>*, and *Ind<sub>CA</sub>*) that included portions of the Trp23 indole anchor (Figure 2.1A). *Ind* includes the double ringed indole portion of the tryptophan side chain alone, *Ind<sub>CB</sub>* includes the atoms from *Ind* as well as the CB atom and *Ind<sub>CA</sub>* contains the atoms from *Ind* in addition to both the CB and CA atoms. In a number of the *apo* NMR structures (PDB ID: 1Z1M), the p53 anchor (Trp23) binding site is partially closed and thus we could not directly insert the indole into the binding site without significant clashes (Figure 2.1B). For each of the bound simulations (*Ind*, *Ind<sub>CB</sub>* and *Ind<sub>CA</sub>*), we first ran a short (5 ns) simulation wherein we manually position the indole “above” the binding site and then use a Steered Molecular Dynamics (SMD) simulation to bring the indole into the binding site using distance



constraint between atom CZ2 of the indole and the beta carbon of Leu57. At the end of this simulation, the indole is located approximately in the bound position. We then restarted the simulation from the final frame (maintaining the velocities) to initiate a 200 ns simulation with position restraints on the heavy atoms of the indole and the backbone atoms N, C and O of helix  $\alpha 2$  (residues 52 to 62, Figure 2.1B) with force constant of 10.0 kcal/(mol $\cdot$ Å<sup>2</sup>). The indole was restrained to its position in 1YCR relative to helix  $\alpha 2$ . The *apo* simulation was simulated unconstrained for 200 ns.

### 2.2.1 Simulation of Inhibitor-Bound Structures

MD simulations of several inhibitor-bound structure of MDM2 inhibitors were performed as described above. Additionally, the backbone atoms of MDM2 and the heavy-atoms of the inhibitors were held in place with position restraints.

### 2.2.2 Simulation Parameters

MD simulations were carried out using the pmemd.cuda[33] module of AMBER14[34]. The force fields AMBER ff12SB and gaff[35] (general amber force fields) were used. Ligand fragments were parameterized using *antechamber*[36] at a neutral charge and default parameters. An octahedral TIP3P water box was constructed with 12 Å from the edge of the box to the solute and the total system charge was neutralized by adding Cl<sup>-</sup> ions. The non-bonded cutoff was specified at 10 Å. In the first energy minimization run, the solute was held fixed and the solvent was relaxed through 500 cycles of steepest descent followed by 500 cycles of conjugate gradient minimization.

Subsequently, the system was minimized while restraining the protein backbone atoms through 2,000 cycles of steepest descent followed by 3,000 cycles of conjugate gradient minimization. Following the energy minimization, a 50,000 step MD equilibration simulation was used to raise the system temperature to 300 K while holding the solute fixed with weak ( $10.0 \text{ kcal}/(\text{mol}\cdot\text{\AA}^2)$ ) restraints. Bonds involving hydrogens were held at a fixed length and the integration step was set to 2 fs. For the Steered Molecular Dynamics (SMD) simulations, we brought the indole into the pocket using a sliding distance constraint starting at  $\sim 11 \text{ \AA}$  and ending at  $4.0 \text{ \AA}$ , the distance observed in the PPI crystal structure (PDB ID: 1YCR). The 200 ns MD simulations were held at 300 K under constant pressure with the constraints as listed above for each system. For all analyses, the first 20 ns of the production simulations were discarded as system equilibration. To validate the importance of MDM2 Val93 coordination in binding site opening, we ran simulations of the *apo* structure with four distance constraints on the  $\alpha$ -carbon of Val93. The restraints were between the  $\alpha$ -carbon of Val93 and the  $\alpha$ -carbons of Met50, Leu57, Gly58 and Ile61, respectively (Figure 2.8). The distance constraints were set to  $\pm 0.1 \text{ \AA}$  of those measured in the bound structure (PDB ID: 1YCR). The distance restraint weights were set to  $10.0 \text{ kcal}/(\text{mol}\cdot\text{\AA}^2)$ .

### 2.2.3 Analysis Tools

The PyMOL Molecular Graphics System v1.7.4.0 was used for structure preparation and analysis[37]. Trajectories were also analyzed using VMD v1.9.2[38]. The LOOS (Lightweight Object-Oriented Structure) library v2.1.3[39] was used for root-mean-square deviation (RMSD), distance and dihedral calculations.

The atom overlaps of the p53 binding interface residues were calculated using a custom Python script that iterates through the frames of an aligned MD simulation and counts the number

of frames where the receptor atoms overlap with these atoms. An atom is considered to be overlapped if there is a receptor atom center within its the van der Waals radii[40]. The percent of simulation frames for which an atom is overlapped by at least one protein atom is reported as a measure of binding pocket occlusion. This script is provided in APPENDIX C.

#### 2.2.4 Preparation of Known Inhibitor Structures

We examined the 32 currently available inhibitor-bound structures of MDM2 from the Protein Data Bank listed in Table 2.4. We excluded several structures from our analysis, due to the non-canonical nature of their binding (indicated in the table). These ligands bind to MDM2 such that the Leu pocket is vacant. For each co-crystal structure, we first superimposed it on the PPI structure (PDB ID: 1YCR) by aligning to helix  $\alpha 2$ . For structures with more than one ligand, the ligand in the PPI binding site was retained and the rest discarded. While there are many more known inhibitors of MDM2 and p53[12], we limited ourselves to those that have been co-crystallized in order to make direct structural comparisons of the binding mode.

### 2.3 RESULTS

Here we show the molecular binding mechanism of the MDM2-p53 complex. We ran MD simulations with different portions of the Trp23 anchor residue from p53 (see Figure 2.1A), *Ind* (indole alone), *Ind<sub>CB</sub>* (indole + CB), *Ind<sub>CA</sub>* (indole + CB + CA), as well the *apo* structure for comparison. We then measured the effects of the residue fragments on the MDM2 interface structure.

### 2.3.1 RMSD Calculations

We computed the backbone RMSD of the MDM2 residues in the binding pocket defined as the MDM2 residues within 6 Å of the p53 binding residues in the bound structure (PDB ID: 1YCR). The simulation results (Figure 2.1C) show that the *apo* simulation does not closely sample the bound state, open-pocket configuration. The *Ind* simulation is very similar to the *apo* simulation in that the MDM2 structure does not resemble the bound state. In the *Ind<sub>CB</sub>* simulation we can observe that the CB atom is the critical factor in binding-site opening as when it is added to the indole, we see that the pocket adopts the open configuration; however, it does not fully stabilize the native p53-bound open state. Finally, when we add the CA atom as in the *Ind<sub>CA</sub>* simulation, the simulations adopt a near native state for the duration of the simulation. These observations are qualitatively consistent with the simulations started from the second and third NMR structures (Figure 2.9).

### 2.3.2 Binding Pocket Distance Measurements

To more directly quantify the extent of opening or closing of the p53 binding site, we measured the distance across the binding pocket at the locations of the three p53 interface residues (Phe19, Trp23, Leu26). For each simulation, we measured the distances between the  $\alpha$ -carbons of three pairs of the MDM2 atoms, each pair is representative of the openness of one of the p53 interface residue binding sites (Phe pocket: His73-Ile61, Trp pocket: Val93-Leu57, and Leu pocket: His96-Leu54) shown in Figure 2.1D. As these are measurements of backbone-backbone distances, they are designed to capture larger movements of the  $\beta$ -sheets and  $\alpha$ -helices of the binding interface as opposed to smaller side chain movements. The measurements for the simulations of the first NMR

structure are shown in Figure 2.1E-G and for the simulations of the second and third NMR structure are shown in Figure 2.10.

For all three measured distances, those measured in the *apo* simulation generally fall within the range of the NMR structures. The *Ind* simulation measurements also generally fall into the range of the NMR structures, with the exception of the distance across the Trp pocket (Figure 2.1F) where the presence of the indole in the simulation stabilizes the central pocket in the open state, although it is not sufficient to open the other two pockets. The *Ind<sub>CB</sub>* simulation is able to stabilize the bound state, however when the CA atom is added in the *Ind<sub>CA</sub>* simulation, the pocket stays open with longer dwell times. This indicates that the CB atom and to a greater extent the CA atom are responsible for opening the Phe19 and Leu26 binding pockets. In these simulations, the Leu26 binding site does not achieve the fully open state ("Bound" line in Figure 2.1E), indicating that the anchor residue (with CB and CA) is able to partially open the pocket, but further induced fit rearrangements will be necessary to achieve the bound-state high-affinity complex, the nature of which is beyond the scope of this paper.

### 2.3.3 Free Energy Calculations

By measuring the number of frames that are in the "open" and "closed" state of the pocket, we are able to calculate the free energy required to open the binding pocket of MDM2. To calculate the free energy difference between the "open" and "closed" pocket in the simulations we used the following equation.

$$\frac{\langle N_{open} \rangle}{\langle N_{closed} \rangle} = e^{\frac{-\Delta G}{K_b T}}$$

In each frame, we measured the distance across each pocket (Figure 2.1E-G) to determine whether the pocket is in the “open” or “closed” state, based on whether the distance across each pocket is above or below a critical threshold, defined for each pocket as the maximum distance observed in the 24 *apo* NMR structures (“NMR max” lines in Figure 2.1E-G). A pocket is considered to be in the “open” state if the cross-pocket distance is larger the maximum cross-pocket distance in the *apo* NMR ensemble. For all three pockets, the  $\frac{\Delta G}{K_b T}$ , which is a measurement of the energy required to open the pocket, decreases with increasing indole group size (*apo*  $\rightarrow$  *Ind*  $\rightarrow$  *Ind<sub>CB</sub>*  $\rightarrow$  *Ind<sub>CA</sub>*), indicating that the open pocket is being stabilized (Table 2.1).

Table 2.1 Free energy difference between open and closed binding pockets. Values shown are the average and standard deviation of the simulations of the first three NMR structures and represent  $\frac{\Delta G}{K_b T}$ . More negative values indicate stabilization of the open pocket.

	Phe	Trp	Leu
<i>apo</i>	0.16 $\pm$ 1.41	2.97 $\pm$ 4.30	1.68 $\pm$ 1.70
<i>Ind</i>	-1.84 $\pm$ 2.21	-2.49 $\pm$ 1.49	0.04 $\pm$ 2.65
<i>Ind<sub>CB</sub></i>	-2.57 $\pm$ 0.74	-3.80 $\pm$ 1.25	-0.99 $\pm$ 1.82
<i>Ind<sub>CA</sub></i>	-3.77 $\pm$ 1.23	-3.32 $\pm$ 2.14	-1.07 $\pm$ 1.48

### 2.3.4 Residue Overlaps

In order to further quantify the opening of the MDM2 pocket, we computed the fractional overlap of the native p53 interface residues with the receptor if they had been in their bound-state positions throughout the simulations. Figure 2.2 reports the average fractional overlap from the simulations starting from the first three NMR structures. The interaction residues become less overlapped as

the size of the indole in the simulation is increased (*apo* → *Ind* → *Ind<sub>CB</sub>* → *Ind<sub>CA</sub>*). The *apo* simulations show a high degree of overlap and are in general agreement with the overlap produced from the 24 *apo* NMR structures (Figure 2.1F). Likewise, the *Ind* simulations show high overlap on the Phe19 and Leu26 residues. For the *Ind* simulations (as well as the *Ind<sub>CB</sub>* and *Ind<sub>CA</sub>* simulations) the presence of the indole in the simulation precludes overlaps with the Trp. The *Ind<sub>CB</sub>* simulations and to a greater extent the *Ind<sub>CA</sub>* simulations open up both sides of the pocket, as evidenced by the decreased amount of clashing observed.

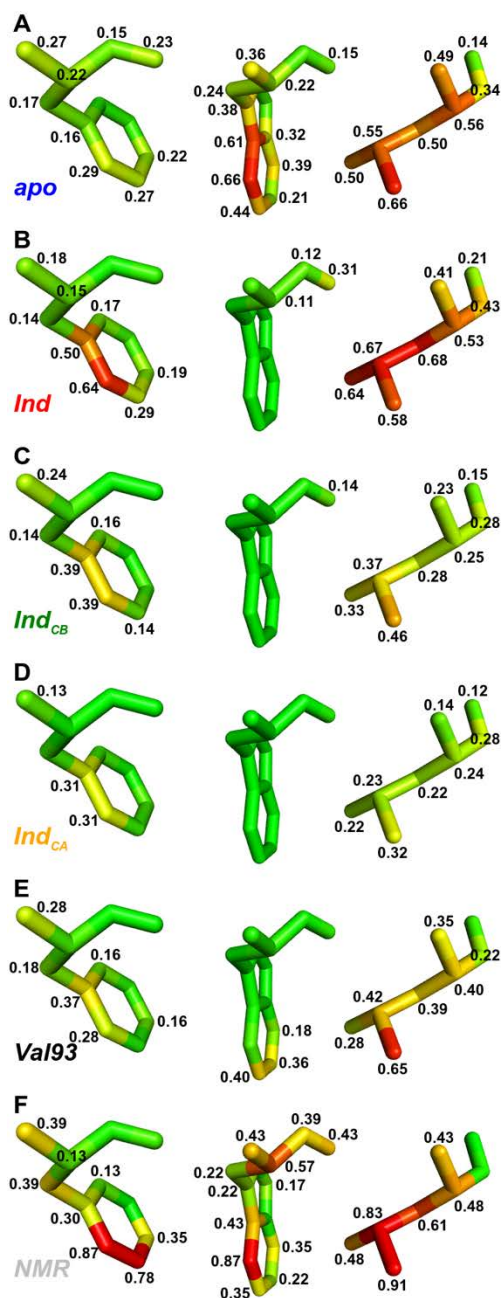


Figure 2.2 Fractional overlap of p53 interface residues under different simulation conditions. Values are averaged from simulations starting from the first three unbound MDM2 NMR structures. Shown are A) *apo*, B) *Ind*, C) *Ind<sub>CB</sub>*, D) *Ind<sub>CA</sub>*, E) *Val93* simulations, F) average of all NMR structures. Green colors indicate little or no overlap and red colors indicate high degree of overlap. In all overlap plots, average overlap values under 10% are not shown for clarity.



All these data show that in the *apo* simulations, the binding site is occluded and full complexation of p53 is not possible. The initiation of binding by the indole of the Trp23 anchor is modeled in the *Ind* simulations. While the presence of the isolated indole group opens the Trp23 binding pocket, it is not sufficient to open the Phe19 and Leu26 binding sites. It is only when we introduce the beta carbon of the Trp23 in the *Ind<sub>CB</sub>* simulations that the Phe19 and Leu26 pockets start to open. The full binding site opens in the *Ind<sub>CB</sub>* simulations, but it is not fully stabilized in the open configuration. The addition of the alpha carbon in the *Ind<sub>CA</sub>* simulations leads to full stabilization the open pocket. The opening of the binding site allows for complexation of the Phe19 and Leu26 residues, which will lead to further rearrangements as the two proteins form the high-affinity complex.

### **2.3.5 Val93 Rotamer Analysis**

A visual inspection of the simulations indicated that Val93 and His96 often occlude the Trp23 and Leu26 binding sites in the *apo* simulations. This indicates that their dynamics may play important roles in opening the binding site in the presence of the CB and CA atoms. To validate these observations we first conducted an analysis of the rotamers of the Val70 side chain of MDM2 in the PPI crystal and in the NMR structures, which revealed three dominant rotamer states, the PPI state and two rotamers that each represent about half of the NMR structures (Figure 2.3B-D and Figure 2.11A). We examined the rotamers of Val93 in the MD simulations and found that in the *Ind<sub>CB</sub>* and *Ind<sub>CA</sub>* simulations the valine is in the bound-state to a much higher degree than in the *apo* and *Ind* simulations (Figure 2.3A, Table 2.1). The results that we observed in the simulations of the first NMR structure shown in Figure 2.3 are consistent with the replicated simulations of the second and third NMR structures shown in Figure 2.12. This indicates that the CB and CA

atoms of Trp23 play an important role in controlling the orientation of the Val93 side chain at the binding interface.

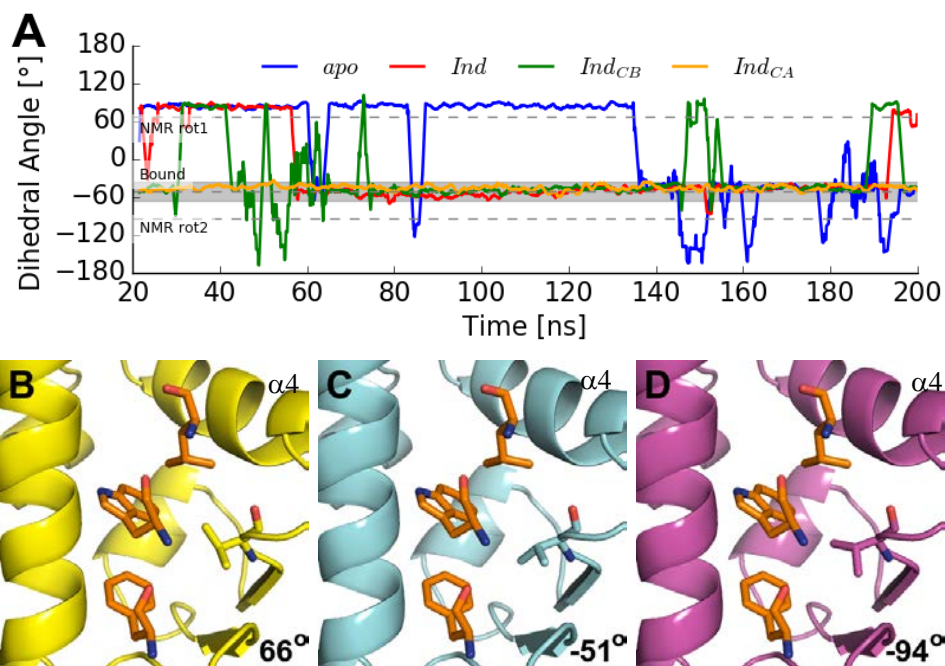


Figure 2.3 Rotational coordination of MDM2 residue Val93 A) Dihedral angle of Val93 side chain in the simulations starting from the first NMR structure. The average dihedral angle of the two dominant unbound rotamer states and the rotamer angle of the bound state are shown as dashed grey lines. The shaded area indicates  $\pm 15^\circ$  of the bound state rotamer. Lines are colored as follows: *apo*: blue, *Ind*: red, *Ind<sub>CB</sub>*: green, *Ind<sub>CA</sub>*: yellow. Corresponding plots for the simulations starting from the second and third NMR structure are shown in Figure 2.12. The rotamers of Val93 in the B) NMR rot1, C) bound and D) NMR rot2 positions.

Table 2.2 Percent of frames within 15 degrees of the bound state rotamer of Val93 (-51°)

	NMR 1	NMR 2	NMR 3	Average
<i>apo</i>	21.4	43.0	24.1	29.5
<i>Ind</i>	74.7	31.9	55.0	53.9
<i>Ind<sub>CB</sub></i>	72.4	72.1	60.3	68.3
<i>Ind<sub>CA</sub></i>	99.0	94.0	90.7	94.6

### 2.3.6 Val93 Position and Orientation

An examination of the simulations indicates that the CB and CA atoms in the *Ind<sub>CB</sub>* and *Ind<sub>CA</sub>* simulations also help to coordinate the position of Val93 (Figure 2.1F, Figure 2.3A). We hypothesize that pushing Val93 back into the “open” state will also push on helix  $\alpha$ 4 and sheets  $\beta$ 3 and  $\beta$ 4 thus opening the binding sites for Phe19 and Leu26.

To computationally validate this hypothesis, we conducted another set of ligand-free simulations starting from the closed-pocket, *apo* state. We used harmonic distance restraints to move and hold the position of Val93 in its bound-like position. The purpose of these simulations was to determine if the coordination of Val93 is sufficient to open the Phe19 and Leu26 binding pockets. The average overlap of the p53 interface residues in the coordinated Val93 simulations (Figure 2.2E) and the *Ind<sub>CB</sub>* simulations (Figure 2.2D), are comparable, although  $\delta$ -carbons of Leu26 in the Val93 simulation has a higher degree of overlap. This may be due to coordination of the indole to residues deeper in the binding pocket that is absent in the Val93 simulations. The Val93 simulations indicate that the indole + CB help to coordinate Val93, but are not enough to fully open the Phe19 and Leu26 pockets. A visual examination of the simulations and the NMR structures (Figure 2.1B) indicates that the major contribution to overlapping the Leu pocket comes from His96. The simulation data suggests that the added size of the alpha carbon in the *Ind<sub>CA</sub>*

simulations helps to coordinate His96 and guide it, and the rest of helix  $\alpha 4$ , into the bound configuration.

### **2.3.7 His96 Rotamer Analysis**

We examined His96 throughout the simulations and concluded that its position and rotamer relative to the bound state are important in the opening of the Leu binding pocket. An analysis of the rotamers of His96 present in the available crystallographic and NMR structures (Figure 2.11B) indicate that His96 is almost always in the same orientation (Figure 2.4B) with one notable exception (the fourth NMR structure), which is a rotamer wherein the His96 side chain is flipped out of the pocket (the ‘open’ rotamer; Figure 2.4C). We propose that this conformation of His96 in the non-native, flipped outward pose represents an important transition state in the binding process specifically in its role in the movement of helix  $\alpha 4$ , as discussed below.

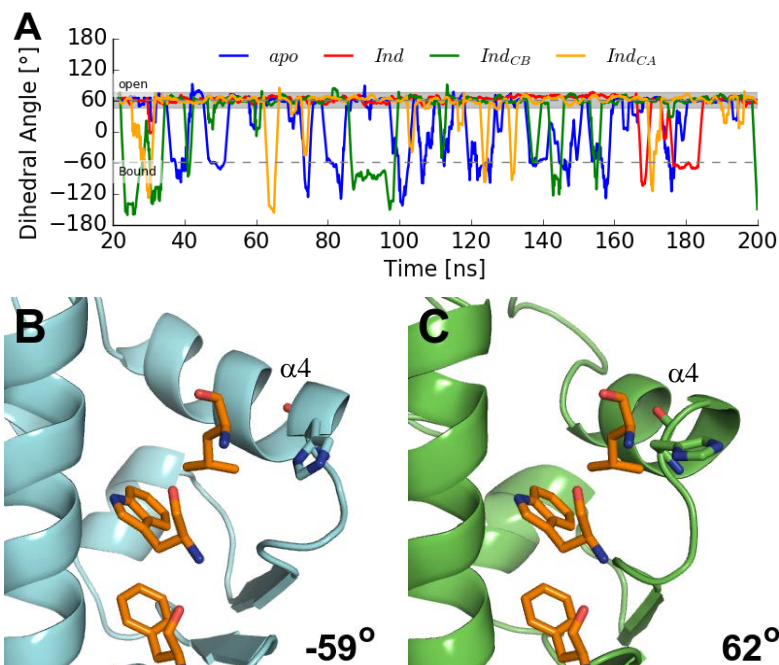


Figure 2.4 The rotamer of His96 plays a key role in the transition state of p53 binding to MDM2

A) Dihedral angle of His96 side chain in the 4 simulation types of the first NMR structure. Lines are colored as follows, *apo*: blue, *Ind*: red, *Ind<sub>CB</sub>*: green, *Ind<sub>CA</sub>*: yellow. Corresponding plots for the simulations starting from the second and third NMR structure are shown in Figure 2.13. Rotamers of His96 side chain in the B) closed state and the C) open state observed in the 4<sup>th</sup> NMR structure (PDB ID: 1Z1M).

Table 2.3 Percent of frames within 15 degrees of the rotated outward rotamer of His96 (62°)

	NMR 1	NMR 2	NMR 3	Average
<i>apo</i>	0.469	0.428	0.548	0.482
<i>Ind</i>	0.904	0.148	0.839	0.630
<i>Ind<sub>CB</sub></i>	0.718	0.758	0.785	0.754
<i>Ind<sub>CA</sub></i>	0.821	0.721	0.850	0.797

### 2.3.8 His96 Position and Orientation

In Figure 2.5, we constructed a heat map of the distance across the Leu binding pocket (as in Figure 2.1E) and the dihedral angle of the histidine side chain (Figure 2.4A). In the *apo* and *Ind* simulations, helix  $\alpha 4$  moves inward, closing the pocket (dark colors near the “NMR Min” line) and the side chain is largely in the flipped out (“NMR open”) state observed in the fourth NMR structure. However, as the size of the indole is increased as in the *Ind<sub>CB</sub>* and *Ind<sub>CA</sub>* simulations, the histidine shifts almost entirely into the “open and out” state where the helix is pushed out further than the unbound states (but less than the fully bound state) and the side chain is flipped into the open position (Table 2.3). We propose that this “open and out” state is a transition state that is required for binding to the Leu26 binding site. As we see here and previously in Figure 2.1E, the indole, even with the CA and CB atoms, does not sufficiently fully push the helix into the bound state. Therefore, the side chain of His96 must also flip out of the way until Leu26 is able to bind; after which His96 will flip back into place.

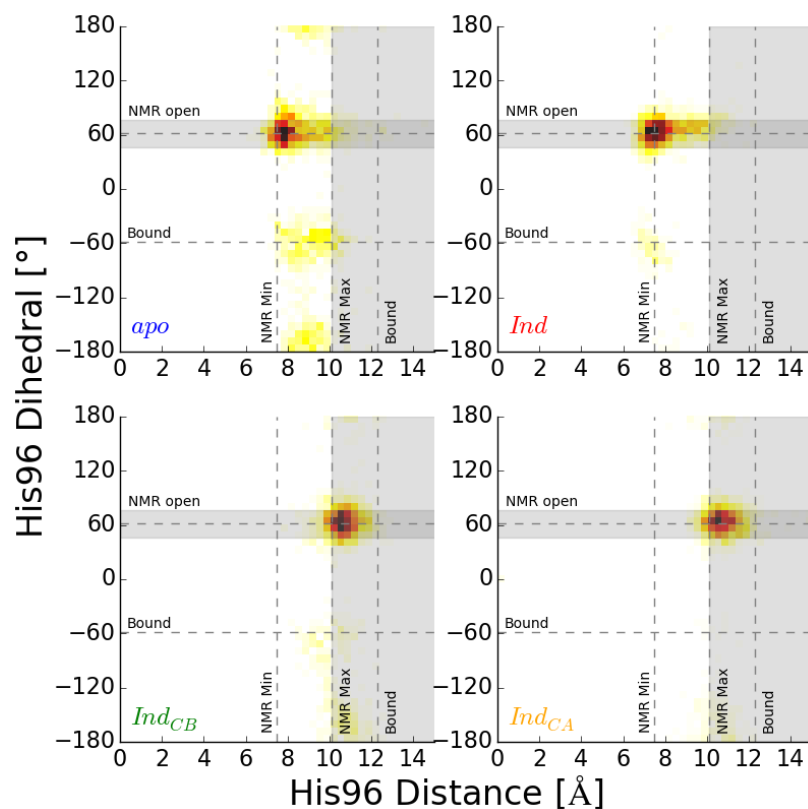


Figure 2.5 Heat map of the distance from the helix  $\alpha 2$  to the  $\alpha$ -carbon of His96 and the His96 dihedral angle. Data shown for the four simulation types of the first NMR structure. Horizontal grey bars indicate values within  $15^\circ$  of the dihedral angle observed in the fourth NMR structure ( $62^\circ$ ). Vertical grey bars indicate distances greater than the largest distance observed in the 24 NMR structures. Dark and light colors indicate high and low populations, respectively. Corresponding plots for the simulations starting from the second and third NMR structure are shown in Figure 2.14.

To validate this transitory model, we ran MD simulations to show that when MDM2 adopts the bound state in complex with its binding partner (small molecule or protein) that the His96 will flip back into the closed state. We conducted MD simulations of the full MDM2-p53 PPI complex

(PDB ID: 1YCR) in its native closed state and after manually rotating the His96 side chain into the open state. Figure 2.6A shows that, in the simulations starting in the open and closed states, His96 is in the closed state for 46% and 71% of the time, respectively, indicating that His96 can go from the open to the closed state when bound to a ligand (e.g. p53). To further confirm that His96 will sample the bound (closed) state we ran MD simulations of three randomly chosen inhibitor bound crystal structures of MDM2 (PDB IDs: 1T4E, 3VZV, 3LBK). Before starting the simulation, the rotamer of His96 was set to the open position. These simulations also demonstrate that His96 samples the closed state in the presence of a small molecule inhibitor (Figure 2.6B).

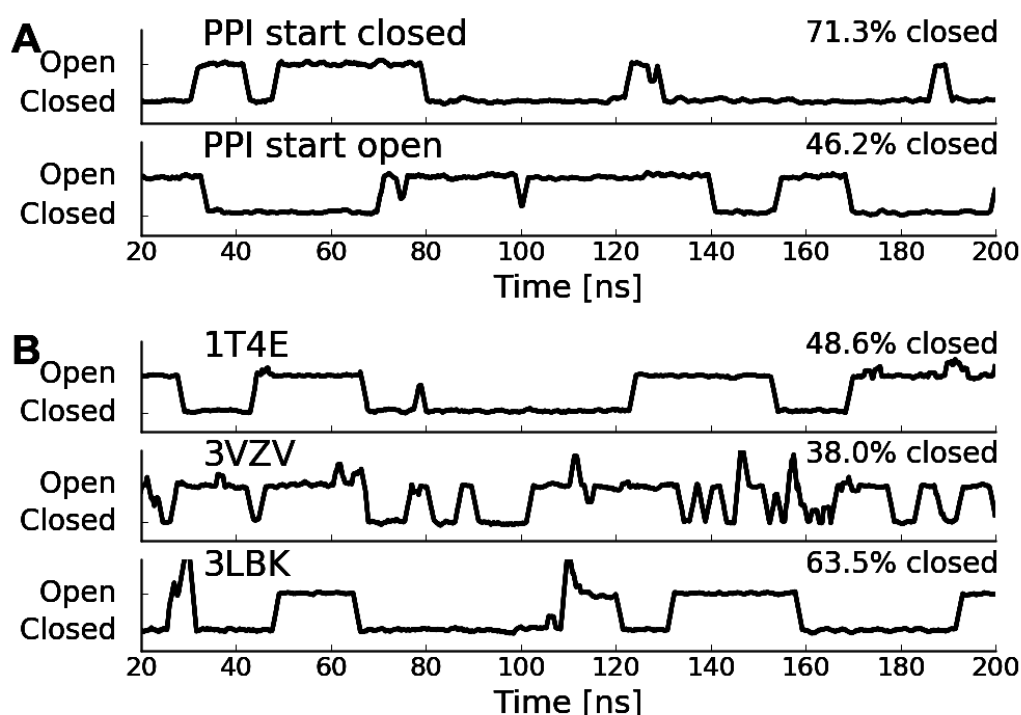


Figure 2.6 Following final complexation, His96 rotamer returns to the closed state. A) Dihedral angle of His96 in simulations of the PPI structure of MDM2-p53 (PDB ID: 1YCR) starting from the (top) native closed or (bottom) open rotamer of His96. Shown are the percentage of frames that His96 is in the closed state. B) Dihedral angle of His 96 in simulations of inhibitor bound (His96 closed) crystal structures of MDM2.



### 2.3.9 Examination of co-crystal structures of MDM2 inhibitors

To validate the importance of the CB and CA atoms in opening the Phe19 and Leu26 pockets with experimental data, we examined how many of the co-crystallized structures of known MDM2 inhibitors (N=28) contained a heavy atom in the same position (within 1.25 Å) as the CA and CB atoms of Trp23 (see Methods and Table 2.4). Figure 2.7 shows that 100% and 90% of the analyzed ligands have a heavy atom (typically carbon or nitrogen) in the same position as the CB and CA atoms, respectively. This agrees with our findings of the importance of the CB and CA atoms because only compounds with heavy atoms in those positions are able to open the binding site, as the Trp23 anchor does, successfully bind to the protein and be crystallized. This provides support for our simulation results and indicates that the coordination of the position and orientation of Val93 and His96 are critical for successful binding of small molecule inhibitors.

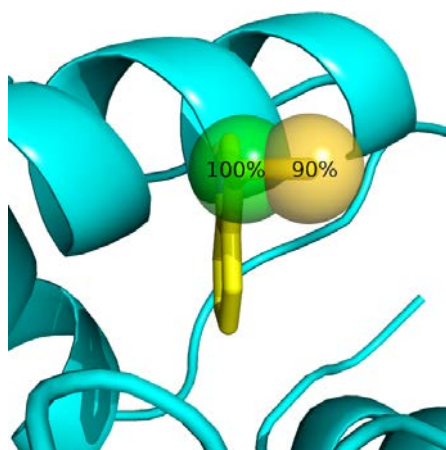


Figure 2.7 Co-crystallized inhibitors have CB and CA analogs. Percent of co-crystallized MDM2 inhibitors with heavy atoms within 1.25 Å of the CB (green sphere) and CA (orange sphere) atoms of Trp23. The radius of the spheres indicate the 1.25 Å region of overlap.

## 2.4 DISCUSSION AND CONCLUSION

The interaction between tumor-suppressor p53 and its inhibitor MDM2 have been shown to play a critical role in many human cancers[41]. Understanding, with atomic level detail, how these two proteins interact will provide useful insight into rational drug design efforts. When unbound, the MDM2 binding interface is closed, and the hydrophobic binding site is largely occluded (Figure 2.1B) and the p53 peptide is disordered. The binding site of the Trp23 anchor of p53 is transitively open allowing for initial recognition of the binding partners[30]. In this work, we have used molecular dynamics simulations to explore the exact factors of the Trp23 anchor of p53 that are important to fully opening the binding site and allowing for binding to occur.

Our MD simulations suggest that the binding of the Trp23 anchor residue is necessary to open the binding pocket by coordinating the position and orientation of Val93 and His96 through hydrophobic contacts of the Trp23 CB and CA carbons. These atoms regulate both the position and orientation of Val93 by pushing it into its bound-state position and orientation. We also have identified a putative transition state mechanism by which helix  $\alpha$ -4 shifts towards the bound-like state and the side chain of His96 rotates outwards, thus opening the pocket for binding. These interactions open the binding cavity enough such that the Phe19 and Leu26 residues can bind and further induce the structural transition to the high-affinity complex, the full extent of which is beyond the scope of this paper. We have computationally verified that coordinating the position and orientation of Val93 to its bound state and His96 to its open, transition state is critical for opening the MDM2 binding cleft and allowing for full binding of p53 and small molecule inhibitors.

We also examined the available crystallographic data of MDM2 inhibitors and have found that nearly all of them have a heavy atom in the same position as the CB and CA atoms.

While the sample size of the crystallized inhibitors is relatively small, this provides evidence for the importance of Val93 and His96 positioning. Thus, when designing small molecule inhibitors of MDM2 one must incorporate this critical interaction.

The impact of this work is two-fold. First, it provides a deeper understanding of the dynamic binding interaction of this important pair of proteins. Second, it provides insight into the rational design of small molecule inhibitors, of which the potential benefit to human health cannot be understated.

## 2.5 SUPPLEMENTARY FIGURES AND TABLES

Table 2.4 List of PDB IDs of inhibitor bound structures of MDM2. Dots indicate whether the aligned ligand has a heavy atom within 1.25 Å of the CB or CA atoms of Trp23 of p53 from PDB 1YCR. Three structures have been excluded from the analysis (indicated by an \* ) as they bind to MDM2 in a non-canonical (non “three-fingered”) fashion; we have listed them here for completeness.

PDB ID	CB	CA
1RV1[42]	•	•
1T4E[43]	•	
2LZG[44]	•	•
3JZK[45]	•	•
3LBJ[26]	•	•
3LBK[26]	•	•
3LBL[26]	•	•
3TJ2[46]	•	•
3TU1[47]	•	•
3VZV[25]	•	•
3W69[25]	•	•
4DIJ[48]	•	•
4ERE[27]	•	
4ERF[27]	•	•
4HBM[49]	•	•
4HG7[50]	•	•
4JV7*[51]		
4JV9*[51]		
4JVE*[51]		
4JVR[51]	•	•
4JWR[51]	•	•
4MDN[24]	•	•
4MDQ[24]	•	•
4OAS[52]	•	•
4OBA[53]	•	•
4OCC[54]	•	•
4ODE[54]	•	•
4ODF[54]	•	•
4OGN[54]	•	•
4OGT[54]	•	•
4OGV[54]	•	•
4OQ3[55]	•	•

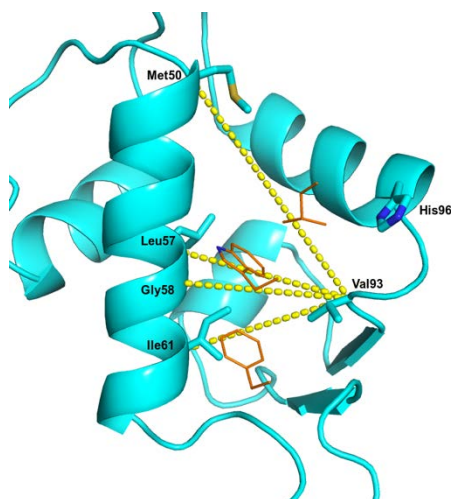


Figure 2.8 Distance restraints used in Val93 coordination simulation. Yellow dashed lines indicate the harmonic distance restraints used between the  $\alpha$ -carbons of the labeled residues.

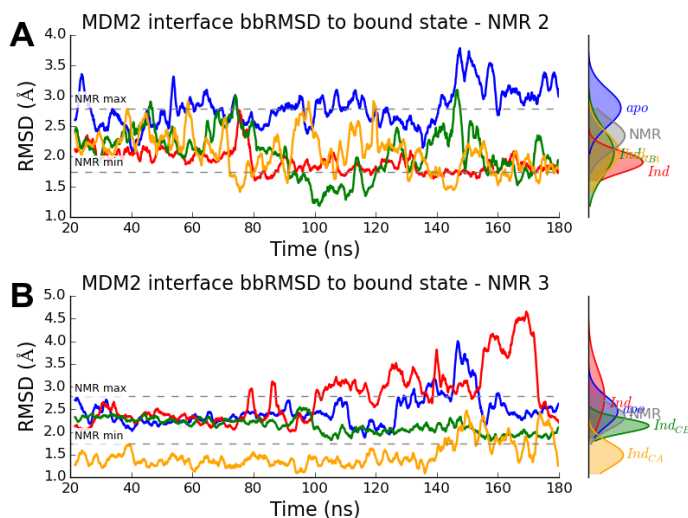


Figure 2.9 RMSD of backbone atoms of binding pocket residues to the bound state for the NMR2 and NMR3. RMSD results of simulations starting from the **A**) second and **B**) third NMR structure and distribution of the NMR structures. The lines are colored (here and throughout) as follows: *apo*: blue, *Ind*: red, *Ind<sub>CB</sub>*: green, *Ind<sub>CA</sub>*: yellow, NMR: grey. The horizontal dashed lines indicate the minimum and maximum RMSDs of the 24 NMR structures. The right side of the plot shows Gaussians fit to the data.

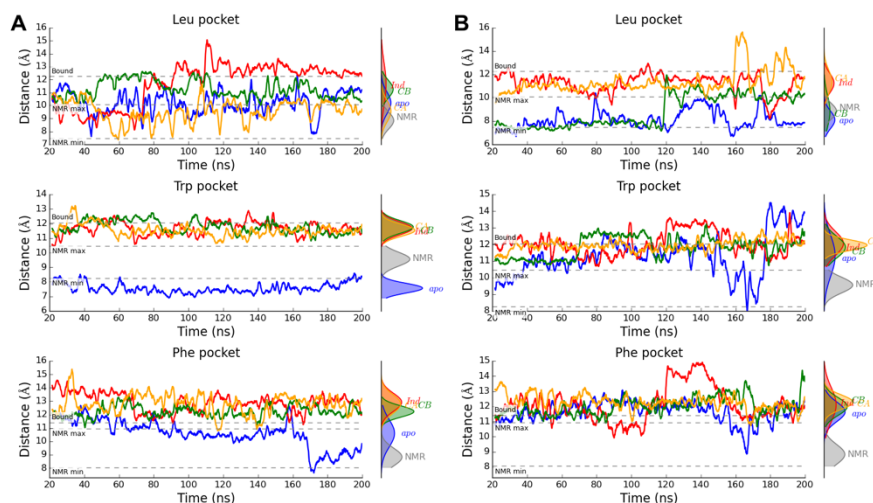


Figure 2.10 The distances measured for the Leu, Trp and Phe pockets NMR2 and NMR3 simulations. Measurements of the simulations of the A) second and B) third NMR structures. The horizontal dashed lines indicate the distances in the bound structure and the minimum and maximum distances observed in the 24 unbound NMR structures.

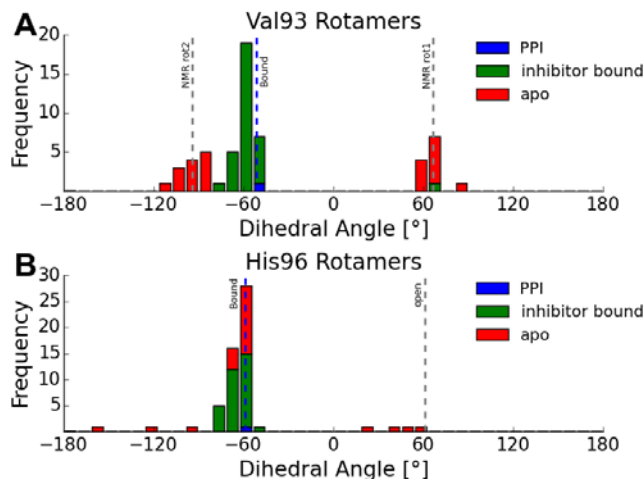


Figure 2.11 Distribution of dihedral angles of side chains of critical interaction residues. Distribution of dihedral angles observed for A) Val93 and B) His96 in the PPI structure, inhibitor bound and *apo* NMR structures. Vertical dashed lines indicate rotamer states referenced in the main text. In A) the “NMR rot1” and “NMR rot2” lines indicate the average values of the *apo*

structures whose dihedral angles are above or below 0, respectively. In B) the “Bound” line indicates the dihedral angle of the PPI structure and the “open” link indicates the dihedral angle of the fourth NMR

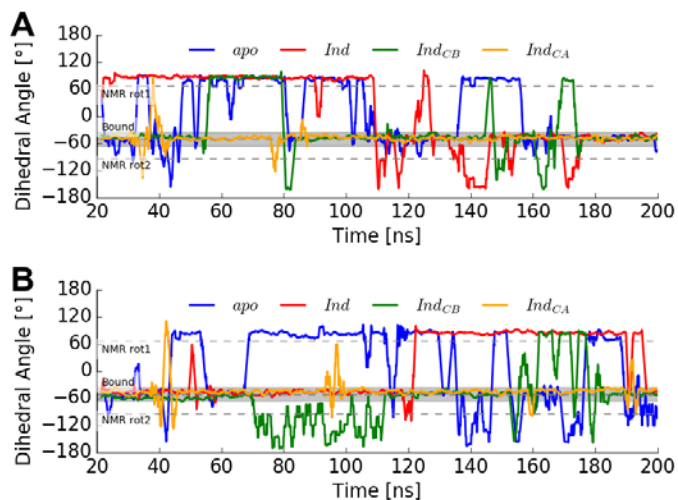


Figure 2.12 Dihedral angle of Val93 side chain in the NMR2 and NMR3 simulations. Val93 dihedral angle in simulations starting from the A) second and B) third NMR structure. The average dihedral angle of the two dominant unbound rotamer states and the rotamer angle of the bound state are shown as dashed grey lines. The shaded area indicates  $\pm 15^\circ$  of the bound state rotamer.

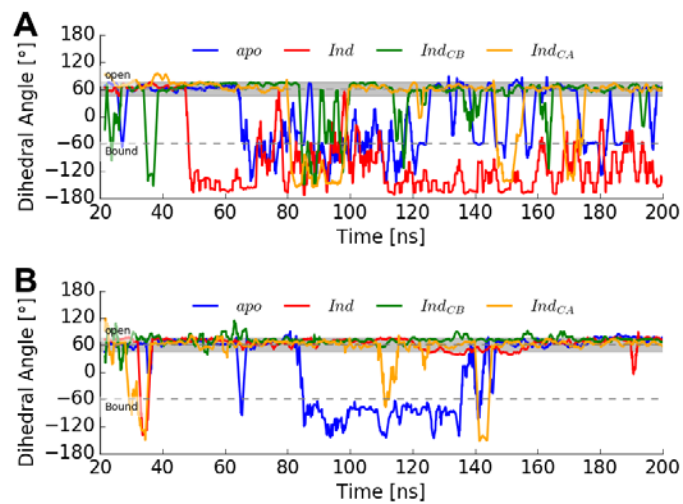


Figure 2.13 Dihedral angle of His96 side chain in the NMR2 and NMR3 simulations. His96 side chain dihedral angle in the simulations starting from the A) second and B) third NMR structure. The dihedral angles of the bound “closed” state and the “open” state are shown as dashed grey lines. The shaded area indicates  $\pm 15^\circ$  of the open state rotamer.



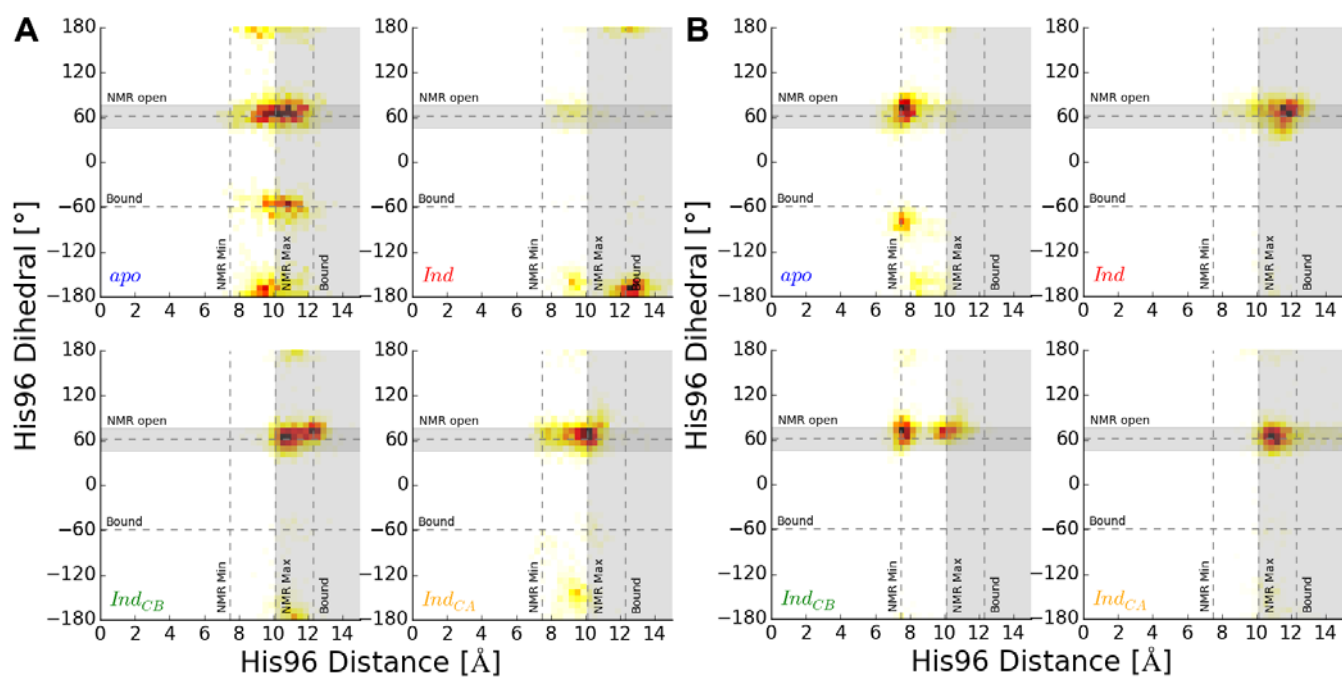


Figure 2.14 Heat map of the distance from the helix  $\alpha 2$  to the  $\alpha$ -carbon of His96 and the His96 dihedral angle. Data shown form the four simulation types of the (A) second and (B) third NMR structure (NMR2 and NMR3). Horizontal grey bars indicate values within  $15^\circ$  of the dihedral angle observed in the fourth NMR structure ( $62^\circ$ ). Vertical grey bars indicate distances greater than the largest distance observed in the 24 NMR structures. Dark and light colors indicate high and low frequencies, respectively.

### 3.0 KNOWLEDGE BASED INSIGHT OF LEVERAGING CRYSTAL WATER MOLECULES IN RATIONAL DRUG DESIGN

#### 3.1 BACKGROUND

Drug discovery efforts have traditionally relied on high throughput screening (HTS) experiments but HTS has several major drawbacks that are increasingly motivating other approaches[2]. Although HTS has been successfully used in a number of drug discovery efforts, it can be an extremely expensive approach with a fairly low accuracy for new drug targets[56]. In recent years, *in silico* virtual screening methods have become popular complements to HTS due to their many advantages including speed and relatively low cost[29, 56-61]. Virtual screening methods can provide great benefits as a prescreening approach to drastically reduce the size of a compound library to a small fraction that is highly enriched for compounds that are likely to be effective, thus resulting in savings in time, resources and money.

There are many different virtual screening approaches that have been developed but they can be generally categorized into ligand-based methods or structure-based methods[62]. Ligand-based methods characterize given ligands based on similarity to known active and inactive compounds[61] and typically do not require knowledge of the protein structure. Examples of ligand-based virtual screening methods include ligand-based pharmacophore modeling, quantitative structure-activity relationships (QSAR) approaches and molecular fingerprinting methods[56]. In contrast, structure-based methods incorporate information about the 3D protein structure. Examples of structure-based methods include molecular docking, structure-based pharmacophore models and scoring and energy minimization[62]. Ligand-based methods are

particularly powerful when there are a large number of active and inactive compounds for the target that are already known. However, if there are few or no known actives or inactives, as is the case for novel targets, structure-based methods are more applicable.

X-ray crystallography is often used to determine the three dimensional (3D) structure of macromolecules such as protein and DNA. Briefly, X-ray crystallography is conducted by exposing a purified crystal of the protein of interest to a beam of X-rays that diffract off the atoms of the protein and hit a detector. The position of the atoms of the protein can be calculated based on the pattern of the diffracted X-rays. In addition to determining the position of the atoms of the protein, often times X-ray crystallography is able to resolve the position of individual water molecules which are interacting with the protein through one or more hydrogen bonds. These water molecules are often referred to as “crystal waters.”

We will describe our efforts to leverage information of these water molecules in the context of virtual screening efforts. Briefly, we examine pairs of crystal structures of proteins that are bound to a small molecule inhibitor and in their unbound, or *apo*, state. In the first phase, we classify clusters of crystal waters based on their geometry and interactions with the protein. In the second phase, we show that certain types of water clusters often overlap specific ligand functional groups observed in the inhibitor-bound crystal structures. From this we can create a mapping between water cluster types and ligand chemistries. This has useful applications in structure-based virtual screening efforts particularly in fragment-based methods. While conducting a virtual screen, if a given water cluster is observed in a crystal structure of the protein of interest, the compound database that is being screened can be limited to just those that have the corresponding functional group.

## 3.2 DATASET ASSEMBLY

A database of over 23,000 pairs of unbound (*apo*) and ligand-bound (*holo*) protein structures was assembled from the AH-DB (Apo and Holo structures DataBase; <http://ahdb.ee.ncku.edu.tw/>)[63]. The following options were used for searching the AH-DB: Target proteins: all proteins in all organisms, Added molecules: without protein and without nucleic acid and with ligand, Miscellaneous: *apo* single protein and exclude NMR and technology consistency and completeness  $\geq 95\%$  and ML-superimposition. A permanent link to the search results is available here: <http://ahdb.ee.ncku.edu.tw/53099>. These filters ensure that the resulting pairs of structures only contain a single (and the same) protein without any other biomolecules such as DNA or other proteins.

The initial search resulted in 47,202 pairs of structures. This initial set was subjected to a series of filters which eliminated pairs of structures for a variety of reasons, such as when the bound structure contained multiple copies of the ligand or protein or where the ligand name was the same as a standard amino acid. Pairs of structures were also eliminated if the *apo* structure failed to align to the *holo* structure or if there were no water molecules resolved in the binding site of the *apo* structure. After filtering, we are left with 23,654 pairs of structures which consisted 1,729 *holo* and 1,424 *apo* structures.

## 3.3 WATER CLUSTER CHARACTERIZATION

For each *apo* structure, we examined all of the crystallographic water molecules observed in the entire protein structure. For each water molecule, we identified and counted the number of polar

contacts with the protein or other water molecules. Polar contacts were defined as polar atoms on the protein (e.g. oxygen and nitrogen) that are within specified distance and angle constraints (Figure 3.1). In this case, the distance cutoff was set at 4.0 Å and the angle cutoff (donor – donor hydrogen – acceptor) was set to 100°. Due to limitations in X-ray crystallography, the position of the hydrogen atoms of the protein or water are unable to be determined. There are a number of methods for adding hydrogen to protein structures that are quite reliable[64], however, because of the inherent number of degrees of freedom, most methods perform rather poorly at adding hydrogens to water molecules. Due to this limitation, we chose to not add hydrogens to the water molecules and assume that if a bond is otherwise reasonable, we accepted it. Because we can only calculate bonding angles when the donor is on the protein, we call this definition “semi-directional” as opposed to “directional” or “non-directional” bonds. Later we will use the number of polar contacts with the protein as a primary characteristic of the clusters. Figure 3.2 shows the distribution of the number of polar contacts in all of the waters in the dataset.

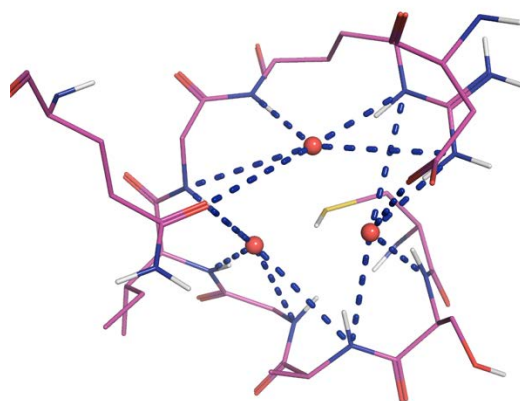


Figure 3.1 Polar contacts between crystal waters and the protein surface. Each crystal water (red spheres) has zero or more polar contacts (blue dashed lines) with the protein surface (purple lines).

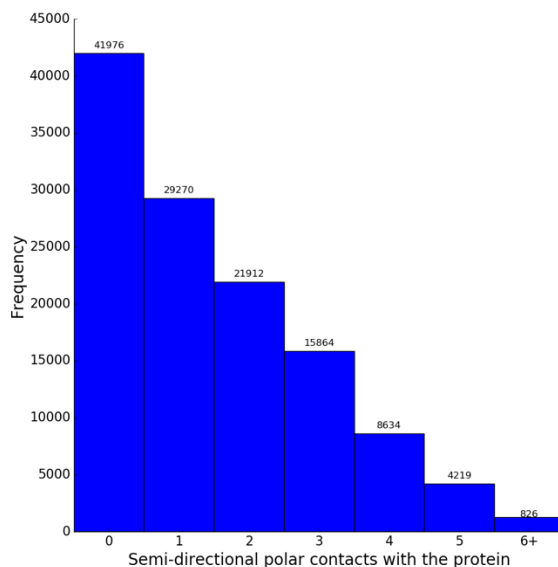


Figure 3.2 Distribution of the number of polar contacts with the protein by binding site waters. Note that these do not include polar contacts with other water molecules.

Clusters of waters were assembled by grouping all sets of water molecules which are within a specified cutoff distance of each other (4.0 Å). If there are no other water molecules within the specified cutoff, the water is put into its own cluster (Figure 3.3). The “size” of the water cluster is defined as the number of water molecules in it, typically 1-3. Figure 3.4 shows the distribution of water cluster sizes.

In the future analysis, we will use these to general descriptors to classify the water clusters. In particular, the water clusters will be classified based on their size and average number of hydrogen bonds with the protein. To simplify the discussion, we will use a parenthetical notation to describe a water cluster such as (2, 3), meaning the cluster contains two water molecules which have an average of three polar contacts with the protein. Note that because this is an average, the number of bonds may not be evenly distributed between the water molecules, i.e. one water might

have four bonds and the other might have two bonds. Also worth emphasizing is the fact that the average number of hydrogen bonds is *with the protein only*. Water clusters of size two or more have bonds between the waters in the cluster, but those bonds are excluded in this analysis.

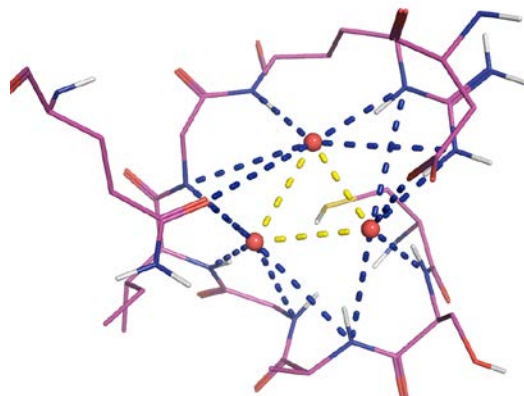


Figure 3.3 Defining water clusters on the protein surface. Each cluster is defined as water molecules within hydrogen bonding range of each other, if any. Yellow dashed lines indicate intra-water hydrogen bonds, blue dashed lines indicate hydrogen bonds to the protein (shown in purple lines).

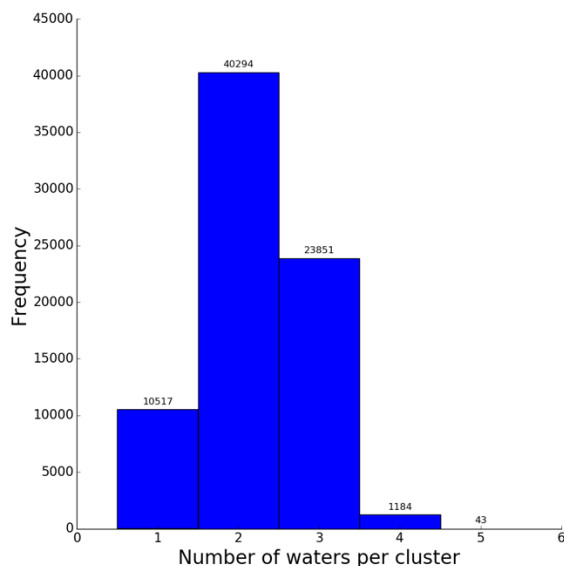


Figure 3.4 Distribution of the sizes of binding site water clusters

### 3.4 CHARACTERIZING LIGAND FUNCTIONAL GROUPS

For each ligand in the bound structures of the dataset, we determined what functional groups that it consisted of and which atoms contributed to each functional group. We compiled a list of functional group definitions starting from a list of SMARTS patterns[65] obtained from the ‘maccs.txt’ file provided in OpenBabel v2.3.90[66]. SMARTS patterns can be thought of regular expressions for small molecules. For example, the SMARTS pattern '[N,n]=\*' looks for a nitrogen ‘[N,n]’ that is double bonded ‘=’ to anything ‘\*’. Ligands that contains a nitrogen double-bonded to a carbon (N=C) would match this SMARTS pattern as would a ligand that contains a nitrogen double-bonded to another nitrogen (N=N). The full list of 50 functional groups that we used for the initial set this work are listed in Table 3.1. We excluded a number of the 166 functional groups originally defined in the ‘maccs.txt’ file that were noted in the original file as being incorrect or those that did not match any of the ligands in our dataset.



Table 3.1 List of functional groups considered for ligand characterization. In the comment section are human readable descriptions of the SMARTS pattern. Parentheses indicate branches, ‘Q’ indicates non-carbon or hydrogen atoms, ‘A’ indicates any atom, ‘X’ indicates a halogen, ‘-‘ indicates a single bond, ‘=’ indicates a double bond and ‘T’ indicates a triple bond. Bolded functional groups indicate those that had at least 100 representatives in the dataset.

ID	SMARTS Pattern	Comment	ID	SMARTS Pattern	Comment
1	'[O,o]~[N,n](~[C,c])~[C,c]'	ON(C)C	26	'[NH2]'	NH2
2	'[O,o]~[C,c](~[O,o])~[O,o]'	OC(O)O	27	'[C,c]~[N,n](~[C,c])~[C,c]'	CN(C)C
3	'[N,n]~[C,c](~[O,o])~[O,o]'	NC(O)O	<b>28</b>	'[S,s]'	<b>S</b>
4	'[N,n]-[O,o]'	N-O	29	'[C,c]=[C,c]'	C=C
<b>5</b>	'[N,n]~[C,c](~[N,n])~[N,n]'	<b>NC(N)N</b>	<b>30</b>	'[F,Cl,Br,I]~*(~*)~*'	<b>XA(A)A</b>
6	'[!C;!c;!#1]~[CH2]~[!C;!c;!#1]'	QCH2Q	31	'[N,n]~[C,c]~[O,o]'	NCO
<b>7</b>	'[C,c]~[S,s]~[N,n]'	<b>CSN</b>	32	'[CH3]~[CH2]~*'	CH3CH2A
<b>8</b>	'[N,n]~[S,s]'	<b>NS</b>	<b>33</b>	'[N,n]~*~[O,o]'	<b>NAO</b>
9	'[N,n]~[C,c](~[O,o])~[N,n]'	NC(O)N	<b>34</b>	'[N,n]=*'	<b>N=A</b>
<b>10</b>	'[N,n]~[C,c](~[C,c])~[N,n]'	<b>NC(C)N</b>	35	'[O,o]~[C,c]~[O,o]'	OCO
<b>11</b>	'[O,o]~[S,s](~[O,o])~[O,o]'	<b>OS(O)O</b>	<b>36</b>	'[F,Cl,Br,I]'	<b>X (HALOGEN)</b>
<b>12</b>	'[S,s]-[O,o]'	<b>S-O</b>	37	'[!C;!c;R]'	Heterocycle
13	'[C,c]#[N,n]'	CTN	38	'[OH,OH2,OH3]'	OH
14	'[S,s]~*~[N,n]'	SAN	<b>39</b>	'[N!H0]'	<b>NH</b>
15	'[C,c]~[S,s]~[O,o]'	CSO	40	'[O,o]~[C,c](~[C,c])~[C,c]'	OC(C)C
<b>16</b>	'[O,o]~[S,s]~[O,o]'	<b>OSO</b>	<b>41</b>	'[C,c]=[O,o]'	<b>C=O</b>
17	'[O,o]~[N,n](~[O,o])~[C,c]'	ON(O)C	<b>42</b>	'[C,c]-[O,o]'	<b>C-O</b>
<b>18</b>	'[\$(O@*),\$(o@*)]'	<b>O Heterocycle</b>	43	'[C,c]-[N,n]'	C-N
<b>19</b>	'[S,s]=[O,o]'	<b>S=O</b>	<b>44</b>	'[CH3]'	<b>CH3</b>
<b>20</b>	'*~[S,s](~*)~*'	<b>AS(A)A</b>	<b>45</b>	'[N,n]'	<b>N</b>
<b>21</b>	'[C,c]~[C,c](~[C,c])(~[C,c])~*'	<b>CC(C)(C)A</b>	<b>46</b>	'a'	<b>Aromatic</b>
22	'[N,n]~[O,o]'	NO	<b>47</b>	'[O,o]'	<b>O</b>
23	'[CH3]~*~[CH3]'	CH3ACH3	<b>48</b>	'[C,c]'	<b>any carbon</b>
24	'[C,c]=[C,c](~*)~*'	C=C(A)A	49	'[C,c]~[C,c](~[O,o])~[N,n]'	CC(O)N
<b>25</b>	'[C,c]=[N,n]'	<b>C=N</b>	<b>50</b>	'[O,o]~[P,p](~[O,o])(~[O,o])~[O,o]'	<b>OP(O)(O)O</b>

For each atom in the ligand, we are able to determine which of the functional groups listed in Table 3.1 it is a part of. From this information, we are able to create correlations between the water clusters and the ligand functional groups that displace them.

### 3.5 WATER CLUSTER DISPLACEMENT BY LIGAND FUNCTIONAL GROUPS

For each functional group in a ligand, we can determine what water clusters in the corresponding *apo* structures overlap with it. In doing so, we can characterize water clusters found in the *apo* structure as being in the binding site or not. Figure 3.5A shows the distribution of all of the water clusters in the dataset based on the size of the cluster (number of waters) and the average number of polar contacts with the protein. Comparatively, Figure 3.5B shows the distribution of only binding site water clusters and Figure 3.5C shows the proportional difference between the water clusters in the binding site compared to the water clusters in the entire protein structure. We observe that there is a striking difference in the distribution of water clusters inside and outside the binding site. Highly coordinated pairs of waters i.e. (2, 2.5) and (2, 3) are much more prevalent in the binding site relative to throughout the entire protein. Additionally, pairs and trio of water with no (or few) polar contacts are more common in the binding site. These “unstable” waters are typically in hydrophobic pockets and are generally considered prime targets for displacement by an inhibitor[67]. Conversely, single waters, of all coordination states are more common outside the binding site.

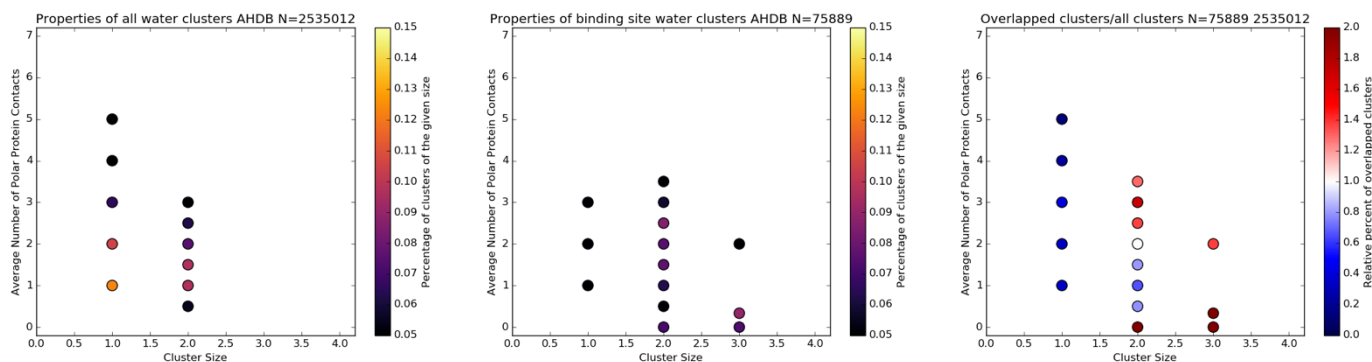


Figure 3.5 Distribution of water cluster sizes and average polar contacts. Distribution of A) all or B) binding site water clusters based on cluster size and the average number of polar contacts with the protein. Lighter colors indicate higher proportion of water clusters. Clusters that represent less than 2.5% of all clusters in each set are hidden for clarity. C) The proportional difference between binding site water clusters and all water clusters. Deeper red colors indicate clusters that are more prevalent in the binding site and deeper blue colors indicate clusters that are more prevalent outside the binding site.

For each of the functional groups listed in Table 3.1, we analyzed the distribution of water clusters that displace it in the dataset. From this list of 50 functional groups, 30 were represented at least 100 times (shown in bold in the table) and were further analyzed. The water clusters that most commonly displaced each functional group is shown in Table 3.2. Interestingly, all of the most commonly displaced water clusters are of size 1 or 2, despite the relative abundance of clusters of size 3 (Figure 3.4).

Table 3.2 Most common water clusters that are displaced by functional groups with at least 100 examples

Functional Group	Size	Avg. Contacts	% represented	Total
AS(A)A	2	3.0	21.89%	3851
Aromatic	1	0.0	24.10%	1311
C-O	2	0.0	31.68%	2386
C=N	2	1.5	27.70%	473
C=O	1	1.0	25.83%	751
CC(C)(C)A	2	1.5	21.78%	101
CH3	1	2.0	21.55%	297
CSN	2	3.0	24.27%	2678
CSO	2	2.5	22.07%	2633
CTN	1	1.0	54.47%	123
N	1	3.0	20.83%	1805
N=A	2	1.5	26.36%	497
NAO	2	3.0	20.46%	3529
NC(C)N	2	1.5	28.45%	116
NC(N)N	1	1.0	24.77%	109
NH	1	3.0	27.83%	1247
NH2	1	3.0	33.54%	957
NS	2	3.0	26.05%	2668
O	1	2.0	21.17%	3434
O Heterocycle	2	1.5	25.00%	100
OP(O)(O)O	1	2.0	26.72%	524
OS(O)O	2	2.5	21.12%	805
OSO	2	2.5	22.85%	3335
S	2	0.0	27.13%	4427
S-O	2	2.5	22.15%	727
S=O	2	2.5	22.86%	3150
X (HALOGEN)	1	0.0	32.69%	104
XA(A)A	1	0.0	30.95%	294
any carbon	2	0.0	31.35%	5972

An inspection of the ligands that contribute to each functional group revealed that several functional groups were identifying the same chemistry. For example, the functional groups AS(A)A, CSO, CSN, NAO, OSO, S=O, and S-O can all match a common sulfonamide chemistry: CS(O)(O)N (as in Figure 3.6B). Due to these overlapping descriptors, we manually selected a

handful of representative functional groups for further analysis: AS(A)A, NC(C)N, Aromatic, and X (Halogen).

An analysis of the water clusters that overlap the functional group AS(A)A revealed a large propensity for displacement by (2, 2.5) and (2, 3) water clusters (Figure 3.6A). Figure 3.6B and C show two examples of ligand chemistries that are displaced by (2, 3) water clusters. The geometry of the water clusters is well matched to the chemistry of the ligands that displace them in the bound structure. These highly coordinated waters form many well defined hydrogen bonds that are well suited to replacement by the AS(A)A functional group.

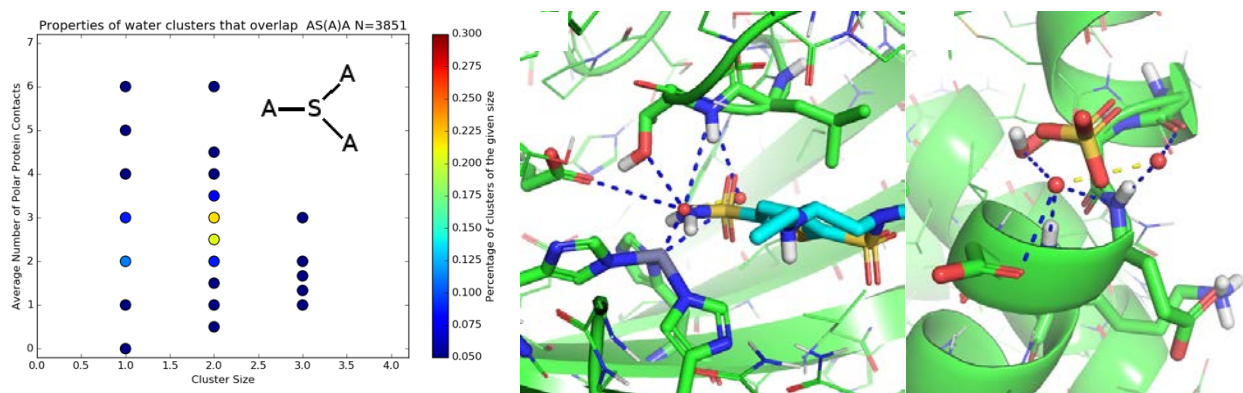


Figure 3.6 Distribution and examples of water clusters that overlap the AS(A)A functional group

A) Heat map of the water clusters that overlap the AS(A)A functional group where “A” is any atom; (inset) chemical sketch of the functional group. Warmer colors indicate higher percentages.

B) Inhibitor from a *holo* structure (light blue sticks; PDB ID: 1A42[68]) overlaid on the *apo* crystal structure of carbonic anhydrase II and a (2, 3) water cluster (PDB ID: 1FQL[69]). C) Small molecule from a *holo* structure (PDB ID: 1VXA[70]) overlaid on the *apo* crystal structure of myoglobin and a (2, 3) water cluster (PDB ID: 3H57[71]).

The NC(C)N functional revealed a high propensity for displacing (2, 1.5) water clusters (Figure 3.7A). An analysis the pairs of crystal structures revealed a variety of ligand chemistries that displaced the (2, 1.5) water clusters. Figure 3.7B shows one prime example of the NC(C)N functional group displacing the water cluster. In this case, the nitrogen atoms overlap the positions of the waters and the carbons form the connecting bond between the two fused aromatic rings. This overlap and replacement of the loosely coordinated (2, 1.5) water clusters is similarly observed in other protein families such as protein kinase CK2 shown in Figure 3.7C.

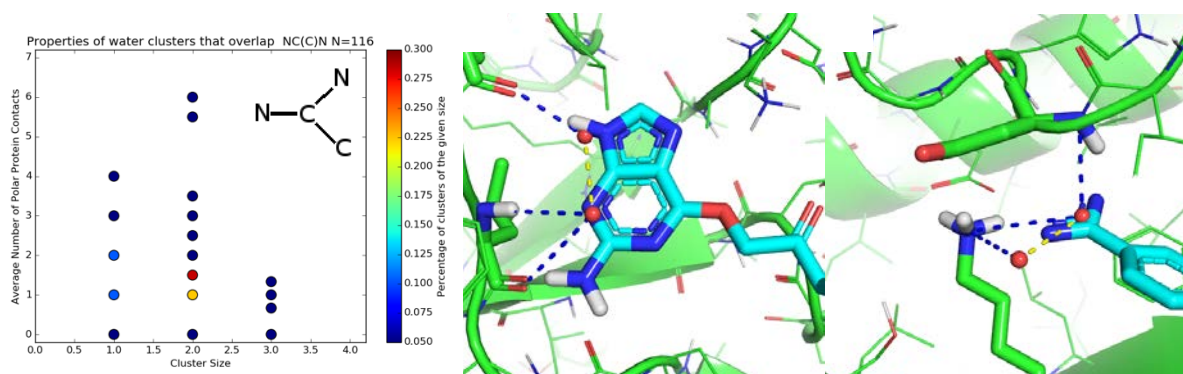


Figure 3.7 Distribution and examples of water clusters that overlap the NC(C)N functional group. A) Heat map of the water clusters that overlap the NC(C)N functional group; (inset) chemical sketch of the functional group. Warmer colors indicate higher percentages. B) Inhibitor from a *holo* structure (light blue sticks; PDB ID: 1GZ8[72]) overlaid on the *apo* crystal structure of cyclin dependent kinase 2 (CDK2) and a (2, 1.5) water cluster (PDB ID: 1HCL[73]). C) Small molecule from a *holo* structure (light blue sticks; PDB ID: 1LPU[74]) overlaid on the *apo* crystal structure of protein kinase CK2 and a (2, 1.5) water cluster (PDB ID: 1JAM[75]).

In addition to the polar functional groups mentioned above, (AS(A)A and NC(C)N) we also observed displacement of water molecules by hydrophobic functional groups (e.g. aromatics and halogens). Perhaps non-surprisingly, these hydrophobic functional groups most often displace

single water molecules that have very few or no hydrogen bonds with the protein (Table 3.2). These water molecules are often found in hydrophobic cavities where they are entropically unfavorable. Thus their displacement from the non-polar cavity by a hydrophobic functional group can be easily rationalized. Figure 3.8A shows that aromatic functional groups most often displace (1, 0) water clusters. This is closely followed by (1, 1) and (2, 1.5) both of which are poorly coordinated and ripe for displacement by an aromatic group. Figure 3.8B and Figure 3.8C show two prime examples of unstable water molecules in hydrophobic pockets being displaced by aromatic groups on the ligand.

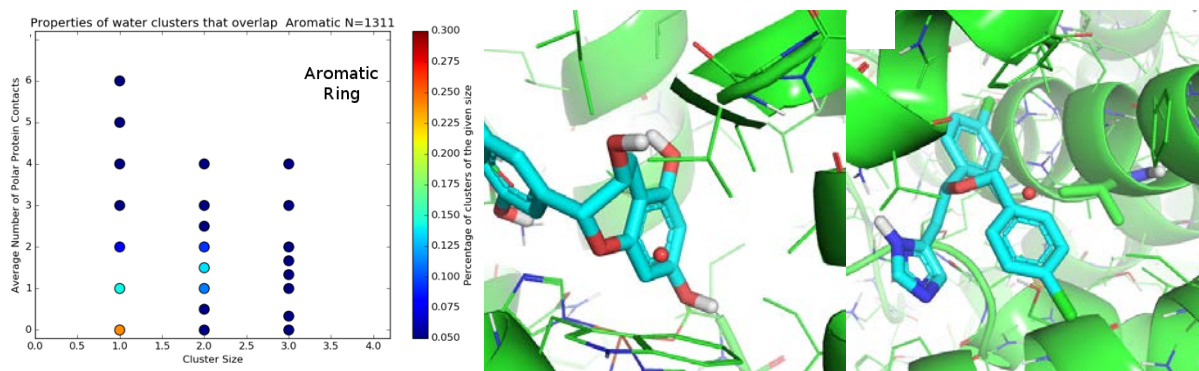


Figure 3.8 Distribution and examples of water clusters that overlap the aromatic functional group. A) Heat map of the water clusters that overlap the aromatic functional group; (inset) chemical sketch of the functional group. Warmer colors indicate higher percentages. B) Inhibitor from a *holo* structure (light blue sticks; PDB ID: 1JNQ[76]) overlaid on the *apo* crystal structure of lipoxygenase-3 and a (1, 0) water cluster (PDB ID: 1LNH[77]). C) Small molecule from a *holo* structure (light blue sticks; PDB ID: 3IW2[78]) overlaid on the *apo* crystal structure of cytochrome P450 and a (1, 0) water cluster (PDB ID: 3IVY[78]).

Finally, we examined the water clusters that displace halogens (F, Cl, Br and I). Much like aromatic groups, halogens preferentially displace single water molecules (cluster size = 1), however the water clusters exhibit higher variability in their number of polar contacts (Figure 3.9A). Unmatched water molecules are the most common followed by waters that have 1, 2 and 3 contacts with the protein. Examples of the ligands that displace the most common (1, 0) water clusters is shown in Figure 3.9B and Figure 3.9C. The hydrophobic halogens are able to displace the unstable water molecules and fill the non-polar cavities.

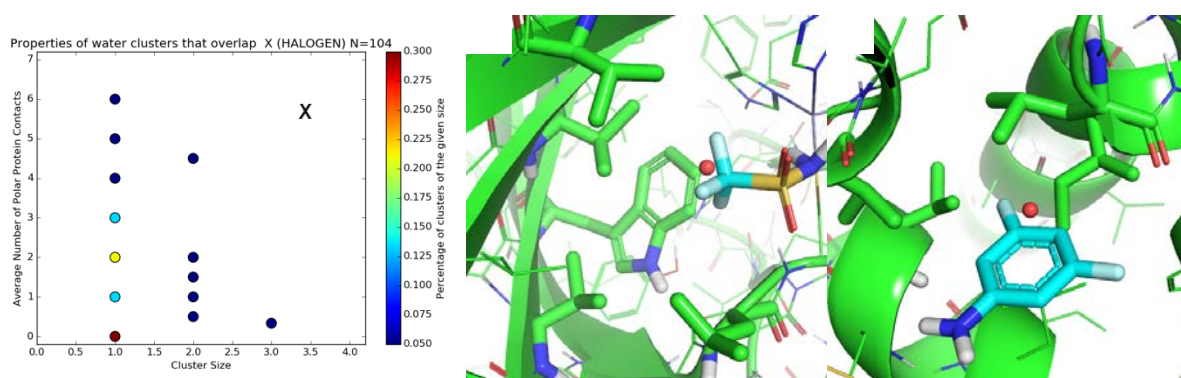


Figure 3.9 Distribution and examples of water clusters that overlap the X (Halogen) functional group. A) Heat map of the water clusters that overlap the X (Halogen) functional group; (inset) chemical sketch of the functional group. Warmer colors indicate higher percentages. B) Inhibitor from a *holo* structure (light blue sticks; PDB ID: 1BCD[79]) overlaid on the *apo* crystal structure of carbonic anhydrase II and a (1, 0) water cluster (PDB ID: 1DCA[80]). C) Small molecule from a *holo* structure (light blue sticks; PDB ID: 1LGX[81]) overlaid on the *apo* crystal structure of T4 lysozyme and a (1, 0) water cluster (PDB ID: 157L[82]).



### 3.6 SUMMARY

We have combined data mining and protein structural analysis to discover clusters of water molecules in X-ray crystal structures that are commonly displaced upon binding of a small molecule inhibitor. We compiled a dataset of 23,654 pairs of crystal structures in their unbound (*apo*) and bound (*holo*) states. We grouped water molecules in the *apo* structure into clusters and characterized them according to their size (the number of waters) and the number of semi-directional hydrogen bonds observed with the protein. Clusters of water in inhibitor binding sites are generally larger and more coordinated compared to clusters of water found outside the binding site.

We used generalized descriptors of ligand chemistry (SMARTS patterns) to categorize the functional groups present in the ligands. By comparing the functional groups of the ligand in the *holo* structure with the directly overlapping water clusters from the *apo* structure, we were able to draw correlations between the two. Here we discussed four general functional groups: two polar (AS(A)A and NC(C)N) and two non-polar (Aromatic and X (Halogen)). All four functional groups have large numbers of representative examples in the dataset (>100 each). This allows us to draw strong associations between the geometry of the water clusters in the *apo* structure with the ligand functional groups that are likely to displace them.

These findings have a number of applications in the realm of structure-based and fragment-based drug design. For example, if a certain type of water cluster, which is associated with a functional group as described above, is observed in an existing crystal structure of the target, you can limit the search results of the virtual screen to only include ligands with the functional group of interest in that site. Additionally, these finding might lead to the indentification of previously unknown allosteric binding sites, however more investigation is needed.

There are a number of improvements of this method planned or under way. Expansion of the initial set of queried functional groups beyond those found in the MACCS keys (see Table 3.1) might allow us to discover more nuanced functional groups. The four functional groups that we have analyzed in detail do not constitute an exhaustive list of functional groups which are amenable to this analysis methods.

Currently, the water clusters are being characterized by two simple descriptors: the number of waters in the cluster and the average number of hydrogen bonds that they make with the protein. The creation of more complex descriptors, such as description of the distances and angles between the protein contacts, would allow us to perform a finer grain dissection of the types of water clusters that can be displaced by certain functional groups. This is also likely to be necessary to support the ongoing efforts provide the predictive model described above in a virtual screening setting.

## 4.0 DEVELOPMENT OF TOOLS TO INCORPORATE EXPERT KNOWLEDGE IN VIRTUAL SCREENING

### 4.1 BACKGROUND

#### 4.1.1 High-Throughput Screening

Structure-based drug design is an efficient and effective method for computationally predicting compounds that are likely to bind to and modulate the activity of a target protein[5, 29, 57, 58, 62, 83]. In contrast to ligand-based drug design methods, structure-based methods use information from the target protein structure to predict compounds that are likely inhibitors[3, 59, 61]. Structure-based methods do not require prior knowledge of active compounds, which makes structure-based drug design the preferred method for predicting modulators of targets with no previously known inhibitors/activators.

Historically, pharmaceutical companies have relied on high-throughput screening (HTS) to discover novel inhibitors (or activators) of a protein of interest. In general, HTS works as follows: a proprietary library of compounds (~1 million) are assayed against a protein *in vitro* by adding the drug to a vial containing the protein (typically using liquid handling robots) and measuring the effect either biochemically or in cellular assays[84]. Compounds that produce the desired effect are called “hits”. While HTS has worked successfully in the past[84], the cost often reaches into the millions of dollars[85].

The cost of performing a high-throughput screen comes from a number of sources. The first is producing and maintaining the libraries of compounds, which often number in the hundreds

of thousands or millions[84]. Each compound requires a significant investment in time, effort and money to create (or recreate if you run out). Because of this cost, the compound libraries created by companies are considered proprietary information.

The second major cost of HTS is acquiring sufficient quantities of the substrate, typically purified protein. Obtaining purified protein in the quantities required for HTS is non-trivial for most proteins and in some cases, such as membrane proteins or proteins that are hard to separate from similar proteins, purification is currently impossible[86]. These challenges limit the types of proteins that can be targeted with *in vitro* HTS methods and lend a great deal of cost to the process, especially early in the discovery process.

Another major cost driver is in the equipment required to perform the HTS. Typically, HTS is performed by one or more robotic platforms that include functions for sample preparation, mixing and result detection. Some of these integrated platforms are able to test up to 100,000 compounds per day. However, these robotic platforms are extremely expensive to purchase up front which typically limits smaller academic researchers from harnessing them.

One of the main advantages of high-throughput screening is the speed by which they can screen compounds. However, being able to screen hundreds of thousands of compounds per day often comes at the cost of the fidelity of the assay. Simpler assays that provide less information (or less accurate information) but can be greatly scaled up are often favored. This tradeoff of speed versus accuracy is required in order to efficiently screen the millions of compounds in a company's chemical library. Unfortunately, this may come at the cost of both false positives and false negatives, meaning that time might be wasted on compounds that are not actually hits or that an actual hit may have been accidentally discarded.

In all, high-throughput screening has historically been useful in identifying “hits,” but at a great financial cost due to the need to test the entirety (or a sizable fraction) of the company’s chemical library. Creating and maintaining the chemical library, purifying sufficient protein and buying and maintaining robotics systems all lead to the often multi-million dollar cost of performing a high-throughput screen. This cost strains pharmaceutical and biotech companies that are already facing the increasing cost of drug discovery[84]. Fortunately, there are well-established and developing computational methods such as virtual screening, that allow for computationally testing the compounds before actually performing the expensive experiments, thus providing a much smaller set of compounds enriched for potential inhibitors that can be experimentally tested with higher fidelity assays at a greatly reduced overall cost.

#### **4.1.2 Virtual Screening**

Over the last five to ten years, the Camacho Lab at the University of Pittsburgh has developed a comprehensive computational pipeline for structure-based drug discovery. These open-access tools represent a start-to-finish method of identifying potential inhibitors for novel protein-protein interactions (PPIs)[29, 57, 58].

Figure 4.1 illustrates the structure-based virtual screening pipeline that we employ in our drug discovery efforts. Given the structure of a PPI of therapeutic interest, key interaction residues are identified using the PocketQuery web service[58]. PocketQuery analyzes the interacting residues of the PPI and predicts a set of residues that are likely to be good starting points for a structure-based drug discovery effort. The identified residues can be exported with one click to ZincPharmer[57] or AnchorQuery[29]. ZincPharmer allows for extremely efficient searching of large databases of commercially available compounds (i.e. the ZINC database[87],

~35 million compounds) using a set of pharmacophores. Pharmacophores are general descriptors of a molecule such as hydrogen bond donors/acceptors, hydrophobic groups and positive/negative charges. In ZincPharmer, a database of compounds is searched using a set of pharmacophore features and compounds matching those features are returned. AnchorQuery is similar to ZincPharmer in that it also uses pharmacophores, but it searches a database of compounds (~31 million) that are biased to targeting protein-protein interactions. This is accomplished by incorporating the chemistry of an “anchor” residue in the compound. Anchor residues (typically, tryptophan, phenylalanine or leucine) play an important role in protein-protein binding by binding deeply in the binding interface. Additionally, all of the compounds in the AnchorQuery database are synthetically available as a one-step, multi-component chemical reaction[29]. Both AnchorQuery and ZincPharmer allow for extremely fast searches of billions of conformers of 30+ million compounds through the use of specialized data structures[29, 57]. In most cases, the searches complete in under a minute, and are often much faster.

## A Validated Virtual Screening Approach

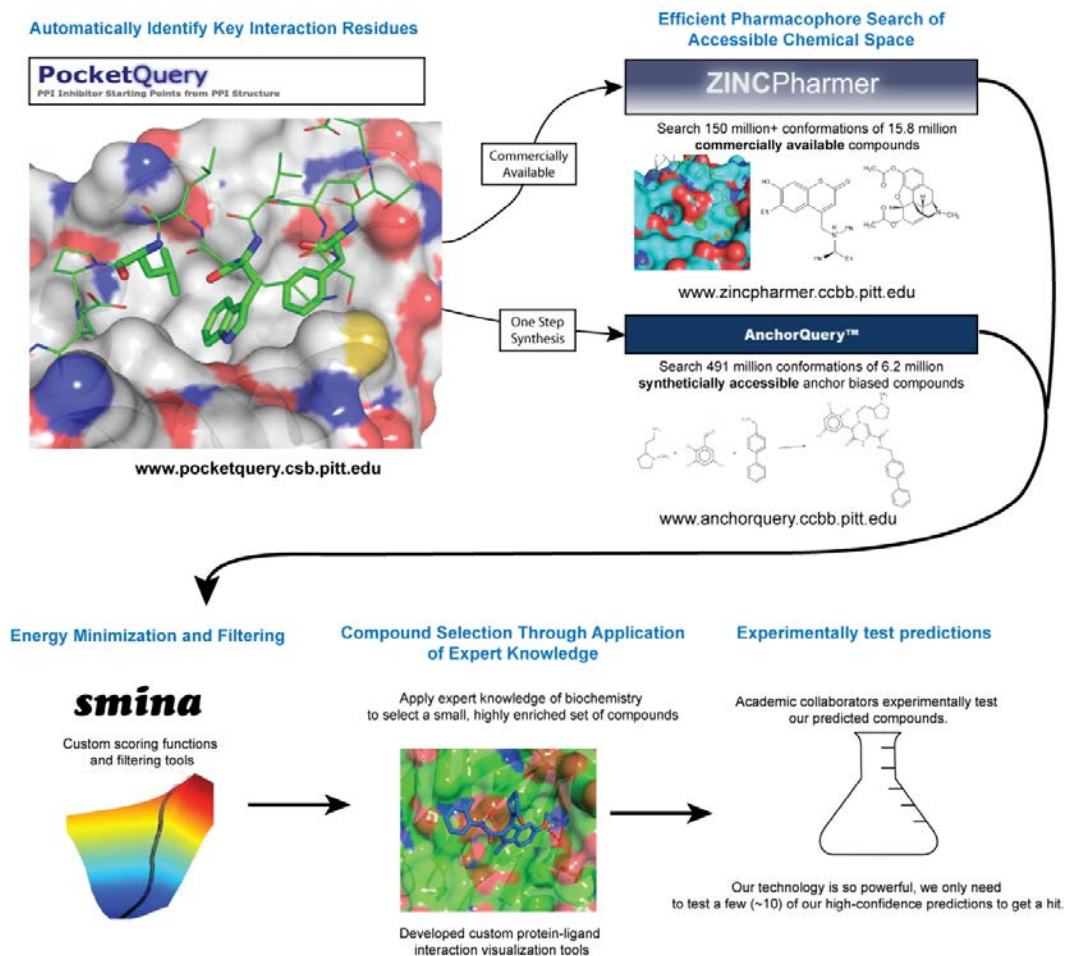


Figure 4.1 Overview of our structure-based virtual screening pipeline.

Key interaction residues are identified using PocketQuery. These residues are transferred with one click to either ZincPharmer or AnchorQuery depending on whether commercially available compounds or new chemistry is desired, respectively. The results of the pharmacophore screen are energy minimized with respect to the protein surface using smina followed by simple filters based on the predicted score. Finally, the work presented here, ClusterMols, is used to allow for efficient visualization and selection of a small set of compounds that are purchased (or synthesized) for experimental validation.

Following the initial pharmacophore screen we are generally left with approximately 1,000-20,000 compounds out of the initial set of 30+ million compounds. The first part of the secondary screening procedure is energy minimization with respect to the protein structure. There are numerous methods for accomplishing this[88-91], however we have found that smina, a tool developed in our lab which is based on AutoDock Vina[88], performs very well in terms of both efficiency and accuracy. After minimizing the compounds, smina predicts the energy of the final pose; we will refer to this as the “score” of the compound or pose. Following minimization, we filter out those that have very poor scores or that moved a significant amount during the minimization indicating that they no longer match the initial pharmacophore query. This results in 500-2,500 compounds that are presumably enriched for actual inhibitors. However, in many cases such as in pilot screens or in academic settings, this many compounds cannot be tested. Therefore, these 500-2,500 compounds must be further filtered down to a short list of likely inhibitors. Here we will discuss three ways that one could obtain this final set of compounds: sorting by predicted energy, a “linear manual selection” method and the method presented here, a software package that allows for efficient navigation and selection of compounds by substructure clustering.

The first possible method for limiting the ~1,000 compounds which passed all of the previous filters is to simply rank them based on their predicted score and take the compounds with the highest score. For a number of reasons, this is likely the least effective method in finding likely inhibitors. First, the scoring functions used in smina, as well as in all other currently available methods, are not accurate enough to finely discriminate binders from non-binders[92-111]. Even the best available methods are unable to accurately rank-order sets of congeneric compounds. In blinded competitions[93, 106, 107], the top methods are only able to



achieve correlation coefficients of around  $R^2 = 0.65$ . This means that when choosing the top ten compounds for testing, we cannot be very confident that the top scoring compounds will actually be inhibitors. A second problem with this method is that when sorting by predicted affinity and taking the top N compounds, it is entirely possible that very similar or even identical compounds will be in the top N. Testing very similar compounds can be wasteful in initial screens where chemical diversity is preferred.

The second method for selecting small sets of compounds for experimental testing is the “linear manual selection” method. Prior to the work presented here, this was the preferred method of selecting compounds at the end of a virtual screening effort. In this method, the minimized compounds are manually examined one at a time by a scientist and compounds of interest are noted (often on pencil and paper) for later analysis or selection. This method has a number of advantages and drawbacks. The main advantage of this method is that a scientist is able to evaluate the compounds based on both their prior knowledge and the compound’s predicted contacts with the protein. This allows for well-informed, high-quality predictions by leveraging the biochemical insight of the person performing the virtual screen, who just might be the world’s foremost expert on this particular target. However, there are several drawbacks to this method, the largest being the amount of time required to manually examine 100s or 1000s of compounds, which is often on the order of several full days. Additionally, the compounds are often sorted by their predicted affinity (if they are sorted at all), which makes it difficult to compare two similar compounds and choose the best if they are separated in the ranking. A third drawback of this method is that the user must manually visualize the contacts between each compound and the protein. Examining the contacts (hydrogen bonds, hydrophobic interactions, etc.) between the ligand and protein allows the user to understand the complementarity (or lack

thereof) between the two binding partners. This is often accomplished by using the measurement tool in PyMOL to visualize hydrogen bonds between the ligand and protein. While ultimately effective in identifying likely inhibitors, this method is also very time consuming.

The method presented here is software called “ClusterMols,” which attempts to improve on the benefits of the linear manual selection method while addressing all of the drawbacks. This software product allows for extremely efficient examination of the set of compounds generated by previous virtual screening by grouping similar compounds together such that the most desirable compound in a scaffold can be chosen easily or, alternatively, the entire scaffold can be discarded. In addition, this software provides a number of tools for automatic visualization of polar contacts and allows for easy selection and extraction of compounds of interest. Overall, this tool allows users to apply their expert knowledge to select a small set of compounds much more rapidly than before. This reduces the required time for selecting a small set of compounds from the initial set of ~1000 from over a week to under a day in most cases. The software is implemented as a plugin in PyMOL with a full graphical interface (explained in Section 4.2) and extended with a set of keyboard controls.

## **4.2 GENERAL METHOD**

ClusterMols has two major phases of calculation: pairwise similarity calculation and compound clustering. After loading the clustering results, ClusterMols provides a number of controls for efficiently navigating the search results, including advanced visualization of the protein-ligand contacts and keyboard shortcuts for selecting and extracting compounds of interest. As an

illustrative example, we will use a virtual screen of the XRCC1/DNA Pol $\beta$  PPI interaction shown in Figure 4.2.

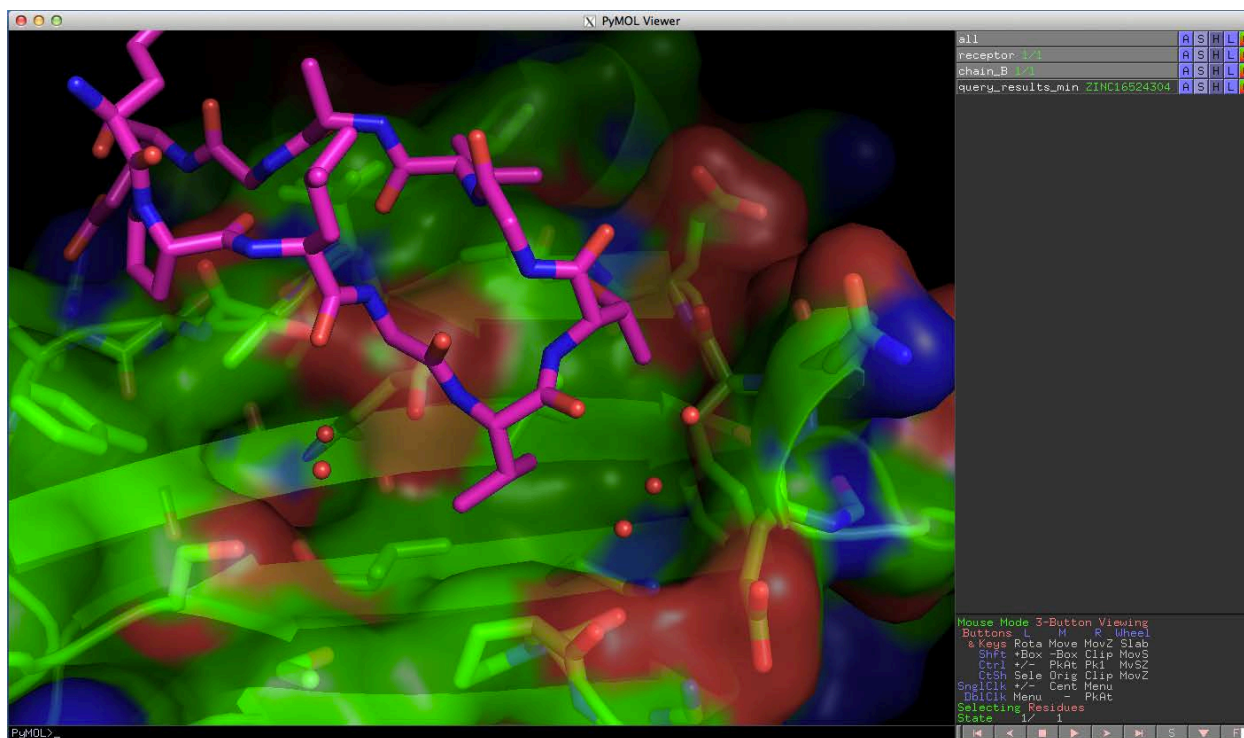


Figure 4.2 Structure of a protein-protein interaction. Example PPI structure shown in green surface and purple sticks in the PyMOL user interface. Water molecules are shown as small red spheres.

#### 4.2.1 Similarity Calculations

In the first phase of the calculation, the user selects the compounds to be analyzed: either a file (e.g. .sdf or .mol2) or an object that is already loaded into PyMOL (Figure 4.3). These compounds represent the ~1000 resulting from the pharmacophore search and the initial structural filtering. After choosing the compounds, the calculations and visualization proceed automatically with

default parameters. The user is then presented with the options pages shown in Figure 4.4 which allow for modifying the parameters of the similarity calculations and clustering. To perform the substructure clustering, a pairwise similarity matrix is computed using the chemfp[112] package from OpenBabel[66] using the FP3 fingerprint. Depending on the number of compounds in the input dataset, this calculation can take several minutes. Therefore, after computing the pairwise similarity matrix, the results are saved so future recalculation is not necessary.

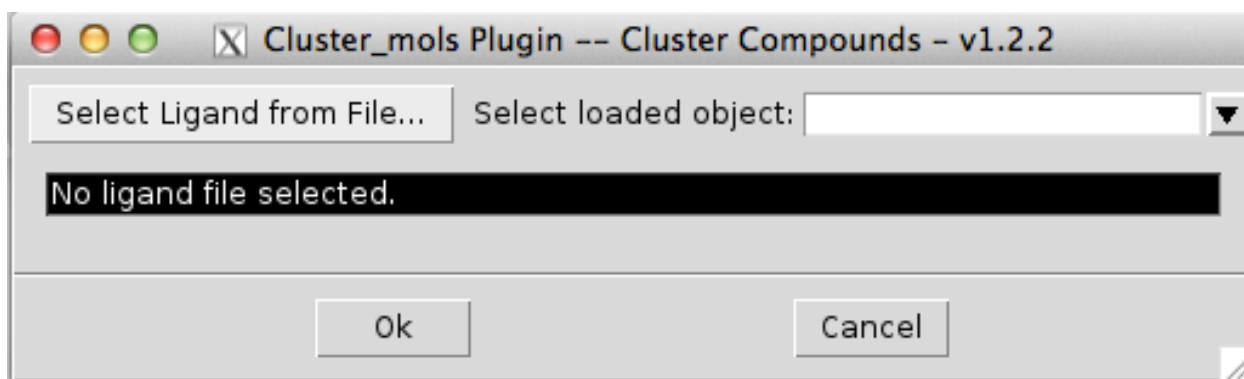


Figure 4.3 Load file dialog of ClusterMols. This pop-up window of ClusterMols allows the user to open a multi-compound file (e.g. .sdf or .mol2) or select previously loaded compounds from within PyMOL. Pressing “Ok” automatically starts the calculations.

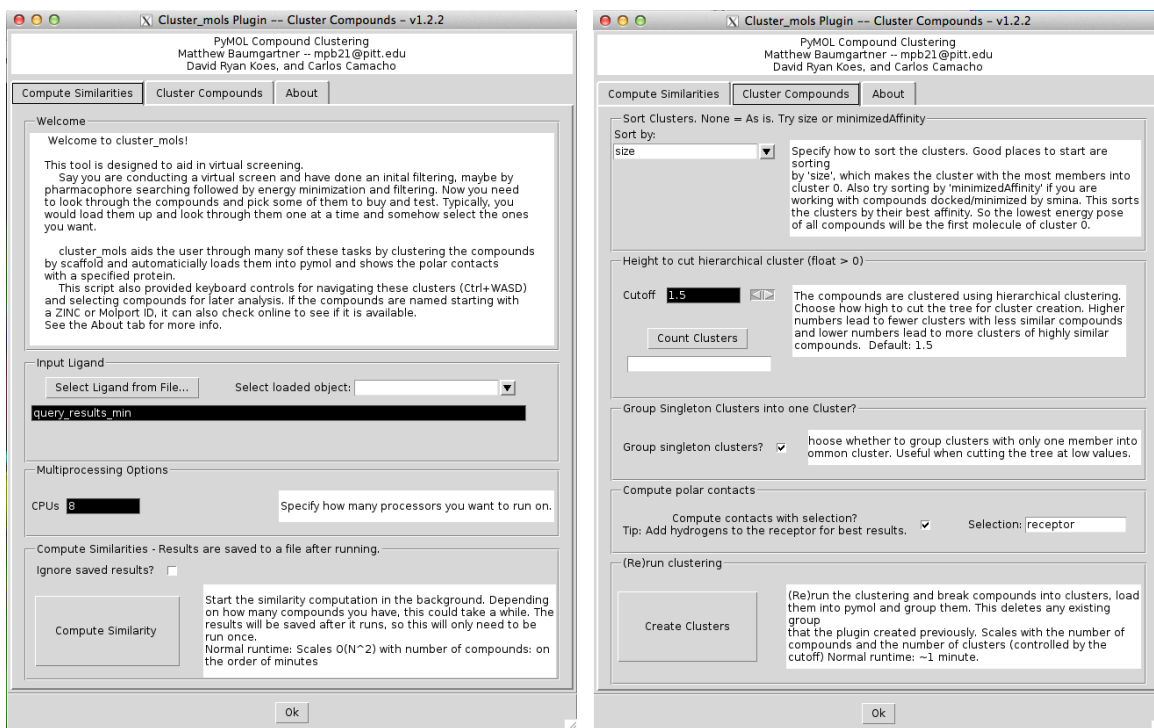


Figure 4.4 Graphical User Interface (GUI) of ClusterMols. After loading the input file, the user is presented with a GUI that contains two pages of options that are used to control the respective phases of the program (left: similarity calculation, right: clustering). Each option is accompanied by a short description and some suggested values. There is also an “About” tab that contains information about how the program works and a description of the keyboard controls.

## 4.2.2 Clustering

In the second phase of calculations, hierarchical clustering is performed on the similarity matrix grouping similar compounds together. Hierarchical clustering is a machine learning method by which items start in their own cluster and pairs of clusters are merged based on the similarity until all items have been linked into one cluster. The advantage of using hierarchical clustering for this

problem is that clusters are created by “cutting” the tree at a specified height; which has been made a user tunable parameter such that the user can create clusters of compounds that are as similar or diverse as they desire. Figure 4.5 shows an example of hierarchical clustering. Increasing the cutoff parameter (dashed line in Figure 4.5) leads to fewer clusters with less internal similarity, decreasing the cutoff parameter leads to more clusters which are more internally similar.

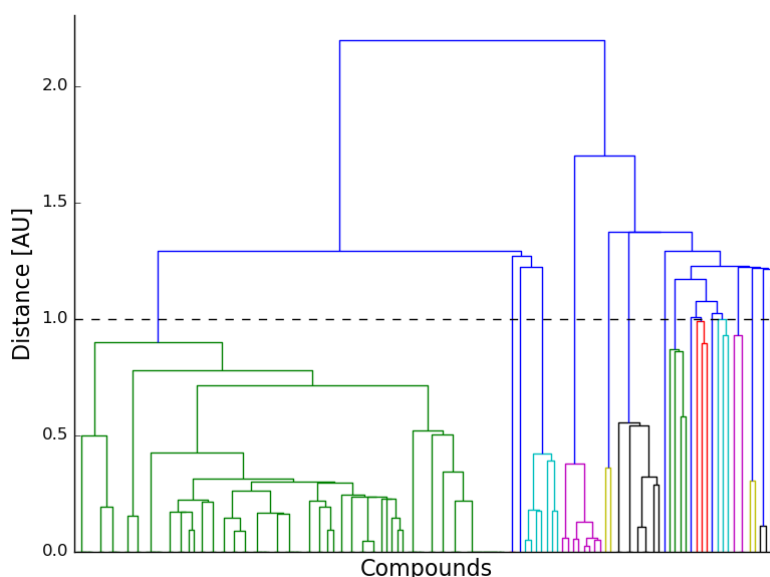


Figure 4.5 Hierarchical clustering of compounds based on their similarity. The dashed line indicates the user-tunable cut-off height. Colored clades beneath the cutoff height are loaded into PyMOL as clusters. The y-axis is in arbitrary units [AU].

### 4.3 USER INTERFACE

The clusters are then loaded into PyMOL and the software automatically displays both favorable and unfavorable polar interactions with a specified protein structure (Figure 4.6). The clusters are

loaded into PyMOL as groups that make it easy for the user to show or hide clusters and their corresponding contacts (left side of Figure 4.6). In addition to these features, the user is also able to easily navigate the list of clusters using a set of keyboard commands listed in Table 4.1. One of the most useful keyboard shortcuts in ClusterMols is the ability to automatically extract compounds of interest from the list of clusters. For various reasons, this seemingly simple task is actually non-trivial in PyMOL without using ClusterMols. Selected compounds are added to a new object with the suffix “\_selected,” which itself is a first-class PyMOL object that can be enabled or disabled, or more importantly, saved to a file for further analysis.

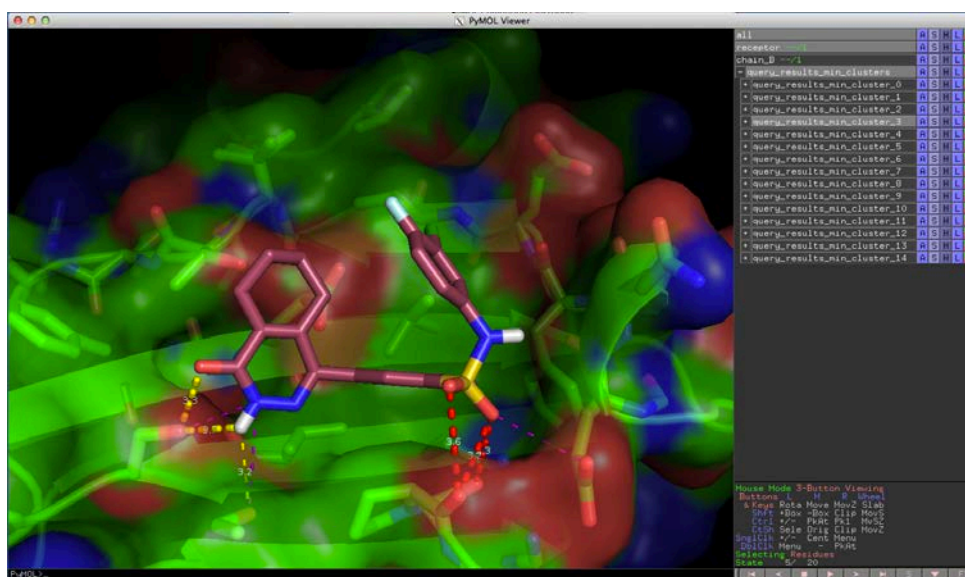


Figure 4.6 Clustered virtual screening results showing polar contacts. The user interface of PyMOL and ClusterMols showing the clusters (15 in this example) loaded into PyMOL on the right side. In the structure viewer, the receptor structure is shown as green cartoon and surface and one of the virtual screening results is shown as sticks. The software automatically highlights both favorable (shown in yellow) polar contacts and unfavorable polar contacts (red). These colored dashed lines allow for the user to quickly evaluate the quality of the pose and either accept or reject it.

Table 4.1. ClusterMols keyboard commands. ClusterMols has a number of keyboard controls for navigating inside and between clusters, selecting compounds for later analysis and for getting vendor information. \*Vendor information is only available if the compound names are ZINC[87] or MolPort IDs. In the latter case, price information is also reported.

<b>Cluster Movement</b>	
<i>Key</i>	<i>Action</i>
Ctrl+W	Move up one cluster
Ctrl+S	Move down one cluster
Ctrl+D	Go to next compound in cluster
Ctrl+A	Go to previous compound in cluster
<b>Compound Selection</b>	
<i>Key</i>	<i>Action</i>
F1	Print currently selected molecule
F2	Remove most recently added compound
F3	Add current compound to list
F4	Print list
<b>Vendor Information</b>	
<i>Key</i>	<i>Action</i>
Ctrl+F	Get purchasability information*

As listed in Table 4.1, the user is also able to check for up-to-date information on the commercial availability of a compound of interest using the keyboard shortcut Ctrl+F. If the compound of interest originates from the ZINC database (i.e. results from ZincPharmer), the ZINC compound webpage is checked for available vendors. Example output is shown in Figure 4.7. Additionally, if the compound of interest originates from MolPort [www.molport.com], the prices and availability are extracted from the MolPort compound webpage.



```
Looking for vendors of: MolPort-007-696-472
Availability: In stock
ChemDiv, Inc.
  5 mg      63.60 USD
  S&H to USA: 120.00 USD
ZINC15669378 has 7 listed
Vendors: ['eMolecules', 'Molport', 'Mcoule', 'AKOS (make-on-demand)', 'AKOS BB (make-on-demand)',
'Ambinter', 'Mcoule Make-on-demand']
```

Figure 4.7. Vendor information and prices of compounds of interest in ClusterMols.

#### 4.4 SUMMARY

Structure-based virtual screening is a credible alternative to conventional high-throughput screening. It provides savings in time, resources and money by computationally predicting a small subset of compounds to experimentally test. The structure-based virtual screening pipeline that has been developed in the Camacho Laboratory at the University of Pittsburgh provides a proven pathway for discovering novel modulators of protein-protein interactions[113, 114].

Presented here is a software solution allowing for extremely efficient selection of a highly enriched set of compounds which are most likely to bind to a protein target by leveraging a user's expert knowledge. This software, called ClusterMols, is a plugin for the PyMOL molecular graphics viewer which automatically clusters compounds based on chemical similarity and displays them in PyMOL in a user-friendly manner. Additionally, ClusterMols provides a number of additional useful features such as automatic visualization of polar contacts/clashes, one-button extraction of compounds of interest and easy-to-use keyboard controls.

ClusterMols vastly speeds up the time required for manually selecting compounds for purchase/synthesis by allowing the user to examine compounds on the scaffold level which allows

for easy exploring or discarding of the cluster. The software is designed to provide tools to the user that allow them to easily examine the compound-protein interactions and use their knowledge of the specific protein target to select the best set of compounds for experimental validation. This software has been successfully used as a part of a number of ongoing virtual screening efforts on a number of therapeutically relevant targets.

This software is in active development and a number of features are planned or underway. This includes support for offloading the computationally expensive portions of the program to a remote server. This has a number of advantages including allowing the end user to avoid installing a number of dependencies as well as enabling analysis of large datasets on commodity hardware. Additionally, advancements are planned for improving the visualization of protein-ligand interactions, in particular, highlighting unmatched polar atoms on the protein and ligand.

## 5.0 CONCLUSIONS AND FUTURE RESEARCH

The research presented in this dissertation has focused on applying biophysical insight to improving rational drug design. We have focused on three major challenges in computational drug discovery: protein flexibility, the role of water in ligand binding and using and understanding large virtual screening datasets. Using computational approaches, we have presented novel insights into these important areas. Below, we will briefly recap the main contributions of this research which were described in the previous chapters.

In Chapter 2.0, we described our work in understanding the induced-fit binding mechanism of the tumor suppressor p53 and its inhibitor, MDM2. Using molecular dynamics simulations, we have determined the precise chemistry of the indole side chain of the Trp23 anchor residue of p53 which is responsible for opening the binding cleft on MDM2. The indole group plus the beta-carbon of the side chain proved to be the critical elements required for opening the binding pocket. This opening occurs through coordination of the position and orientation of MDM2 residues Val93 and His96, which must first adopt a previously undescribed “open and out” transition state wherein helix  $\alpha 4$  must be pushed partially outward and the side chain of His96 must adopt a rotated outward conformation. We have verified the importance of the indole chemistry by demonstrating its presence in all available inhibitor-bound co-crystal structures of MDM2.

In Chapter 3.0, we examined the issue of water in ligand binding sites. We analyzed over 23,000 pairs of protein structures in their *apo* and *holo* states. By comparing the clusters of water molecules in the binding site of *apo* structure to the overlapping ligand chemistries in the *holo* structure, we were able to draw associations between ligand functional groups and types of water clusters that they can displace. We identified four generalized functional groups which commonly

displace characteristic water clusters upon binding: two polar (AS(A)A and NC(C)N) and two non-polar (aromatic rings and halogens). In general, the polar functional groups displace larger water clusters with higher levels of polar coordination with the protein and the non-polar functional groups displace smaller water clusters (mostly lone waters) with few or no polar interactions with the protein. We have presented structural examples of each functional group demonstrating that the polar atoms of AS(A)A and NC(C)N align quite well with the water molecules in the *apo* structure, which validates our hypothesis of replacement of the polar interactions of the water molecules with those of the ligand. Additionally, we found that the non-polar functional groups often displaced poorly-coordinated, lone waters which are often found in hydrophobic cavities. These findings have applications in a structure-based virtual screening effort wherein compounds can be filtered out based on whether they have a functional group which can displace an observed water cluster. Future work will focus on enhancing the descriptors of water clusters beyond the simple descriptors of their size and the count of their polar contacts. These descriptors might include description of the angles and distances of their hydrogen bonding network. Increasing the diversity and specificity of the functional groups analyzed is an important next step as well. Additionally, virtual waters will be explored wherein, short molecular dynamics simulations can be run to solvate the system to better understand water clusters which might not be resolved in the crystal structure.

In Chapter 4.0, we describe our approach for intelligently handling large amounts of virtual screening data. Our software, called “ClusterMols,” is a PyMOL plugin which clusters input compounds based on a user-tunable level of chemical similarity and presents the clusters to the user in the familiar PyMOL interface in an easy to understand manner. This allows the user to navigate the results on the level of the chemical scaffold. This has many improvements over

previous methods as the user can quickly skip undesirable scaffolds or choose the best representative from a scaffold of interest. Included in the plugin are a number of advanced methods for visualizing the predicted polar interactions with the protein, both hydrogen bonds and polar clashes (i.e. acceptor-acceptor). These tools allow for rapid evaluation of a compound by drawing the attention of the user to problem areas as well as positive interactions. Finally, ClusterMols provides a number of keyboard controls for quickly navigating the clusters and selecting compounds of interest for further analysis as well as the ability to check online for commercial availability and prices of compounds of interest with the press of a button. This software provides a significant savings in time and effort in the virtual screening process and has been utilized in a number of ongoing virtual screening efforts. Future work will focus on expanding the visualization of the protein-ligand interactions, particularly in highlighting unpaired polar atoms on the protein and ligand.

## **APPENDIX A APPLICATIONS OF MODELING PROTEIN-LIPID INTERACTIONS**

In this appendix, we will present two published applications of modeling small molecule inhibitors to proteins. In Appendix A.1, we describe how we predicted the binding mode of a selective inhibitor of the c-Src kinase[115]. Then in Appendix A.2, we present how we predicted the binding site and pose of an allosteric activator of the ABL kinase[116]. Both are reproduced with permission.

### **A.1 MODELING SELECTIVE INHIBITORS OF C-SRC**



# A Discovery Strategy for Selective Inhibitors of c-Src in Complex with the Focal Adhesion Kinase SH3/SH2-binding Region

Jamie A. Moroco<sup>1,†</sup>, Matthew P. Baumgartner<sup>2</sup>, Heather L. Rust<sup>1</sup>, Hwan Geun Choi<sup>3,4</sup>, Wooyoung Hur<sup>3,4</sup>, Nathanael S. Gray<sup>3,4</sup>, Carlos J. Camacho<sup>2</sup> and Thomas E. Smithgall<sup>1,\*</sup>

<sup>1</sup>Department of Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine, 450 Technology Drive, Pittsburgh, PA 15219, USA

<sup>2</sup>Department of Computational and Systems Biology, University of Pittsburgh School of Medicine, 3501 Fifth Avenue, Pittsburgh, PA 15260, USA

<sup>3</sup>Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 250 Longwood Avenue, Boston, MA 02115, USA

<sup>4</sup>Department of Cancer Biology, Dana-Farber Cancer Institute, 250 Longwood Avenue, Boston, MA 02115, USA

\*Corresponding author: Thomas E. Smithgall, tsmithga@pitt.edu

<sup>†</sup>Present address: Department of Chemistry and Chemical Biology, Northeastern University, 360 Huntington Avenue, Boston, MA 02115, USA

The c-Src tyrosine kinase co-operates with the focal adhesion kinase to regulate cell adhesion and motility. Focal adhesion kinase engages the regulatory SH3 and SH2 domains of c-Src, resulting in localized kinase activation that contributes to tumor cell metastasis. Using assay conditions where c-Src kinase activity required binding to a tyrosine phosphopeptide based on the focal adhesion kinase SH3-SH2 docking sequence, we screened a kinase-biased library for selective inhibitors of the Src/focal adhesion kinase peptide complex versus c-Src alone. This approach identified an aminopyrimidinyl carbamate compound, WH-4-124-2, with nanomolar inhibitory potency and fivefold selectivity for c-Src when bound to the phospho-focal adhesion kinase peptide. Molecular docking studies indicate that WH-4-124-2 may preferentially inhibit the 'DFG-out' conformation of the kinase active site. These findings suggest that interaction of c-Src with focal adhesion kinase induces a unique kinase domain conformation amenable to selective inhibition.

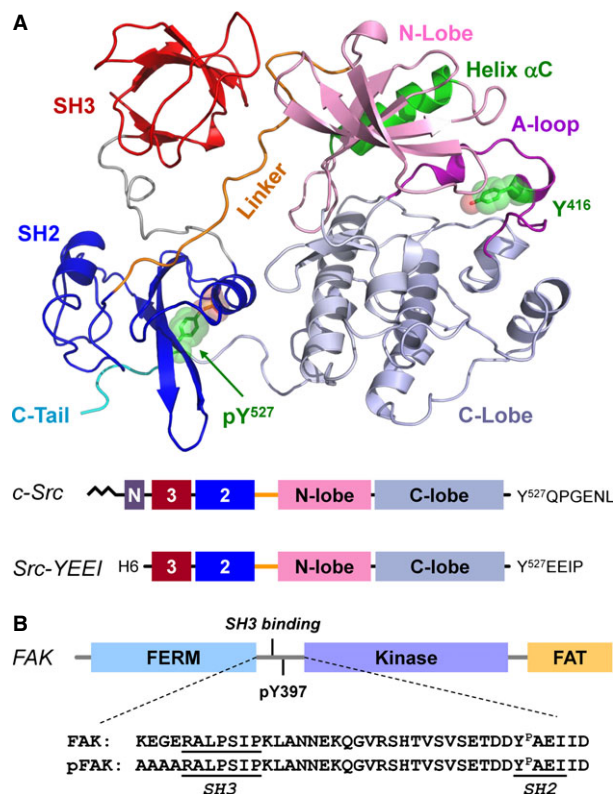
**Key words:** Src kinase, focal adhesion kinase, SH3 domain, SH2 domain, kinase inhibitors, cancer drug discovery

Received 10 June 2014, revised 29 September 2014 and accepted for publication 29 October 2014

The c-Src protein-tyrosine kinase is the prototype of the Src kinase family and is broadly expressed in virtually every mammalian cell type (1). All eight mammalian Src-family kinases (SFKs) share a common regulatory and kinase domain organization (Figure 1A). At the N-terminus is a signal sequence for myristoylation, which localizes c-Src to the cell membrane. Membrane association facilitates interactions of c-Src with binding partners and is essential for function (2). The myristoylation signal is followed by a unique domain, modular SH3 and SH2 domains, an SH2-kinase linker, the kinase domain, and a C-terminal tail with a conserved tyrosine residue essential for negative regulation of kinase activity. The SH3 and SH2 domains, which recognize proline-rich and phosphotyrosine-containing sequences, respectively, are important for kinase regulation as well as interactions with other cellular proteins.

SFKs maintain their inactive state through two intramolecular interactions involving the SH2 and SH3 domains. The SH3 domain interacts with a polyproline type II helix formed by the SH2-kinase linker while the SH2 domain binds to the tyrosine-phosphorylated tail (3). Tail phosphorylation is mediated by the independent regulatory kinase Csk as well as the related kinase, Chk (4,5). These intramolecular regulatory interactions are present in the X-ray crystal structure of c-Src in the downregulated conformation (modeled in Figure 1A) (6).

Multiple mechanisms of SFK activation have been reported, including dephosphorylation of the C-terminal tail tyrosine and subsequent release from the SH2 domain (7) as well as displacement of one or both of the intramolecular interactions in *trans* (8,9). Based on their amino acid sequences, the SH2-kinase linker and phosphorylated C-terminal tail represent low affinity ligands for their respective target domains. Displacement of these interactions by proteins with higher affinities for the SH3 and/or SH2 domains provides a mechanism for SFK activation by both physiological substrates as well as exogenous proteins expressed as a result of microbial infection. For example, the Nef protein encoded by HIV-1 binds to the SH3 domain of the Src-family member Hck, resulting in linker displacement and constitutive kinase activation (10,11). Alternatively, juxtamembrane autophosphorylation sites on active receptor tyrosine kinases may recruit c-Src



**Figure 1:** Structure and domain organization of c-Src and focal adhesion kinase (FAK). (A) Crystal structure of c-Src in the downregulated conformation (PDB ID: 2SRC) showing the SH3 domain, SH3-SH2 connector, the SH2 domain, the SH2-kinase linker, the kinase domain, and the C-terminal tail. Key kinase domain features include the N-lobe, activation loop,  $\alpha$ C-helix, and C-lobe. The side chains of the conserved tyrosine residues in the activation loop (Y416) and the tail (pY527) are also shown. Intramolecular interactions between the SH3 domain and the SH2-kinase linker as well as the SH2 domain and the tyrosine-phosphorylated tail are required to maintain the inactive state. The modular domain organization of c-Src is shown below the structure, which includes a myristoylated N-terminal domain not present in the crystal structure. The organization of the recombinant Src-YEEI protein used for this study is also shown, in which the N-terminal unique domain is replaced with a His-tag and the C-terminal tail sequence is modified to encode YEEI. (B) FAK consists of an N-terminal FERM domain, central kinase domain, and a C-terminal FAT domain. The linker connecting the FERM and kinase domains encompasses the binding site for c-Src, consisting of a PxxP motif (SH3 binding) and an autophosphorylation site in the sequence context pYAEI (SH2 binding). The sequence of the pFAK peptide based on this region and used in the inhibitor screen is also shown, with the Src SH3 and SH2 binding sites underlined.

through its SH2 domain, resulting in kinase activation through tail displacement. Other cellular protein partners for c-Src contain both SH3- and SH2-binding sequences believed to displace both intramolecular interactions, such as p130Cas (12) and the focal adhesion kinase (FAK) (13).

Focal adhesion kinase is a non-receptor protein-tyrosine kinase that localizes to focal adhesions, the intracellular

structures formed at sites of cell adhesion to the extracellular matrix (ECM). The domain organization of FAK consists of a protein 4.1/ezrin/radixin/moesin (FERM) domain, followed by binding sites for the c-Src SH3 and SH2 domains (Figure 1B). The kinase domain of FAK is located in the center of the protein, followed by a proline-rich region with binding sites for the SH3 domains of p130Cas and other components of the focal adhesion complex. The C-terminal end of FAK encompasses the focal adhesion targeting (FAT) domain which localizes FAK to focal adhesions in response to integrin stimulation (14–16). Upon recruitment to focal adhesions, FAK undergoes autophosphorylation on Y397, creating a high-affinity binding site for the c-Src SH2 domain. Additionally, the c-Src SH3 domain binds to the proline-rich sequence adjacent to the SH2 binding site. These tandem binding events displace both intramolecular regulatory interactions, leading to c-Src kinase activation. This mechanism integrates relocalization of c-Src to focal contacts with kinase activation, providing an elegant mechanism for spatial and temporal regulation of c-Src function.

Active c-Src phosphorylates FAK at multiple tyrosine residues, including kinase domain activation loop tyrosines important for full FAK kinase activity (17). Coordinated activation of c-Src and FAK stimulates multiple signaling pathways linked to focal adhesion turnover and cell migration, an essential process in embryonic development, wound healing, and the immune response (17). Cells lacking c-Src and FAK show a dramatically reduced rate of migration (18,19) while hyperactivation of these kinases leads to an increased rate of migration that has been implicated in tumorigenesis. Indeed, overexpression and increased activity of both c-Src and FAK have been reported in multiple tumor sites (17,20).

The relationship of the c-Src:FAK complex to cancer progression highlights the potential for c-Src and FAK as targets for cancer therapy (21–24). Multiple inhibitors of both c-Src and FAK have been discovered, and some have progressed into clinical trials. However, the success of these inhibitors, especially as monotherapy, has been limited (20,25). In this study, we hypothesized that binding of FAK to c-Src induces a unique disease-associated conformation of the Src active site that may be amenable to selective inhibitor targeting. In an effort to find a selective inhibitor of c-Src in the FAK-bound state, we first developed screening assay conditions where c-Src activity was entirely dependent on the presence of a phosphopeptide based on the FAK sequences for c-Src SH3/SH2 binding. We then screened a small, kinase-biased inhibitor library and identified several compounds selective for the c-Src:pFAK peptide complex versus c-Src alone. The most promising compound showed a fivefold preference for the active complex in both end-point and kinetic kinase assays and inhibited the complex with nanomolar potency. Computational docking studies suggest that this compound prefers the ‘DFG-out’ conformation of the kinase active



site, suggesting that binding of the pFAK peptide induces this c-Src kinase domain conformation. Our results provide an important proof-of-concept result that state-selective ATP-site inhibitors for c-Src can be identified under appropriate screening assay conditions. This approach may provide a new path to the discovery of state and context-selective inhibitors for large kinase families with highly homologous active sites.

## Materials and Methods

### Expression and purification of recombinant Src-YEEI

A human c-Src cDNA clone was modified on its C-terminal tail to encode the sequence Tyr-Glu-Glu-Ile-Pro ('YEEI') as described elsewhere (26). In addition, the N-terminal unique domain was replaced with a hexa-histidine tag. The resulting sequence was used to produce a recombinant baculovirus in Sf9 insect cells using BaculoGold DNA and the manufacturer's protocol (BD Pharmingen) as previously described (27). Src-YEEI was co-expressed with the *Yersinia pestis* YopH phosphatase to promote dephosphorylation of the activation loop tyrosine and maintain the downregulated state (28,29). Sf9 cells were grown in monolayer cultures and co-infected with the Src-YEEI and YopH baculoviruses. Cells were harvested 72 h after infection, and Src-YEEI was purified as previously described (27). Purified Src-YEEI protein was stored in 20 mM Tris-HCl, pH 8.3, containing 100 mM NaCl. Kinase protein used in the Z'Lyte *in vitro* kinase assays also contained 3 mM DTT. The molecular weight of purified Src-YEEI was confirmed by mass spectrometry and determined to be phosphorylated on the YEEI tail but not the activation loop (26), consistent with the structure of the downregulated conformation (6).

### Chemical library screen

A library of 586 kinase-biased inhibitors (30,31) was screened using the FRET-based Z'Lyte *in vitro* kinase assay and Tyr2 peptide substrate (Life Technologies) (32). Assays were performed in quadruplicate in 384-well low-volume, non-binding, black polystyrene microplates (Corning) according to the manufacturer's instructions and as described previously (27,33,34). Briefly, the assay measures phosphorylation of the Tyr2 FRET-peptide substrate which is labeled with coumarin and fluorescein on its N- and C-termini, respectively, which form a FRET pair. After the kinase reaction, a development step involves site-specific proteolytic cleavage of the unphosphorylated but not the phosphorylated peptide. Peptide cleavage results in loss of the FRET signal. Src-YEEI was first titrated into the assay over a concentration range of 0.5–500 ng/well (0.908–908 nM). Kinase activity was measured in the absence and presence of a 10-fold molar excess of a pFAK peptide with the sequence AAAARALPSIPKLAN NEKQGVRSHTVSVSETDDY<sup>P</sup>AEIID (13), as well as the

control peptides pYEEI (EPQY<sup>P</sup>EEIPIYL) (35), and VSL12 (VSLARRPLPLP) (36). All peptides were synthesized by the University of Pittsburgh Genomics and Proteomics Core Laboratories, and the mass and purity of each compound was confirmed by LC-MS. The kinase was preincubated with or without the peptides for 15 min, and the reaction was initiated by the addition of ATP (100  $\mu$ M) and Tyr2 peptide substrate (1  $\mu$ M). The kinase reaction was incubated 1 h, followed by addition of the development protease to cleave the unphosphorylated substrate. Coumarin and fluorescein fluorescence were measured 1 h later on a Molecular Devices SpectraMax M5 plate reader. Results are expressed as percent of maximum kinase activity relative to a stoichiometrically phosphorylated positive control peptide and negative control wells where no ATP is present.

Screening assays were performed with Src-YEEI (30 ng) in the presence of a 10-fold molar excess of the pFAK peptide (545 nM pFAK versus 54.5 nM Src-YEEI) or with Src-YEEI alone (125 ng; 227 nM). Each compound was assayed at a final concentration of 10  $\mu$ M with carrier solvent (DMSO) at 2.5%. Src-YEEI was preincubated with the pFAK peptide for 15 min. Compounds were then added and the mixtures incubated for an additional 30 min before initiating the kinase reaction by the addition of substrate and ATP. The IC<sub>50</sub> values for the hit compounds were determined over a concentration range of 3 nM to 30  $\mu$ M, and the resulting concentration–response curves were best-fit by nonlinear regression analysis to obtain the IC<sub>50</sub> values (GraphPad Prism). Results are expressed as percent inhibition, calculated relative to negative control wells without ATP and positive control wells without inhibitor (DMSO only).

### Kinetic kinase assays

These studies used the fluorescence-based ADP Quest Assay (DiscoverRx), which monitors the production of ADP in kinase reactions over time (37). All kinase assays were performed in quadruplicate in black 384-well microplates (Corning # 3571), in a final assay volume of 50  $\mu$ L/well at 25 °C. ATP stocks were prepared in 10 mM Tris-HCl, pH 7.0, and the ATP concentration in each assay was held constant at 100  $\mu$ M. The SFK substrate peptide, YIYGSKF (Anaspec) (38) was prepared in ADP Quest assay buffer (15 mM HEPES, pH 7.4, 20 mM NaCl, 1 mM EGTA, 0.02% Tween-20, 10 mM MgCl<sub>2</sub>, 0.1 mg/mL bovine  $\gamma$ -globulin). The substrate concentration in the assay was set to the K<sub>m</sub> value (162  $\mu$ M). In kinase inhibition assays, the Src-YEEI concentration in the absence of the pFAK peptide was 40 ng/well (14.5 nM) and in the presence of the peptide was 15.6 ng/well (5.7 nM). Kinase reactions were initiated by the addition of ATP. Assay plates were read every 5 min for 3 h on a Molecular Devices SpectraMax M5 microplate reader. Results are expressed as percent of maximum kinase activity relative to the activity of the kinase in the absence of inhibitor. The mass and purity of inhibitors



WH-4-124-2 and WH-4-023 were confirmed by LC-MS (>95%). WH-4-023 has been previously described as a potent inhibitor for c-Src and Lck (39).

### Computational docking

The binding mode of the inhibitor WH-4-124-2 was modeled to the X-ray crystal structure of the Tyr416-dephosphorylated, imatinib-bound Lck kinase domain (PDB ID: 2PL0) by generating conformers of the compound using OMEGA2 (40) (OpenEye) with default settings. A pharmacophore alignment of each WH-4-124-2 conformer was then performed with the piperazine and aliphatic rings of imatinib in this Lck structure using *Pharmer* (41). The aligned conformers were energy-minimized using *smina* (42) with the default minimization parameters and fixed receptor protein. Poses of WH-4-124-2 that showed a low-energy score and a good structural alignment to the carbamate core of a related inhibitor (compound 43) bound in a crystal complex with phosphorylated Lck (PDB ID: 2OFU) (39) in a region with high structural similarity with dephosphorylated Lck were selected for refinement/minimization. Finally, the low-energy-minimized structure, which recovers most of the interactions present in both Lck co-crystal structures, was selected as the best model. The binding mode of WH-4-023 was modeled to the X-ray crystal structures of the Lck kinase domain in the DFG-in (PDB: 2OFU) (39) and DFG-out (PDB: 2PL0) (43) states through pharmacophore alignment and energy minimization.

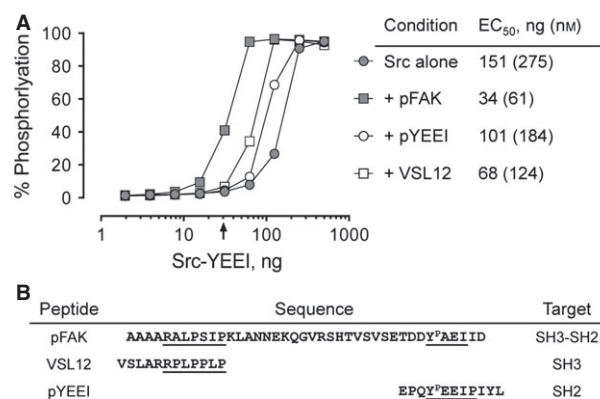
## Results and Discussion

### Development of a screening assay for Src:FAK selective inhibitors

To screen for selective inhibitors of the c-Src:pFAK complex, we used a synthetic peptide based on the c-Src SH3/SH2 docking region from FAK (pFAK phosphopeptide; Figure 2) to activate recombinant, downregulated c-Src. Our goal was to model the conformation of active c-Src that results from interaction with FAK in focal adhesions. Previous studies have shown that this pFAK peptide binds to a tandem c-Src SH3-SH2 protein with low nanomolar affinity and is sufficient to activate c-Src via regulatory domain displacement *in vitro* (13).

Before initiating the chemical library screen, we first established kinase assay conditions under which c-Src activity was wholly dependent upon binding to the pFAK peptide. The form of recombinant, downregulated c-Src used in these studies has a modified C-terminal tail in which the wild-type sequence, Y<sup>527</sup>QPGENL, is replaced with YEEI to create a high-affinity SH2-binding sequence. This modification allows the kinase, termed Src-YEEI, to undergo tail autophosphorylation, enabling expression and purification in the downregulated state (26).

## Selective Inhibitors of a FAK: Src Kinase Complex



**Figure 2:** Activation of Src-YEEI by SH3- and SH2-binding peptides. The activity of Src-YEEI was measured in the Z'Lyte *in vitro* kinase assay in the absence or presence of peptides that bind to the SH3 domain alone (VSL12), the SH2 domain alone (pYEEI), or to both (pFAK). Recombinant Src-YEEI was titrated into the assay over the range of 2-500 ng as shown, and peptides were added at a 10-fold molar excess at each kinase concentration. The concentration of Src-YEEI required to reach 50% maximal activation (EC<sub>50</sub>) for each condition is shown on the right. The arrow indicates the Src-YEEI concentration used for the library screen, where Src-YEEI activity is dependent on the pFAK peptide (30 ng kinase input; 54.5 nM). The sequences of the peptides are shown below the graph with the docking sites for the Src SH3 and SH2 domains indicated.

We first measured purified Src-YEEI activity over a range of kinase concentrations in the absence or presence of the pFAK peptide, using the Z'Lyte kinase assay (44). As shown in Figure 2A, the extent of substrate phosphorylation increased in sigmoidal fashion as a function of input kinase concentration, with an EC<sub>50</sub> value of 151 ng/well (275 nM) in the absence of peptide. When this experiment was repeated in the presence of the pFAK peptide, the activation curve was shifted markedly to the left, yielding an EC<sub>50</sub> value of 34 ng/well (61 nM). This result is consistent with the interaction of the pFAK peptide with the Src-YEEI SH3 and SH2 domains, resulting in displacement of the regulatory apparatus from the back of the kinase domain and subsequent activation.

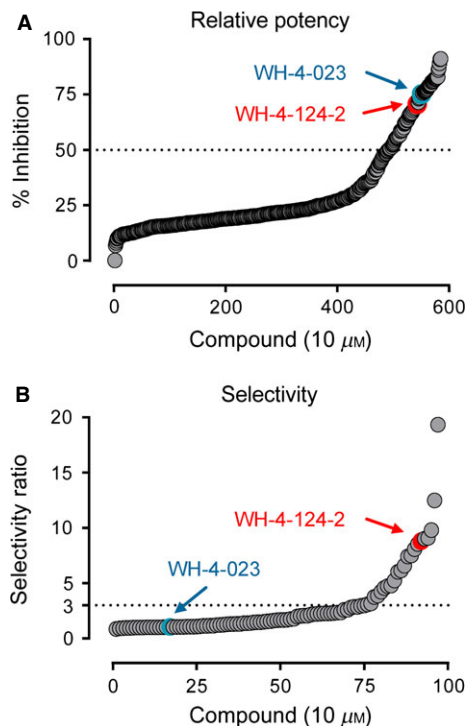
To provide additional evidence that dual engagement of the SH3 and SH2 domains is necessary for maximal activation of Src-YEEI in our assay, we examined the effect of peptides that bind individually to the SH3 domain (VSL12) or the SH2 domain (pYEEI) on kinase activity. The sequences of these peptides are presented in Figure 2B. Both of these peptides also activated Src-YEEI, but to a lesser extent than the pFAK peptide (Figure 2A). The pYEEI peptide, which binds only to the SH2 domain, resulted in an EC<sub>50</sub> value for Src-YEEI activation of 101 ng/well (184 nM), while the VSL12 peptide, which exclusively engages the SH3 domain, yielded an EC<sub>50</sub> value of 68 ng/well (124 nM). Both values are higher than that obtained

with the pFAK peptide, supporting the idea that dual regulatory domain displacement results in maximal kinase activity.

### Identification of inhibitors selective for Src-YEEI in complex with the pFAK peptide

For inhibitor screening, we chose a concentration of Src-YEEI that was dependent on the pFAK peptide for activity so as to model the kinase conformation induced by FAK binding. As indicated by the arrow in Figure 2A, Src-YEEI alone was inactive in the assay at 30 ng/well, while addition of the pFAK peptide resulted in 40% of maximal activation at this Src-YEEI concentration. This condition was therefore used for the library screen. We also identified a higher concentration of Src-YEEI alone (125 ng/well) with activity comparable to that of the Src-YEEI:pFAK peptide complex, as a basis of comparison. The ATP concentration was set to 100  $\mu\text{M}$ , which is in excess of published  $K_m$  values for c-Src (26,45–47) while the peptide substrate concentration was set to 1  $\mu\text{M}$ . Using these assay conditions, we screened a kinase-biased small molecule library of 586 compounds for preferential inhibitors of Src-YEEI in the presence of the pFAK peptide compared with Src-YEEI alone. Most of the compounds in this library were designed to bind kinase domains in the so-called Type II mode and exploit specific ATP-site conformations to impart binding specificity (30,31). Compounds were screened at a final concentration of 10  $\mu\text{M}$ , and the rank order of inhibitory potency for each compound against the Src-YEEI:pFAK peptide complex is presented in Figure 3A.

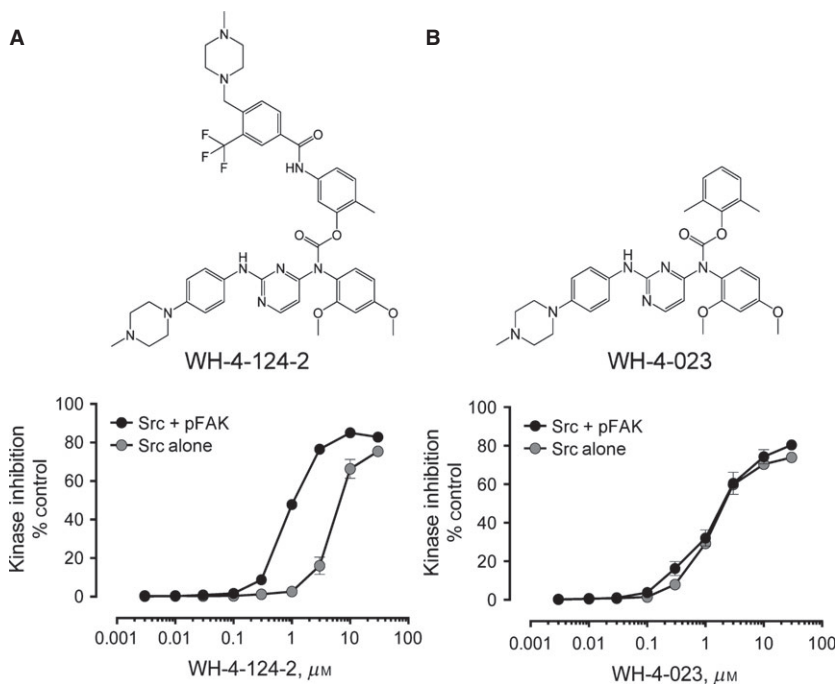
The 97 compounds that showed at least 50% inhibition of the Src-YEEI:pFAK complex at 10  $\mu\text{M}$  were then assessed for inhibition of Src-YEEI alone. Results for each of these compounds were then ranked according to their selectivity ratios, defined as the percent inhibition of the Src-YEEI:pFAK peptide complex divided by the percent inhibition of Src-YEEI alone. Twenty-four compounds with a selectivity ratio >3 (Figure 3B) were selected for further evaluation in concentration–response experiments. Of these, four compounds reproducibly inhibited Src-YEEI in the presence of the pFAK peptide to a greater extent than Src-YEEI alone with selectivity ratios ranging from 5 to almost 20 (data not shown). However, the overall potency for three of these compounds was rather low, with only about 50% inhibition of the kinase:peptide complex at a concentration of 10  $\mu\text{M}$ . For this reason, we did not follow up on these compounds. Instead, we focused on the aminopyrimidinyl carbamate designated WH-4-124-2 (see Figure 4A for structure) because it was the most potent inhibitor of Src-YEEI in the presence of the pFAK peptide. As shown in Figure 4A and Table 1, WH-4-124-2 inhibited the Src-YEEI:pFAK peptide complex with an  $\text{IC}_{50}$  value of 0.89  $\mu\text{M}$  in the Z'Lyte assay, as compared to 4.95  $\mu\text{M}$  for Src-YEEI alone.



**Figure 3:** Chemical library screen for selective inhibitors of pFAK-dependent Src-YEEI activity. A kinase-biased library of 586 compounds was screened for inhibitors of Src-YEEI activity in the presence of the pFAK peptide using the FRET-based Z'Lyte assay as described in the text. (A) Inhibitory activities of all 586 compounds are ranked by percent inhibition relative to the unphosphorylated substrate peptide control. Ninety-seven compounds inhibited kinase activity by at least 50% (dotted line). (B) A selectivity ratio was determined for compounds showing >50% inhibition of the Src-YEEI:pFAK complex in part A. This ratio was calculated as the percent inhibition of the Src-YEEI:pFAK complex divided by the percent inhibition observed with Src-YEEI alone. Twenty-four compounds exhibited a selectivity ratio >3 (dotted line). The hit compound WH-4-124-2 is indicated in each graph by a red dot, while the control compound WH-4-023 is indicated with a blue dot.

### WH-4-124-2 is a selective inhibitor of the Src-YEEI:pFAK complex

We next investigated the specific structural features of WH-4-124-2 responsible for its selectivity toward the Src-YEEI:pFAK complex. The compound WH-4-023, which is a substructure of WH-4-124-2 (Figure 4B), inhibited both the Src-YEEI:pFAK peptide complex and Src-YEEI alone by more than 70% in the screening assay (blue data point in Figure 3A). We therefore performed a concentration–response experiment with this compound and found that it inhibited Src-YEEI alone and in complex with the pFAK peptide with an  $\text{IC}_{50}$  value of about 1.3  $\mu\text{M}$  (Figure 4B). This result strongly suggests that the methylpiperazinyl trifluoromethylbenzamide substituent present in WH-4-124-2 contributes to inhibitor selectivity.



**Figure 4:** Selective inhibition of Src-YEEI by WH-4-124-2 in the presence of the pFAK peptide but not its structural analog, WH-4-023. WH-4-124-2 (A) and WH-4-023 (B) were assayed against Src-YEEI alone (gray circles) and the Src-YEEI:pFAK complex (black circles) over the range of inhibitor concentrations shown using the Z'Lyte end-point assay. All data points were measured in quadruplicate, and the values shown represent the mean  $\pm$  SE. IC<sub>50</sub> values were determined by nonlinear regression analysis and are shown in Table 1.

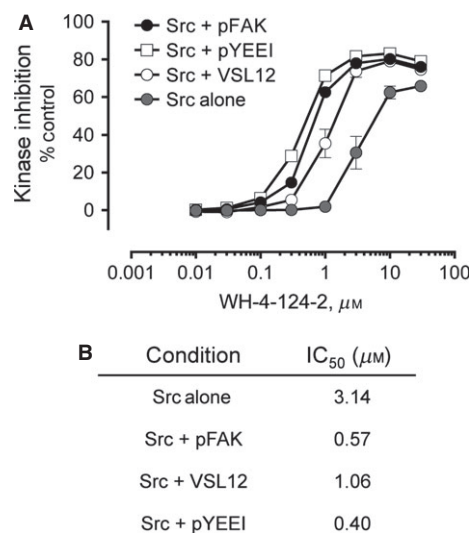
**Table 1:** IC<sub>50</sub> values for kinase inhibitors WH-4-124-2 and WH-4-023 against Src-YEEI in the presence and absence of the pFAK peptide

	WH-4-124-2		WH-4-023	
	Z'Lyte	ADP Quest	Z'Lyte	ADP Quest
Src-YEEI + pFAK <sup>a</sup>	0.89	0.114	1.36	0.025
Src-YEEI alone <sup>a</sup>	4.95	0.531	1.28	0.019
Selectivity ratio <sup>b</sup>	5.56	4.82	0.94	0.76

<sup>a</sup>IC<sub>50</sub> values ( $\mu$ M) were generated in the Z'Lyte end-point assay used for the library screen and in the ADP Quest kinetic kinase assay for Src-YEEI in complex with the pFAK peptide versus Src-YEEI alone.

<sup>b</sup>IC<sub>50</sub> for inhibition of the Src-YEEI:pFAK peptide complex divided by the IC<sub>50</sub> for inhibition of Src-YEEI alone.

We next investigated whether displacement of either the SH3 or SH2 domain alone was sufficient to cause the conformational rearrangement responsible for inhibitor selectivity. As shown in Figure 5, activation of Src-YEEI by displacement of either the SH3 domain (with the VSL12 peptide; see Figure 2B) or the SH2 domain (with the pYEEI peptide) both resulted in enhanced inhibition by WH-4-124-2. Interestingly, inhibition of VSL12-activated Src-YEEI (via SH3 domain displacement) was less



**Figure 5:** Selective inhibition of Src-YEEI by WH-4-124-2 by the pFAK peptide involves displacement of both SH2-tail and SH3-linker interactions. (A) Src-YEEI alone or together with the SH2 peptide ligand pYEEI, the SH3 peptide ligand VSL12, or the pFAK peptide was assayed over the range of inhibitor concentrations shown using the Z'Lyte end-point assay. All data points were measured in quadruplicate, and the values shown represent the mean  $\pm$  SE. (B) IC<sub>50</sub> values were determined by nonlinear regression analysis and are shown in Table 1.



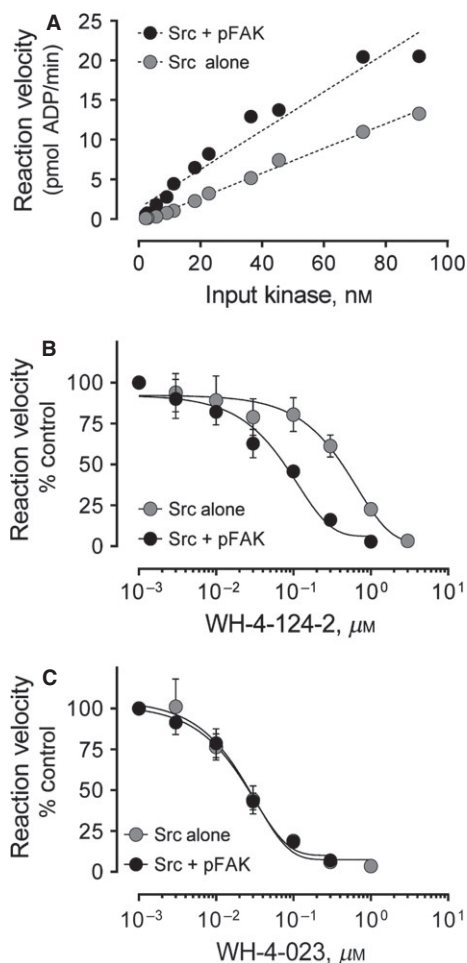
pronounced than inhibition following activation by either the pFAK or pYEEI peptides. These findings suggest that SH3 displacement alone does not allow the kinase domain to adopt the conformation preferred by this inhibitor.

Selective inhibitor discovery and characterization described so far were performed using the Z'Lyte end-point kinase assay, which does not provide direct information about the effect of either the activator peptide or the inhibitors on the kinase reaction rate. We therefore turned to the ADP Quest assay, which measures the progress of the kinase reaction as the accumulation of ADP. We first measured Src-YEEI activity in the absence or presence of the pFAK peptide to establish conditions for testing the compounds (Figure 6A). As with the Z'Lyte assay, these initial experiments identified a Src-YEEI concentration dependent on the presence of the pFAK peptide for activity (15.6 ng/well). ADP production was barely detectable in the absence of the pFAK peptide at this Src-YEEI concentration, but increased to 1.0 pmol ADP produced/min in the presence of the pFAK peptide. Src-YEEI alone at 40 ng/well resulted in the same rate of 1.0 pmol ADP produced/min and was used for comparative purposes.

Both WH-4-124-2 and WH-4-023 inhibited Src-YEEI activity in the ADP Quest assay in a concentration-dependent manner (Figure 6B, C and Table 1). WH-4-124-2 is more potent against Src-YEEI in complex with the pFAK peptide than Src-YEEI alone, with  $IC_{50}$  values of 114 versus 531 nM, respectively. By contrast, WH-4-023 inhibited Src-YEEI alone and in complex with the pFAK peptide with very similar potencies, yielding  $IC_{50}$  values of 19 and 25 nM, respectively (Figure 6C and Table 1). While the  $IC_{50}$  values generated for both inhibitors in the ADP Quest assay are lower than those generated in the Z'Lyte assay, the selectivity for the Src-YEEI:pFAK complex is the same in both. This difference in apparent potency is likely due to the different peptide substrates used as well as the assay types involved (end-point versus kinetic).

### Docking studies are consistent with pFAK peptide binding inducing a DFG flip in the c-Src kinase domain that favors WH-4-124-2 binding

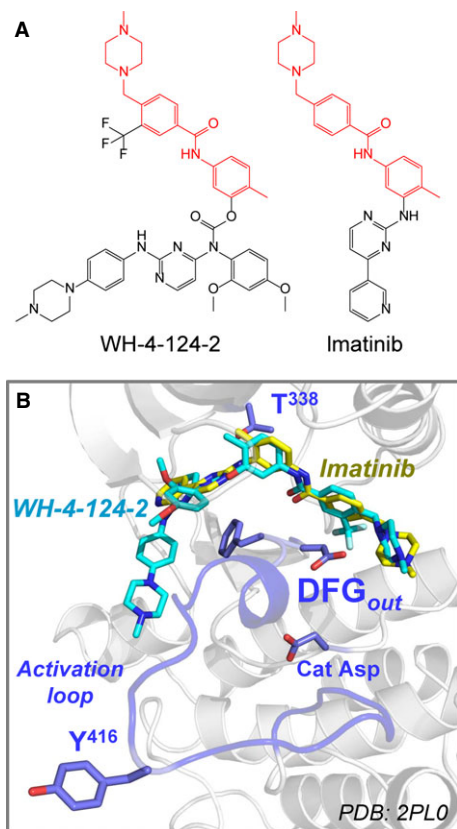
Data presented in the previous sections suggest that binding of pFAK to c-Src may induce a conformation of the kinase active site that is stabilized by WH-4-124-2 binding. One well-known determinant of kinase inhibitor specificity relates to the conformation of the highly conserved aspartate-phenylalanine-glycine ('DFG') motif at the N-terminal end of the activation loop in the kinase active site (48,49). In the active state, the aspartate in the DFG motif moves into the active site where it contributes to the co-ordination of ATP-Mg<sup>++</sup>. In the inactive state, however, this aspartate residue moves away from the active site while the phenylalanine side chain moves inward, resulting in the so-called DFG-out conformation. Previous structural studies have shown that inhibitors such as imatinib prefer the DFG-out



**Figure 6:** WH-4-124-2 exhibits increased potency while maintaining selectivity in a kinetic kinase assay. (A) The activity of Src-YEEI was measured in the ADP Quest kinetic kinase assay in the absence (gray circles) or presence (black circles) of 30 μM pFAK peptide over the range of input kinase concentrations shown. Compounds WH-4-124-2 (B) and WH-4-023 (C) were assayed against Src-YEEI alone (gray circles) and the Src-YEEI:pFAK complex (black circles) in this assay over the range of inhibitor concentrations shown. Reaction rates at each compound concentration were measured in quadruplicate and are presented as the mean ± SE.  $IC_{50}$  values were determined by nonlinear regression analysis and are shown in Table 1.

conformation, which accounts in part for its selectivity toward the Abl kinase domain (48,50). Interestingly, imatinib has also been co-crystallized with both c-Src and Lck, where it binds to a similar DFG-out conformation (43,51). These observations led us to consider whether WH-4-124-2 may exhibit some of its selectivity by inhibiting the DFG-out state of Src-family kinase domains.

Comparison of the chemical structures of WH-4-124-2 and imatinib reveals significant overlap (Figure 7A), allowing us to align WH-4-124-2 with imatinib in the X-ray crystal structure of the Lck kinase domain. This Src-family kinase structure was chosen because all of its kinase



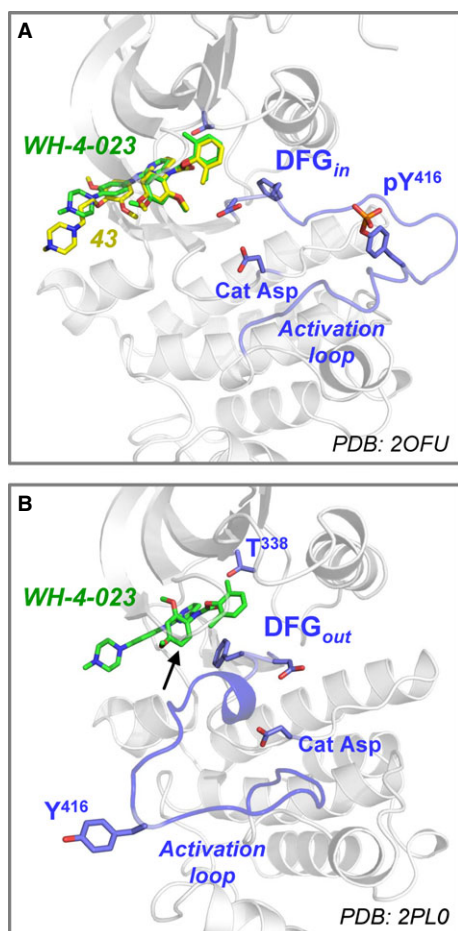
**Figure 7:** Computational docking of WH-4-124-2 to the X-ray crystal structure of the Lck kinase domain bound to imatinib. (A) Comparison of the structures of WH-4-124-2 and imatinib reveals a shared methylpiperazinylmethyl-N-phenylbenzamide backbone (red). (B) Close-up view of the Lck active site with WH-4-124-2 (carbon atoms in cyan) aligned to imatinib (carbon atoms in yellow). WH-4-124-2 was readily accommodated by this imatinib-binding conformation of the Lck kinase domain. In this structure, the N-terminal portion of the activation loop (blue) adopts the DFG-out conformation, with the autophosphorylation site (Y416) extending outward. The side chains of the catalytic aspartate and the gatekeeper residue (T338) are shown for reference. Docking model was produced using the X-ray crystal structure of the Src-family kinase Lck bound to imatinib (PDB: 2PL0). Residue numbering is based on the crystal structure of c-Src (PDB: 2SRC).

domain features are resolved, including the activation loop and autophosphorylation site, which are missing from the c-Src structure with imatinib (51). Binding of imatinib to the kinase domains of both Lck and c-Src stabilizes the DFG-out conformation, similar to what is observed with the Abl kinase domain when bound to this drug (52). In Lck, however, the activation loop and tyrosine autophosphorylation site are extended outward (Figure 7B), whereas in Abl, the equivalent tyrosine residue is hydrogen-bonded to the catalytic aspartate (52). This comparison suggests that in Lck (and by analogy, c-Src), imatinib traps a 'primed' or intermediate state between the down-regulated and fully active forms of the kinase domain.

WH-4-124-2 was aligned with imatinib in the Lck kinase domain crystal structure using the docking routine *smina* (42) a modified version of AutoDock Vina (53) optimized for user-specified custom scoring functions (see Methods for details of the docking approach). WH-4-124-2 was readily accommodated by the crystal conformation of the imatinib-bound Lck active site. Remarkably, no structural rearrangement of the imatinib-bound Lck structure was necessary to obtain the final model presented in Figure 7B, supporting the idea that this inhibitor prefers the DFG-out conformation. A similar docking routine was also performed using the X-ray crystal structure of the c-Src kinase domain bound to imatinib (PDB: 2OIQ) (51). This analysis yielded an almost identical result, although the activation loop is not resolved in this structure (not shown). These docking models suggest that binding of the pFAK peptide and displacement of the SH3-SH2 regulatory subunits allow the kinase to adopt the DFG-out conformation, which in turn is trapped by WH-4-124-2. In contrast, attempts to dock WH-4-124-2 to the kinase domain of Lck in the active conformation (PDB ID: 3LCK) (54) were unsuccessful, resulting in steric clash (not shown). In the active structure, the Lck DFG motif is rotated inward, which is incompatible with compound binding.

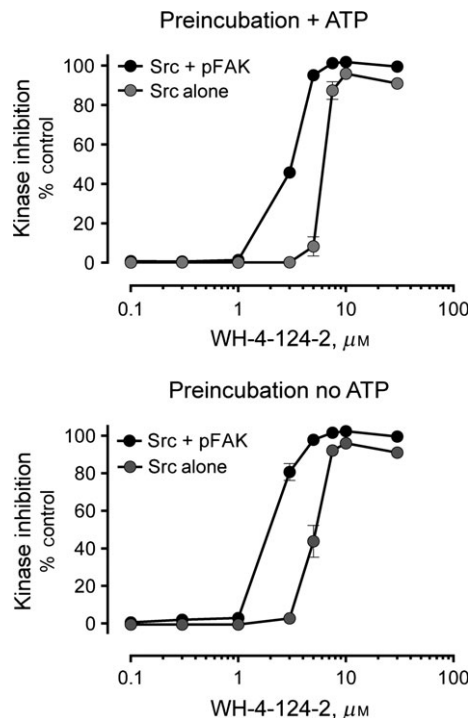
Unlike WH-4-124-2, which selectively inhibits the Src-YEEI:pFAK peptide complex, compound WH-4-023 potently inhibited c-Src activity regardless of pFAK peptide binding. This observation suggests that WH-4-023 may inhibit both the DFG-out and DFG-in conformations of the kinase domain. Molecular docking studies support this idea. The 2-aminopyrimidine carbamate backbone in WH-4-023 closely resembles a similar ligand in an X-ray crystal structure of the Lck kinase domain in the DFG-in conformation (compound 43; PDB: 2OFU) (39). Docking of WH-4-023 to this structure recovered the same binding mode found for compound 43 in the co-crystal (Figure 8A). We then docked WH-4-023 in the same pocket of the DFG-out structure of the Lck kinase domain described above (PDB: 2PL0) (43). The binding orientations of WH-4-023 are very similar in the two models (Figure 8B) and match that of the closely related ligand (compound 43) in structure 2OFU. One minor difference between the two docking models involves the orientation of the WH-4-023 dimethoxyphenyl ring, which is rotated 180° in the DFG-out binding mode to accommodate the side chain of the DFG phenylalanine (Figure 8B, arrow). However, the very similar binding modes for WH-4-023 to both states of the Lck kinase domain may help to explain the almost identical potencies of this compound for c-Src in the presence and absence of the pFAK peptide.

Docking studies presented above suggest that binding of the pFAK peptide to the SH3-SH2 region of Src allows the kinase domain to adopt the DFG-out conformation preferred by WH-4-124-2. However, kinase activation is associated with movement of the DFG motif into the active site, suggesting that preactivation of c-Src by incubation



**Figure 8:** Computational docking of WH-4-023 to the X-ray crystal structures of the DFG-in and DFG-out conformations of the Lck kinase domain. (A) View of the Lck active site with WH-4-023 (carbons in green) aligned to the related 2-aminopyrimidine carbamate (compound 43; carbons in yellow) in the crystal structure of the Lck kinase domain in the DFG-in conformation (PDB: 2OFU). In the DFG-in state, the activation loop tyrosine is phosphorylated (pY416). (B) Crystal structure of the Lck kinase domain in the DFG-out conformation (PDB: 2PLO) accommodates WH-4-023 in the same binding pocket as in panel A, with the ligand dimethoxyphenyl moiety rotated to accommodate the phenylalanine side chain of the DFG motif (arrow). In both models, the activation loop is colored in blue with the side chains of the DFG motif, the catalytic aspartate and the gatekeeper residue (T338) shown. Residue numbering is based on the crystal structure of c-Src (PDB: 2SRC).

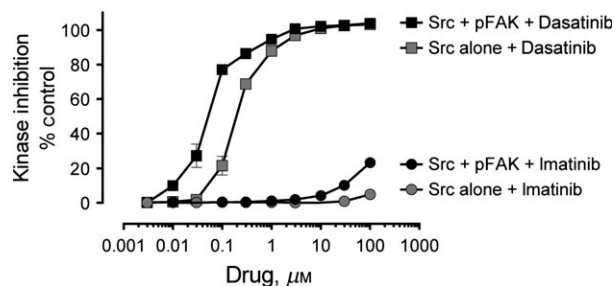
with ATP before inhibitor addition may interfere with WH-4-124-2 action. To test this idea experimentally, we performed a concentration–response experiment with WH-4-124-2 using Src-YEEI that was preincubated with ATP under conditions previously shown to result in stoichiometric autophosphorylation of the activation loop (26). As shown in Figure 9, autophosphorylation of Src-YEEI did not appreciably influence inhibition by this compound compared to a matched control preincubated under identical conditions in the absence of ATP. Addition of the pFAK



**Figure 9:** Preincubation with ATP does not affect WH-4-124-2 potency for Src-YEEI. Src-YEEI was preincubated in the presence (top panel) or absence (bottom panel) of ATP (at the  $K_m$ ) for 3 h at 25 °C to induce autophosphorylation of Tyr416 in the activation loop as described previously (26). Each kinase preparation was then assayed for sensitivity to WH-4-124-2 in the presence or absence of the pFAK peptide using the Z'Lyte end-point assay. All data points were measured in quadruplicate, and the values shown represent the mean  $\pm$  SE.

peptide to the assay enhanced inhibitor sensitivity as observed previously, regardless of ATP preincubation, providing further evidence that peptide engagement of the regulatory domains is largely responsible for the observed enhancement in inhibitor action.

Previous studies have established that the Abl kinase inhibitor imatinib derives its specificity, in part, from its preference for the DFG-out conformation of the Abl kinase domain (48). Docking studies presented in Figure 7 suggest that pFAK peptide binding may induce the DFG-out state of Src, because the preferred docking pose for WH-4-124-2 requires this kinase domain conformation. Taken together, these observations suggest that pFAK peptide binding may also enhance the sensitivity of Src to inhibition by imatinib. To test this idea, we performed concentration–response experiments with imatinib and Src-YEEI in the presence and absence of pFAK as before. As shown in Figure 10, imatinib exhibited very little activity against Src-YEEI alone, consistent with literature reports that imatinib is a very poor Src inhibitor (51). Interestingly, when the experiment was repeated with the Src:pFAK peptide complex, partial inhibition was observed, although only at very



**Figure 10:** Inhibition of Src-YEE1 by imatinib and dasatinib is influenced by pFAK peptide binding. Src-YEE1 alone and the Src-YEE1:pFAK complex were assayed over the range of dasatinib and imatinib concentrations indicated using the Z'Lyte end-point assay. All data points were measured in quadruplicate, and the values shown represent the mean  $\pm$  SE.

high imatinib concentrations (~25% inhibition at 100  $\mu\text{M}$ ). This result suggests that the additional methylphenylpiperazine and/or dimethoxybenzyl moieties present in WH-4-124-2 account for its much greater potency against the pFAK:Src complex relative to imatinib (see Figure 7). Nevertheless, these findings support the idea that pFAK peptide binding induces the DFG-out conformation. Interestingly, the presence of the pFAK peptide also resulted in a modest increase in the sensitivity of Src to inhibition by dasatinib (~4 fold; Figure 10), despite its classification as a 'Type I' kinase inhibitor (i.e., one that does not sense differences in the position of the DFG motif) (30,55). This observation suggests that SH3-SH2 engagement by the pFAK peptide may induce unique conformational changes in the active site in addition to influencing the status of the DFG motif.

## Summary and Conclusions

In this study, we describe a screening strategy that enables discovery of c-Src inhibitors that appear to prefer a specific active site conformation. SFKs are involved in a wide variety of cellular signaling pathways and most cells express multiple members of the Src kinase family, making the search for isoform- and pathway-selective inhibitors difficult. Rather than focusing on the Src kinase domain in isolation, we developed an assay method that models a disease-specific conformation of near-full-length c-Src that is induced by interaction with FAK. Using a synthetic tyrosine phosphopeptide containing the natural FAK sequence known to bind to the tandem SH3/S<sub>H</sub>2 unit of c-Src, we first identified screening assay conditions where Src-YEE1 activity was dependent on the presence of the pFAK peptide. Using pFAK-activated Src-YEE1, we then screened a small library of kinase-biased inhibitors and identified WH-4-124-2, an aminopyrimidinyl carbamate with enhanced potency for Src-YEE1 in the presence of the pFAK peptide. This compound exhibited fivefold greater potency for the Src-pFAK peptide complex relative to Src-YEE1 alone in

## Selective Inhibitors of a FAK: Src Kinase Complex

both end-point and kinetic kinase assays. Interestingly, a smaller structural analog of WH-4-124-2 showed no preference, supporting the idea that WH-4-124-2 stabilizes a specific conformation of the kinase active site that is induced by interaction with the pFAK peptide.

Docking studies with WH-4-124-2 and the imatinib-bound crystal structures of two Src-family kinase domains (c-Src and Lck) suggest that binding to the pFAK peptide may allow a unique DFG-out conformation that is preferentially inhibited by this compound. In this conformation, the  $\alpha$ C-helices of both kinase domains are rotated inward, allowing the formation of the conserved lysine to glutamate salt bridges. The DFG motif at the distal end of the activation loop in each kinase is rotated outward, in the same manner originally observed for imatinib binding to the Abl kinase domain (48). However, in the Lck structure, the autophosphorylation site is extended outward, although the tyrosine is not phosphorylated (this region is not resolved in the c-Src kinase domain structure with imatinib). Using the structural overlap with imatinib as a starting point, WH-4-124-2 was readily docked into this unique Lck kinase domain conformation. This observation with Lck suggests that pFAK peptide binding may induce a similar kinase domain conformation in c-Src, thereby accounting for preferential inhibition by WH-4-124-2 in the presence of the pFAK peptide. Whether or not binding to full-length FAK results in similar changes to the c-Src active site in a cellular context will require further investigation. Nevertheless, these studies support the more general concept that unique disease-state-specific conformations of Src-family kinases exist in cells that may be amenable to selective inhibition with active site inhibitors.

## Acknowledgments

This work was supported by National Institutes of Health grants CA169962 to T.E.S. and GM097082 to C.J.C. J.A.M. was supported by the Pittsburgh AIDS Research Training Program (NIH T32 AI065380).

## References

1. Brown M.T., Cooper J.A. (1996) Regulation, substrates, and functions of Src. *Biochim Biophys Acta*;1287:121–149.
2. Resh M.D. (1994) Myristylation and palmitoylation of Src family members: the fats of the matter. *Cell*;76:411–413.
3. Boggon T.J., Eck M.J. (2004) Structure and regulation of Src family kinases. *Oncogene*;23:7918–7927.
4. Chong Y.P., Ia K.K., Mulhern T.D., Cheng H.C. (2005) Endogenous and synthetic inhibitors of the Src-family protein tyrosine kinases. *Biochim Biophys Acta*;1754:210–220.



5. Chong Y.P., Chan A.S., Chan K.C., Williamson N.A., Lerner E.C., Smithgall T.E., Bjorge J.D., Fujita D.J., Purcell A.W., Scholz G., Mulhern T.D., Cheng H.C. (2006) C-terminal Src kinase-homologous kinase (CHK), a unique inhibitor inactivating multiple active conformations of Src family tyrosine kinases. *J Biol Chem*;281:32988–32999.
6. Xu W., Doshi A., Lei M., Eck M.J., Harrison S.C. (1999) Crystal structures of c-Src reveal features of its autoinhibitory mechanism. *Mol Cell*;3:629–638.
7. Wang D., Esselman W.J., Cole P.A. (2002) Substrate conformational restriction and CD45-catalyzed dephosphorylation of tail tyrosine-phosphorylated Src protein. *J Biol Chem*;277:40428–40433.
8. Lerner E.C., Smithgall T.E. (2002) SH3-dependent stimulation of Src-family kinase autophosphorylation without tail release from the SH2 domain *in vivo*. *Nat Struct Biol*;9:365–369.
9. Lerner E.C., Tribble R.P., Schiavone A.P., Hochrein J.M., Engen J.R., Smithgall T.E. (2005) Activation of the Src family kinase Hck without SH3-linker release. *J Biol Chem*;280:40832–40837.
10. Briggs S.D., Sharkey M., Stevenson M., Smithgall T.E. (1997) SH3-mediated Hck tyrosine kinase activation and fibroblast transformation by the Nef protein of HIV-1. *J Biol Chem*;272:17899–17902.
11. Moarefi I., LaFevre-Bernt M., Sicheri F., Huse M., Lee C.-H., Kuriyan J., Miller W.T. (1997) Activation of the Src-family tyrosine kinase Hck by SH3 domain displacement. *Nature*;385:650–653.
12. Nakamoto T., Sakai R., Ozawa K., Yazaki Y., Harai H. (1996) Direct binding of C-terminal region of p130<sup>Cas</sup> to SH2 and SH3 domains of Src kinase. *J Biol Chem*;271:8959–8965.
13. Thomas J.W., Ellis B., Boerner R.J., Knight W.B., White G.C., Schaller M.D. (1998) SH2- and SH3-mediated interactions between focal adhesion kinase and Src. *J Biol Chem*;273:577–583.
14. Hayashi I., Vuori K., Liddington R.C. (2002) The focal adhesion targeting (FAT) region of focal adhesion kinase is a four-helix bundle that binds paxillin. *Nat Struct Biol*;9:101–106.
15. Arold S.T., Hoellerer M.K., Noble M.E. (2002) The structural basis of localization and signaling by the focal adhesion targeting domain. *Structure*;10:319–327.
16. Lietha D., Cai X., Ceccarelli D.F., Li Y., Schaller M.D., Eck M.J. (2007) Structural basis for the autoinhibition of focal adhesion kinase. *Cell*;129:1177–1187.
17. Zhao X., Guan J.L. (2011) Focal adhesion kinase and its signaling pathways in cell migration and angiogenesis. *Adv Drug Deliv Rev*;63:610–615.
18. Klinghoffer R.A., Sachsenmaier C., Cooper J.A., Soriano P. (1999) Src family kinases are required for integrin but not PDGFR signal transduction. *EMBO J*;18:2459–2471.
19. Ilic D., Furuta Y., Kanazawa S., Takeda N., Sobue K., Nakatsuji N., Nomura S., Fujimoto J., Okada M., Yamamoto T. (1995) Reduced cell motility and enhanced focal adhesion contact formation in cells from FAK-deficient mice. *Nature*;377:539–544.
20. Zhao J., Guan J.L. (2009) Signal transduction by focal adhesion kinase in cancer. *Cancer Metastasis Rev*;28:35–49.
21. Kim L.C., Song L., Haura E.B. (2009) Src kinases as therapeutic targets for cancer. *Nat Rev Clin Oncol*;6:587–595.
22. Zhang S., Yu D. (2012) Targeting Src family kinases in anti-cancer therapies: turning promise into triumph. *Trends Pharmacol Sci*;33:122–128.
23. McLean G.W., Carragher N.O., Avizienyte E., Evans J., Brunton V.G., Frame M.C. (2005) The role of focal-adhesion kinase in cancer – a new therapeutic opportunity. *Nat Rev Cancer*;5:505–515.
24. Lieu C., Kopetz S. (2010) The SRC family of protein tyrosine kinases: a new and promising target for colorectal cancer therapy. *Clin Colorectal Cancer*;9:89–94.
25. Puls L.N., Eadens M., Messersmith W. (2011) Current status of SRC inhibitors in solid tumor malignancies. *Oncologist*;16:566–578.
26. Moroco J.A., Craigo J.K., Iacob R.E., Wales T.E., Engen J.R., Smithgall T.E. (2014) Differential sensitivity of Src-family kinases to activation by SH3 domain displacement. *PLoS ONE*;9:e105629.
27. Tribble R.P., Emert-Sedlak L., Smithgall T.E. (2006) HIV-1 Nef selectively activates SRC family kinases HCK, LYN, and c-SRC through direct SH3 domain interaction. *J Biol Chem*;281:27029–27038.
28. Seeliger M.A., Young M., Henderson M.N., Pellicena P., King D.S., Falick A.M., Kuriyan J. (2005) High yield bacterial expression of active c-Abl and c-Src tyrosine kinases. *Protein Sci*;14:3135–3139.
29. Iacob R.E., Pene-Dumitrescu T., Zhang J., Gray N.S., Smithgall T.E., Engen J.R. (2009) Conformational disturbance in Abl kinase upon mutation and deregulation. *Proc Natl Acad Sci U S A*;106:1386–1391.
30. Liu Y., Gray N.S. (2006) Rational design of inhibitors that bind to inactive kinase conformations. *Nat Chem Biol*;2:358–364.
31. Okram B., Nagle A., Adrian F.J., Lee C., Ren P., Wang X., Sim T., Xie Y., Wang X., Xia G., Spraggon G., Warmuth M., Liu Y., Gray N.S. (2006) A general strategy for creating “inactive-conformation” abl inhibitors. *Chem Biol*;13:779–786.
32. Rodems S.M., Hamman B.D., Lin C., Zhao J., Shah S., Heidary D., Makings L., Stack J.H., Pollok B.A. (2002) A FRET-based assay platform for ultra-high density drug screening of protein kinases and phosphatases. *Assay Drug Dev Technol*;1:9–19.
33. Hellwig S., Miduturu C.V., Kanda S., Zhang J., Filippakopoulos P., Salah E., Deng X., Choi H.G., Zhou W., Hur W., Knapp S., Gray N.S., Smithgall T.E. (2012) Small-molecule inhibitors of the c-Fes protein-tyrosine kinase. *Chem Biol*;19:529–540.
34. Emert-Sedlak L., Kodama T., Lerner E.C., Dai W., Foster C., Day B.W., Lazo J.S., Smithgall T.E. (2009)



- Chemical library screens targeting an HIV-1 accessory factor/host cell kinase complex identify novel antiretroviral compounds. *ACS Chem Biol*;4:939–947.
35. Eck M.J., Shoelson S.E., Harrison S.C. (1993) Recognition of a high-affinity phosphotyrosyl peptide by the Src homology-2 domain of p56lck. *Nature*;362:87–91.
  36. Rickles R.J., Botfield M.C., Zhou X.M., Henry P.A., Brugge J.S., Zoller M.J. (1995) Phage display selection of ligand residues important for Src homology 3 domain binding specificity. *Proc Natl Acad Sci U S A*;92:10909–10913.
  37. Charter N.W., Kauffman L., Singh R., Eglén R.M. (2006) A generic, homogenous method for measuring kinase and inhibitor activity via adenosine 5'-diphosphate accumulation. *J Biomol Screen*;11:390–399.
  38. Lam K.S., Wu J., Lou Q. (1995) Identification and characterization of a novel synthetic peptide substrate specific for Src-family protein tyrosine kinases. *Int J Pept Protein Res*;45:587–592.
  39. Martin M.W., Newcomb J., Nunes J.J., McGowan D.C., Armistead D.M., Boucher C., Buchanan J.L. *et al.* (2006) Novel 2-aminopyrimidine carbamates as potent and orally active inhibitors of Lck: synthesis, SAR, and *in vivo* antiinflammatory activity. *J Med Chem*;49:4981–4991.
  40. Hawkins P.C., Skillman A.G., Warren G.L., Ellingson B.A., Stahl M.T. (2010) Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J Chem Inf Model*;50:572–584.
  41. Koes D.R., Camacho C.J. (2012) ZINCPharmer: pharmacophore search of the ZINC database. *Nucleic Acids Res*;40:W409–W414.
  42. Koes D.R., Baumgartner M.P., Camacho C.J. (2013) Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *J Chem Inf Model*;53:1893–1904.
  43. Jacobs M.D., Caron P.R., Hare B.J. (2008) Classifying protein kinase structures guides use of ligand-selectivity profiles to predict inactive conformations: structure of lck/imatinib complex. *Proteins*;70:1451–1460.
  44. Emert-Sedlak L.A., Narute P., Shu S.T., Poe J.A., Shi H., Yanamala N., Alvarado J.J., Lazo J.S., Yeh J.I., Johnston P.A., Smithgall T.E. (2013) Effector kinase coupling enables high-throughput screens for direct HIV-1 Nef antagonists with antiretroviral activity. *Chem Biol*;20:82–91.
  45. Boerner R.J., Barker S.C., Knight W.B. (1995) Kinetic mechanisms of the forward and reverse pp60c-src tyrosine kinase reactions. *Biochemistry*;34:16419–16423.
  46. Barker S.C., Kassel D.B., Weigl D., Huang X., Luther M.A., Knight W.B. (1995) Characterization of pp60c-src tyrosine kinase activities using a continuous assay: autoactivation of the enzyme is an intermolecular autophosphorylation process. *Biochemistry*;34:14843–14851.
  47. Knight Z.A., Shokat K.M. (2005) Features of selective kinase inhibitors. *Chem Biol*;12:621–637.
  48. Schindler T., Bornmann W., Pellicena P., Miller W.T., Clarkson B., Kuriyan J. (2000) Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science*;289:1938–1942.
  49. Treiber D.K., Shah N.P. (2013) Ins and outs of kinase DFG motifs. *Chem Biol*;20:745–746.
  50. Nagar B., Bornmann W.G., Pellicena P., Schindler T., Veach D.R., Miller W.T., Clarkson B., Kuriyan J. (2002) Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). *Cancer Res*;62:4236–4243.
  51. Seeliger M.A., Nagar B., Frank F., Cao X., Henderson M.N., Kuriyan J. (2007) c-Src binds to the cancer drug imatinib with an inactive Abl/c-Kit conformation and a distributed thermodynamic penalty. *Structure*;15:299–311.
  52. Panjarian S., Iacob R.E., Chen S., Engen J.R., Smithgall T.E. (2013) Structure and dynamic regulation of Abl kinases. *J Biol Chem*;288:5443–5450.
  53. Trott O., Olson A.J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*;31:455–461.
  54. Yamaguchi H., Hendrickson W.A. (1996) Structural basis for activation of human lymphocyte kinase Lck upon tyrosine phosphorylation. *Nature*;384:484–489.
  55. Tokarski J.S., Newitt J.A., Chang C.Y., Cheng J.D., Wittekind M., Kiefer S.E., Kish K., Lee F.Y., Borzilleri R., Lombardo L.J., Xie D., Zhang Y., Klei H.E. (2006) The structure of Dasatinib (BMS-354825) bound to activated ABL kinase domain elucidates its inhibitory activity against imatinib-resistant ABL mutants. *Cancer Res*;66:5790–5797.

## **A.2 MODELING OF ALLOSTERIC ACTIVATORS OF ABL**

RESEARCH ARTICLE

# Fluorescence Polarization Screening Assays for Small Molecule Allosteric Modulators of ABL Kinase Function

Prerna Grover<sup>1</sup>, Haibin Shi<sup>1</sup>, Matthew Baumgartner<sup>2</sup>, Carlos J. Camacho<sup>2</sup>, Thomas E. Smithgall<sup>1\*</sup>

**1** Department of Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, United States of America, **2** Department of Computational and Systems Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, United States of America

\* [tsmithga@pitt.edu](mailto:tsmithga@pitt.edu)



 OPEN ACCESS

**Citation:** Grover P, Shi H, Baumgartner M, Camacho C.J, Smithgall TE (2015) Fluorescence Polarization Screening Assays for Small Molecule Allosteric Modulators of ABL Kinase Function. PLoS ONE 10(7): e0133590. doi:10.1371/journal.pone.0133590

**Editor:** Laszlo Buday, Hungarian Academy of Sciences, HUNGARY

**Received:** March 19, 2015

**Accepted:** June 30, 2015

**Published:** July 29, 2015

**Copyright:** © 2015 Grover et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper.

**Funding:** This work was supported by National Institutes of Health Grants CA185702 (to T.E.S.) and GM097082 (C.J.C.) as well as Commonwealth of Pennsylvania Department of Health Grant SAP 4100062224 (C.J.C.). Computational resources were supported by National Science Foundation Grant CNS1229064 (C.J.C.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

The ABL protein-tyrosine kinase regulates intracellular signaling pathways controlling diverse cellular processes and contributes to several forms of cancer. The kinase activity of ABL is repressed by intramolecular interactions involving its regulatory Ncap, SH3 and SH2 domains. Small molecules that allosterically regulate ABL kinase activity through its non-catalytic domains may represent selective probes of ABL function. Here we report a screening assay for chemical modulators of ABL kinase activity that target the regulatory interaction of the SH3 domain with the SH2-kinase linker. This fluorescence polarization (FP) assay is based on a purified recombinant ABL protein consisting of the N-cap, SH3 and SH2 domains plus the SH2-kinase linker (N32L protein) and a short fluorescein-labeled probe peptide that binds to the SH3 domain. In assay development experiments, we found that the probe peptide binds to the recombinant ABL N32L protein in vitro, producing a robust FP signal that can be competed with an excess of unlabeled peptide. The FP signal is not observed with control N32L proteins bearing either an inactivating mutation in the SH3 domain or enhanced SH3:linker interaction. A pilot screen of 1200 FDA-approved drugs identified four compounds that specifically reduced the FP signal by at least three standard deviations from the untreated controls. Secondary assays showed that one of these hit compounds, the antithrombotic drug dipyridamole, enhances ABL kinase activity in vitro to a greater extent than the previously described ABL agonist, DPH. Docking studies predicted that this compound binds to a pocket formed at the interface of the SH3 domain and the linker, suggesting that it activates ABL by disrupting this regulatory interaction. These results show that screening assays based on the non-catalytic domains of ABL can identify allosteric small molecule regulators of kinase function, providing a new approach to selective drug discovery for this important kinase system.

**Competing Interests:** The authors have declared that no competing interests exist.

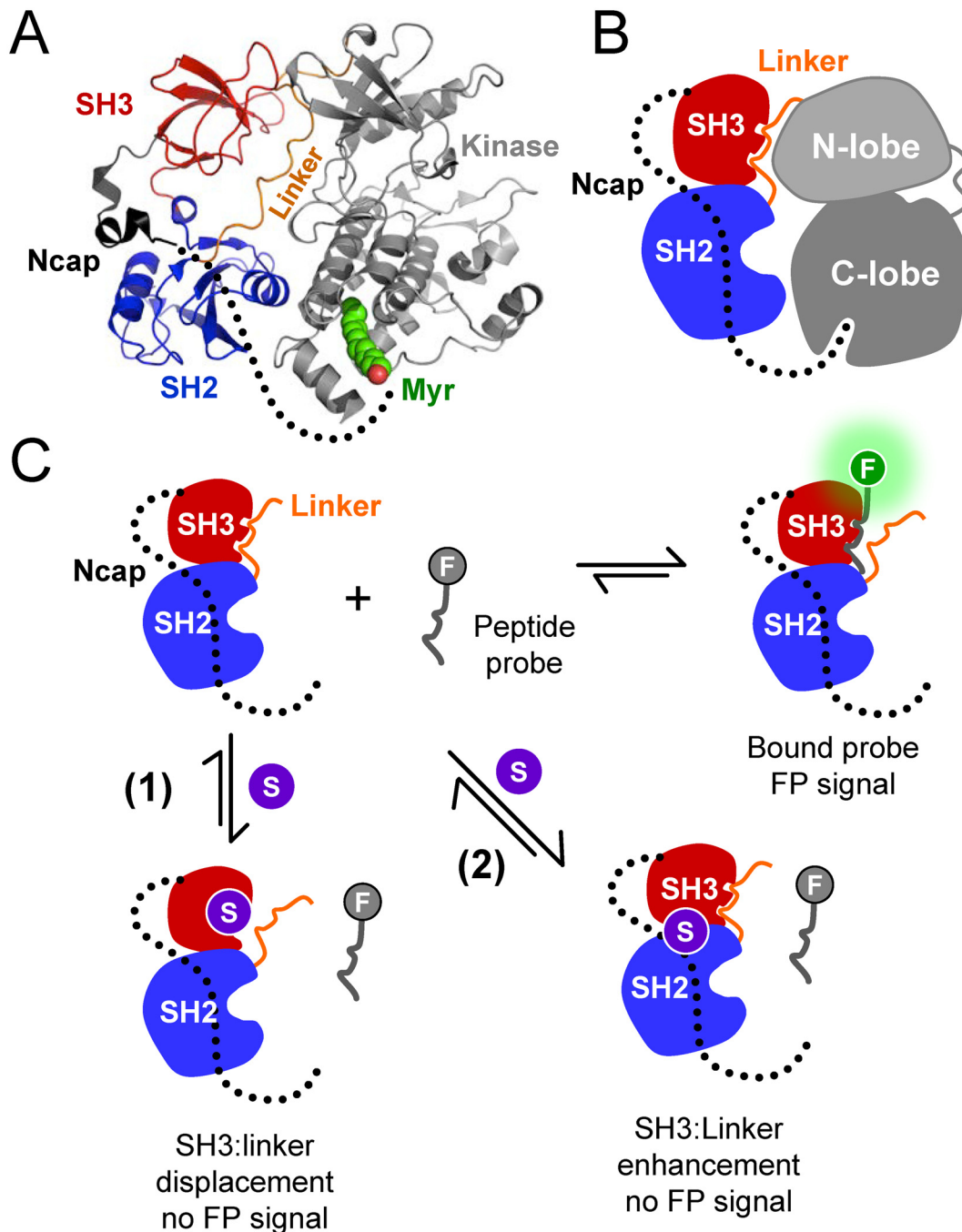
## Introduction

The ABL protein-tyrosine kinase plays diverse roles in the regulation of cell proliferation, survival, adhesion, migration and the genotoxic stress response [1–3]. ABL kinase activity is perhaps best known in the context of BCR-ABL, the translocation gene product responsible for chronic myelogenous leukemia (CML) and some forms of acute lymphocytic leukemia [4,5]. The clinical management of CML has been revolutionized by selective ATP-competitive inhibitors of BCR-ABL, of which imatinib is the prototype [6]. However, chronic use of kinase inhibitors often leads to drug resistance due to selection for mutations that disrupt drug binding or allosterically influence the conformation of the drug binding pocket [7].

The growing problem of imatinib resistance in BCR-ABL has fueled efforts to identify compounds that work outside of the kinase active site. Such compounds offer advantages in terms of enhanced specificity, because they have the potential to exploit non-conserved regulatory features unique to ABL that persist to some extent in BCR-ABL as well [8]. The kinase activity of ABL is tightly regulated *in vivo* by an auto-inhibitory mechanism. The ABL ‘core’ region, which includes a myristoylated N-terminal ‘cap’ (N-cap), SH3 and SH2 domains, an SH2-kinase linker and the kinase domain, is both necessary and sufficient for ABL auto-inhibition [9]. Subsequent X-ray crystal structures of the ABL core revealed three critical intramolecular interactions that regulate kinase activity [10–12] (Fig 1A and 1B). First, the SH2-kinase linker forms a polyproline type II helix that binds to the SH3 domain, forming an interface between the SH3 domain and the N-lobe of the kinase domain. Second, the SH2 domain interacts with the back of the kinase domain C-lobe through an extensive network of hydrogen bonds. Aromatic interactions between the side chains of SH2 Tyr158 and kinase domain Tyr361 also help to stabilize this interaction (see Panjarian *et al.* for an explanation of the ABL amino acid numbering scheme [13]). Finally, the myristoylated N-cap binds a deep hydrophobic pocket in the C-lobe of the kinase domain, clamping the SH3 and SH2 domains against the back of the kinase domain. Small molecules that occupy the myristic acid binding site in the C-lobe of the kinase domain have proven to be effective allosteric inhibitors of BCR-ABL function [14,15].

Mutational analysis demonstrates that intramolecular SH3:linker interaction plays a central role in ABL auto-inhibition. Substitution of linker proline residues at positions 242 and 249 with glutamate disrupts SH3:linker interaction, resulting in ABL kinase activation [16]. In contrast, increasing the proline content of the linker enhances internal SH3 binding and overcomes the activating effects of mutations in the myristic acid binding pocket as well as the kinase domain gatekeeper residue (Thr315) [8]. Remarkably, enhanced SH3:linker interaction also dramatically sensitizes BCR-ABL-transformed cells to inhibition by both imatinib and the allosteric inhibitor, GNF-2, which binds to the myristic acid binding pocket [8]. These findings suggest that small molecules enhancing or disrupting this natural regulatory mechanism may represent selective allosteric modulators of ABL kinase activity.

In contrast to the tremendous research efforts invested in discovering inhibitors for ABL kinase activity, few studies have explored the discovery of small molecules that activate ABL. Selective agonists may represent useful probes to examine the role of ABL kinase activity in normal cellular functions, such as DNA-damage repair pathways. ABL is activated in response to multiple forms of genotoxic stress and interacts with modulators of both DNA damage-induced apoptosis (e.g. p73) and DNA repair (e.g. Rad51), resulting in cell death or survival depending on the cellular environment [2]. Moreover, recent studies have shown that ABL inhibits the growth of breast cancer xenografts and promotes the phenotypic reversion of invasive breast cancer cells [17,18]. Selective and potent agonists would provide a new approach to explore these and other biological roles of ABL *in vivo*.



**Fig 1. FP assay for small molecule modulators of ABL kinase function.** A) Crystal structure of the auto-inhibited ABL core (PDB: 2FO0) [11]. Key features include the N-cap, SH3 and SH2 domains, the SH2-kinase linker, and the kinase domain. The disordered portion of the N-cap is indicated by the dotted line. The N-terminal portion of the N-cap is myristoylated and engages a deep pocket in the kinase domain C-lobe. B) Cartoon depiction of the intramolecular interactions regulating assembly of the downregulated ABL core. Note that the linker forms a polyproline helix that binds in *cis* to the SH3 domain. C) Fluorescence polarization (FP) assay. The FP assay combines a recombinant ABL Ncap-SH3-SH2-linker (N32L) protein and a SH3-binding peptide probe labeled with a fluorescent moiety (F). The probe peptide binds the SH3 domain in the ABL N32L protein, resulting in an FP signal. Small molecules (S) may bind to the SH3 domain and block probe peptide binding directly; such molecules would be expected to disrupt SH3:linker interaction (case 1). Alternatively, small molecules may stabilize SH3:linker interaction, making the SH3 domain inaccessible to the probe peptide (case 2). In either case, small molecule binding is predicted to result in a loss of the FP signal.

doi:10.1371/journal.pone.0133590.g001



In this study, we report the development and validation of a high-throughput screening assay for the identification of small molecules that interact directly with the non-catalytic region of the ABL core. Our assay is based on the interaction of a fluorescent probe peptide with the SH3 domain in the context of a recombinant protein encompassing the regulatory region of the ABL core (Ncap-SH3-SH2-linker; referred to hereafter as the ABL 'N32L' protein). Interaction of the probe peptide with the ABL N32L protein results in fluorescence polarization (FP), providing a convenient assay for SH3 occupancy in a format compatible with high-throughput chemical library screening. In theory, small molecules that bind to SH3 and disrupt probe binding may also disrupt SH3:linker interaction in the context of ABL, resulting in kinase activation. On the other hand, compounds that enhance internal SH3 binding to the natural linker may represent allosteric inhibitors. Using this FP approach, we screened a small library of FDA-approved drugs and identified a compound that specifically inhibited the FP signal from the complex of the ABL N32L protein with the probe peptide, suggesting that it may interfere with SH3:linker interaction. This compound, a substituted pyrimido-pyrimidine known as dipyridamole, was confirmed to bind directly to the ABL N32L protein using both differential scanning fluorimetry and surface plasmon resonance. Dipyridamole was found to enhance the activity of a recombinant downregulated ABL core protein, but had no effect on an ABL core with engineered high-affinity SH3:linker interaction or on the SRC-family kinase, HCK. These observations suggest that dipyridamole binds selectively to the ABL SH3 domain, resulting in linker displacement and kinase activation. This binding model is supported by molecular dynamics (MD) simulations and molecular docking. Our findings provide an important proof-of-concept that small molecules acting on the ABL SH3:linker interface can allosterically modulate ABL kinase activity, and provide a simple yet powerful assay method for their discovery.

## Materials and Methods

### Expression and purification of recombinant kinase proteins

The coding sequence for the ABL Ncap-SH3-SH2-linker region (N32L; corresponding to residues 2–255 with an internal deletion of residues 15–56; numbering based on the crystal structure of the human ABL core; PDB: 2FO0 [11]) was amplified by PCR and subcloned into the bacterial expression vector, pET21a (EMD Millipore). A similar construct was prepared using the sequence of ABL with a high-affinity linker ('HAL9') substitution as described previously [8]. One glycine and six histidine residues (GHHHHHH) were introduced at the N-terminus of the coding sequence of these proteins during sub-cloning. An inactivating mutation of the SH3 domain (W118A) was introduced by site-directed mutagenesis using the QuikChange II method (Stratagene) and the pET21a-ABL N32L WT plasmid as a template. The ABL N32L proteins were expressed in *E. coli* strain Rosetta2(DE3)pLysS (EMD Millipore) and purified using immobilized metal affinity chromatography. The purified proteins were then dialyzed against 20 mM Tris-HCl (pH 8.3) containing 200 mM NaCl and 1 mM DTT.

The wild type and high-affinity linker ('HAL9') ABL core proteins (residues 1–531 with an internal deletion of residues 15–56) were expressed in Sf9 insect cells as previously described [8]. The ABL core proteins were purified using a combination of ion-exchange and affinity chromatography and dialyzed against 20 mM Tris-HCl (pH 8.3) containing 100 mM NaCl and 3 mM DTT. The molecular weight and purity of all recombinant ABL proteins were confirmed by SDS-PAGE and mass spectrometry.

The SRC-family kinase HCK-YEEI was expressed in Sf9 insect cells and purified as previously described [19].

## Peptide synthesis

ABL SH3 domain-binding peptides p41, p40, p8, and 3BP-1 [20,21] were synthesized by the University of Pittsburgh Genomics and Proteomics Core Laboratories. For the FP assay, the peptides were labeled with 6-carboxyfluorescein at their N-termini. Molecular weight and purity of all peptides were verified by mass spectrometry. Stock solutions (10 mM) were prepared in a 1:1 mixture of DMSO and FP assay buffer (20 mM Tris-HCl, pH 8.3) for labeled peptides and neat FP assay buffer for unlabeled peptides. Peptide stock solutions were stored at -20°C.

## Fluorescence polarization assay

Fluorescence Polarization (FP) experiments were performed in quadruplicate in low volume black 384 well plates with a non-binding surface (Corning; catalog # 3676). Peptides and proteins were added to each well in FP assay buffer (20 mM Tris-HCl, pH 8.3) for a final assay volume of 20  $\mu$ L and mixed by shaking for 5 min at room temperature. The FP signal in millipolarization (mP) units was measured at an excitation wavelength of 485 nM and emission wavelength of 515 nM in a SpectraMax M5 microplate reader (Molecular Devices) using the Softmax Pro software (version 5.4.1). Each plate was read three times and the values were averaged prior to analysis. Raw fluorescence intensity was also read at the same wavelengths for each assay.

## Chemical library screening

A collection of 1200 FDA-approved small molecules (Prestwick Chemical, Inc.) was used in a pilot screen with the FP assay. Each library compound was screened at 10  $\mu$ M and a final DMSO concentration of 1%. Compounds were added to 384-well assay plates first, followed by a pre-mixed complex of the ABL N32L protein (25  $\mu$ g) and the p41 probe peptide (50 nM). Each plate also contained twenty-eight wells of the wild type N32L protein plus p41 probe and DMSO as positive controls as well as twenty-eight wells of mutant ABL N32L-W118A protein plus p41 probe peptide and DMSO as negative controls. Each plate was mixed on the shaker for 5 min, read three consecutive times, and the average FP signal for each well was calculated. To identify potential hit compounds, three measures were used: (i) average FP signal; (ii) control normalized percent inhibition, in which the FP signal with each compound was normalized to the mean FP signal of the positive and negative plate controls according to the formula:  $[(\text{sample FP} - \text{mean FP}_{\text{WT}}) / (\text{mean FP}_{\text{W118A}} - \text{mean FP}_{\text{WT}}) \times 100]$  [22,23]; (iii) Z score, a statistical measure of variation of the mean sample FP signal that is independent of plate controls, calculated according to the formula:  $[(\text{sample FP} - \text{mean FP}_{\text{samples}}) / \text{standard deviation}_{\text{samples}}]$  [22,23]. The compounds were then ranked in order of increasing FP signal, decreasing control normalized percent inhibition, and increasing Z score. Potential hit compounds were then retested in quadruplicate using the FP assay under screening assay conditions.

## Differential Scanning Fluorimetry

Hit compounds (100  $\mu$ M) were pre-incubated with the ABL N32L WT protein (1  $\mu$ M) for 30 minutes in bicine assay buffer (10 mM bicine, 150 mM NaCl, pH 8.0). SYPRO Orange (Sigma) was added at 5X final concentration and fluorimetry profiles were acquired with a StepOnePlus real-time quantitative PCR instrument (Applied Biosystems) and software (version 2.3). Assays were performed in duplicate in sealed MicroAmp Fast 96-well qPCR plates (Applied Biosystems), and control reactions without proteins were included to correct for background fluorescence. Assays were equilibrated at 25°C for 2 minutes, followed by an increase in temperature at the rate of 1% (1.6°C/min) to 99°C, with continuous data collection. Mean fluorescence intensities, after subtracting background fluorescence, were plotted against temperature. Non-



linear regression analysis using the Boltzmann sigmoid function in GraphPad Prism 6 was used to determine the  $T_m$  values, the midpoint of the melt curve between the minimum and maximum fluorescence intensities.

## Surface Plasmon Resonance (SPR)

SPR analysis was performed on a BIAcore T100 instrument (GE Healthcare) using four-channel CM5 biosensor chips at 25°C. Recombinant purified ABL proteins were covalently attached to the CM5 chip via standard amine coupling chemistry [24,25]. Compound 142 (dipyridamole; Prestwick Chemical) was prepared in 20 mM Tris-HCl, pH 8.3, 150 mM NaCl and 0.1% DMSO and flowed past the immobilized ABL protein channel and a reference channel on the biosensor at a flow rate of 50  $\mu$ L/min for 3 min over a range of concentrations. The initial binding reaction was followed by dissociation for 5 min, and the chip surface was regenerated using 20 mM Tris-HCl, pH 8.3, 150 mM NaCl, 0.1% DMSO, 0.05% Tween 20 and 1 mM DTT at a flow rate of 50  $\mu$ L/min for 10 min. Sensorgrams were recorded in triplicate, corrected for buffer effects, and fitted with the 1:1 Langmuir binding model using the BIAevaluation software suite version 2.0.4 (GE Healthcare).

## Protein kinase assays

The ADP Quest assay (DiscoverRx) [26], which fluorimetrically measures kinase activity as the production of ADP, was used to determine ABL kinase reaction velocities. Assays were performed in quadruplicate in black 384 well plates (Corning #3571) in reaction volumes of 10  $\mu$ L/well. Recombinant kinase protein concentrations were fixed at 40 ng/well for the wild type ABL core, 9 ng/well for the high affinity linker ABL core, and 250 ng/well for the SRC family kinase HCK. The Tyr2 substrate peptide for ABL (EAIYAAPFAKKK) as well as the SFK substrate peptide (YIYGSEFK) were dissolved in the ADP Quest assay buffer (15 mM HEPES, pH 7.4, 20 mM NaCl, 1 mM EGTA, 0.02% Tween-20, 10 mM MgCl<sub>2</sub>, 0.1 mg/ml bovine  $\gamma$ -globulins), while ATP stocks were prepared in 10 mM Tris-HCl (pH 7.0). Each kinase reaction was initiated by the addition of ATP and read at 5 min intervals for 3 h in a SpectraMax M5 Microplate reader (Molecular Devices). To determine the substrate  $K_m$  for ABL kinases, the ATP concentration was fixed at 50  $\mu$ M and the substrate peptide was serially diluted from 0.2–200  $\mu$ M. For ABL kinase ATP  $K_m$  determination, the substrate concentration was fixed at the respective substrate  $K_m$  for each of the kinases, and the ATP concentration was titrated over the range of 0.2–200  $\mu$ M. The resulting progress curves were analyzed according to the method of Moroco et al. [19]. Briefly, raw fluorescence data were corrected for non-enzymatic ADP production (no kinase or substrate control) and kinase auto-phosphorylation (rate observed in the absence of substrate), and converted to pmol ADP produced using an ADP standard curve generated under the same reaction conditions. The resulting values were normalized to the amount of kinase present in each reaction, and plotted against time. The linear portion of each progress curve was fit by regression analysis to determine the reaction velocity. Substrate and ATP  $K_m$  values were determined by non-linear regression analysis using the Michaelis-Menten equation (GraphPad Prism 6).

Half-maximal effective concentrations ( $EC_{50}$ ) and activation constants ( $K_{act}$ ) were determined for the ABL kinase core by both dipyridamole and the known ABL activator, DPH (5-(1,3-diaryl-1H-pyrazol-4-yl)hydantoin; Sigma-Aldrich) [27]. The kinase was pre-incubated with each compound (10 nM to 100  $\mu$ M) for 30 min at room temperature, followed by kinase assay with the substrate and ATP concentrations fixed at their respective  $K_m$  values. The rate of each reaction was plotted against compound concentration, and analyzed by non-linear regression analysis (GraphPad Prism 6) to determine the  $EC_{50}$  value. To determine the activation constant

$K_{act}$ , the basal rate of kinase activity was subtracted from the rates of reaction in the presence of each compound concentration, and plotted as a function of compound concentration. The resulting curves obeyed saturation kinetics and were best-fit by the following equation [19]:

$$V_a = V_{act}[L]/(K_{act} + [L])$$

where  $V_a$  is the reaction velocity in the presence of each activator concentration,  $V_{act}$  is the maximal reaction velocity,  $L$  is the activator concentration, and  $K_{act}$  is the activator concentration that yields half-maximal reaction velocity.

## Molecular dynamics

To understand the dynamics of the recombinant N32L protein, for which there is no X-ray crystal structure, we ran unconstrained molecular dynamics (MD) simulations of residues 65–254 of the assembled, downregulated ABL core structure (PDB 2FO0). We disrupted the interaction between linker Pro249 and the SH3 domain by rotating the backbone bonds of linker Gly246. This glycine residue was chosen as a pivot point as it is more flexible and is located between Pro249 and the next strongly interacting residue (Val244) based on the predicted interaction energy with the SH3 domain [28]. We also ran MD simulations of the (unmodified) isolated SH3 domain (residues 80–145) using the same parameters described below.

MD simulations were conducted with the *pmemd.cuda* [29] module of AMBER14 [30], using the force fields AMBER ff14SB and gaff (general amber force fields) [31]. An octahedral TIP3P water box was constructed with 12 Å from the edge of the box to the solute and the total system charge was neutralized by adding chloride ions. The non-bonded cutoff was specified at 10 Å. In the first energy minimization run, the solute was held fixed and the solvent was relaxed through 500 cycles of steepest descent followed by 500 cycles of conjugate gradient minimization. Subsequently, the system was minimized again with no constraints through 2,000 cycles of steepest descent followed by 3,000 cycles of conjugate gradient minimization. Following the energy minimization, a 50,000 step MD simulation was used to raise the system temperature to 300 K while holding the solute fixed with weak (10.0 kcal/mol) restraints on the solute atoms. The bonds involving hydrogens were held at a fixed length and an integration step of 2 fs was used. This simulation was followed by a second equilibration simulation at constant pressure for 50,000 steps. The final MD simulation of this equilibrated structure was run with no constraints for 100 ns.

## Computational docking

The binding mode of hit compound 142 (dipyrimadole) was modeled to snapshots of the N32L and SH3 simulations by molecular docking using the program smina [32] with default docking parameters. The box was defined by the coordinates of linker residues 247–251 plus an outer shell of 8 Å after alignment to the SH3 domain of the crystal structure (PDB 2FO0).

## Transient expression of ABL core proteins in 293T cells

ABL core proteins were expressed in human 293T cells and analyzed for kinase activity as described elsewhere [8]. Briefly, cells were grown overnight in 60 mm plates, transfected with an expression vector for the wild type ABL core, and treated with compounds 2 h post-transfection. Cells were lysed by sonication 24 h later, and cell lysates clarified by centrifugation. ABL was immunoprecipitated via the C-terminal His-tag, resolved by SDS-PAGE, and transferred to nitrocellulose membranes for immunoblot analysis with antibodies to the ABL protein and the three autophosphorylation sites: pTyr412 (activation loop), pTyr245 (SH2-kinase linker), and pTyr89 (SH3 domain).

## Results and Discussion

### ABL fluorescence polarization (FP) assay design

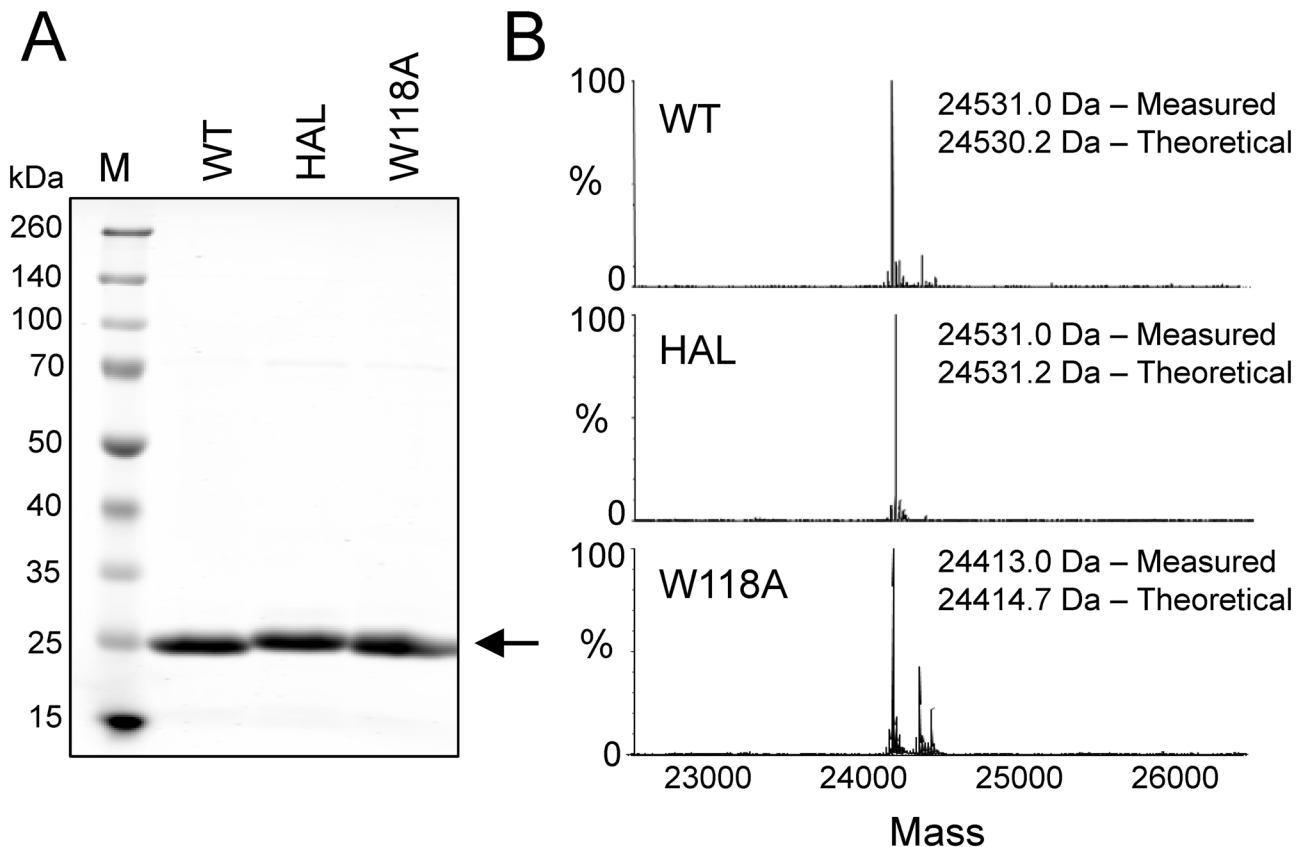
In this study, we developed a screening assay for small molecule allosteric modulators of ABL kinase function. Our goal was to enable discovery of chemical scaffolds that interact with the regulatory region of the ABL kinase core, as opposed to the kinase domain, thereby providing a path to enhanced selectivity and allosteric control of kinase function. In addition, we wanted a flexible assay with the potential to identify both inhibitors and activators of ABL function. To accomplish these goals, we developed a fluorescence polarization (FP) assay based on the N-terminal region of ABL consisting of the Ncap, SH3 and SH2 domains, and the SH2-kinase linker (ABL N32L protein). Binding of a fluorescently labeled probe peptide to the SH3 domain (displacing the linker) should result in an increased FP signal due to the slowed rotation of the N32L target protein-peptide complex (Fig 1C). A small molecule that binds to the ABL N32L protein and enhances SH3 interaction with the linker *in cis* is predicted to prevent probe peptide binding, resulting in a decrease in the FP signal. Molecules in this class are predicted to act as allosteric inhibitors of ABL kinase activity, because they may enhance the natural negative regulatory interaction between the SH3 domain and the linker. Alternatively, compounds that interact with the SH3 domain and block probe peptide binding are also predicted to cause a decrease in the FP signal. By displacing SH3:linker interaction in the context of downregulated ABL, compounds of this type may act as allosteric activators of kinase activity. This assay design therefore has the potential to identify both types of ABL-binding compounds in a single chemical library screen. Their impact on ABL function can be easily distinguished in secondary assays for direct binding to the ABL domains, as well as functional assays.

### Recombinant ABL regulatory proteins for FP assay development

The target protein for the ABL FP assay consists of the first 255 residues of ABL (isoform 1b), and encompasses the Ncap, the SH3 and SH2 domains, as well as the SH2-kinase linker as described above. This ABL N32L protein was expressed in bacteria in soluble form, purified to homogeneity, and its purity and identity were confirmed by SDS-polyacrylamide gel electrophoresis and mass spectrometry, respectively (Fig 2). Previous studies have established that regulatory SH3:linker interaction is maintained in this construct, despite the absence of the kinase domain [8,33]. In addition to the wild type protein, two mutant forms of N32L were produced for use as controls. The first of these has an alanine substitution for a conserved tryptophan on the SH3 domain binding surface (W118A mutant; see Fig 3 for SH3 domain structure), which renders it unable to bind to the probe peptide and thus serves as a negative control. In the second mutant, five linker residues were replaced with prolines to enhance interaction with the SH3 domain [8]. This high-affinity linker (HAL) substitution suppresses the activating effects of kinase domain mutations and influences the conformation of the kinase domain, enhancing both imatinib and allosteric inhibitor action (see Introduction). The HAL protein therefore represents a second negative control for probe peptide binding to the SH3 domain. Both the W118A and HAL forms of the ABL N32L protein were also expressed and purified from bacteria, and yielded soluble purified proteins of the expected mass (Fig 2).

### Structural basis for high affinity probe peptide binding to the ABL SH3 domain

A suitable probe for the ABL N32L FP assay required a short, proline-rich peptide with sequence specificity for the ABL SH3 domain. In addition, the probe peptide needed to bind to the SH3 domain with sufficient affinity to compete for *cis*-interaction of the SH3 domain with



**Fig 2. Recombinant ABL Ncap-SH3-SH2-linker (N32L) proteins.** Wild type ABL N32L protein and the corresponding high-affinity linker (HAL) and W118A mutants were expressed in *E. coli* using the pET system [40] and purified by immobilized metal affinity chromatography. Protein purity and mass were verified by SDS polyacrylamide gel electrophoresis (A) and mass spectrometry (B).

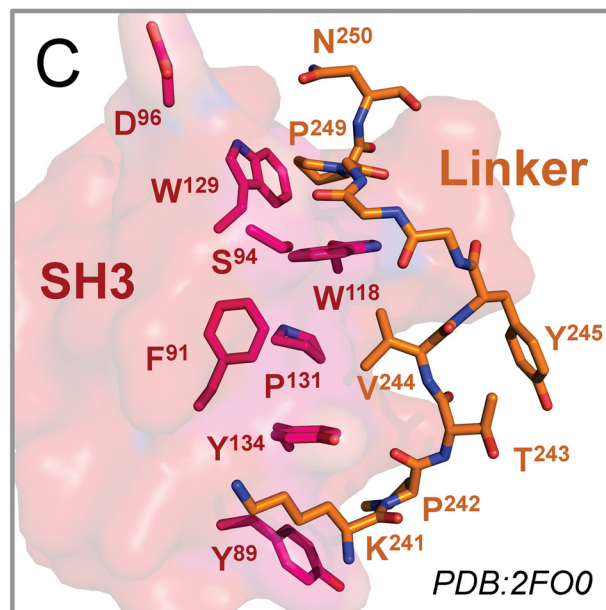
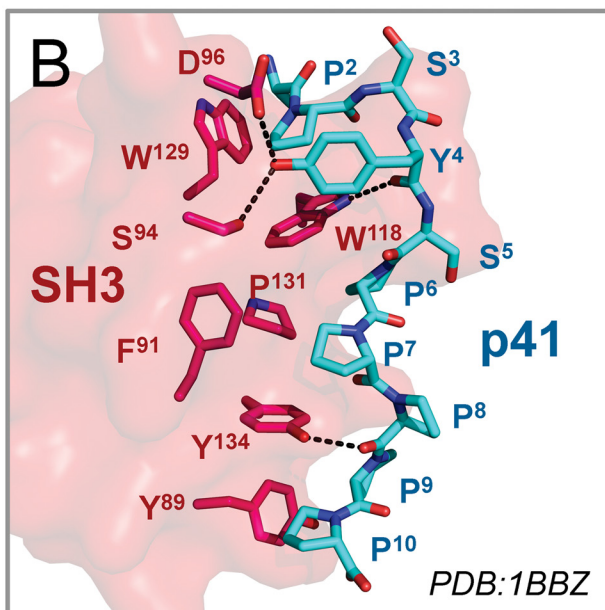
doi:10.1371/journal.pone.0133590.g002

the natural linker (Fig 1). A survey of the literature identified four ABL SH3-binding peptides with the potential to serve as probes [20,21]. These peptides, designated p41, p40, p8, and 3BP-1, have  $K_D$  values for the ABL SH3 domain in the 0.4 to 34  $\mu$ M range. The ABL SH3-binding peptide sequences are presented in Fig 3A, and are aligned with those of the wild type and high-affinity SH2-kinase linkers of ABL.

To explore the potential of known ABL SH3 peptide ligands to compete for natural SH3:linker interaction, we first compared the structure of the ABL SH3:linker interface from the downregulated ABL core (PDB: 2FO0) [11] with the crystal structure of the p41 peptide in complex with the ABL SH3 domain (PDB: 1BBZ) [21]. The C-terminal half of the p41 peptide is comprised exclusively of proline, which facilitates both PPII helix formation as well as tight interaction with the hydrophobic SH3 binding surface (Fig 3B). In contrast, this region of the SH2-kinase linker is comprised of the less favorable SH3-binding sequence, KPTVY (Fig 3C). Specifically, p41 proline residues 9 and 10 fill the hydrophobic groove formed by the aromatic side chains of SH3 tyrosines 89 and 134; the linker is substituted with lysine in this position (Lys241). The main chain carbonyl of p41 Pro8 forms a stabilizing hydrogen bond with Tyr134. This position is substituted with threonine (Thr243) in the linker, which swings away from the SH3 surface. The N-terminal sequence of the p41 peptide forms a network of polar contacts involving SH3 residues Ser94, Asp96, and Trp118. None of these contacts are present in the SH3:linker interface, and the side chain of SH3 Asp96 is rotated away from the linker.

A

Peptide	Orientation	Sequence	SH3 $K_D$
p41	(C-N)	P <sup>10</sup> PPPPSYSP <sup>2</sup> A	1.5 $\mu$ M
p40	(C-N)	P <sup>10</sup> PPPSYTP <sup>2</sup> A	0.4 $\mu$ M
p8	(C-N)	P <sup>10</sup> PAPPYTP <sup>2</sup> A	5.0 $\mu$ M
3BP-1	(C-N)	P <sup>10</sup> PLPPMTP <sup>2</sup> A	34.0 $\mu$ M
ABL WT linker	(N-C)	-RNK <sup>241</sup> PTVYGVSP <sup>249</sup> NYDKWE-	
ABL HAL	(N-C)	-RNP PPPYPPSP NYDKWE-	



**Fig 3. Peptide and linker interactions with the ABL SH3 domain.** A) Sequences of the ABL SH3 binding peptides, p41, p40, p8, and 3BP-1, and their published binding affinities for the ABL SH3 domain [20,21]. Sequences of the wild type (WT) and high-affinity (HAL) SH2-kinase linker sequences are also shown at the bottom. The peptide sequences are presented in the C- to N-terminal orientation to align with those of the linkers. B) Crystal structure of the p41 peptide (cyan) bound to the ABL SH3 domain (PDB: 1BBZ) [21]. The SH3 surface is shown as a space filling model (red) and side chains of residues that interact with the p41 peptide are shown as sticks. C) Crystal structure of the SH2-kinase linker (orange) bound to the ABL SH3 domain (red) from the ABL core (PDB: 2FO0) [11]. Side chains of SH3 domain residues that interact with the p41 peptide as per panel B are shown as sticks. Note the lack of hydrophobic interactions and hydrogen bonds between the SH3 domain and the linker in comparison to the p41 peptide.

doi:10.1371/journal.pone.0133590.g003

Taken together, these structural features strongly suggested that p41, or one of the closely related peptides (p40, p8, and 3BP-1), may interact with the ABL N32L target protein with sufficient affinity to displace the wild type linker and provide a stable FP signal.

### Selection of a probe peptide for the ABL N32L FP assay

To evaluate the suitability of the four ABL SH3 peptide ligands (p41, p40, p8, 3BP-1; Fig 3A) as FP probes, each peptide was synthesized and labeled with 6-carboxyfluorescein on its N-terminus. We first examined the baseline FP signal as well as the fluorescence intensity exhibited by each labeled peptide over a broad concentration range (1–1,000 nM) in the absence of the ABL



N32L target protein. As shown in [Fig 4A](#), probe peptide concentrations greater than 50 nM exhibited stable baseline FP readings with minimal well-to-well variation.

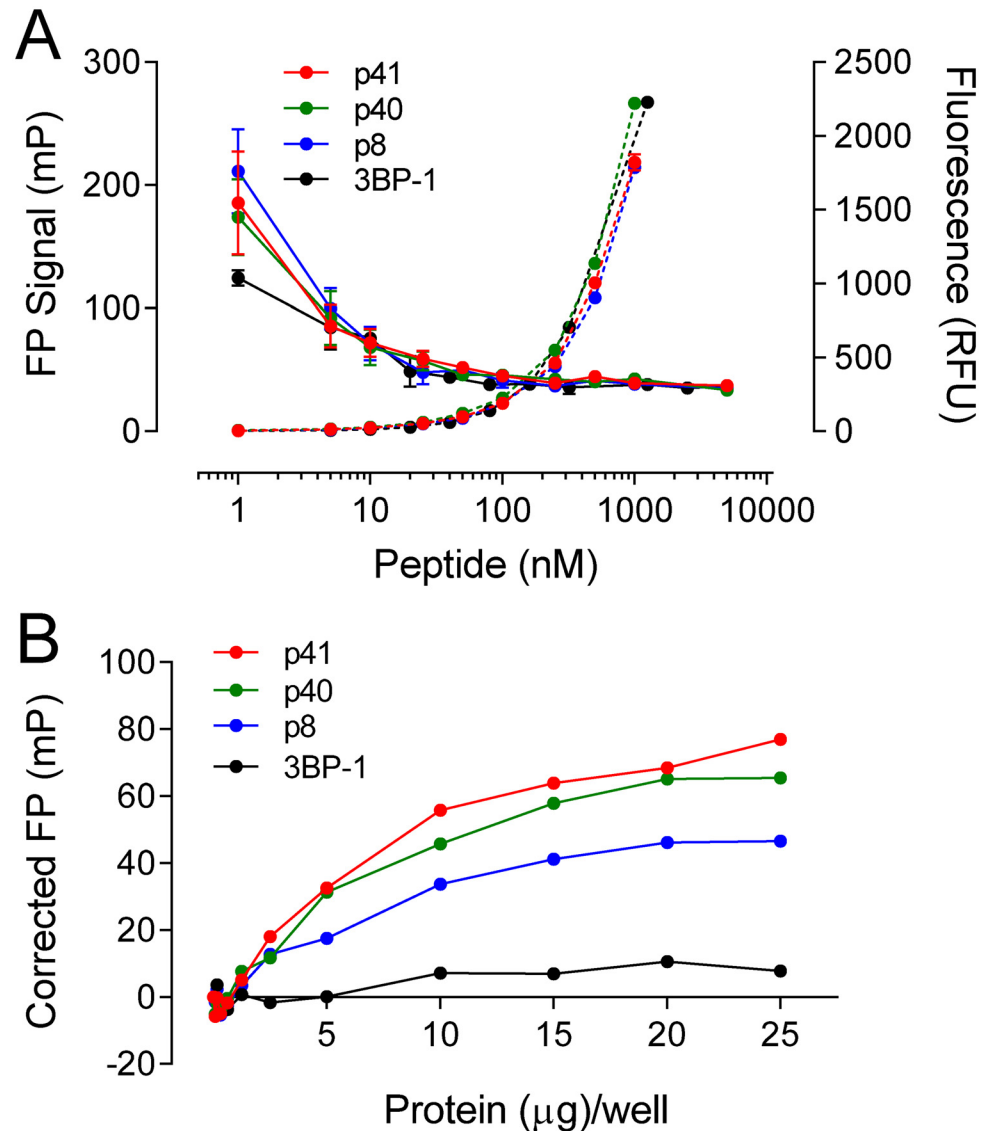
To test for ABL N32L protein interaction with each peptide in the FP assay, we held each probe peptide concentration at 50 nM and added the wild type N32L protein over a range of concentrations. As shown in [Fig 4B](#), both the p40 and p41 probe peptides produced a strong, saturable FP signal as a function of the N32L protein concentration. The p8 peptide also produced an FP response, albeit somewhat lower than that observed with p40 and p41, while the 3BP-1 peptide was inactive. These FP results correspond to the rank order of binding affinities previously reported for these peptides with the isolated SH3 domain [20,21]. Since the structure of the ABL SH3 domain in complex with p41 is known ([Fig 3B](#)), we chose the p41 peptide for FP assay optimization.

### ABL N32L FP assay development and optimization

We next investigated whether the FP signal obtained with the p41 probe peptide was due to interaction with SH3 domain of the recombinant ABL N32L target protein. For these experiments, we compared the FP signal produced from the wild type ABL N32L protein with the SH3 domain mutant (W118A) as well as the high-affinity linker (HAL) protein. As shown in [Fig 5A](#), the wild type ABL N32L protein produced a concentration-dependent increase in the FP signal as observed previously. In contrast, the N32L W118A mutant failed to produce an FP signal with the p41 peptide over the same concentration range, indicating that the peptide requires this conserved SH3 domain tryptophan residue for binding as predicted from the crystal structure (see [Fig 3](#)). On the other hand, the ABL N32L HAL protein showed a greatly reduced FP signal in comparison to the wild type protein with the p41 probe. This result is consistent with enhanced *cis*-interaction of the linker with SH3 domain in this protein as a result of the higher linker proline content (see [Fig 3A](#) for HAL sequence). Results with these control proteins demonstrate that the p41 probe peptide interacts with the ABL N32L target protein through its SH3 domain. FP experiments with the recombinant purified ABL SH3 domain alone also produced a very similar FP response, supporting this conclusion (data not shown). Findings with these ABL N32L mutants support the idea that small molecules that disrupt or stabilize intramolecular interaction between the SH3 domain and linker will also reduce probe peptide binding and loss of the FP signal.

We next tested the stability of the FP signal as a function of time ([Fig 5B](#)). For this experiment, the p41 probe peptide (50 nM) and ABL N32L protein (12.8  $\mu\text{g}/\text{well}$ ) concentrations were held constant. Under these conditions, no significant variation in the FP signal was observed up to 10 hours. We also found that DMSO, the carrier solvent for the screening library compounds, did not influence the FP signal or the negative controls even at the highest concentration tested (1%; [Fig 5C](#)).

In a final validation experiment, we tested the effect of unlabeled p41 peptide on the FP signal ([Fig 5D](#)). For this study, we fixed the p41 probe peptide concentration at 50 nM and the wild type ABL N32L protein concentration at 20  $\mu\text{g}/\text{well}$ . Unlabeled p41 peptide was added to the assay over the concentration range of 2.5 to 100  $\mu\text{M}$ . The FP signal decreased as a function of unlabeled p41 peptide concentration, demonstrating competition for the labeled probe peptide binding to the N32L protein. As a negative control, the peptide competition experiment was repeated with a non-specific peptide of similar length. This peptide had no effect on the FP signal, even at a concentration of 100  $\mu\text{M}$ , demonstrating the specificity of p41 peptide recognition by the SH3 domain in the N32L target protein.

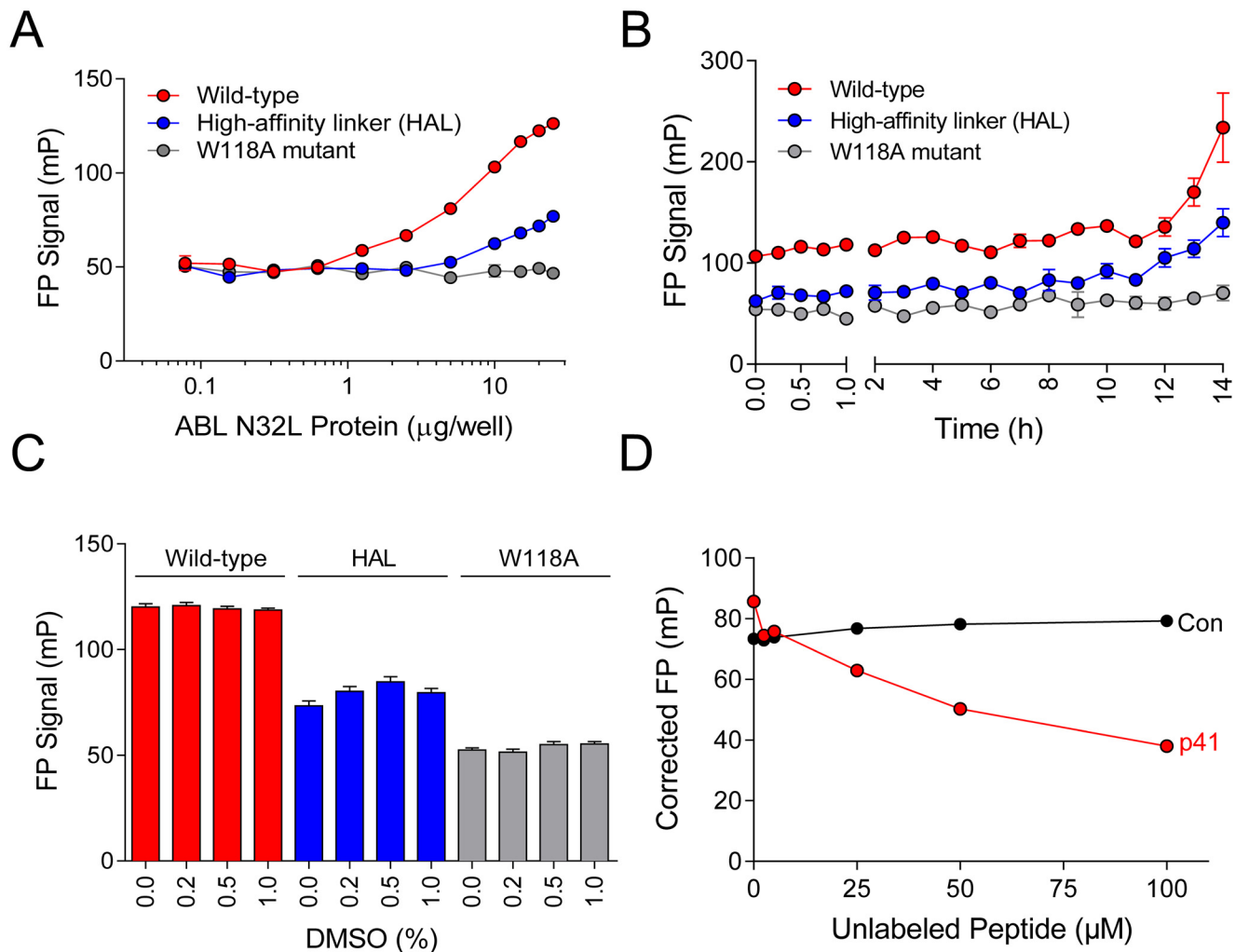


**Fig 4. Identification of p41 as optimal probe peptide for the ABL N32L FP assay.** A) To characterize the baseline FP signal, the 6-carboxy-fluorescein labeled probe peptides p41 (red), p40 (green), p8 (blue), and 3BP-1 (black) were serially diluted in the concentration range of 1–1000 nM. The FP signals (solid lines, left Y axis) and corresponding fluorescence intensities (dashed lines, right Y axis) were measured and plotted as a function of peptide concentration. Average values are shown  $\pm$  SE from four measurements per condition. B) To test for probe peptide interaction with ABL N32L by FP, each peptide (50 nM) was incubated with the ABL N32L protein over the range of 0.08–25  $\mu$ g/well. The resulting FP signals were corrected for baseline FP signal recorded in the absence of the N32L protein and plotted against the N32L protein concentration. Average FP values are shown  $\pm$  SE from four measurements per condition; error bars are smaller than the diameter of some data points.

doi:10.1371/journal.pone.0133590.g004

### Identification of inhibitors of p41 interaction with ABL N32L

To test the performance of the ABL N32L FP assay under screening conditions, we performed a pilot screen of 1200 FDA-approved compounds. The wild type ABL N32L protein (25  $\mu$ g) was added to each well together with the p41 probe peptide (50 nM). The compounds were then added to a final concentration of 10  $\mu$ M in 1% DMSO. Each plate contained twenty-eight wells with the wild type N32L target protein plus DMSO as positive controls, and twenty-eight



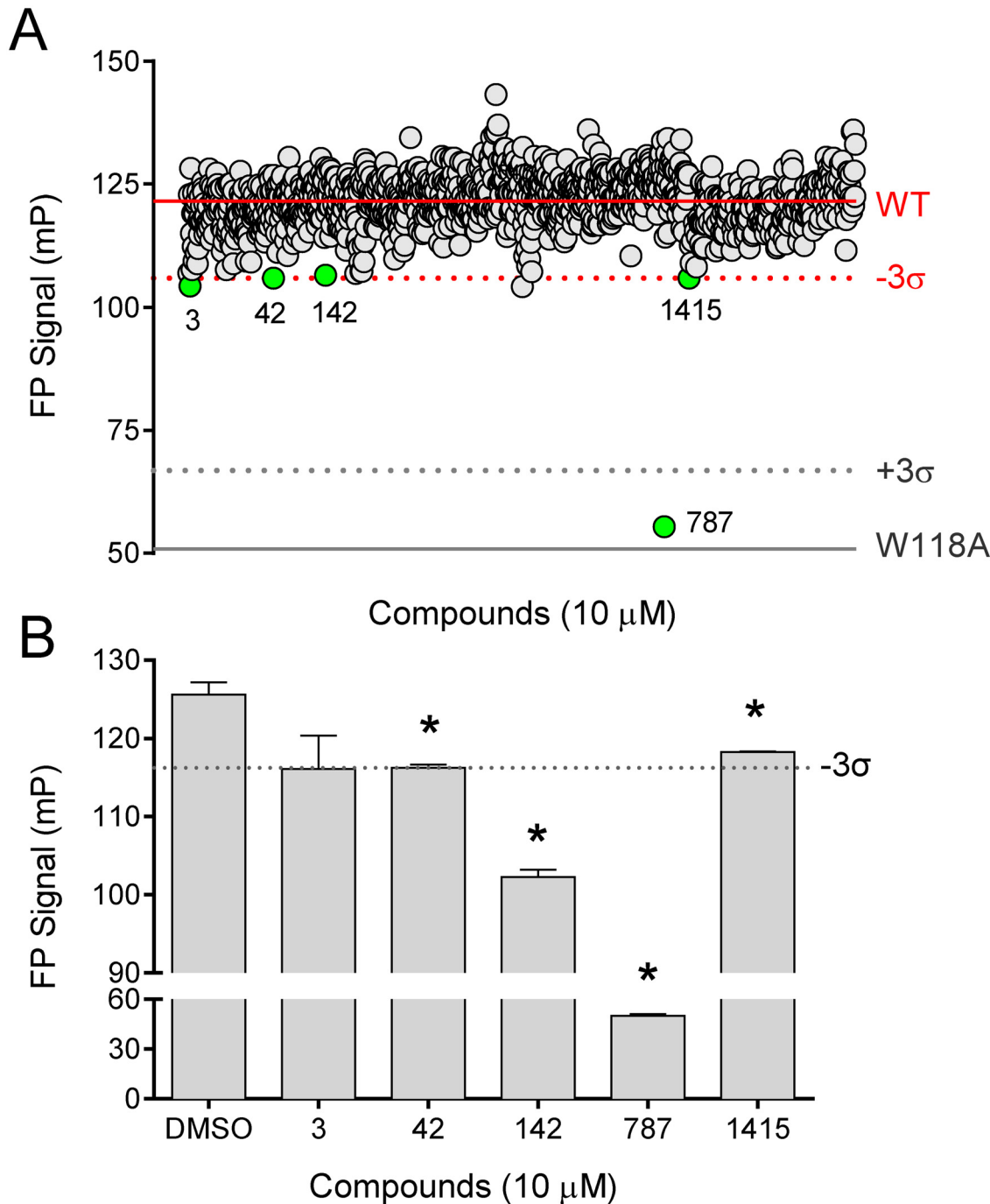
**Fig 5. ABL N32L FP assay development and optimization.** A) The p41 FP probe binds the ABL N32L protein through the SH3 domain. The p41 probe peptide (50 nM) was combined with wild type, HAL, and W118A ABL N32L proteins over the range of concentrations shown. The resulting FP signals were measured and plotted as a function of N32L protein concentration. B) FP assay stability. The p41 probe peptide (50 nM) was combined with the three ABL N32L proteins (12.8 μg/well) and FP signals were recorded over the time course shown. C) DMSO tolerance. FP assays consisting of the p41 probe peptide (50 nM) and each ABL N32L protein (25 μg/well) were incubated with the DMSO concentrations shown, and FP signals were recorded 1 h later. D) Unlabeled peptide competition. For the competition assay, the p41 probe peptide (50 nM) was mixed with unlabeled p41 peptide or a negative control peptide of unrelated sequence (QKEGERALPSIP) and similar length (Con) over the range of concentrations shown. The ABL N32L protein (20 μg/well) was then added, and FP signals were recorded. FP signals were corrected for the background p41 peptide FP signal and plotted as a function of the unlabeled peptide concentration. In all experiments (A through D), average FP values are shown ± SE from four measurements per condition.

doi:10.1371/journal.pone.0133590.g005

wells with the non-binding W118A mutant protein plus DMSO as negative controls. The overall Z factor for the pilot screen was 0.57, indicative of a reliable screening assay [34]. The average FP signals observed with the controls as well as the readings observed with each of the test compounds are presented in Fig 6A.

The FP signals for each compound were ranked by three different methods: increasing raw FP signal, decreasing normalized percent inhibition, and increasing Z score (see Materials and Methods). We then compared the top 1% of compounds present in each of these three rankings. Five compounds were present in at least two of these rankings, and were selected for follow-up assays (compound numbers 3, 42, 142, 787, and 1415; Fig 6A). Each of the raw hit compounds was then retested in multiple wells under screening assay conditions (Fig 6B). Four





**Fig 6. Pilot screen identifies inhibitors of p41 interaction with the ABL N32L protein.** A) A library of 1200 FDA-approved compounds was screened using the ABL N32L protein (25  $\mu$ g/well) and the p41 probe peptide (50 nM) in the FP assay. The solid lines correspond to the mean FP signals for the wild type (WT) and SH3 mutant (W118A) control N32L proteins across all assay plates, with the dotted lines indicating three standard deviations from the means ( $\pm 3\sigma$ ). Compounds were screened at 10  $\mu$ M and each FP signal is represented as an individual circle. Five putative hit compounds were identified (green circles). B) The five potential hit compounds were re-tested in quadruplicate at 10  $\mu$ M vs. the DMSO control under FP screening assay conditions, and the mean FP values are shown  $\pm$  SE. Four of these compounds significantly inhibited the FP signal relative to the DMSO control as indicated by the asterisk ( $p < 0.05$ ; 2-tailed t-test). The dotted line shows the FP value three standard deviations below the DMSO control FP signal ( $-3\sigma$ ).

doi:10.1371/journal.pone.0133590.g006

out of the five compounds produced a significant inhibition of the FP signal relative to the DMSO controls (compounds 42, 142, 787, and 1415). We then performed control FP experiments with each of these compounds under the same conditions but in the absence of the N32L target protein. This counter-screen showed that compound 787 also inhibited the baseline FP signal produced by the p41 probe peptide, indicating non-specific quenching of the FP signal (data not shown). None of the other compounds affected baseline p41 probe peptide fluorescence, and were therefore moved forward into secondary assays.

### Compounds identified in the ABL N32L FP screen interact directly with the ABL N32L protein in orthogonal assays

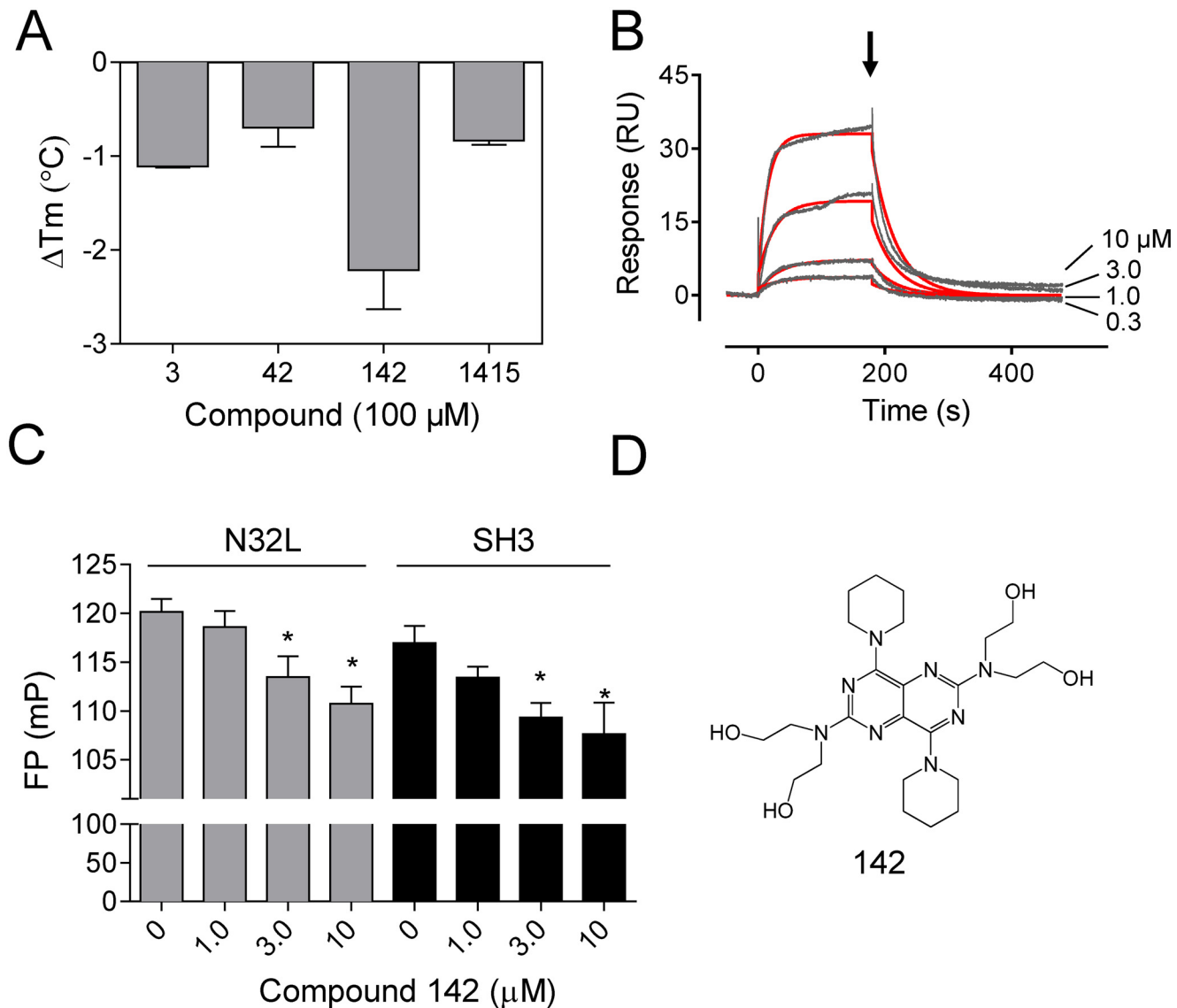
As an independent measure of hit compound interaction with the ABL N32L target protein, we performed differential scanning fluorimetry assays [35]. For these experiments, the ABL N32L protein was heated with a molar excess of each compound in the presence of the reporter dye, SYPRO orange. As the temperature rises and the N32L protein unfolds, the reporter dye accesses the hydrophobic interior of the protein, resulting in an increase in dye fluorescence. The resulting protein 'melt curve' is then fit by regression analysis to obtain a  $T_m$  value, the temperature at which half-maximal thermal denaturation is observed. Small molecule binding to a target protein can either increase or decrease the  $T_m$  value, depending upon the effect of the compound on protein stability. Differential scanning fluorimetry was performed with the wild type N32L protein in the presence of each of the hit compounds from the FP assay, and the change in  $T_m$  value ( $\Delta T_m$ ) was determined compared to DMSO as the reference control. As shown in Fig 7A, all four compounds produced a significant decrease in the  $T_m$  value. Compound 142 had the largest impact on N32L thermal stability, producing a decrease of more than 2°C in the  $T_m$ , consistent with its effect in the FP assay (Fig 6B).

To confirm direct interaction of compound 142 with the ABL N32L protein and explore the binding kinetics, we next performed surface plasmon resonance assays. For these experiments, the ABL N32L protein was immobilized on the biosensor surface while compound 142 was flowed past the immobilized protein over a range of concentrations. As shown in Fig 7B, concentration-dependent interaction of compound 142 with ABL N32L was readily detected by this approach, yielding an association rate constant of  $3.87 \pm 0.59 \times 10^3 \text{ M}^{-1}\text{s}^{-1}$  and a dissociation rate constant of  $2.70 \pm 0.72 \times 10^{-2} \text{ s}^{-1}$ . The equilibrium dissociation constant ( $K_D$ ) for this interaction, calculated as the ratio of  $k_d/k_a$ , is  $6.90 \pm 0.78 \times 10^{-6} \text{ M}$ .

Compounds that inhibit the FP signal in the ABL N32L assay may either interfere directly with probe peptide binding to the SH3 domain or allosterically tighten the cis-interaction of the SH3 with the linker, indirectly reducing probe peptide interaction. To distinguish between these two possibilities with compound 142, we performed FP assays with the N32L protein as well as the isolated ABL SH3 domain. As shown in Fig 7C, Compound 142 resulted in a concentration-dependent decrease in the FP signal with both the ABL N32L and SH3 proteins, suggesting that this compound binds directly to the ABL SH3 domain.

### Allosteric activation of ABL kinase by compound 142

Compound 142 reproducibly scored as a hit in the ABL N32L FP assay and demonstrated direct interaction with ABL N32L protein by both differential scanning fluorimetry and SPR. This compound, a symmetrically substituted pyrimido-pyrimidine known as dipyridamole (Fig 7D), is a selective inhibitor of phosphodiesterase V and also an adenosine transport inhibitor used clinically for its antithrombotic activity [36]. However its potential impact on protein kinase function has not been reported. Because compound 142 interacts with the regulatory region of ABL, we investigated its effects on ABL kinase activity using two purified



**Fig 7. Hit compound 142 interacts directly with the ABL N32L protein.** A) Differential scanning fluorimetry assays were performed on the ABL N32L protein in the presence of the four confirmed hit compounds as described under Materials and Methods. The average change in the mid-point of the thermal melt profile ( $\Delta T_m$ ) relative to the  $T_m$  obtained with the N32L protein in the presence of the DMSO carrier solvent is plotted on the Y-axis  $\pm$  SE ( $n = 2$ ). B) Surface plasmon resonance was performed with the ABL N32L protein immobilized on the biosensor chip and compound 142 as analyte. Responses were recorded for the four compound concentrations shown, and the flow path was switched back to buffer after 180 s to induce dissociation (arrow). The resulting sensorgrams (black lines) were fit by a 1:1 Langmuir binding model (red lines) to generate kinetic constants. C) Compound 142 inhibits p41 peptide binding to the ABL N32L and SH3 proteins in the FP assay. Compound 142 was added to N32L and SH3 FP assays over the range of concentrations shown, and the resulting FP signals are presented as the mean  $\pm$  SE. Significant inhibition for both N32L and SH3 was observed at 3 and 10  $\mu$ M ( $*p < 0.05$  by 2-tailed t-test). D) Chemical structure of compound 142 (dipyridamole).

doi:10.1371/journal.pone.0133590.g007

recombinant forms of ABL in a kinetic kinase assay. These included the wild type ABL core region, consisting of the Ncap, the SH3 and SH2 domains, the SH2-kinase linker, and the kinase domain. This ABL protein was produced in Sf9 insect cells, which results in myristoylation of the N-cap and interaction with the C-lobe of the kinase domain, thereby assembling the downregulated state (Fig 1A). In addition to the wild type core, we also tested a high-affinity linker (HAL) mutant version of the ABL protein, which has a modified proline-rich linker that packs more tightly against the SH3 domain [8].

Baseline kinase activities and kinetic parameters of each recombinant ABL kinase core protein were determined first using a fluorimetric assay that measures the kinase reaction rate as the generation of ADP [26]. For wild type ABL, we obtained  $K_m$  values of  $9.78 \pm 0.14 \mu\text{M}$  and  $144.65 \pm 1.64 \mu\text{M}$  for ATP and peptide substrate, respectively. The ABL HAL core yielded a similar  $K_m$  value for substrate ( $150.25 \pm 5.35 \mu\text{M}$ ), with a higher value for ATP ( $21.24 \pm 1.6 \mu\text{M}$ ). In subsequent experiments, the ATP and peptide substrate concentrations were set to their respective  $K_m$  values, and input kinase concentrations were adjusted to yield the same basal reaction rates (7 pmol ADP produced per minute).

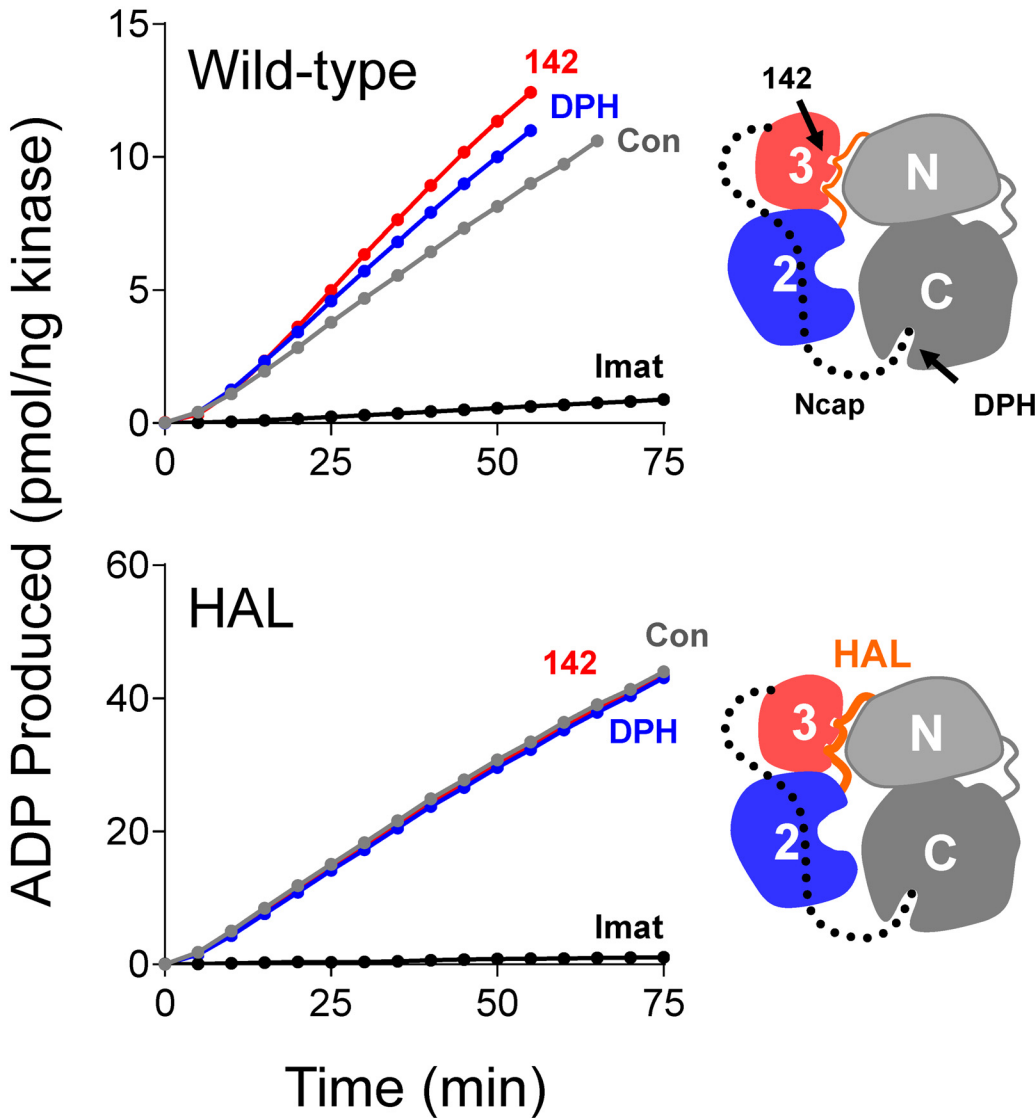
We first examined the effect of compound 142 on the activity of the wild type ABL kinase core protein. As shown in Fig 8, compound 142 stimulated wild type ABL kinase activity by about 40% at a concentration of 10  $\mu\text{M}$  relative to the DMSO control in this assay. As a positive control, we also assayed ABL core activity in the presence of the same concentration of a previously described ABL activator, DPH, and observed a similar degree of activation. DPH, unlike compound 142, stimulates ABL through the kinase domain via the myristic acid binding pocket in the C-lobe [27].

The mechanism of ABL activation by compound 142 may involve binding to the SH3 domain and subsequent displacement of its regulatory interaction with the SH2-kinase linker. Indeed, mutations that disrupt SH3:linker interaction also have a stimulatory effect on ABL kinase activity (see Introduction). To test this idea, we next examined the effect of this compound on the ABL core mutant with enhanced SH3:linker interaction. Unlike wild type ABL, compound 142 did not affect the kinase activity of the ABL core with the HAL substitution (Fig 8), consistent with the idea that enhanced SH3:linker interaction prevents compound 142 access to the ABL SH3 domain. Interestingly, DPH did not activate the ABL HAL core protein either, consistent with previous results showing that enhanced SH3:linker interaction overcomes ABL core activation by mutations in the myristic-acid binding pocket [8].

To further characterize ABL activation by compound 142, we repeated kinetic kinase assays with the wild type ABL core over a range of compound concentrations. As shown in Fig 9, compound 142 activates the ABL core in a concentration-dependent manner, with an  $EC_{50}$  value of  $0.63 \pm 0.07 \mu\text{M}$ . This value compares favorably to that obtained with DPH, the myristic acid binding pocket agonist ( $EC_{50} = 1.11 \pm 0.5 \mu\text{M}$ ). We also calculated the activation constant ( $K_{act}$ ) for each compound from these data, which is defined as the concentration at which the reaction rate reaches half-maximum velocity ( $V_{act}$ ). For compound 142, the  $K_{act}$  was calculated as  $0.4 \pm 0.02 \mu\text{M}$ , while DPH yielded a value of  $1.02 \pm 0.07 \mu\text{M}$ . This analysis also showed that the extent of ABL activation by compound 142 was higher than that observed for DPH, based on the  $V_{max}$ .

## Molecular dynamics simulations and docking studies predict binding of compound 142 to the SH3:linker interface in the ABL kinase core

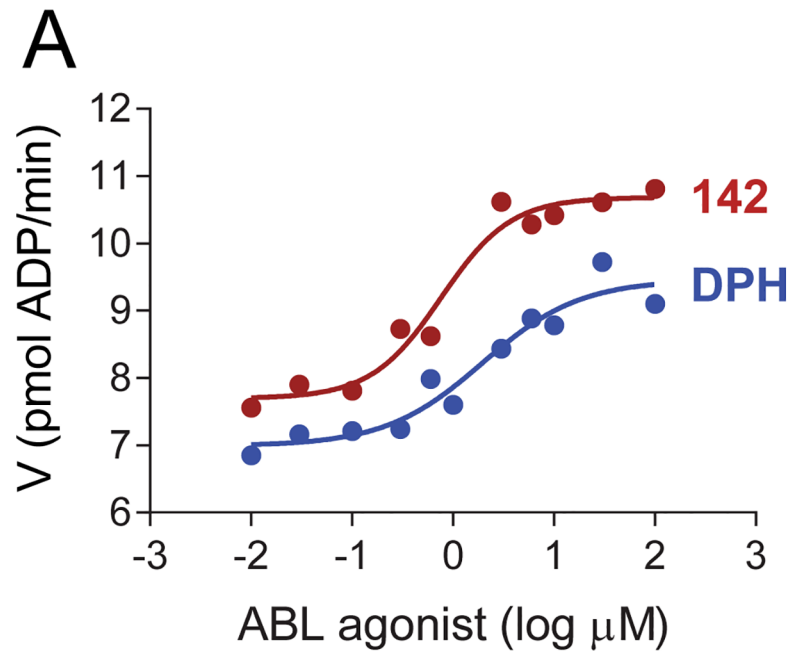
Data presented in the previous sections demonstrate that compound 142 interacts with the regulatory N32L region of ABL, resulting in a decrease in thermal stability and a concomitant increase in kinase activity. We used molecular dynamics (MD) simulations to explore the dynamics of the N32L region used in the assays. To model the effect of the linker being displaced from the SH3 domain, we manually pulled the linker a short distance away from the SH3 domain prior to the simulation (see Materials and Methods). After approximately 20 ns, the linker reconnected with the SH3 domain through the interaction of linker Pro249 and SH3 Trp118. To explore possible binding sites for this compound on the N32L region of the ABL kinase core, we used the computational docking tool *smina* [32] to dock 142 to snapshots of the simulation prior to the reconnection of the SH3:linker interface. As shown in Fig 10 (top



**Fig 8. Compound 142 activates the ABL kinase core in vitro.** *Top:* The recombinant ABL core protein, consisting of the Ncap, SH3, SH2 and kinase domains, was assayed in the presence of compound 142 (10  $\mu$ M), the known ABL activator DPH (10  $\mu$ M), and imatinib (1  $\mu$ M) or with DMSO as control (Con) using a kinetic kinase assay (see [Materials and Methods](#)). Data are plotted as pmol ADP produced per ng kinase as a function of time. The cartoon (right) depicts the domain organization of the wild type ABL core, and indicates the binding site for DPH (myristic acid binding pocket) as well as the predicted binding site for compound 142 (SH3 domain). *Bottom:* Kinase assays were performed using an ABL core protein with a high-affinity linker (HAL) in the presence of the same three compounds; the cartoon indicates the position of the modified linker (HAL). In both cases, the ATP and peptide substrate concentrations were set to their respective  $K_m$  values (wild type ABL:  $9.78 \pm 0.14 \mu$ M for ATP and  $144.65 \pm 1.64 \mu$ M for substrate; ABL HAL:  $21.24 \pm 1.6 \mu$ M for ATP and  $150.25 \pm 5.35 \mu$ M for substrate).

doi:10.1371/journal.pone.0133590.g008

panel), compound 142 fits into a surface pocket defined by the SH3:linker interface in the N32L protein. This predicted binding site involves an aromatic interaction between the pyrimido-pyrimidine moiety of 142 and the indole side chain of SH3 Trp118, as well as polar contacts involving all four hydroxyl groups on the ligand. This aromatic interaction is consistent with probe peptide displacement as well as the observed decrease in the FP signal produced by the W118A mutation ([Fig 5C](#)). Two of the hydroxyl groups of 142 make potential hydrogen bonds with the side and main chains of SH3 Asn97 as well as the side chain of Thr98. The other two hydroxyl groups of 142 form hydrogen bonds with the main chain carbonyls of



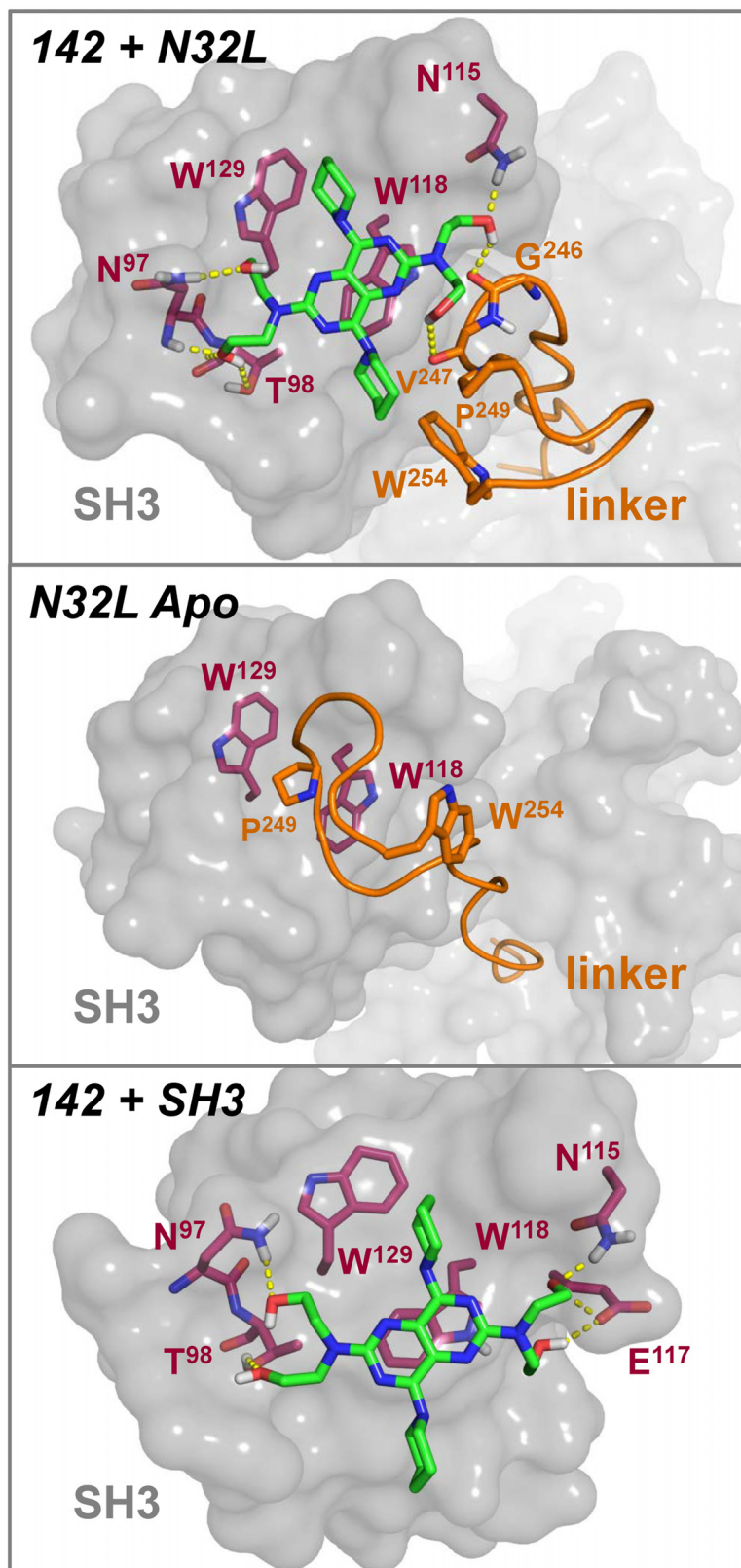
	142	DPH
$EC_{50}$ ( $\mu\text{M}$ )	$0.63 \pm 0.07$	$1.11 \pm 0.5$
$K_{\text{act}}$ ( $\mu\text{M}$ )	$0.4 \pm 0.02$	$1.02 \pm 0.07$
$V_{\text{max}}$ ( $\mu\text{M}$ )	$3.2 \pm 0.28$	$1.90 \pm 0.4$
$V_{\text{max}}/K_{\text{act}}$	8.00	1.86

**Fig 9. Concentration-dependent activation of the ABL kinase core protein by compound 142.** The wild type ABL kinase core was assayed in the presence of compound 142 and DPH at the indicated concentrations using a kinetic kinase assay (see [Materials and Methods](#)). Reaction velocities are plotted as a function of compound concentrations. The resulting data were curve-fit to determine the  $EC_{50}$ ,  $K_{\text{act}}$  and  $V_{\text{max}}$  for each activator as described under Materials and Methods. Each of these parameters was determined in triplicate, and the mean values  $\pm$  SE are presented in the table below the graph. The table also provides the ratio  $V_{\text{max}}/K_{\text{act}}$  as a measure of overall enhancement of catalytic efficiency in the presence of each of these two ABL agonists.

doi:10.1371/journal.pone.0133590.g009

linker Gly246 and Val247 as well as the side chain of SH3 Asn115. In addition, one of the piperidine groups of compound 142 makes hydrophobic contacts with SH3 Trp129, while the other approaches the side chains of linker residues Pro249 and Trp254. Note that in the crystal structure of the fully assembled, downregulated conformation of the ABL core, linker Pro249





**Fig 10. Molecular dynamics (MD) and molecular docking predict binding of compound 142 to the ABL SH3:linker interface.** *Top:* The lowest energy pose of the ligand (compound 142; carbon atoms rendered in

green) is shown docked to a snapshot of an MD simulation of the ABL N32L structure. SH3 domain residues predicted to contribute to ligand binding include Asn97, Thr98, Asn115, Trp118, and Trp129 (carbons in red). The backbone of the linker is shown as an orange ribbon, with Gly246, Val247, Pro249 and Trp254 predicted to contribute to the binding pocket. One of the piperidine groups of compound 142 makes hydrophobic contacts with linker Pro249 and Trp254, while the pyrimido-pyrimidine scaffold of compound 142 is  $\pi$ -stacking with Trp118. *Middle panel*: Model of the SH3:linker interface in the N32L region based on the crystal structure of the downregulated ABL core (PDB:2FO0), highlighting the interaction of linker Pro249 with SH3 Trp118 and Trp129. Ligand binding (top panel) is predicted to displace this regulatory interaction, leading to kinase activation. *Lower panel*: The lowest energy pose of compound 142 is shown docked to a snapshot of an MD simulation of the SH3 domain in the absence of the linker. The position of the 142 ligand is similar (within 1.5 Å RMSD) to that in the SH3 domain of N32L (top), except that the ligand contacts Glu117 rather than linker residues Gly246 and Val247. Without the linker, the potential hydrophobic stabilization of the 142 piperidine group is also lost.

doi:10.1371/journal.pone.0133590.g010

inserts between SH3 Trp118 and Trp129 (Fig 10, middle panel); displacement of this regulatory contact by compound 142 binding may contribute to kinase activation.

For comparison, we also ran unconstrained MD simulations of the SH3 domain in the absence of the linker. Our docked model to a snapshot from this simulation is shown in Fig 10 (bottom panel). The overall position of the ligand in this model is quite similar to that observed with the N32L snapshot (within 1.5 Å RMSD), and includes the potential stacking interaction with SH3 Trp118 and hydrogen bonding to SH3 Asn97, Thr98 and Asn115. Polar contacts of compound 142 with linker Gly246 and Val247 as well as hydrophobic interactions with linker Trp254 and Pro249 are not possible. However, additional hydrogen bonds are observed with SH3 Glu117. Loss of these hydrophobic interactions in the SH3-only model helps to explain the lower binding affinity of compound 142 for the isolated SH3 domain ( $K_D = 3.00 \pm 0.40 \times 10^{-5}$  M) relative to the N32L protein ( $K_D = 6.90 \pm 0.78 \times 10^{-6}$  M) as determined by SPR (data not shown and Fig 7B).

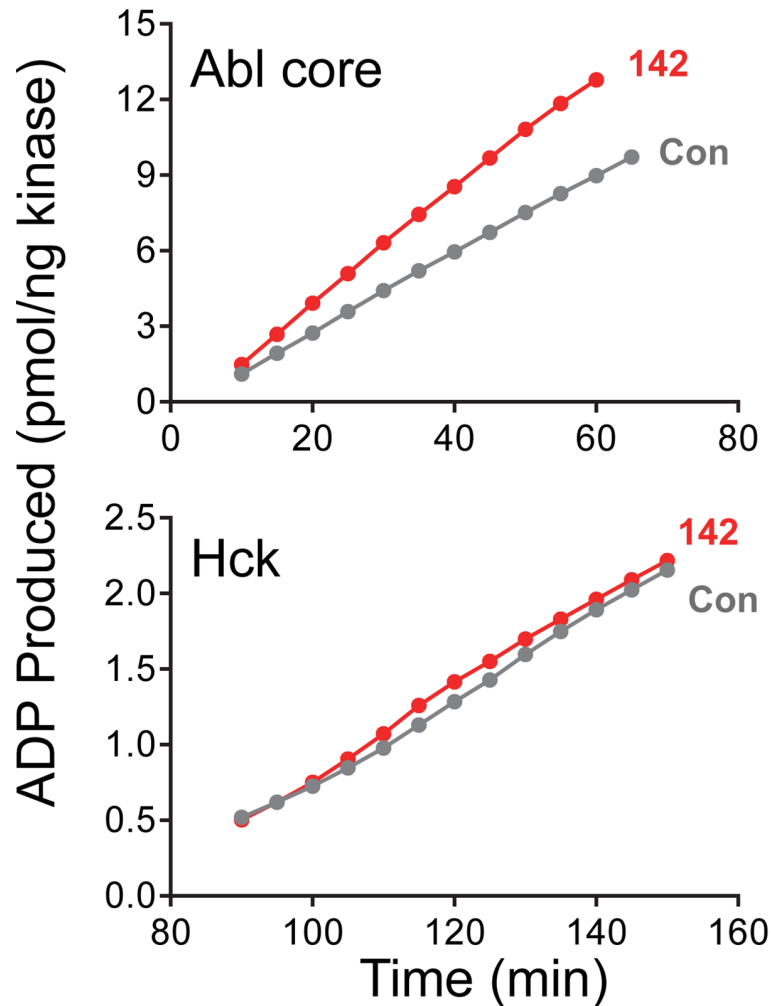
## Compound 142 does not activate the SRC-family kinase, HCK

Like ABL, SRC-family kinases exhibit a very similar arrangement of SH3, SH2, and kinase domains in the downregulated conformation, and are also susceptible to activation by mutations and binding partners that disrupt intramolecular SH3:linker interaction [37,38]. To determine whether compound 142 also activates SRC-family kinases by a similar SH3:linker displacement mechanism, we performed kinetic kinase assays on the SRC-family member HCK in the presence and absence of this ABL activator. As shown in Fig 11, addition of compound 142 to recombinant near-full-length HCK had no effect the reaction velocity, suggesting that it is selective for ABL. Alignment of the SH3 domain and SH2-kinase linker sequences supports this view (Fig 11, bottom). Five of the seven residues predicted by the docking model to participate in compound 142 binding to the ABL SH3:linker surface are either substituted with different amino acids or are missing from the HCK sequence. These observations suggest that even subtle differences in the SH3 and linker sequences of ABL and SRC-family kinases can be exploited for the development of selective agonists or antagonists.

## Compound 142 cooperates with DPH to activate ABL in cells

In a final series of studies, we evaluated the effect of compound 142 on the activity of the ABL core in cells. For these studies, we expressed the wild type ABL core in 293T cells, and then treated the cells over a range of compound 142 concentrations (1–10  $\mu$ M). We then immunoprecipitated ABL and analyzed the level of phosphorylation at three important regulatory tyrosine sites by immunoblotting with phosphospecific antibodies [8,39]: pY412 (activation loop), pY245 (SH2-kinase linker), and pY89 (SH3 domain). As shown in Fig 12, treatment with





**SH3 Domains**

	sm	s	a
Ab1	LFVALYDFVASDGN <b>NTLSITKGEKLRVLGYNHNGEW</b>		
Hck	IVVALYDYEAIIH <b>HE</b> DL <b>S</b> FQKGDQMVVL-- <b>EESGEW</b>		

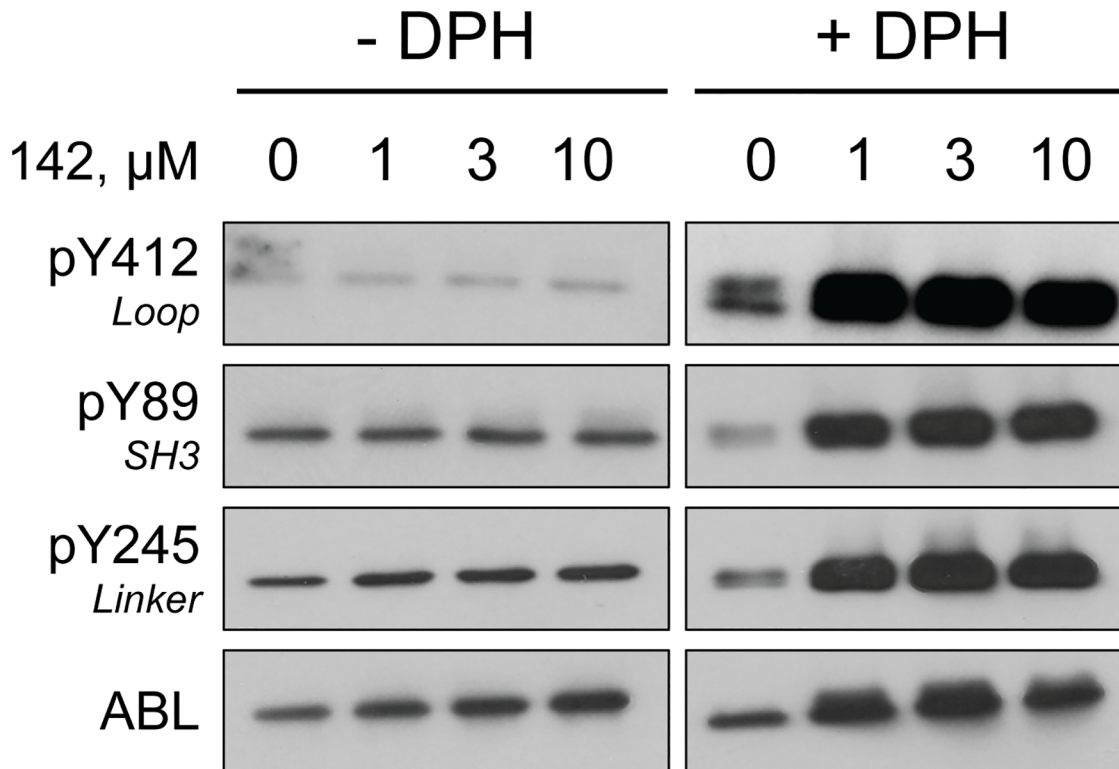
	a
Ab1	CEA <b>Q</b> TKNG-- <b>QGW</b> VPSNYITPVNS
Hck	WKARSLATRKE <b>G</b> YIPSNYVARVDS

**Linkers**

	mm
Ab1	RNKPTVY <b>GV</b> SP-NYDKWE
Hck	SSKPQ--- <b>KPWEKDAWE</b>

**Fig 11. Compound 142 fails to activate downregulated HCK in vitro.** *Top panels:* Recombinant, downregulated wild type ABL core and near full-length HCK proteins were assayed in the presence of compound 142 (1 μM) or with DMSO alone as control (Con) using a kinetic kinase assay (see [Materials and Methods](#)). Data are plotted as pmol ADP produced per ng kinase as a function of time. *Bottom:* Amino acid sequence alignment of the ABL and HCK SH3 domains and SH2-kinase linkers. SH3 and linker residues predicted to contribute to compound 142 binding are highlighted in red (see [Fig 10](#)). The type of interaction is indicated as side chain (s), main chain (m), or aromatic (a).

doi:10.1371/journal.pone.0133590.g011



**Fig 12. Compound 142 cooperates with DPH to stimulate ABL core autophosphorylation in cells.** Human embryonic kidney 293T cells were transfected with an expression vector for the wild type ABL core followed by treatment with compound 142 overnight over the range of concentrations shown. The experiment was performed in the absence or presence of the myristic acid binding pocket agonist DPH at a final concentration of 10  $\mu$ M. ABL proteins were immunoprecipitated from clarified cell lysates, and analyzed by immunoblotting with phosphospecific antibodies to three regulatory sites: pTyr412 in the activation loop, pTyr89 in the SH3 domain, and pTyr245 in the SH2-kinase linker, as well as an ABL antibody to control for protein recovery. This experiment was repeated twice with comparable results.

doi:10.1371/journal.pone.0133590.g012

compound 142 alone did not affect phosphorylation of these sites. We then repeated this experiment in the presence of DPH, the myristic acid binding pocket activator described above, at a concentration of 10  $\mu$ M. In the absence of compound 142, DPH had a very small effect on ABL phosphorylation. However, the combination of compound 142 and DPH led to remarkable enhancement of phosphorylation on all three regulatory tyrosines, providing strong evidence for an activating effect of compound 142 on ABL in cells. These observations suggest that in the cellular environment, unlike *in vitro*, compound 142 alone may be not sufficient to disturb SH3:linker interaction and stimulate ABL activity. However, when the regulatory interaction of the myristoylated Ncap with the kinase domain C-lobe is perturbed by DPH, then the activating effect of SH3:linker displacement becomes readily apparent.

## Summary and Conclusions

In this study, we developed a screening strategy to identify allosteric small molecule modulators of ABL kinase activity that work outside of the kinase domain. Our FP-based assay targets the regulatory domains of ABL that control its kinase activity through intramolecular interactions. Specifically, this assay is based on a recombinant ABL protein comprising the complete regulatory apparatus (Ncap-SH3-SH2-linker) and a synthetic polyproline probe peptide (p41) that selectively binds the ABL SH3 domain. Interaction of the probe peptide with the ABL N32L protein results in a robust and reproducible FP signal. Mutation of the SH3 binding site

(W118A) or introduction of a high-affinity linker both resulted in loss of the FP signal, demonstrating that probe access requires an intact and accessible SH3 domain. A small-scale pilot screen of 1200 FDA-approved compounds identified dipyridamole (compound 142) as an inhibitor of the FP signal observed with the N32L:p41 complex, and direct interaction of this compound with the ABL N32L protein was confirmed by SPR and DSF assays. Dipyridamole was observed to stimulate the kinase activity of downregulated ABL kinase *in vitro*, and was more potent than the previously described ABL agonist DPH which targets the myristic acid binding pocket in the kinase domain. In contrast to wild type ABL, dipyridamole had no effect on a modified ABL core protein with a high-affinity linker, suggesting that it works by binding to the SH3 domain and disrupting the SH3:linker interface. Molecular dynamics simulations in combination with molecular docking support this proposed mechanism of action: dipyridamole was predicted to interact with the ABL core through a pocket defined by the SH3:linker interface. This interaction involved Trp118 on the SH3 domain binding surface, potentially disrupting a key regulatory interaction with Pro249 in the linker. Discovery of dipyridamole as an ABL agonist provides an important proof of concept that small molecules altering SH3:linker interaction represent allosteric modulators of ABL kinase activity. Selective agonists of ABL function have potential as chemical probes to better understand the role of ABL kinase activity in solid tumors and in response to genotoxic stress. Conversely, allosteric antagonists may also be discovered by this approach and have the potential to complement current ATP-competitive inhibitors of BCR-ABL in the context of CML and other cancers. The allosteric inhibitor discovery concept may also be extended to the discovery of allosteric modulators of other kinases systems with multi-domain regulatory interactions, including members of the SRC and TEC kinase families.

## Acknowledgments

The authors wish to thank Roxana Iacob and John Engen of Northeastern University for performing mass spectrometry on the purified recombinant ABL proteins used in this study.

## Author Contributions

Conceived and designed the experiments: PG HS MB CC TS. Performed the experiments: PG HS MB. Analyzed the data: PG HS MB CC TS. Contributed reagents/materials/analysis tools: PG HS MB CC TS. Wrote the paper: PG MB CC TS.

## References

1. Pendergast AM (2002) The Abl family kinases: mechanisms of regulation and signaling. *Adv Cancer Res* 85: 51–100. PMID: [12374288](#)
2. Maiani E, Diederich M, Gonfloni S (2011) DNA damage response: the emerging role of c-Abl as a regulatory switch? *Biochem Pharmacol* 82: 1269–1276. doi: [10.1016/j.bcp.2011.07.001](#) PMID: [21763684](#)
3. Colicelli J (2010) ABL tyrosine kinases: evolution of function, regulation, and specificity. *Sci Signal* 3: re6. doi: [10.1126/scisignal.3139re6](#) PMID: [20841568](#)
4. Wong S, Witte ON (2004) The BCR-ABL story: bench to bedside and back. *Annu Rev Immunol* 22: 247–306. PMID: [15032571](#)
5. Hantschel O, Superti-Furga G (2004) Regulation of the c-Abl and Bcr-Abl tyrosine kinases. *Nat Rev Mol Cell Biol* 5: 33–44. PMID: [14708008](#)
6. Druker BJ (2004) Imatinib as a paradigm of targeted therapies. *Adv Cancer Res* 91: 1–30. PMID: [15327887](#)
7. Azam M, Latek RR, Daley GQ (2003) Mechanisms of autoinhibition and STI-571/imatinib resistance revealed by mutagenesis of BCR-ABL. *Cell* 112: 831–843. PMID: [12654249](#)
8. Panjarian S, Iacob RE, Chen S, Wales TE, Engen JR, Smithgall TE (2013) Enhanced SH3/linker interaction overcomes Abl kinase activation by gatekeeper and myristic acid binding pocket mutations and

- increases sensitivity to small molecule inhibitors. *J Biol Chem* 288: 6116–6129. doi: [10.1074/jbc.M112.431312](https://doi.org/10.1074/jbc.M112.431312) PMID: [23303187](https://pubmed.ncbi.nlm.nih.gov/23303187/)
9. Pluk H, Dorey K, Superti-Furga G (2002) Autoinhibition of c-Abl. *Cell* 108: 247–259. PMID: [11832214](https://pubmed.ncbi.nlm.nih.gov/11832214/)
  10. Hantschel O, Nagar B, Guettler S, Kretzschmar J, Dorey K, Kuriyan J et al. (2003) A myristoyl/phosphotyrosine switch regulates c-Abl. *Cell* 112: 845–857. PMID: [12654250](https://pubmed.ncbi.nlm.nih.gov/12654250/)
  11. Nagar B, Hantschel O, Seeliger M, Davies JM, Weis WI, Superti-Furga G et al. (2006) Organization of the SH3-SH2 unit in active and inactive forms of the c-Abl tyrosine kinase. *Mol Cell* 21: 787–798. PMID: [16543148](https://pubmed.ncbi.nlm.nih.gov/16543148/)
  12. Nagar B, Hantschel O, Young MA, Scheffzek K, Veach D, Bornmann W et al. (2003) Structural basis for the autoinhibition of c-Abl tyrosine kinase. *Cell* 112: 859–871. PMID: [12654251](https://pubmed.ncbi.nlm.nih.gov/12654251/)
  13. Panjarian S, Iacob RE, Chen S, Engen JR, Smithgall TE (2013) Structure and dynamic regulation of Abl kinases. *J Biol Chem* 288: 5443–5450. doi: [10.1074/jbc.R112.438382](https://doi.org/10.1074/jbc.R112.438382) PMID: [23316053](https://pubmed.ncbi.nlm.nih.gov/23316053/)
  14. Adrian FJ, Ding Q, Sim T, Velentza A, Sloan C, Liu Y et al. (2006) Allosteric inhibitors of Bcr-abl-dependent cell proliferation. *Nat Chem Biol* 2: 95–102. PMID: [16415863](https://pubmed.ncbi.nlm.nih.gov/16415863/)
  15. Zhang J, Adrian FJ, Jahnke W, Cowan-Jacob SW, Li AG, Iacob RE et al. (2010) Targeting Bcr-Abl by combining allosteric with ATP-binding-site inhibitors. *Nature* 463: 501–506. doi: [10.1038/nature08675](https://doi.org/10.1038/nature08675) PMID: [20072125](https://pubmed.ncbi.nlm.nih.gov/20072125/)
  16. Barila D, Superti-Furga G (1998) An intramolecular SH3-domain interaction regulates c-Abl activity. *Nat Genet* 18: 280–282. PMID: [9500553](https://pubmed.ncbi.nlm.nih.gov/9500553/)
  17. Greuber EK, Smith-Pearson P, Wang J, Pendergast AM (2013) Role of ABL family kinases in cancer: from leukaemia to solid tumours. *Nat Rev Cancer* 13: 559–571. doi: [10.1038/nrc3563](https://doi.org/10.1038/nrc3563) PMID: [23842646](https://pubmed.ncbi.nlm.nih.gov/23842646/)
  18. Allington TM, Gallier-Beckley AJ, Schiemann WP (2009) Activated Abl kinase inhibits oncogenic transforming growth factor-beta signaling and tumorigenesis in mammary tumors. *FASEB J* 23: 4231–4243. doi: [10.1096/fj.09-138412](https://doi.org/10.1096/fj.09-138412) PMID: [19690215](https://pubmed.ncbi.nlm.nih.gov/19690215/)
  19. Moroco JA, Craigo JK, Iacob RE, Wales TE, Engen JR, Smithgall TE (2014) Differential sensitivity of Src-family kinases to activation by SH3 domain displacement. *PLoS One* 9: e105629. doi: [10.1371/journal.pone.0105629](https://doi.org/10.1371/journal.pone.0105629) PMID: [25144189](https://pubmed.ncbi.nlm.nih.gov/25144189/)
  20. Pisabarro MT, Serrano L (1996) Rational design of specific high-affinity peptide ligands for the Abl-SH3 domain. *Biochemistry* 35: 10634–10640. PMID: [8718852](https://pubmed.ncbi.nlm.nih.gov/8718852/)
  21. Pisabarro MT, Serrano L, Wilmanns M (1998) Crystal structure of the abl-SH3 domain complexed with a designed high-affinity peptide ligand: implications for SH3-ligand interactions. *J Mol Biol* 281: 513–521. PMID: [9698566](https://pubmed.ncbi.nlm.nih.gov/9698566/)
  22. Dragiev P, Nadon R, Makarenkov V (2012) Two effective methods for correcting experimental high-throughput screening data. *Bioinformatics* 28: 1775–1782. doi: [10.1093/bioinformatics/bts262](https://doi.org/10.1093/bioinformatics/bts262) PMID: [22563067](https://pubmed.ncbi.nlm.nih.gov/22563067/)
  23. Shun TY, Lazo JS, Sharlow ER, Johnston PA (2011) Identifying actives from HTS data sets: practical approaches for the selection of an appropriate HTS data-processing method and quality control review. *J Biomol Screen* 16: 1–14. doi: [10.1177/1087057110389039](https://doi.org/10.1177/1087057110389039) PMID: [21160066](https://pubmed.ncbi.nlm.nih.gov/21160066/)
  24. Murphy M, Jason-Moller L, Bruno J (2006) Using Biacore to measure the binding kinetics of an antibody-antigen interaction. *Curr Protoc Protein Sci Chapter* 19: Unit 19.14.
  25. Jason-Moller L, Murphy M, Bruno J (2006) Overview of Biacore systems and their applications. *Curr Protoc Protein Sci Chapter* 19: Unit 19.13.
  26. Charter NW, Kauffman L, Singh R, Eglen RM (2006) A generic, homogenous method for measuring kinase and inhibitor activity via adenosine 5'-diphosphate accumulation. *J Biomol Screen* 11: 390–399. PMID: [16751335](https://pubmed.ncbi.nlm.nih.gov/16751335/)
  27. Yang J, Campobasso N, Biju MP, Fisher K, Pan XQ, Cottom J et al. (2011) Discovery and characterization of a cell-permeable, small-molecule c-Abl kinase activator that binds to the myristoyl binding site. *Chem Biol* 18: 177–186. doi: [10.1016/j.chembiol.2010.12.013](https://doi.org/10.1016/j.chembiol.2010.12.013) PMID: [21338916](https://pubmed.ncbi.nlm.nih.gov/21338916/)
  28. Champ PC, Camacho CJ (2007) FastContact: a free energy scoring tool for protein-protein complex structures. *Nucleic Acids Res* 35: W556–W560. PMID: [17537824](https://pubmed.ncbi.nlm.nih.gov/17537824/)
  29. Gotz AW, Williamson MJ, Xu D, Poole D, Le GS, Walker RC (2012) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J Chem Theory Comput* 8: 1542–1555. PMID: [22582031](https://pubmed.ncbi.nlm.nih.gov/22582031/)
  30. Case DA, Babib V, Berryman JT, Betz RM, Cai Q et al. (2014) AMBER 14, University of California San Francisco, version 14 [computer program].
  31. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. *J Comput Chem* 25: 1157–1174. PMID: [15116359](https://pubmed.ncbi.nlm.nih.gov/15116359/)

32. Koes DR, Baumgartner MP, Camacho CJ (2013) Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *J Chem Inf Model* 53: 1893–1904. doi: [10.1021/ci300604z](https://doi.org/10.1021/ci300604z) PMID: [23379370](https://pubmed.ncbi.nlm.nih.gov/23379370/)
33. Chen S, Dumitrescu TP, Smithgall TE, Engen JR (2008) Abl N-terminal cap stabilization of SH3 domain dynamics. *Biochemistry* 47: 5795–5803. doi: [10.1021/bi800446b](https://doi.org/10.1021/bi800446b) PMID: [18452309](https://pubmed.ncbi.nlm.nih.gov/18452309/)
34. Zhang JH, Chung TDY, Oldenburg KR (1999) A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J Biomol Screen* 4: 67–73. PMID: [10838414](https://pubmed.ncbi.nlm.nih.gov/10838414/)
35. Niesen FH, Berglund H, Vedadi M (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* 2: 2212–2221. PMID: [17853878](https://pubmed.ncbi.nlm.nih.gov/17853878/)
36. Balakumar P, Nyo YH, Renushia R, Raaginey D, Oh AN, Varatharajan R et al. (2014) Classical and pleiotropic actions of dipyridamole: Not enough light to illuminate the dark tunnel? *Pharmacol Res* 87: 144–150. doi: [10.1016/j.phrs.2014.05.008](https://doi.org/10.1016/j.phrs.2014.05.008) PMID: [24861566](https://pubmed.ncbi.nlm.nih.gov/24861566/)
37. Engen JR, Wales TE, Chen S, Marzluff EM, Hassell KM, Weis DD et al. (2013) Partial cooperative unfolding in proteins as observed by hydrogen exchange mass spectrometry. *Int Rev Phys Chem* 32: 96–127. PMID: [23682200](https://pubmed.ncbi.nlm.nih.gov/23682200/)
38. Engen JR, Wales TE, Hochrein JM, Meyn MA III, Banu OS, Bahar I et al. (2008) Structure and dynamic regulation of Src-family kinases. *Cell Mol Life Sci* 65: 3058–3073. doi: [10.1007/s00018-008-8122-2](https://doi.org/10.1007/s00018-008-8122-2) PMID: [18563293](https://pubmed.ncbi.nlm.nih.gov/18563293/)
39. Meyn MA III, Wilson MB, Abdi FA, Fahey N, Schiavone AP, Wu J et al. (2006) Src family kinases phosphorylate the Bcr-Abl SH3-SH2 region and modulate Bcr-Abl transforming activity. *J Biol Chem* 281: 30907–30916. PMID: [16912036](https://pubmed.ncbi.nlm.nih.gov/16912036/)
40. Chen S, Brier S, Smithgall TE, Engen JR (2007) The Abl SH2-kinase linker naturally adopts a conformation competent for SH3 domain binding. *Protein Sci* 16: 572–581. PMID: [17327393](https://pubmed.ncbi.nlm.nih.gov/17327393/)

## **APPENDIX B APPLICATION OF MODELING PROTEIN-LIPID INTERACTIONS**

In this appendix, we will present an application of modeling the interactions of small molecules and lipids to the lipoxygenase h-5-LOX[117]. Reproduced with permission.

### **B.1 MODELING LIPID INTERACTIONS WITH H-5-LOX**



## Discovery of a novel activator of 5-lipoxygenase from an anacardic acid derived compound collection



Rosalina Wisastra<sup>a</sup>, Petra A. M. Kok<sup>a</sup>, Nikolaos Eleftheriadis<sup>a</sup>, Matthew P. Baumgartner<sup>b</sup>, Carlos J. Camacho<sup>b</sup>, Hidde J. Haisma<sup>a</sup>, Frank J. Dekker<sup>a,\*</sup>

<sup>a</sup> Department of Pharmaceutical Gene Modulation, Groningen Research Institute of Pharmacy, University of Groningen, Antonius Deusinglaan 1, 9713 AV Groningen, The Netherlands

<sup>b</sup> Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA 15260, United States

### ARTICLE INFO

#### Article history:

Received 5 August 2013

Revised 8 October 2013

Accepted 12 October 2013

Available online 23 October 2013

#### Keywords:

Human 5-lipoxygenase

Anacardic acid

Cyclooxygenase-2

Enzyme kinetics

Allosteric binding

### ABSTRACT

Lipoxygenases (LOXs) and cyclooxygenases (COXs) metabolize poly-unsaturated fatty acids into inflammatory signaling molecules. Modulation of the activity of these enzymes may provide new approaches for therapy of inflammatory diseases. In this study, we screened novel anacardic acid derivatives as modulators of human 5-LOX and COX-2 activity. Interestingly, a novel salicylate derivative **23a** was identified as a surprisingly potent activator of human 5-LOX. This compound showed both non-competitive activation towards the human 5-LOX activator adenosine triphosphate (ATP) and non-essential mixed type activation against the substrate linoleic acid, while having no effect on the conversion of the substrate arachidonic acid. The kinetic analysis demonstrated a non-essential activation of the linoleic acid conversion with a  $K_A$  of 8.65  $\mu$ M,  $\alpha K_A$  of 0.38  $\mu$ M and a  $\beta$  value of 1.76. It is also of interest that a comparable derivative **23d** showed a mixed type inhibition for linoleic acid conversion. These observations indicate the presence of an allosteric binding site in human 5-LOX distinct from the ATP binding site. The activatory and inhibitory behavior of **23a** and **23d** on the conversion of linoleic compared to arachidonic acid are rationalized by docking studies, which suggest that the activator **23a** stabilizes linoleic acid binding, whereas the larger inhibitor **23d** blocks the enzyme active site.

© 2013 The Authors. Published by Elsevier Ltd. Open access under [CC BY-NC-SA license](http://creativecommons.org/licenses/by-nc-sa/4.0/).

### 1. Introduction

The lipoxygenases (LOXs) and the cyclooxygenases (COXs) are enzymes that play key roles in the major metabolic pathways of poly-unsaturated fatty acids (PUFAs). These pathways convert PUFAs into inflammatory mediators such as lipoxins, leukotrienes, and prostaglandins, which play a regulatory role in numerous inflammatory and proliferative diseases including asthma, arthritis and cancer.<sup>1</sup> The biological roles of LOXs and COXs in metabolic pathways towards inflammatory mediators and signaling molecules demonstrate the utility of these enzymes as therapeutic targets.

Lipoxygenases are non-heme iron containing enzymes. There are four types of lipoxygenases with different positional specificities for arachidonic acid oxidation present in mammalian tissues; 5-LOX, 8-LOX, 12-LOX and 15-LOX. In general, lipoxygenases comprise of two domains in which the N-terminal domain is the regu-

latory domain and the C-terminal domain is the catalytic domain.<sup>2</sup> Although, human lipoxygenases show about 60% sequence similarities,<sup>3</sup> their regulatory mechanisms are variable. Unlike other LOXs, 5-LOX activity requires the presence of activators such as  $Ca^{2+}$  and adenosine triphosphate (ATP). In addition, different types of LOXs have been identified based on their distribution over different tissues. For example, 12-LOX is subdivided in platelet 12-LOX (p12-LOX) and 12R-LOX. As its name implies, platelet 12-LOX (p12-LOX) is mostly found in platelets, whereas 12R-LOX is most frequently found in skin cells.<sup>4,5</sup> Similarly, 15-LOX is subdivided in 15-LOX-1 and 15-LOX-2. 15-LOX-1 is highly expressed in leukocytes and airways endothelial cells,<sup>6,7</sup> whereas 15-LOX-2 is expressed in multiple tissues such as prostate, lung, cornea, liver, colon, kidney, and brain but not in leukocytes.<sup>8</sup> Moreover, stimulation of cells by interleukin (IL)-4 and IL-13 shows selective enhancement of the 15-LOX-1 expression and not 15-LOX-2 expression.<sup>9</sup> In addition, 15-LOX-1 and 15-LOX-2 lack similarity at the primary sequence level, which also hints at distinct biological roles and the possibility to develop selective inhibitors or activators.<sup>10</sup>

The overexpression of certain lipoxygenases has been reported in numerous inflammatory diseases. For example, increased 15-LOX-1 expression in bronchial epithelium, which is accompanied by elevated concentrations of 15-hydroxyeicosatetraenoic acid

\* Corresponding author. Tel.: +31 50 3638030.

E-mail address: [f.j.dekker@rug.nl](mailto:f.j.dekker@rug.nl) (F.J. Dekker).



(15-HETE), has been observed in patients with asthma and chronic bronchitis.<sup>11,12</sup> In addition, activity of other LOXs isoenzymes has been reported to be significantly increased in disease conditions. For example, upregulation of 12R-LOX activity has been observed in the development of skin tumor and melanoma.<sup>13,14</sup> In addition, overexpression of 5-LOX has been reported to contribute to airway inflammation in asthma and to the growth of some tumors in, for example, lung, prostate, brain and colon.<sup>15–19</sup> In contrast, lipoxins, another type of arachidonic acid metabolites, have been reported to have important roles as anti-inflammatory and pro-resolution mediators,<sup>20</sup> which play an important role in the termination of immune responses. Production of lipoxins upon inflammatory stimulation involves two main lipoxygenase-mediated pathways. The first pathway involving 15-LOX and 5-LOX occurs in mucosal tissues, such as the respiratory tract, gastrointestinal tract and the oral cavity. The second pathway involves 5-LOX and p12-LOX, in which 5-LOX produces the Leukotriene A4 (LTA4) and p12-LOX converts the LTA4 to lipoxin A4 (LXA4).<sup>21</sup> Lipoxin A4 has important roles in stopping neutrophil migration, stimulation of monocyte and macrophage activation, and inhibition of leukotriene B4 formation.<sup>20</sup> Induction of 15-LOX-1 by IL-13 in synovial tissues of patients with rheumatoid arthritis enhances the level of LX4 in a negative feedback loop to limit inflammatory responses induced by pro-inflammatory mediators such as Leukotriene B4.<sup>22</sup> Another study showed that prior to induction with peroxisome proliferator-activated receptors gamma (PPAR $\gamma$ ) agonists in brain, the 5-LOX expression was enhanced in association with the increase of cerebral LX4 production and inhibition of Leukotriene B4 production.<sup>23</sup> The association of 5-lipoxygenase activity in the production of both pro- and anti-inflammatory mediators indicates its crucial role in pathophysiological processes of inflammation. Based on these findings we speculate that activators of 5-lipoxygenase might also have beneficial effects in inflammation by triggering the termination of immune responses. In addition, it might be speculated that the redirection of lipoxygenase metabolism by 5-LOX activators towards other substrates might also lead to the formation of less harmful pro-inflammatory products. Enzyme activation as a concept in drug discovery has been shown to be successful for glucokinase activators.<sup>24</sup> We stress, however, that lipoxygenase activation by small molecules with the aim to limit inflammation is a novel concept and lacks experimental proof.

Prostaglandin endoperoxide H synthases (PGHSs), are also known as cyclooxygenases (COXs), catalyze the formation of prostaglandins (PGs). COXs are comprised of two main sites, a heme site with peroxidase activity and a bis-oxygenase site. COXs convert arachidonic acid into prostaglandin endoperoxide H<sub>2</sub> (PGH<sub>2</sub>) in a two steps process. The first step is enzymatic catalysis of the insertion of two molecules O<sub>2</sub> into arachidonic acid to generate prostaglandin G<sub>2</sub> (PGG<sub>2</sub>). The second step is the peroxidase reaction in which the PGG<sub>2</sub> is oxidized to PGH<sub>2</sub> by a two electron abstraction. The peroxidase reaction occurs at a heme-containing active site located near the protein surface, whereas the oxygenation reaction occurs in a hydrophobic channel in the core of the enzyme.<sup>25</sup> The COX isoenzymes COX-1 and COX-2 are differentially expressed in various tissues. Under normal conditions, COX-1 is expressed in most mammalian cells, whereas COX-2 is not expressed. The COX-2 expression is increased under inflammatory conditions. Animal models of inflammatory arthritis indicate that overexpression of COX-2 is responsible for the increase of PGs in joint tissue and cause acute and chronic inflammation.<sup>26</sup> Elevated levels of PGs were observed in inflammatory bowel diseases (IBD).<sup>27,28</sup> A similar behavior was observed in cancers. For example, the expression of COX-2, but not COX-1, was observed in various regions of the colon in human colorectal cancer.<sup>29,30</sup> Therefore,

COX-2 inhibition is considered to be a relevant therapeutic strategy for treatment of cancer and inflammatory diseases.

Anacardic acid is a naturally occurring 6-alkyl substituted salicylate that is identified from cashew nutshells. Previously, we found that this salicylate and its derivatives are allosteric regulators of lipoxygenase activity. We identified an allosteric inhibitor of soybean lipoxygenase-1 (SLO-1) (**19**) and an allosteric activator of potato 5-lipoxygenase (5-LOXs) (**21**).<sup>31</sup> It should, however, be noted that the lack of similarity between the human and plant enzymes prohibits extrapolation or these data towards biological effects in cell-based studies.<sup>32</sup> Therefore, we continue our previous studies by exploring the inhibitory potency of a salicylate-based compound collection for inhibition of human 5-lipoxygenase and cyclooxygenase-2. In this study, we describe the synthesis of a novel focused collection of salicylate-based compounds and we investigate their influence on the enzyme activity of human 5-LOX and human COX-2. This enabled the identification of a novel activator **23a** and a mixed inhibitor **23d** of the conversion of the linoleic acid by h-5-LOX. Interestingly, no effect on the arachidonic acid conversion was observed. Enzyme kinetic investigations demonstrated that activator **23a** binds to an allosteric binding site on h-5-LOX that is different from the ATP regulatory site. Modeling studies suggest that **23a** binds close to the substrate linoleic acid, thereby enhancing the  $K_m$  and  $k_{cat}$  for its conversion.

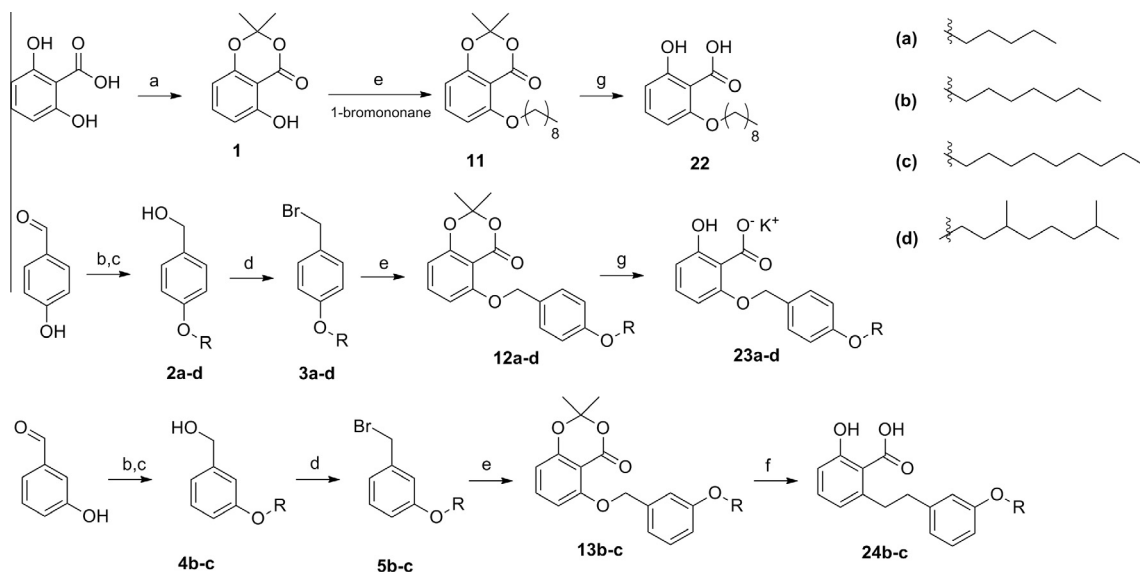
## 2. Results and discussion

### 2.1. Synthesis

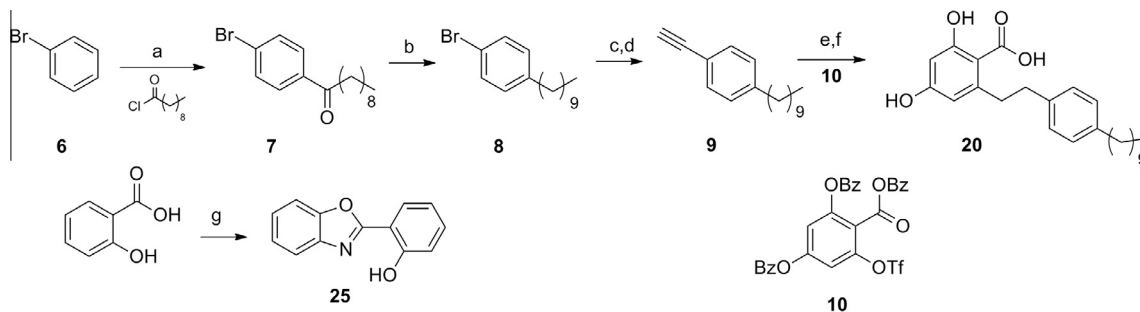
Our salicylate-based compound collection that is inspired by anacardic acid (**14**) comprises of the previously published salicylates **14–19**, and **21** and the newly synthesized salicylates **20**, **22**, **23a–d**, **24b–c** and benzoxazole **25**. The synthesis of anacardic acid (**14**) and its derivatives **15–19**, and **21** were described previously by Ghizzoni et al.<sup>33,34</sup> This collection includes the previously described compounds that were reported as inhibitors or activator of soybean lipoxygenase-1 (SLO1) (**19**) and potato 5-LOX (**21**).<sup>31</sup> The newly synthesized compounds were inspired by compound **19** and **21**, which contain either a 6-phenylethyl substitution or a 6-alkoxy substitution. This was merged into a *para*-alkoxybenzyl-oxy substitution, which can be synthesized using convenient methods.

The starting materials, acetone **1** (Scheme 1) and triflate **10** (Scheme 2), were synthesized as described by Uchiyama et al.<sup>35</sup> 2-Hydroxy-6-(nonyloxy)benzoic acid (**22**) was synthesized from commercially available 1-bromononane and acetone **1** following synthetic procedure described by Ghizzoni et al.<sup>33</sup> (Scheme 1). *p*-Alkyloxybenzylbromides (**3a–d**) and *m*-alkyloxybenzylbromides (**5b–c**) were prepared in three steps from *p*-hydroxybenzaldehyde or *m*-hydroxybenzaldehyde. The first step involves the coupling of alkylbromides with the hydroxybenzaldehydes to give the corresponding ethers. Subsequently, the benzaldehydes were reduced to the corresponding benzyl alcohols **2a–d** and **4b–c**<sup>36</sup> in high yields. Direct coupling of benzyl alcohols to phenol **1** by Mitsunobu reaction did not provide the desired ethers. Therefore, the benzylalcohols **2a–d** and **4b–c** were brominated using PBr<sub>3</sub> to give the corresponding benzylbromides **3a–d** and **4b–c**. The benzylbromides were coupled to acetone **1** to give the corresponding products **12a–d** and **13b–c** in high yields (>70%). Subsequently, the acetone protecting group was cleaved using potassium hydroxide in THF to provide the final products **23a–d** as potassium salts (Scheme 1). Isolation of salicylates **23a–d** as salicylate salts proved to be necessary to avoid degradation of the *p*-alkoxybenzylethers by the strongly acidic salicylic acids. Acidification results a slow hydrolysis of the product to give 2,6-dihydroxybenzoate and the





**Scheme 1.** Reagents and conditions: (a)  $\text{SOCl}_2$ , DMAP, DME, acetone,  $0^\circ\text{C}$  for 1 h, then rt overnight; (b) 1-bromo alkane,  $\text{K}_2\text{CO}_3$ , DMF, rt overnight; (c)  $\text{NaBH}_4$ , MeOH,  $0^\circ\text{C}$  for 1 h, then rt for 30 min; (d)  $\text{PBr}_3$ ,  $\text{CH}_2\text{Cl}_2$ ,  $0^\circ\text{C}$  for 1.5 h; (e) 1-bromononane,  $\text{K}_2\text{CO}_3$ , DMF, rt overnight; (f) compound **1**,  $\text{K}_2\text{CO}_3$ , DMF, rt overnight; (g) 5 M KOH, THF,  $60^\circ\text{C}$  overnight.



**Scheme 2.** Reagents and conditions: (a) decanoylchloride,  $\text{AlCl}_3$ ,  $\text{CH}_2\text{Cl}_2$ ,  $60^\circ\text{C}$  for 1 h; (b)  $\text{NH}_2\text{NH}_2 \cdot \text{H}_2\text{O}$ , KOH, 1-octanol, reflux, 3 h; (c) trimethylsilyl-acetylene, CuI,  $\text{PdCl}_2(\text{PPh}_3)_2$ ,  $\text{Et}_2\text{NH}$ ,  $\text{PPh}_3$ ,  $\text{CH}_3\text{CN}$ , (MW,  $120^\circ\text{C}$ , 95 W, 35 min); (d) TBAF, THF,  $0^\circ\text{C}$ , 10 min; (e) CuI,  $\text{PdCl}_2(\text{PPh}_3)_2$ ,  $\text{Et}_2\text{NH}$ , triflate **10**,  $\text{CH}_3\text{CN}$  (MW,  $120^\circ\text{C}$ , 70 W, 35 min); (f)  $\text{H}_2$ , Pd/C, MeOH,  $45^\circ\text{C}$ , 24 h; (g) 2-aminophenol, polyphosphoric acid,  $180^\circ\text{C}$  for 6.5 h.

corresponding (4-(alkoxy)benzyl) cation, which is stabilized by the electron-donating alkoxy substituent on the *para*-position. As expected, for the *meta*-substituted benzyl substituents **24b–c** this problem was not observed.

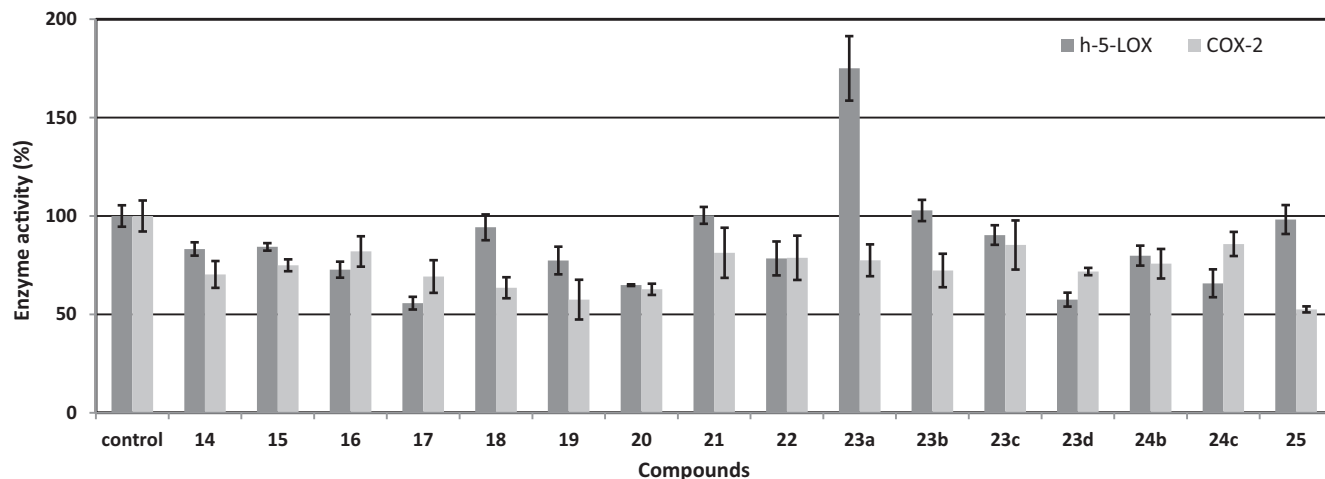
*p*-Decylphenyl bromide (**8**) (Scheme 2) was prepared from bromobenzene **6** and decanoylchloride following two reaction steps. The first step was the formation of 1-(4-bromophenyl)decan-1-one **7** through Friedel–Craft acylation using  $\text{AlCl}_3$ .<sup>37</sup> The second step was the Wolff–Kishner reduction of 1-(4-bromophenyl)decan-1-one **7** using hydrazine monohydrate and potassium hydroxide to give product **8**.<sup>37</sup> Subsequently, arylbromide **8** was converted to alkyne **9** by coupling with trimethylsilyl (TMS) acetylene by a Sonogashira reaction and subsequent removal of the TMS protection using tetra butyl ammonium fluoride (TBAF). The resulting alkyne **9** was coupled to triflate **10**<sup>38</sup> to give the product in high yield (80%), which was subsequently converted into product **20** by catalytic hydrogenation.

## 2.2. Enzyme inhibition

LOXs are non-heme iron dependent enzymes. Benzoxazole **25** was designed based on the presumption that iron-binding is a key property for the inhibition of lipoxygenases and cyclooxygen-

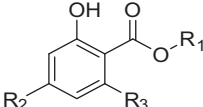


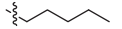
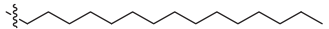




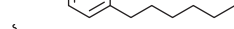
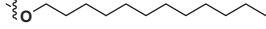
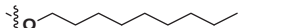


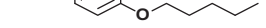

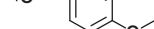
ases. Benzoxazole **25** was synthesized from salicylic acid and 2-amino phenol through direct cyclization using polyphosphoric acid (Scheme 2).<sup>39</sup> Compound **25** did not inhibit h-5-LOX, however it gives 50% COX-2 inhibition at 50  $\mu\text{M}$  inhibitor concentration. This indicates that **25** is a potential starting point for development of selective inhibitors of COX-2 (Fig. 1).

The influence of the salicylates (Table 1) on the enzyme activity of human 5-lipoxygenase and/or cyclooxygenase-2 was investigated. The h-5-LOX activity was monitored in real time by detection of the formation of the UV absorbance of the conjugated diene 13-hydroperoxy-*cis*-9-*trans*-11-octa decanoic acid (13-HPOD) from linoleic acid.<sup>40,41</sup> The residual h-5-LOX activity was monitored after 10 min pre-incubation with 50  $\mu\text{M}$  of the compound of interest. The COX-2 activity was determined by measuring the conversion of arachidonic acid, which conversion is monitored by the oxidation of *N,N,N,N*-tetramethyl-*p*-phenylenediamine (TMPD).<sup>42</sup> The residual enzyme activity was measured after 5 min pre-incubation in the presence of 50  $\mu\text{M}$  of the compound of interest. The enzyme activity without inhibitor present was taken as control and set to 100% and the activity without enzyme present was set to 0%. The residual enzyme activities of h-5-LOX and COX-2 in the presence of different salicylates are shown in Figure 1.



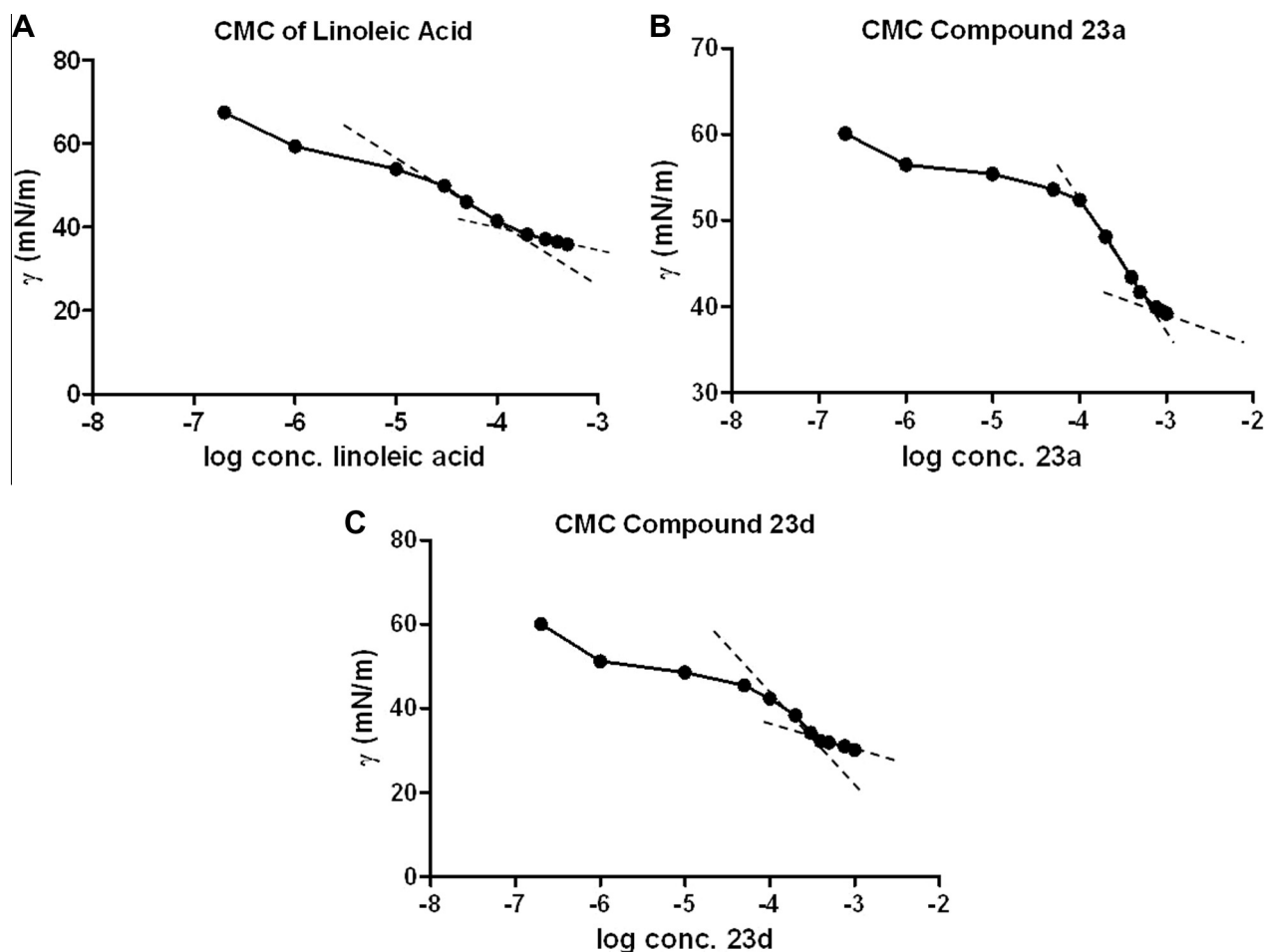
**Figure 1.** Residual enzyme activity that was observed for the screening of a salicylate-based compound collection for inhibition of h-5-LOX and COX-2 activity in the presence of 50  $\mu$ M of the respective compounds. The h-5-LOX activity was determined in presence of 100  $\mu$ M linoleic acid as a substrate. The COX-2 activity was determined in presence of 2 mM arachidonic acid as a substrate. The results presented were the average of three independent experiments and the standard deviations are shown.

**Table 1**  
Focused compound collection

Compounds	R1	R2	R3
			
<b>14</b>	H	H	
<b>15</b>	H	H	
<b>16</b>	H	H	
<b>17</b>	H	OH	
<b>18</b>	H	H	
<b>19</b>	H	OH	
<b>20</b>	H	OH	
<b>21</b>	H	H	
<b>22</b>	H	H	
<b>23a</b> (PK131)	K	H	
<b>23b</b>	K	H	
<b>23c</b>	K	H	
<b>23d</b> (PK147)	K	H	
<b>24b</b>	H	H	
<b>24c</b>	H	H	
<b>25</b>			

Anacardic acid **14**, compounds **15**, and **16** show little inhibition of both h-5-LOX and COX-2 at 50  $\mu$ M. In comparison to anacardic acid **14**, a slight improvement in the inhibitory potency of h-5-LOX was observed for compound **17**. Importantly, neither inhibition nor activation was observed from compound **21**, which in our previous report<sup>31</sup> shows a strong activation on potato 5-LOX, which demonstrates the difference between plant and human enzymes.

Compounds **23a–d** with alkoxy substituents in the *para*-positions of the benzylethers, show an interesting structure activity relationship for modulation of h-5-LOX activity. Compound **23a** shows strong activation of h-5-LOX at 50  $\mu$ M, whereas compound **23b** and **23c** with a longer side chain show no or little inhibition of h-5-LOX. Furthermore, compound **23d** with a branched side chain provides almost 50% inhibitions of h-5-LOX at the same concentration. The compound is racemic at its asymmetric carbon.



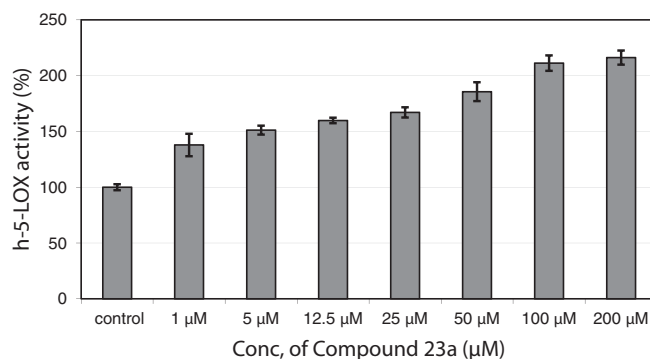
**Figure 2.** Surface tensions of (A) anacardic acid **14** and (B) compound **23a** (C) compound **23d** against the logarithm of concentration. CMC values (A–C) were measured in human 5-LOX assay conditions, Tris buffer (50 mM), 2 mM EDTA and 2 mM CaCl<sub>2</sub>, pH 7.5, rt ( $t = 19^\circ\text{C}$ ).

The data in Figure 1 demonstrate that the salicylate-based compounds do not inhibit COX-2 activity more than 50% at 50  $\mu\text{M}$ , which indicates that this type of compounds has a tendency for selective modulation of h-5-LOX compared to COX-2. Interestingly, benzoxazole **25** inhibits the COX-2 activity by about 50%, whereas no inhibition was observed on h-5-LOX, which indicates that further optimization of this class of compound may be feasible.

In order to assure that the assays are performed in homogenous solutions the critical micelle concentrations (CMCs) of the different compounds were determined and the assays were performed at concentration below the CMC. Homogenous solutions are required for proper analysis of the enzyme kinetics. The CMCs for **23a** and **23d** at the assay conditions (50 mM Tris buffer, 2 mM EDTA and 2 mM CaCl<sub>2</sub>, pH 7.5, rt) are, respectively, 628  $\mu\text{M}$  and 395  $\mu\text{M}$  (Fig. 2B and C) which indicate that micelle formation does not occur at concentrations employed for inhibition or activation of h-5-LOX. In addition, we found a CMC value for linoleic acid of 185  $\mu\text{M}$  (Fig. 2A),<sup>43</sup> which demonstrates that the substrate concentration in the inhibition studies (100  $\mu\text{M}$ ) was below the CMC.

### 2.3. Enzyme kinetic studies for human 5-lipoxygenase

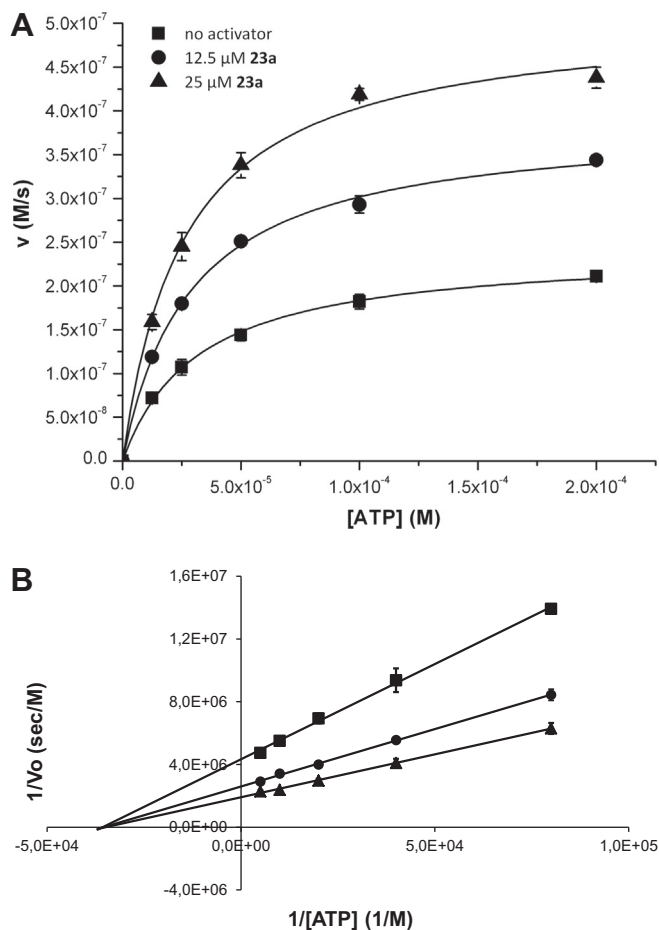
A steady-state enzyme kinetic analysis of h-5-LOX in the presence of the activator **23a** was performed in order to unravel the activation mechanism. The kinetic characterization of h-5-LOX activity versus linoleic acid in the presence of activator **23a** shows a concentration dependent activation (Fig. 3). The activation was



**Figure 3.** Concentration dependent activation of the linoleic acid conversion by h-5-LOX in the presence of various concentrations of compound **23a** and in its absence (control). The results were the average of three independent experiments with error bars ( $\pm\text{SD}$ ).

determined in the presence of various concentrations of the natural substrate linoleic acid or the h-5-LOX activator ATP. The initial velocities of h-5-LOX were determined at various concentration of ATP, various concentrations **23a** (0, 12.5 and 25  $\mu\text{M}$ ) and a fixed concentration of linoleic acid (100  $\mu\text{M}$ ). The velocities were plotted in the Michaelis–Menten and the Lineweaver–Burk plot as shown in Figure 4.

The Michaelis–Menten data show that activator **23a** causes an increase in  $V_{\text{max}}$ , whereas the  $K_{\text{m}}$  remains constant (Table 2).



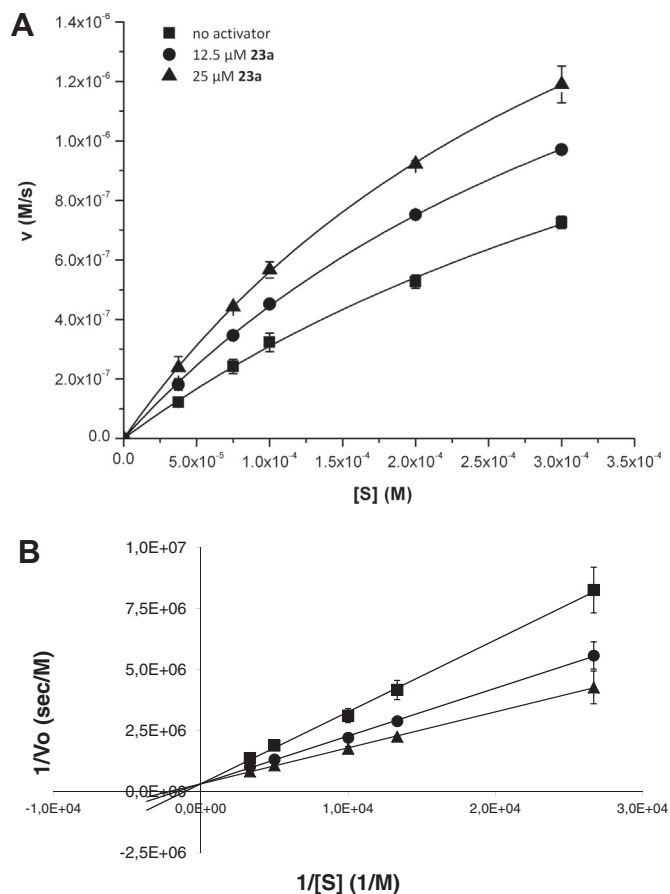
**Figure 4.** Steady-state kinetic characterization of the linoleic acid conversion versus the ATP concentration of h-5-LOX activator **23a**. (A) The Michaelis–Menten plots and (B) the Lineweaver–Burk plots show the relation of h-5-LOX activity versus the ATP concentration at three selected concentration of **23a** (■) 0  $\mu\text{M}$ , (●) 12.5  $\mu\text{M}$ , and (▲) 25  $\mu\text{M}$  in the presence of 100  $\mu\text{M}$  linoleic acid substrate.

**Table 2**  
Enzyme kinetic parameter for activation of h-5-LOX by activator **23a** versus ATP

<b>23a</b> ( $\mu\text{M}$ )	$K_m^{\text{app}}$ ( $\mu\text{M}$ )	$V_{\text{max}}^{\text{app}}$ (nM/s)	$R^2$
0	$27.9 \pm 1.7$	$230.9 \pm 4.2$	0.998
12.5	$28.3 \pm 1.7$	$386.1 \pm 7.3$	0.999
25	$28.6 \pm 2.3$	$523.6 \pm 13.7$	0.999

Lineweaver–Burk analysis provided similar values. The values demonstrate non-competitive activation of linoleic acid conversion by h-5-LOX by compound **23a** compared to ATP. The fact that binding of **23a** does not influence the binding constant of ATP to h-5-LOX indicates that **23a** does not compete for binding in the ATP binding pocket that allosterically regulates h-5-LOX activity. This indicates the presence of an allosteric binding pocket that is distinct from the ATP binding pocket and regulates the h-5-LOX enzyme activity.

The enzyme kinetics of **23a** were further investigated through the Michaelis–Menten and the Lineweaver–Burk plots (Fig. 5). The initial velocities of h-5-LOX at different concentration of linoleic acid and a fixed concentration of ATP (100  $\mu\text{M}$ ), and at several selected concentrations of **23a** (0, 12.5 and 25  $\mu\text{M}$ ) were measured. Activator **23a** causes an increase in  $V_{\text{max}}$  and a decrease in  $K_m$  (Table 3), which indicates non-essential mixed type activation. This

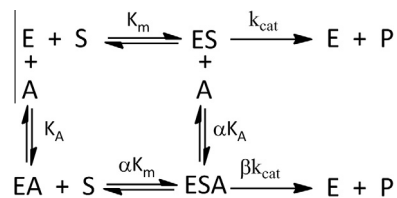


**Figure 5.** Steady-state kinetic characterization of the linoleic acid conversion by h-5-LOX activation by activator **23a**. (A) Michaelis–Menten plots and (B) Lineweaver–Burk plots show the relation of h-5-LOX activity versus linoleic acid concentrations at three selected concentrations of **23a** (■) 0  $\mu\text{M}$ , (●) 12.5  $\mu\text{M}$ , and (▲) 25  $\mu\text{M}$ .

**Table 3**

Enzyme kinetic parameters for activation of h-5-LOX by activator **23a** versus linoleic acid

<b>23a</b> ( $\mu\text{M}$ )	$K_m^{\text{app}}$ (mM)	$V_{\text{max}}^{\text{app}}$ ( $\mu\text{M/s}$ )	$R^2$
0	$0.590 \pm 0.080$	$2.14 \pm 0.21$	0.999
12.5	$0.438 \pm 0.020$	$2.39 \pm 0.07$	0.999
25	$0.383 \pm 0.011$	$2.70 \pm 0.05$	0.999



**Scheme 3.** Kinetic model for non-essential activation.

indicates that the activator can bind to the free enzyme (E, Scheme 3) and also to the substrate bound enzyme (ES, Scheme 3). In this model binding of the substrate and **23a** mutually influence each other. According to this model the activation of h-5-LOX by **23a** can be described by Eq. 1 (Fig. 7).

The enzyme kinetic parameters were derived as described in the material and methods using methods from Leskovac.<sup>44</sup> The activator dissociation constant ( $K_A$ ), the change in the affinity of

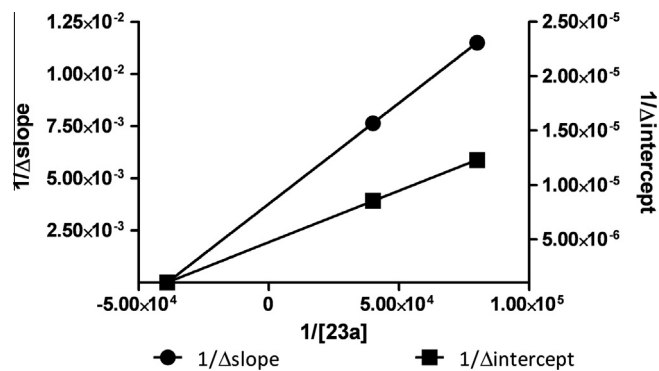


Figure 6. Re-plot of  $1/\Delta$  slopes and  $1/\Delta y$  intercept versus concentration of **23a**.

$$v = \frac{V_{\max} \times [S]}{K_m \left( \frac{1 + [A]}{K_A} \right) + [S] \left( \frac{1 + [A]}{\alpha K_A} \right)} \quad \text{Equation 1}$$

Figure 7. Equation 1 for the enzyme kinetics according to the model in Scheme 2.<sup>44</sup>  $v$  is the reaction velocity,  $V_{\max}$  is the maximal reaction velocity,  $[S]$  is the substrate concentration and  $K_m$  is the Michaelis–Menten constant,  $[A]$  is the activator concentration.  $\alpha$  and  $\beta$ , respectively, are the parameters to describe the change in the affinity of substrate binding and the change in the catalytic constant.

substrate binding ( $\alpha$ ) and the change in the catalytic constant ( $\beta$ ) were determined by re-plotting the slopes and the  $y$ -interceptions derived from the Lineweaver–Burk plot as a function of activator **23a** concentration (Fig. 6).<sup>44</sup> The  $K_A$  value was determined from the intersection of  $1/\Delta$ slopes plot and  $1/\Delta y$  intercept plot. The intersection of  $1/\Delta y$  intercept plot with the ordinate is defined as  $\beta \cdot V_{\max}/(\beta - 1)$  from which the  $\beta$  value was derived. Subsequently, the resulting  $\beta$  value was substituted to the equation  $\beta \cdot V_{\max}/K_m(-\beta - \alpha)$ , which is the intersection of  $1/\Delta$ slopes plot with the ordinate, to acquire the  $\alpha$  value.

The analysis shows that  $\alpha$  is 0.0438,  $\beta$  is 1.76 and  $K_A$  is 8.65  $\mu\text{M}$ . The  $\alpha$  value, which indicates the change in substrate binding, gives about 50-fold affinity enhancement of the substrate binding to the activator bound enzyme compared to the free enzyme. Compound **23a** shows a high potency as an activator of h-5-LOX with the  $K_A$  value in the micromolar range. Furthermore, the affinity of **23a** to the substrate bound enzyme is about 50-fold enhanced, which indicates that activation of h-5-LOX at high substrate concentrations ( $>10 \mu\text{M}$ ) takes place at low activator **23a** concentrations ( $<10 \mu\text{M}$ ) (Fig. 3). The catalytic constant for substrate conversion reaction was enhanced by 1.76-fold in the presence of the activator **23a** compared to the normal reaction conditions. In conclusion, compound **23a** gives a non-essential mixed-type activation of h-5-LOX and binds with high affinity ( $\alpha K_A = 390 \text{ nM}$ ) to the substrate bound enzyme and enhances the conversion of the substrate linoleic acid.

The observed results from the enzyme kinetic studies that give non-essential mixed-type activation versus linoleic acid and non-essential non-competitive activation versus ATP demonstrate that there is an allosteric regulatory site in h-5-LOX that is distinct from the ATP binding site. In order to get a better understanding of this allosteric site in the regulation of the enzyme activity, we also studied the inhibition mechanism of h-5-LOX by compound **23d**.

Interestingly, compound **23d** has a structure that is closely related to **23a** and this compound inhibits h-5-LOX at 50  $\mu\text{M}$ . We analyzed the underlying enzyme kinetic parameters. The Michaelis–Menten plots and the Lineweaver–Burk plots show an increase

Table 4

Enzyme kinetic parameter for inhibition of h-5-LOX by inhibitor **23d** versus linoleic acid

<b>23a</b> ( $\mu\text{M}$ )	$K_m^{\text{app}}$ (mM)	$V_{\max}^{\text{app}}$ ( $\mu\text{M/s}$ )	$R^2$
0	$0.624 \pm 0.085$	$2.24 \pm 0.23$	0.999
25	$0.666 \pm 0.121$	$1.84 \pm 0.25$	0.998
50	$0.690 \pm 0.101$	$1.47 \pm 0.16$	0.999

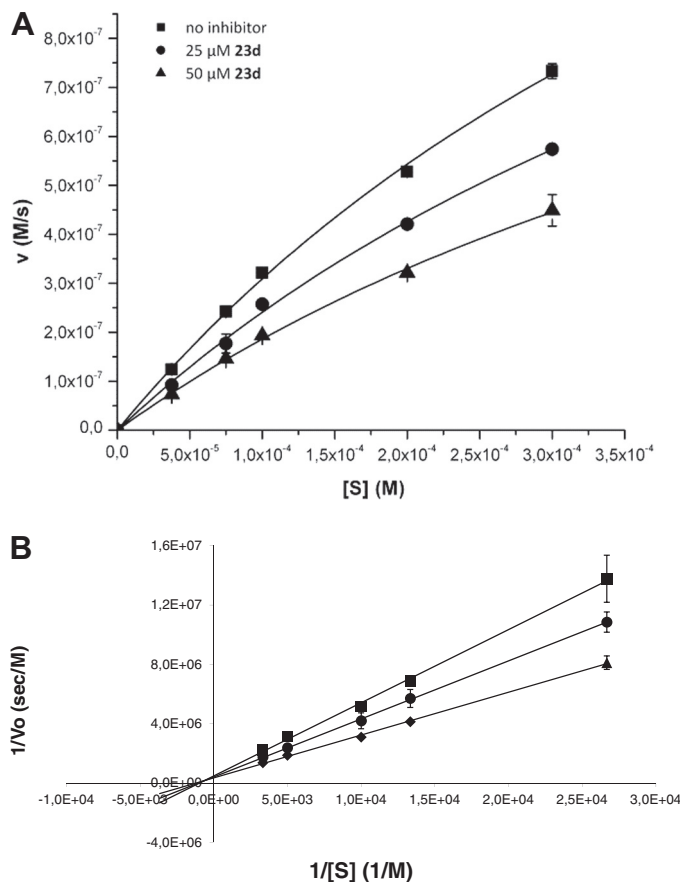
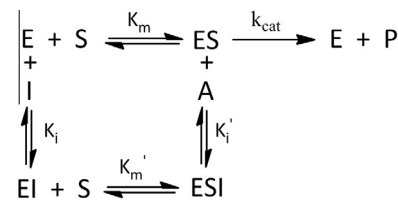


Figure 8. Steady-state kinetic characterization of linoleic acid conversion by h-5-LOX by inhibitor **23d**. (A) Michaelis–Menten plots and (B) Lineweaver–Burk plots show the relation of h-5-LOX activity versus linoleic acid concentration at three selected concentrations of **23d** (■) 0  $\mu\text{M}$ , (●) 25  $\mu\text{M}$ , and (▲) 50  $\mu\text{M}$ .



Scheme 4. Kinetic model for mixed type enzyme inhibition.

of the  $K_m^{\text{app}}$  and reduction of the  $K_m^{\text{app}}$  (Table 4), which is a mixed-type inhibition of h-5-LOX (Fig. 8). This indicates that inhibitor **23d** can bind to the free enzyme as well as to the substrate bound enzyme following the inhibition model in Scheme 4 from which the kinetic parameters  $V_{\max}$  and  $K_m$  can be derived by non-linear curve fitting of equation 2 (Fig. 9). Furthermore, the changes in the  $K_m$  and  $V_{\max}$  are  $\alpha/\alpha'$ .  $K_m$  and  $V_{\max}/\alpha'$  are used to determine the  $K_i$  and  $K_i'$  values for inhibitor **23d** using equations 3 and 4 (Fig. 9).  $K_i$  and  $K_i'$  values for inhibitor **23d** respectively are

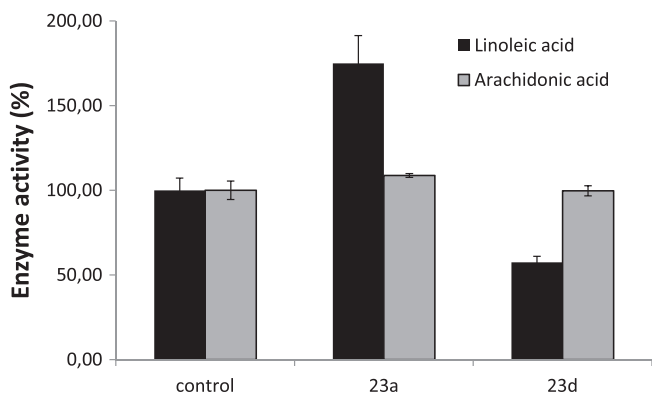


$$v = \frac{V_{\max} \times [S]}{\alpha K_m + \alpha' [S]} \quad \text{equation 2}$$

$$\alpha = 1 + \frac{[I]}{K_i} \quad \text{equation 3}$$

$$\alpha' = 1 + \frac{[I]}{K_i'} \quad \text{equation 4}$$

**Figure 9.** Equations 2(a), 3(b) and 4(c) for the enzyme kinetics according to the model in Scheme 2.  $v$  is the reaction velocity,  $V_{\max}$  is the maximal reaction velocity,  $[S]$  is the substrate concentration and  $K_m$  is the Michaelis–Menten constant.  $\alpha$  and  $\alpha'$ , respectively, are the parameters to describe the change of substrate binding affinity to the enzyme and the change of the maximum velocities.  $K_i$  is the dissociation constant of the inhibitor to the free enzyme and  $K_i'$  is the dissociation constant of the inhibitor to the enzyme–substrate complex.



**Figure 10.** Residual enzyme activity that was observed for compounds **23a** and **23d** for inhibition of h-5-LOX in the presence of 50  $\mu\text{M}$  of the respective compounds with linoleic acid or arachidonic acid as a substrate both at final concentrations of 100  $\mu\text{M}$ . The results presented were the average of three independent experiments and the standard deviations are shown.

70.3  $\mu\text{M}$  and 140  $\mu\text{M}$ . The fact that inhibitor **23d** shows a mixed type inhibition also indicates the presence of an allosteric site in human 5-lipoxygenase. We presume that compound **23d** binds to the same allosteric binding pocket as **23a** but that the difference in the alkyl chain causes a difference in affinity as well as a change from activation to inhibition.

Taking this together, the kinetic studies on activation or inhibition of human 5-lipoxygenase by either **23a** or **23d** indicate the presence of an allosteric binding site that influences the enzyme activity. In addition, both activation and inhibition obey a comparable kinetic model, which suggests that both effects originate from the same binding site in the enzyme. Presumably, depending on the structure of the compounds that bind to this allosteric site the enzyme would either be activated or inhibited. This is in line with the literature where the presence of an allosteric binding site in h-5-LOX that regulates its activity has been described frequently.<sup>45–48</sup> We conclude that anacardic acid derivatives can be used as tools to modulate the h-5-LOX activity toward linoleic acid and that these compounds modulate the activity of this enzyme by binding to an allosteric binding site.

We continued our studies with investigation of the modulation of **23a** and **23d** of the h-5-LOX activity for arachidonic acid (Fig. 10). Interestingly, no modulation of the h-5-LOX activity toward arachidonic acid was observed in presence of **23a** or **23d**. These results demonstrate that **23a** and **23d** are selective modulators of linoleic acid conversion and not arachidonic acid conversion by h-5-LOX.

#### 2.4. Molecular modeling

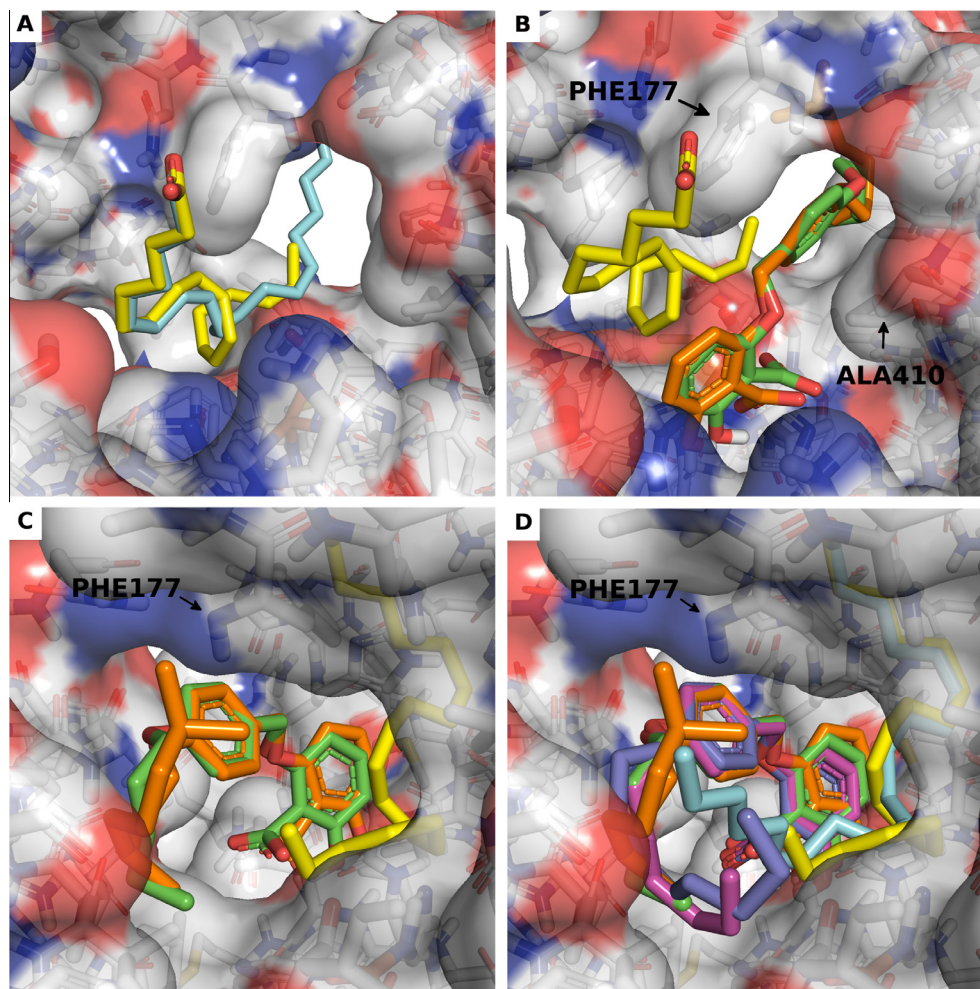
These results justify the hypothesis that the allosteric regulators **23a** and **23d** bind to an allosteric regulatory site in h-5-LOX that is

close to the active site. The hypothesis is supported by a previous report, which stated that h-5-LOX has a wide active site compared to the others LOXs, which could harbor two substrate molecules.<sup>49</sup> The observation that **23a** and **23d** modulate the conversion of linoleic acid, which has 18 carbon atoms, and not arachidonic acid, which has 20 carbon atoms, is in line with this hypothesis, because it is an indication for steric interactions between the substrate and **23a** or **23d**. Therefore, we hypothesize that the substrates together with either **23a** or **23d** bind in the active site, thus resulting in a different binding constant and a different turnover rate of the substrate linoleic acid. We employed this presumption as a basis for the modeling studies.

The binding models of the activator **23a** and the inhibitor **23d** in the linoleic acid bound as well as the arachidonic acid bound h-5-LOX active site were made based on the recently published crystal structure of h-5-LOX (PDB code 3V99).<sup>50</sup> The model of h-5-LOX was further refined using the related PDB structures 3V98 and 3V92 to model Ile673 and its carboxylic acid of the C-terminal (missing in 3V99), which coordinates the catalytic iron in the active site. We note that the model is not meant to capture the role of the phosphomimic S663D, but only as the structural basis for substrate binding. Additionally, the active site is in an open configuration accessible to a second small molecule. The above notwithstanding the structure is nearly indistinguishable from the wild type structure.<sup>50</sup>

The arachidonic acid substrate molecule lacks visible specific interactions with the protein except for the interaction of the *cis* double bond at carbon 11 to the catalytic iron. This lack of visible specific interactions renders the virtual re-docking of arachidonic acid and de novo docking of linoleic acid in the h-5-LOX enzyme active site futile. Therefore, we modeled a bound structure of linoleic acid by manually threading the lipid into the crystal structure of arachidonic acid using PYMOL.<sup>51</sup> The threaded linoleic acid was then minimized using default settings in UCSF Chimera (version 1.6)<sup>52</sup> to produce the final model of linoleic acid bound to the h-5-LOX active site as shown in Figure 11A. This model is justified by the high similarity of linoleic acid to arachidonic acid, which is present in the crystal structure. This model demonstrates that arachidonic acid protrudes further into the active site than linoleic acid, which support the idea that steric interaction between **23a** and **23d** with the lipid substrate plays a role in the observed enzyme kinetics.

Compound **23a** and **23d** were docked in the free enzyme, which can be justified by the observations that their affinity for the enzyme is higher (**23a**  $K_A$  8.6  $\mu\text{M}$ , **23d**  $K_i$  70  $\mu\text{M}$ ) than the substrate linoleic acid ( $K_m$  600  $\mu\text{M}$ ). Using the structure of h-5-LOX with the arachidonic acid substrate removed, we modeled the binding mode of **23a** and **23d** by molecular docking using *smina*<sup>55</sup> and the resulting poses were reranked using the *vina* scoring function. Despite the large binding site (relative to the size of the compound), we observe a common binding mode for compounds **23a** and **23d** (as well as **23b** and **23c**, not shown). Subsequently, the substrate linoleic acid or arachidonic acid were put into the enzyme as described above. This binding mode of **23a** and **23d** is characterized by a strong *pi*–*pi* stacking interaction between the central phenyl ring of the compound and Phe177 on one side and the hydrophobic side chain of Ala410 on the other (Fig. 11B). The buried polar functionalities of the planar salicylate head group may be stabilized by a hydrogen bond between the carboxylic –OH group of the head group and the side chain of His372. Also of importance is the hydrogen bond between the backbone nitrogen of Phe177 and the ether oxygen of the hydrophobic tail (Fig. 11C). The dominant docking poses of compounds **23a–d** share a similar geometry. A few of the low score poses have the salicylate head group rotated by 180 degrees. Compounds **23a–d** are identical except for the length of their alkoxy tails (Table 1). The only



**Figure 11.** Docking models of Compounds **23 a,b,c,d** in a model of linoleic acid bound to h-5-LOX. (A) Model of bound linoleic acid (yellow) based on a superposition to the configuration of arachidonic acid in the crystal structure (blue). Front (B) and back (C) view of the highest ranking binding mode of compound **23a** (green) and compound **23d** (orange) in the linoleic acid bound active site (A). Yellow dashed lines indicate the hydrogen bond between the ether oxygen and the backbone nitrogen of Phe177. Black dashed lines indicate hydrophobic interactions. (D) Superposition of highest ranked docked configurations of compounds **23a** (green), **23b** (purple), **23c** (dark blue), **23d** (orange) show that larger carbon tails block larger portions of the entrance to the catalytic site. Linoleic acid (yellow) and arachidonic acid (blue) are also shown as references.

differences in the full series of docked conformations is in the long flexible alkoxy tails of **23c–d** that can fit in a variety of conformations with almost identical energy scores (Fig. 11D).

The models of **23a–d** (Fig. 11D) are consistent with our experimental observations. Figure 1 and Table 1 demonstrate that the h-5-LOX enzymatic activity decreases with an increase in the alkoxy tail length. In addition, it was observed that **23a** and **23d** influence the linoleic acid conversion and not the arachidonic acid conversion by h-5-LOX (Fig. 10).

Our docking experiments suggest that **23a** binds alongside the substrate linoleic acid and stabilizes its binding and enhances the enzyme activity. Compound **23d** has a longer tail with multiple binding modes that are likely to block the end of the binding site, hindering the entry of linoleic acid and thus reducing enzymatic activity (Fig. 11C). Thus, the shorter linoleic acid substrate leaves space for the alkoxy tails of **23a** and **23d** (Fig. 11C). This model would explain the high affinity of **23a** for the linoleic acid bound enzyme ( $\alpha K_A$  0.38  $\mu\text{M}$ ) compared to the free enzyme ( $K_A$  8.7  $\mu\text{M}$ ) and it is also in line with a higher conversion rate ( $\beta$  1.76). Perhaps more interestingly, the model of **23d** indicates that an increase in 'bulk' for the alkoxy tail provides inhibition of the enzyme activity for linoleic acid in contrast to the observed activation by **23a**. Namely, the increased 'bulk' in **23d** must interfere with the

enzymatic activity for linoleic acid by disrupting the necessary binding pose or by blocking the entrance to the active site, which limits the access of linoleic acid. Conversely, the larger arachidonic acid might not permit binding of **23a** or **23d** and thus its enzymatic activity is unaffected.

Taking all these data together we propose a model to explain our observations. We conclude that compound **23a** is an activator for linoleic acid conversion by h-5-LOX, presumably because it stabilizes the substrate–enzyme complex without blocking the entrance to the active site. On the other hand, compound **23d** could bind in a comparable configuration but its larger carbon tail, presumably, blocks one end of the active site which interferes with the diffusion of the substrate and molecular oxygen thus overriding the stabilization effect. Although these models provide plausible explanations for the observed behavior of this compound class, we realize crystallography studies would provide the ultimate experimental proof for the mode of action of these molecules.

### 3. Conclusion

Anacardic acid derivatives show interesting structure–activity relationships that demonstrate that these compounds can either activate or inhibit the activity of human-5-LOX. Compound **23a**

and **23d** either activate or inhibit the activity of human 5-LOX toward linoleic acid as a substrate depending on the substitution in *para*-position of the 2-(benzyloxy)-6-hydroxybenzoic acid core. Compound **23a** was identified as a non-essential activator of human 5-lipoxygenase with  $K_A$  is 8.65  $\mu$ M,  $\alpha$  is 0.0438, and  $\beta$  is 1.76. The affinity of activator **23a** to the substrate bound enzyme ( $\alpha K_A$ ) is 0.38  $\mu$ M, which is 50 times higher than to the free enzyme. Furthermore, activator **23a** acts as non-competitive activator against ATP, which indicates the presence of an allosteric site that is different from the ATP binding site. Kinetic studies on inhibitor **23d** show a mixed type inhibition, which indicates the presence of an allosteric regulatory site for human 5-LOX activity. Both modulators **23a** and **23d** demonstrated no effect on the conversion of the substrate arachidonic acid by h-5-LOX. Molecular modeling studies indicate that **23a** and **23d** bind close to the substrate linoleic acid in the active site of h-5-LOX, but that these modulators do not fit simultaneously in the active site with the arachidonic acid, which is slightly larger than linoleic acid.

## 4. Experimental

### 4.1. Chemistry

#### 4.1.1. General

All reagents and solvents were purchased from commercial suppliers (Fluka, Sigma–Aldrich, Acros Organics) and were used without further purification. Dichloromethane was distilled over CaH<sub>2</sub> before use. Merck silica gel 60 F<sub>254</sub> plates were used for analytical thin layer chromatography (TLC) and spots were detected by UV light, or stained using KMnO<sub>4</sub> solution. Column chromatography was performed with MP Ecochrom silica gel 32–63, 60 Å using the flash chromatography technique. <sup>1</sup>H (200 MHz) and <sup>13</sup>C (50 MHz) NMR spectra were recorded on a Varian Gemini 200 spectrometer. <sup>13</sup>C NMR spectra were recorded using the attached proton test (APT). Chemical shifts are reported in ppm ( $\delta$ ) relative to the solvent. Atmospheric Pressure Photoionization mass spectra (APPI-MS) and electrospray ionization (ESI) were recorded on an Applied Biosystems/SCIEX API3000-triple quadrupole mass spectrometer. High-resolution mass spectra (HR-MS) were recorded using a flow injection method on a LTQ-Orbitrap XL mass spectrometer (Thermo Electron, Bremen, Germany) with a resolution of 60,000 at  $m/z$  400. Protonated testosterone (lock mass  $m/z$  = 289.2162) was used for internal recalibration in real time.

*N,N,N,N'*-Tetramethyl-*p*-phenylenediamine (TMPD) hydrochloride, human 5-lipoxygenase enzyme, and COX-2 (human recombinant) were obtained from Cayman Chemicals. Adenosine triphosphate (ATP), arachidonic acid, and linoleic acid were obtained from Sigma–Aldrich. Raw Blue cells assay was obtained from InvivoGen. The MTS assay kit was obtained from Promega. Interferon  $\gamma$  (IFN- $\gamma$ ) was obtained from PeproTech.

#### 4.1.2. Synthetic procedure 1: synthesis of alkoxy benzyl alcohol

1-Bromo alkane (13 mmol) was added to a light yellow suspension of hydroxybenzaldehyde (1.1 g, 9.0 mmol) and K<sub>2</sub>CO<sub>3</sub> (5.2 g, 37 mmol) in 45 mL DMF under nitrogen atmosphere. The suspension was stirred overnight at room temperature. The reaction mixture was diluted with demiwat (100 mL) and extracted with ethylacetate (3  $\times$  80 mL). The combined organic layers were washed with demiwat (4  $\times$  80 mL), 0.1 M aqueous HCl (100 mL), brine (2  $\times$  30 mL), dried with MgSO<sub>4</sub>, filtered and concentrated in vacuo.

NaBH<sub>4</sub> (0.17 g, 4.5 mmol) was added to a solution of benzaldehyde (8.0 mmol) in 30 mL methanol at 0 °C under nitrogen atmosphere. The reaction mixture was stirred for 1 h at 0 °C and then

warmed to room temperature. The reaction mixture was diluted with demiwat (80 mL) and extracted with ethylacetate (2  $\times$  60 mL). The combined organic layers were washed with demiwat (2  $\times$  40 mL), brine (2  $\times$  40 mL), dried with MgSO<sub>4</sub>, filtered and concentrated in vacuo. The product was obtained after crystallization or purification using column chromatography.

#### 4.1.3. Synthetic procedure 2: bromination of alkoxy benzyl alcohol

Phosphorus tribromide (PBr<sub>3</sub>) (0.34 mL, 3.6 mmol) was added slowly to a suspension of alkoxy benzyl alcohol (2.5 mmol) in 10 mL CH<sub>2</sub>Cl<sub>2</sub> under nitrogen atmosphere at 0 °C to give an orange suspension. The reaction mixture was stirred for 1.5 h at 0 °C. The reaction mixture was poured into 80 mL ice water, and extracted with ethylacetate (3  $\times$  50 mL). The combined organic layers were washed with demiwat (50 mL), 0.1 M aqueous HCl (2  $\times$  30 mL), brine (50 mL), dried with MgSO<sub>4</sub>, filtered and concentrated in vacuo. The aryl bromide product was obtained after purification using column chromatography.

#### 4.1.4. Synthetic procedure 3: Williamson ether synthesis coupling of aryl bromide

5-Hydroxy-2,2-dimethyl-4*H*-benzo[*d*][1,3]dioxin-4-one **1** (0.22 g, 1.1 mmol) and K<sub>2</sub>CO<sub>3</sub> (0.63 g, 4.6 mmol) were suspended in 8 mL DMF under nitrogen atmosphere. A suspension of aryl bromide in 7 mL DMF was added to the suspension. The mixture was stirred overnight at room temperature and was diluted with demiwat (60 mL) and extracted with ethylacetate (3  $\times$  45 mL). The combined organic layers were washed with demiwat (2  $\times$  40 mL), 0.1 M aqueous HCl (40 mL), brine (2  $\times$  20 mL), dried over MgSO<sub>4</sub>, filtered and concentrated in vacuo. The product was obtained after purification using column chromatography.

#### 4.1.5. Synthetic procedure 4: saponification of the acetone

5 M KOH (0.16 mL, 0.80 mmol) was added to a solution of 2,2-dimethyl-5-((4-(alkoxy)benzyl)oxy)-4*H*-benzo[*d*][1,3]dioxin-4-one (0.15 g, 0.40 mmol) in 10 mL THF. The reaction mixture was heated till 60 °C and stirred overnight affording an orange suspension. The reaction mixture was concentrated in vacuo yielding the product as potassium salt.

#### 4.1.6. Synthetic procedure 5: hydrolysis of acetone

5 M KOH (1.2 mL, 6 mmol) was added to a colorless solution of acetone (0.60 mmol) in 10 mL THF. The reaction mixture was stirred for 24 h at 62 °C. The reaction mixture was diluted and acidified with 1 N aqueous HCl (2.6 mL) and extracted with ethylacetate (2  $\times$  40 mL). The combined organic layers were washed with water (50 mL), brine (2  $\times$  30 mL), dried over MgSO<sub>4</sub>, filtered and concentrated in vacuo.

#### 4.1.7. (4-(Pentyloxy)phenyl)methanol (2a)

The product was obtained using Williamson ether synthesis followed by aldehyde reduction of 1-bromopentane, and 4-hydroxybenzaldehyde using synthetic procedure 1. The product was purified using column chromatography with heptane/EtOAc 5:1 (v/v) as an eluent. Yield 73%. Light orange solid.  $R_f$  = 0.41 (heptane/EtOAc 1:1). <sup>1</sup>H NMR (200 MHz, CDCl<sub>3</sub>)  $\delta$  7.40–7.18 (m, 2H), 6.97–6.79 (m, 2H), 4.59 (s, 2H), 3.95 (t,  $J$  = 6.6, 2H), 1.87 (s, 1H), 1.86–1.72 (m, 2H), 1.46–1.34 (m, 4H), 0.94 (t,  $J$  = 7.0, 3H). <sup>13</sup>C NMR (50 MHz, CDCl<sub>3</sub>)  $\delta$  158.47, 133.10, 128.80, 114.73, 68.24, 65.20, 29.14, 28.38, 22.65, 14.20. MS (APPI):  $m/z$  194.2 [M<sup>+</sup>].

#### 4.1.8. 1-(Bromomethyl)-4-(pentyloxy)benzene (3a)

The product was obtained from (4-(pentyloxy)phenyl)methanol **2a** using synthetic procedure 2. The product was obtained in high purity and no further purification was required. Yield 97%. Yellow



solid.  $R_f = 0.72$  (heptane/EtOAc 1:1).  $^1\text{H}$  NMR (200 MHz,  $\text{CDCl}_3$ )  $\delta$  7.38–7.22 (m, 2H), 6.94–6.74 (m, 2H), 4.51 (s, 2H), 3.94 (t,  $J = 6.2$ , 2H), 1.91–1.65 (m, 2H), 1.59–1.21 (m, 4H), 0.94 (t,  $J = 7.1$ , 3H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CDCl}_3$ )  $\delta$  159.47, 130.60, 129.88, 114.94, 68.26, 34.30, 29.11, 28.38, 22.65, 14.22. MS (ESI):  $m/z$  177.1  $[\text{M}-\text{Br}]^+$ .

#### 4.1.9. 2,2-Dimethyl-5-((4-(pentyloxy)benzyl)oxy)-4H-benzo[d][1,3]dioxin-4-one (12a)

The product was obtained from 5-hydroxy-2,2-dimethyl-4H-benzo[d][1,3]dioxin-4-one **1**, and 1-(bromomethyl)-4-(pentyloxy)benzene **3a** using synthetic procedure 3. The product was crystallized from iso-propanol. Yield 70%. White solid.  $R_f = 0.6$  (heptane/EtOAc 1:1).  $^1\text{H}$  NMR (200 MHz,  $\text{CDCl}_3$ )  $\delta$  7.50–7.32 (m, 3H), 6.90 (d,  $J = 8.3$ , 2H), 6.65 (d,  $J = 8.5$ , 1H), 6.54 (d,  $J = 8.2$ , 1H), 5.18 (s, 2H), 3.95 (t,  $J = 6.6$ , 2H), 1.81–1.70 (m, 8H), 1.46–1.40 (m, 4H), 0.93 (t,  $J = 7.1$ , 3H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CDCl}_3$ )  $\delta$  160.69, 159.10, 157.99, 136.40, 128.53, 128.26, 114.82, 109.57, 107.62, 105.41, 105.22, 104.17, 70.89, 68.23, 29.17, 28.41, 25.86, 22.67, 14.23. MS (APPI):  $m/z$  370.2  $[\text{M}^+]$ .

#### 4.1.10. Potassium 2-hydroxy-6-((4-(pentyloxy)benzyl)oxy)benzoate (23a)

The product was obtained from 2,2-dimethyl-5-((4-(pentyloxy)benzyl)oxy)-4H-benzo[d][1,3]dioxin-4-one **12a** using synthetic procedure 4. Yield quantitatively determined from the final yield of product in excess of KOH. Orange gum.  $R_f = 0.16$  (heptane/EtOAc 1:1 containing 1% acetic acid). The product was more than 95% pure as judged from TLC and NMR.  $^1\text{H}$  NMR (200 MHz,  $\text{D}_2\text{O}$ )  $\delta$  7.15 (d,  $J = 8.5$ , 2H), 6.89 (t,  $J = 8.3$ , 1H), 6.68 (d,  $J = 8.5$ , 2H), 6.34 (d,  $J = 8.3$ , 1H), 6.27 (d,  $J = 8.3$ , 1H), 4.91 (s, 2H), 3.68 (t,  $J = 6.5$ , 2H), 1.50–1.47 (m, 2H), 1.14–1.11 (m, 4H), 0.70 (t,  $J = 7.0$ , 3H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{D}_2\text{O}$ )  $\delta$  195.90, 159.11, 158.08, 157.85, 131.54, 129.50, 128.81, 114.48, 111.31, 109.15, 105.52, 70.26, 68.25, 28.21, 27.56, 21.89, 13.38. HRMS:  $m/z$   $[\text{M}-\text{H}]^-$  calcd for  $\text{C}_{19}\text{H}_{21}\text{O}_5$  329.13945, found 329.13995.

#### 4.1.11. (4-(Heptyloxy)phenyl)methanol (2b)

The product was obtained using Williamson ether synthesis followed by aldehyde reduction of 1-bromoheptane, and 4-hydroxybenzaldehyde using synthetic procedure 1. The product was crystallized from heptane. Yield 87%. Colorless to white crystal.  $R_f = 0.44$  (heptane/EtOAc 1:1).  $^1\text{H}$  NMR (200 MHz,  $\text{CDCl}_3$ )  $\delta$  7.49–7.16 (m, 2H), 6.99–6.79 (m, 2H), 4.59 (s, 2H), 3.95 (t,  $J = 6.6$ , 2H), 1.83 (s, 1H), 1.81–1.71 (m, 2H), 1.55–1.19 (m, 8H), 0.89 (t,  $J = 6.4$ , 3H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CDCl}_3$ )  $\delta$  158.97, 133.09, 128.81, 114.75, 68.27, 65.22, 31.98, 29.46, 29.26, 26.20, 22.80, 14.28. MS (APPI):  $m/z$  222.2  $[\text{M}^+]$ .

#### 4.1.12. 1-(Bromomethyl)-4-(heptyloxy)benzene (3b)

The product was obtained from (4-(heptyloxy)phenyl)methanol **2b** using synthetic procedure 2. The product was obtained in high purity and no further purification was required. Yield 97%. Yellow oil.  $R_f = 0.71$  (heptane/EtOAc 1:1).  $^1\text{H}$  NMR (200 MHz,  $\text{CDCl}_3$ )  $\delta$  7.37–7.27 (m, 2H), 6.91–6.75 (m, 2H), 4.51 (s, 2H), 3.95 (t,  $J = 6.5$ , 2H), 1.91–1.68 (m, 2H), 1.57–1.18 (m, 8H), 0.91 (t,  $J = 6.5$ , 3H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CDCl}_3$ )  $\delta$  159.47, 130.60, 129.87, 114.94, 68.27, 34.30, 31.98, 29.42, 29.25, 26.19, 22.81, 14.29. MS (ESI):  $m/z$  205.1  $[\text{M}-\text{Br}]^+$ .

#### 4.1.13. 5-((4-(Heptyloxy)benzyl)oxy)-2,2-dimethyl-4H-benzo[d][1,3]dioxin-4-one (12b)

The product was obtained from 5-hydroxy-2,2-dimethyl-4H-benzo[d][1,3]dioxin-4-one **1**, and 1-(bromomethyl)-4-(heptyloxy)benzene **3b** using synthetic procedure 3. The product was purified using column chromatography with heptane/EtOAc 11:1 (v/v) as an eluent. Yield 78%. White solid.  $R_f = 0.41$  (heptane/EtOAc

3:1).  $^1\text{H}$  NMR (200 MHz,  $\text{CDCl}_3$ )  $\delta$  7.50–7.32 (m, 3H), 6.90 (d,  $J = 8.6$ , 2H), 6.90 (d,  $J = 8.4$ , 1H), 6.65 (d,  $J = 8.4$ , 1H), 5.18 (s, 2H), 3.94 (t,  $J = 6.5$ , 2H), 1.85–1.65 (m, 8H), 1.40–1.30 (m, 8H), 0.89 (t,  $J = 6.6$ , 3H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CDCl}_3$ )  $\delta$  160.68, 159.11, 158.23, 157.99, 136.40, 128.53, 128.25, 114.82, 109.57, 107.61, 105.41, 104.17, 70.88, 68.25, 31.99, 29.48, 29.27, 26.21, 25.85, 22.82, 14.29. MS (APPI):  $m/z$  398.2  $[\text{M}^+]$ .

#### 4.1.14. Potassium 2-((4-(heptyloxy)benzyl)oxy)-6-hydroxybenzoate (23b)

The product was obtained from 5-((4-(heptyloxy)benzyl)oxy)-2,2-dimethyl-4H-benzo[d][1,3]dioxin-4-one **12b** using synthetic procedure 4. Yield quantitatively determined from the final yield of product in excess of KOH. Orange solid.  $R_f = 0.49$  (heptane/EtOAc 2:1 containing 1% acetic acid). The product was more than 95% pure as judged from TLC and NMR.  $^1\text{H}$  NMR (200 MHz,  $\text{D}_2\text{O}$ )  $\delta$  6.93 (m, 2H), 6.60 (m, 1H), 6.42 (m, 2H), 6.18 (m, 1H), 5.94 (m, 1H), 4.74 (s, 2H), 3.38 (m, 2H), 1.31–1.06 (m, 10H), 0.68 (t,  $J = 6.6$ , 3H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  202.2, 179.61, 161.89, 161.16, 160.88, 133.19, 131.99, 130.65, 116.15, 111.26, 106.62, 105.29, 72.48, 69.85, 33.85, 31.30, 31.08, 28.01, 24.53, 15.27. HRMS:  $m/z$   $[\text{M}-\text{H}]^-$  calcd for  $\text{C}_{21}\text{H}_{25}\text{O}_5$  357.17075, found 357.17109.

#### 4.1.15. (4-(Nonyloxy)phenyl)methanol (2c)

The product was obtained using Williamson ether synthesis followed by aldehyde reduction of 1-bromononane, and 4-hydroxybenzaldehyde using synthetic procedure 1. The product was crystallized from heptane. Yield 73%. White crystal.  $R_f = 0.44$  (heptane:ethylacetate 1:1).  $^1\text{H}$  NMR (200 MHz,  $\text{CDCl}_3$ )  $\delta$  7.35–7.20 (m, 2H), 6.95–6.82 (m, 2H), 4.60 (s, 2H), 3.95 (t,  $J = 6.6$ , 2H), 1.85–1.74 (m, 2H), 1.71 (s, 1H), 1.42–1.29 (m, 12H), 0.89 (t,  $J = 7.0$ , 3H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CDCl}_3$ )  $\delta$  158.99, 133.09, 128.82, 114.76, 68.28, 65.27, 32.08, 29.74, 29.61, 29.47, 26.24, 22.87, 14.31. MS (APPI):  $m/z$  250.2  $[\text{M}^+]$ .

#### 4.1.16. 1-(Bromomethyl)-4-(nonyloxy)benzene (3c)

The product was obtained from (4-(nonyloxy)phenyl)methanol **2c** using synthetic procedure 2. The product was obtained in high purity and no further purification was required. Yield 94%. Light yellow oil.  $R_f = 0.74$  (heptane/EtOAc 1:1).  $^1\text{H}$  NMR (200 MHz,  $\text{CDCl}_3$ )  $\delta$  7.39–7.20 (m, 2H), 6.94–6.76 (m, 2H), 4.50 (s, 2H), 3.95 (t,  $J = 6.5$ , 2H), 1.88–1.66 (m, 2H), 1.54–1.14 (m, 12H), 0.89 (t,  $J = 6.4$ , 3H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CDCl}_3$ )  $\delta$  159.48, 130.61, 129.88, 114.96, 68.29, 34.31, 32.09, 29.74, 29.59, 29.47, 29.42, 26.23, 22.88, 14.32. MS (ESI):  $m/z$  233.1  $[\text{M}-\text{Br}]^+$ .

#### 4.1.17. 2,2-Dimethyl-5-((4-(nonyloxy)benzyl)oxy)-4H-benzo[d][1,3]dioxin-4-one (12c)

The product was obtained from 5-hydroxy-2,2-dimethyl-4H-benzo[d][1,3]dioxin-4-one **1**, and 1-(bromomethyl)-4-(nonyloxy)benzene **3c** using synthetic procedure 3. The product was purified using column chromatography with heptane/EtOAc 15:1 (v/v) as an eluent. Yield 85%. White solid.  $R_f = 0.48$  (heptane/EtOAc 3:1).  $^1\text{H}$  NMR (200 MHz,  $\text{CDCl}_3$ )  $\delta$  7.53–7.29 (m, 3H), 6.90 (d,  $J = 8.4$ , 2H), 6.65 (d,  $J = 8.4$ , 1H), 6.54 (d,  $J = 8.4$ , 1H), 5.18 (s, 2H), 3.94 (t,  $J = 6.5$ , 2H), 1.88–1.64 (m, 8H), 1.36 (d,  $J = 30.5$ , 12H), 0.88 (t,  $J = 6.2$ , 3H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CDCl}_3$ )  $\delta$  202.3, 160.67, 159.09, 158.20, 157.98, 136.39, 128.52, 128.24, 114.81, 109.56, 107.60, 105.40, 102.26, 70.87, 68.24, 32.08, 29.74, 29.61, 29.47, 26.25, 25.85, 22.87, 14.35, 14.31. MS (APPI):  $m/z$  426.2  $[\text{M}^+]$ .

#### 4.1.18. Potassium 2-hydroxy-6-((4-(nonyloxy)benzyl)oxy)benzoate (23c)

The product was obtained from 2,2-dimethyl-5-((4-(nonyloxy)benzyl)oxy)-4H-benzo[d][1,3]dioxin-4-one **12c** using syn-

thetic procedure 4. Yield quantitatively determined from the final yield of product in excess of KOH.  $R_f = 0.18$  (heptane/EtOAc 1:1 containing 1% acetic acid). The product was more than 95% pure as judged from TLC and NMR.  $^1\text{H}$  NMR (200 MHz,  $\text{D}_2\text{O}$ )  $\delta$  6.95 (d,  $J = 8.3$ , 2H), 6.57 (t,  $J = 8.5$ , 1H), 6.43 (d,  $J = 8.3$ , 2H), 6.16 (d,  $J = 8.5$ , 1H), 5.91 (d,  $J = 8.5$ , 1H), 4.75 (s, 2H), 3.38 (m, 2H), 1.34 (m, 2H), 1.06 (m, 12H), 0.71 (t,  $J = 7.0$ , 3H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  202.2, 164.24, 160.76, 156.88, 133.10, 131.98, 130.67, 116.19, 111.48, 109.33, 106.36, 72.41, 69.89, 61.77, 33.83, 31.47, 31.31, 31.18, 28.10, 27.95, 24.52, 14.22. HRMS:  $m/z$   $[\text{M}-\text{H}]^-$  calcd for  $\text{C}_{23}\text{H}_{29}\text{O}_5$  385.20205, found 385.20248.

#### 4.1.19. (4-((3,7-Dimethyloctyl)oxy)phenyl)methanol (2d)

The product was obtained using Williamson ether synthesis followed by aldehyde reduction of 1-1-bromo-3,7-dimethyloctane, and 4-hydroxybenzaldehyde using synthetic procedure 1. The product was used for the next reaction step without further purification. Yield 92%. Yellow oil.  $R_f = 0.48$  (heptane/EtOAc 1:1).  $^1\text{H}$  NMR (200 MHz,  $\text{CDCl}_3$ )  $\delta$  7.27 (d,  $J = 8.6$ , 2H), 6.88 (d,  $J = 8.6$ , 2H), 4.59 (s, 2H), 3.98 (t,  $J = 6.5$ , 2H), 1.88–1.46 (m, 5H), 1.32–1.14 (m, 6H), 0.93 (d,  $J = 6.2$ , 3H), 0.87 (d,  $J = 6.5$ , 6H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CDCl}_3$ )  $\delta$  158.98, 133.10, 128.82, 114.77, 66.59, 65.27, 39.45, 37.49, 36.40, 30.06, 28.17, 24.86, 22.91, 22.81, 19.86. MS (APPI):  $m/z$  264.1  $[\text{M}^+]$ .

#### 4.1.20. 1-(Bromomethyl)-4-((3,7-dimethyloctyl)oxy)benzene (3d)

The product was obtained from (4-((3,7-dimethyloctyl)oxy)phenyl)methanol **2d** using synthetic procedure 2. The product was used for the next reaction step without further purification. Yield quantitative. Light yellow oil.  $R_f = 0.75$  (heptane/EtOAc 1:1).  $^1\text{H}$  NMR (200 MHz,  $\text{CDCl}_3$ )  $\delta$  7.30 (d,  $J = 8.6$ , 2H), 6.85 (d,  $J = 8.6$ , 2H), 4.49 (s, 2H), 3.98 (t,  $J = 6.5$ , 2H), 1.87–1.46 (m, 4H), 1.32–1.10 (m, 6H), 0.93 (d,  $J = 6.2$ , 3H), 0.87 (d,  $J = 6.4$ , 6H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CDCl}_3$ )  $\delta$  159.46, 130.60, 129.88, 114.95, 66.59, 39.45, 37.47, 36.35, 34.30, 30.04, 28.18, 24.86, 22.92, 22.82, 19.85. MS (ESI):  $m/z$  247.1  $[\text{M}-\text{Br}]^+$ .

#### 4.1.21. 5-((4-((3,7-Dimethyloctyl)oxy)benzyl)oxy)-2,2-dimethyl-4H-benzo[d][1,3]dioxin-4-one (12d)

The product was obtained from 5-hydroxy-2,2-dimethyl-4H-benzo[d][1,3]dioxin-4-one **1**, and 1-(bromomethyl)-4-((3,7-dimethyloctyl)oxy)benzene **3d** using synthetic procedure 3. The product was purified using column chromatography with heptane/EtOAc 19:1 (v/v) as eluent. Yield 68%. White solid.  $R_f = 0.35$  (heptane/EtOAc 4:1).  $^1\text{H}$  NMR (200 MHz,  $\text{CDCl}_3$ )  $\delta$  7.46–7.26 (m, 3H), 6.92 (d,  $J = 8.6$ , 2H), 6.65 (d,  $J = 8.5$ , 1H), 6.54 (d,  $J = 8.5$ , 1H), 5.18 (s, 2H), 3.98 (t,  $J = 6.5$ , 2H), 1.84–1.5 (m, 10H), 1.32–1.14 (m, 6H), 0.93 (d,  $J = 6.2$ , 3H), 0.87 (d,  $J = 6.4$ , 6H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CDCl}_3$ )  $\delta$  160.69, 159.10, 158.23, 157.99, 136.40, 128.53, 128.25, 114.82, 109.57, 107.61, 105.41, 70.89, 66.56, 39.45, 37.49, 36.40, 30.05, 28.17, 25.85, 24.85, 22.91, 22.81, 19.86. MS (APPI):  $m/z$  440.4  $[\text{M}^+]$ .

#### 4.1.22. Potassium 2-((4-((3,7-dimethyloctyl)oxy)benzyl)oxy)-6-hydroxybenzoate (23d)

The product was obtained from 5-((4-((3,7-dimethyloctyl)oxy)benzyl)oxy)-2,2-dimethyl-4H-benzo[d][1,3]dioxin-4-one **12d** using synthetic procedure 4. The compound is racemic at its asymmetric carbon. Yield quantitatively determined from the final yield of product in excess of KOH. Orange solid.  $R_f = 0.54$  (heptane/EtOAc 1:1 containing 1% acetic acid). The product was more than 95% pure as judged from TLC and NMR.  $^1\text{H}$  NMR (200 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  7.42 (d,  $J = 8.7$ , 2H), 7.05 (t,  $J = 8.2$ , 1H), 6.95–6.85 (m, 2H), 6.43–6.38 (m, 2H), 5.03 (s, 2H), 3.99 (t,  $J = 6.7$ , 2H), 1.92–1.48 (m, 4H), 1.40–1.16 (m, 6H), 0.95 (d,  $J = 6.2$ , 3H), 0.89 (d,  $J = 6.5$ , 6H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  202.3, 164.42, 161.42, 160.86, 132.74,

132.06, 130.69, 116.17, 111.76, 105.95, 72.47, 68.11, 41.28, 39.28, 38.27, 31.89, 29.99, 26.67, 23.95, 23.86, 20.90. HRMS:  $m/z$   $[\text{M}-\text{H}]^-$  calcd for  $\text{C}_{24}\text{H}_{31}\text{O}_5$  399.2177, found 399.21844.

#### 4.1.23. (3-(Heptyloxy)phenyl)methanol (4b)

The product was obtained using Williamson ether synthesis followed by aldehyde reduction of 1-bromoheptane, and 3-hydroxybenzaldehyde using synthetic procedure 1. The product was used for the next reaction step without further purification. Yield 94%. Brown solid.  $R_f = 0.48$  (heptane/EtOAc 1:1).  $^1\text{H}$  NMR (200 MHz,  $\text{CDCl}_3$ )  $\delta$  7.26–7.18 (m, 1H), 6.96–6.75 (m, 3H), 4.64 (s, 2H), 3.95 (t,  $J = 6.5$ , 2H), 1.88–1.67 (m, 3H), 1.55–1.17 (m, 8H), 0.89 (t,  $J = 6.5$ , 3H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CDCl}_3$ )  $\delta$  159.62, 142.67, 129.74, 119.09, 114.01, 113.09, 68.18, 65.48, 31.99, 29.49, 29.27, 26.22, 22.82, 14.34. MS (APPI):  $m/z$  222.2  $[\text{M}^+]$ .

#### 4.1.24. 1-(Bromomethyl)-3-(heptyloxy)benzene (5b)

The product was obtained from (3-(heptyloxy)phenyl)methanol **4b** using synthetic procedure 2. Therefore product was used for the next reaction step without further purification. Yield 86%. Brown solid.  $R_f = 0.71$  (heptane/EtOAc 1:1).  $^1\text{H}$  NMR (200 MHz,  $\text{CDCl}_3$ )  $\delta$  7.32–6.27 (m, 4H), 4.76 (s, 2H), 3.92 (t,  $J = 6.4$ , 2H), 1.72–1.65 (m, 2H), 1.34–1.15 (m, 8H), 0.89 (t,  $J = 6.5$ , 3H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CDCl}_3$ )  $\delta$  159.41, 129.63, 119.81, 114.36, 105.23, 68.05, 32.65, 32.01, 29.53, 29.36, 26.24, 22.84, 14.31. MS (ESI):  $m/z$  285.0  $[\text{M}^+]$ .

#### 4.1.25. 5-((3-(Heptyloxy)benzyl)oxy)-2,2-dimethyl-4H-benzo[d][1,3]dioxin-4-one (13b)

The product was obtained from 5-hydroxy-2,2-dimethyl-4H-benzo[d][1,3]dioxin-4-one **1**, and 1-(bromomethyl)-3-(heptyloxy)benzene **5b** using synthetic procedure 3. The product was purified using column chromatography with heptane/EtOAc 16:1 (v/v) as eluent. Yield 58%. White solid.  $R_f = 0.29$  (heptane/EtOAc 5:1).  $^1\text{H}$  NMR (200 MHz,  $\text{CDCl}_3$ )  $\delta$  7.42–7.35 (m, 2H), 7.25 (d,  $J = 7.5$ , 1H), 7.04 (d,  $J = 7.5$ , 1H), 6.83 (d,  $J = 7.9$ , 1H), 6.70–6.49 (m, 2H), 5.22 (s, 2H), 3.99 (t,  $J = 6.4$ , 2H), 1.87–1.68 (m, 2H), 1.56 (s, 6H), 1.42–1.12 (m, 8H), 0.89 (t,  $J = 6.5$ , 3H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CDCl}_3$ )  $\delta$  160.53, 159.77, 158.23, 158.01, 138.02, 136.48, 129.73, 118.57, 114.61, 112.43, 109.67, 107.36, 105.46, 70.74, 68.19, 32.00, 29.50, 29.29, 26.26, 25.85, 22.82, 14.30. MS (APPI):  $m/z$  399.2  $[\text{M}+\text{H}]^+$ .

#### 4.1.26. 2-((3-(Heptyloxy)benzyl)oxy)-6-hydroxybenzoic acid (24b)

The product was obtained from 5-((3-(heptyloxy)benzyl)oxy)-2,2-dimethyl-4H-benzo[d][1,3]dioxin-4-one **13b** using synthetic procedure 5. Yield quantitative. Yellow oil.  $R_f = 0.42$  (heptane/EtOAc 1:1 containing 1% acetic acid). The product was more than 95% pure as judged from TLC and NMR.  $^1\text{H}$  NMR (200 MHz,  $\text{DMSO}-d_6$ )  $\delta$  7.26–7.19 (m, 2H), 7.08 (d,  $J = 7.5$ , 1H), 7.04 (d,  $J = 7.5$ , 1H), 6.79 (d,  $J = 8.11\text{H}$ ), 6.31 (d,  $J = 8.2$ , 2H), 5.04 (s, 2H), 3.95 (t,  $J = 6.4$ , 2H), 1.73–1.63 (m, 2H), 1.36–1.18 (m, 8H), 0.87 (t,  $J = 7.0$ , 3H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{DMSO}-d_6$ )  $\delta$  170.28, 164.46, 159.34, 158.66, 139.68, 130.82, 129.05, 118.77, 113.03, 110.05, 109.85, 102.77, 69.61, 67.26, 31.26, 28.75, 28.48, 25.55, 22.07, 13.96. HRMS:  $m/z$   $[\text{M}-\text{H}]^-$  calcd for  $\text{C}_{21}\text{H}_{25}\text{O}_5$  357.17075, found 357.17114.

#### 4.1.27. (3-(Nonyloxy)phenyl)methanol (4c)

The product was obtained using Williamson ether synthesis followed by aldehyde reduction of 1-bromoheptane, and 3-hydroxybenzaldehyde using synthetic procedure 1. The product was used for the next reaction step without further purification. Yield 89%.  $R_f = 0.54$  (heptane/EtOAc 1:1).  $^1\text{H}$  NMR (200 MHz,  $\text{CDCl}_3$ )  $\delta$  7.22 (t,  $J = 7.8$ , 1H), 6.89 (s, 1H), 6.85–6.72 (m, 2H), 4.59 (s, 2H), 3.92 (t,  $J = 6.5$ , 2H), 2.32 (s, 1H), 1.79–1.69 (m, 2H), 1.42–1.28 (m,

12H), 0.88 (t,  $J = 6.3$ , 3H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CDCl}_3$ )  $\delta$  159.52, 142.67, 129.62, 119.04, 113.88, 113.02, 68.11, 65.25, 32.05, 29.72, 29.58, 29.44, 26.22, 22.84, 14.26. MS (APPI):  $m/z$  250.2 [ $\text{M}^+$ ].

#### 4.1.28. 1-(Bromomethyl)-3-(nonyloxy)benzene (5c)

The product was obtained from 3-(nonyloxy)phenylmethanol **4c** using synthetic procedure 2. Therefore the product was used for the next reaction step without further purification. Yield 78%. Yellow oil.  $R_f = 0.71$  (heptane/EtOAc 1:1).  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  7.22 (t,  $J = 7.8$ , 1H), 6.96–6.92 (m, 2H), 6.84–6.81 (m, 1H), 4.46 (s, 2H), 3.95 (t,  $J = 7.0$ , 2H), 1.87–1.75 (m, 2H), 1.47–1.28 (m, 12H), 0.89 (t,  $J = 6.8$ , 3H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CDCl}_3$ )  $\delta$  159.33, 139.04, 129.73, 121.03, 115.03, 114.67, 68.02, 33.55, 31.86, 29.52, 29.38, 29.24, 26.02, 22.65, 14.09. MS (ESI):  $m/z$  313.0 [ $\text{M}^+$ ].

#### 4.1.29. 2,2-Dimethyl-5-((3-(nonyloxy)benzyl)oxy)-4H-benzo[d][1,3]dioxin-4-one (13c)

The product was obtained from 5-hydroxy-2,2-dimethyl-4H-benzo[d][1,3]dioxin-4-one **1**, and 1-(bromomethyl)-3-(nonyloxy)benzene **5c** using synthetic procedure 3. The product was purified using column chromatography with heptane/EtOAc 16:1 (v/v) as eluent. Yield 88%. Yellow oil.  $R_f = 0.29$  (heptane/EtOAc 5:1).  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  7.41–7.35 (m, 2H), 7.25–7.20 (m, 2H), 7.04 (d,  $J = 7.0$ , 1H), 6.82 (d,  $J = 6.9$ , 1H), 6.59 (d,  $J = 8.0$ , 1H), 5.21 (s, 2H), 3.99 (t,  $J = 6.0$ , 2H), 1.89–1.62 (m, 8H), 1.51–1.13 (m, 12H), 0.87 (d,  $J = 6.3$ , 3H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CDCl}_3$ )  $\delta$  160.46, 159.71, 157.95, 137.97, 136.45, 129.66, 118.51, 114.55, 112.37, 109.62, 107.30, 105.39, 103.97, 70.68, 68.13, 32.04, 29.71, 29.60, 29.43, 26.26, 25.79, 22.83, 14.28. MS (ESI):  $m/z$  427.3 [ $\text{M}+\text{H}^+$ ].

#### 4.1.30. 2-Hydroxy-6-((3-(nonyloxy)benzyl)oxy)benzoic acid (24c)

The product was obtained from 5-((3-(heptyloxy)benzyl)oxy)-2,2-dimethyl-4H-benzo[d][1,3]dioxin-4-one **13c** using synthetic procedure 5. The product was obtained in high purity and no further purification was required. Yield 58%. Yellow oil.  $R_f = 0.60$  (heptane/EtOAc 1:1 containing 1% acetic acid). The product was more than 95% pure as judged from TLC and NMR.  $^1\text{H}$  NMR (200 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  7.27–7.20 (m, 2H), 7.11–7.01 (m, 2H), 6.81 (d,  $J = 7.8$ , 1H), 6.49 (d,  $J = 7.5$ , 2H), 5.16 (s, 2H), 3.97 (t,  $J = 6.2$ , 2H), 1.75–1.71 (m, 2H), 1.41–1.23 (m, 12H), 0.90 (t,  $J = 6.5$ , 3H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  185.96, 161.65, 156.96, 141.16, 131.25, 121.03, 115.76, 115.03, 107.29, 94.05, 72.58, 69.81, 33.92, 31.56, 31.43, 31.30, 28.07, 24.59, 15.30. HRMS:  $m/z$  [ $\text{M}-\text{H}$ ] $^-$  calcd for  $\text{C}_{23}\text{H}_{29}\text{O}_5$  385.20205, found 385.20261.

#### 4.1.31. 2,2-Dimethyl-5-(nonyloxy)-4H-benzo[d][1,3]dioxin-4-one (11)

The product was obtained from 5-hydroxy-2,2-dimethyl-4H-benzo[d][1,3]dioxin-4-one **1**, and 1-bromononane using synthetic procedure 3. The product was purified using column chromatography with heptane as eluent followed by heptane/EtOAc 50:1 (v/v) as eluent until the starting material was completely removed from the column and the product was obtained by eluting the column with EtOAc as eluent. Yield 53%. Light yellow solid.  $R_f = 0.64$  (heptane/EtOAc 1:1).  $^1\text{H}$  NMR (200 MHz,  $\text{CDCl}_3$ )  $\delta$  7.40 (t,  $J = 8.4$ , 1H), 6.59 (d,  $J = 8.1$ , 1H), 6.51 (d,  $J = 8.2$ , 1H), 4.06 (t,  $J = 6.7$ , 2H), 1.95–1.81 (m, 2H), 1.69 (s, 6H), 1.50–1.26 (m, 12H), 0.85 (t,  $J = 6.6$ , 3H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CDCl}_3$ )  $\delta$  161.30, 158.20, 157.96, 136.41, 109.03, 106.62, 105.28, 103.67, 69.60, 32.07, 29.69, 29.55, 29.46, 29.14, 26.03, 25.84, 22.86, 14.31. MS (APPI):  $m/z$  321.0 [ $\text{M}+\text{H}^+$ ].

#### 4.1.32. 2-Hydroxy-6-(nonyloxy)benzoic acid (22)

The product was obtained from 5-((3-(heptyloxy)benzyl)oxy)-2,2-dimethyl-4H-benzo[d][1,3]dioxin-4-one **11** using synthetic procedure 5. The product was obtained in high purity and no

further purification was required. Yield quantitative. Light brown solid.  $R_f = 0.60$  (heptane/EtOAc 1:1 containing 1% acetic acid). The product was more than 95% pure as judged from TLC and NMR.  $^1\text{H}$  NMR (200 MHz,  $\text{DMSO}-d_6$ )  $\delta$  7.02 (t,  $J = 8.2$ , 1H), 6.28 (d,  $J = 8.0$ , 1H), 6.24 (d,  $J = 8.1$ , 1H), 3.87 (t,  $J = 6.5$ , 2H), 1.68–1.58 (m, 2H), 1.40–1.17 (m, 12H), 0.86 (t,  $J = 6.5$ , 3H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{DMSO}-d_6$ )  $\delta$  164.31, 160.21, 131.19, 110.51, 109.82, 103.07, 69.00, 31.74, 29.48, 29.34, 29.14, 25.92, 22.54, 14.40. HRMS:  $m/z$  [ $\text{M}-\text{H}$ ] $^-$  calcd for  $\text{C}_{16}\text{H}_{23}\text{O}_4$  279.16018, found 279.16061.

#### 4.1.33. 1-(4-Bromophenyl)decan-1-one (7)

The product was obtained using Friedel–Craft acylation of bromobenzene **6** with decanoylchloride.  $\text{AlCl}_3$  (0.81 g, 6.1 mmol) was added slowly to a solution of bromobenzene (0.54 mL, 5.1 mmol) and decanoylchloride (1.2 mL, 5.8 mmol) in 20 mL  $\text{CH}_2\text{Cl}_2$  in a 50 mL flask at 0 °C.  $\text{AlCl}_3$  slowly dissolved in the mixture and the reaction mixture was stirred for 1 hour at 60 °C. The reaction mixture was poured into ice-water (25 mL) and extracted with  $\text{CH}_2\text{Cl}_2$  (2  $\times$  30 mL). The combined organic layers were dried over brine and concentrated in vacuo. The product was dissolved in  $\text{CH}_2\text{Cl}_2$  and washed with 2 N aqueous HCl (2  $\times$  15 mL) and dried over brine (15 mL) and concentrated in vacuo. The product was crystallized from heptane in a yield 62% (0.49 mg, 1.5 mmol) as a white solid.  $R_f = 0.69$  (heptane/EtOAc 1:1).  $^1\text{H}$  NMR (200 MHz,  $\text{CDCl}_3$ )  $\delta$  7.89–7.75 (m, 2H), 7.65–7.53 (m, 2H), 2.92 (t,  $J = 7.5$ , 2H), 1.75–1.64 (m, 2H), 1.32–1.27 (m, 12H), 0.87 (t,  $J = 6.4$ , 3H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CDCl}_3$ )  $\delta$  199.72, 136.01, 132.06, 129.82, 128.19, 38.81, 32.08, 29.67 (2C), 29.54, 29.49, 24.50, 22.88, 14.32. MS (APPI):  $m/z$  311.2 [ $\text{M}^+$ ].

#### 4.1.34. 1-Bromo-4-decylbenzene (8)

The product was obtained using a Wolff–Kishner reaction to reduce 1-(4-bromophenyl)decan-1-one **7**. 1-(4-Bromophenyl)decan-1-one **7** (1.3 g, 4.2 mmol), hydrazine monohydrate (1.2 mL, 16 mmol) and KOH (1.8 g, 32 mmol) were dissolved in 25 mL of 1-octanol to yield a green suspension in a two-necked 50 mL flask. The reaction mixture was stirred at reflux at 135 °C for 2 $\frac{3}{4}$  h to give an almost colorless solution. The reaction mixture was then cooled to room temperature and diluted with ether (50 mL), washed with 1 N aqueous HCl (50 mL), 2 N aqueous HCl (20 mL), brine (2 times, 30 mL), dried over  $\text{MgSO}_4$ , filtered, concentrated in vacuo. 1-Octanol was removed by vacuum distillation to yield about 2.7 g of the crude product. The product was purified using column chromatography with heptane as eluent. Yield 40%. Colorless oil.  $R_f = 0.61$  (100% heptane).  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  7.37 (d,  $J = 8.3$ , 2H), 7.03 (d,  $J = 8.1$ , 2H), 2.54 (t,  $J = 7.7$ , 2H), 1.60–1.54 (m, 2H), 1.28–1.25 (m, 14H), 0.88 (t,  $J = 6.8$ , 3H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CDCl}_3$ )  $\delta$  142.05, 131.46, 130.37, 119.47, 35.59, 32.14, 31.56, 29.85, 29.82, 29.71, 29.57, 29.43, 22.93, 14.34. MS (APPI):  $m/z$  296.2 [ $\text{M}^+$ ].

#### 4.1.35. ((4-Decylphenyl)ethynyl)trimethylsilane

Into a dried 10–20 mL microwave vial,  $\text{PdCl}_2(\text{PPh}_3)$  (52 mg, 74  $\mu\text{mol}$ ),  $\text{CuI}$  (18 mg, 95  $\mu\text{mol}$ ), and  $\text{PPh}_3$  (80 mg, 0.30 mmol) were suspended in  $\text{CH}_3\text{CN}$  (2 mL) under nitrogen atmosphere. Subsequently, 1-bromo-4-decylbenzene **8** (0.46 g, 1.5 mmol), diethylamine ( $\text{Et}_2\text{NH}$ ) (2.4 mL, 23 mmol), and trimethylsilyl acetylene (TMSA) (0.24 mL, 1.7 mmol), were added to the mixture. The mixture was heated at 120 °C for 35 min via microwave irradiation (95 W). The reaction mixture was diluted with EtOAc (50 mL) and washed with water (3  $\times$  50 mL). The water layer was washed with EtOAc (2  $\times$  50 mL). The combined organic phases were extracted with brine (1  $\times$  50 mL) and dried over  $\text{Mg}_2\text{SO}_4$ . The solvent was evaporated under reduced pressure and the residue was purified by column chromatography with heptane as eluent. Yield 65%. Brown oil.  $R_f = 0.41$  (100% heptane).  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  7.37 (d,  $J = 8.0$ , 2H), 7.10 (d,  $J = 7.9$ , 2H), 2.58 (t,  $J = 7.6$ , 2H), 1.58–

1.54 (m, 2H), 1.28–1.25 (m, 14H), 0.88 (t,  $J = 6.8$ , 3H), 0.24 (s, 9H);  $^{13}\text{C}$  NMR (50 MHz,  $\text{CDCl}_3$ )  $\delta$  143.89, 132.08, 128.52, 120.44, 105.63, 93.45, 36.11, 32.12, 31.42, 29.82, 29.69 (2C), 29.55, 29.42, 22.91, 14.34, 0.26. MS (APPI):  $m/z$  314.2 [ $\text{M}^+$ ].

#### 4.1.36. 1-Decyl-4-ethynylbenzene (9)

((4-Decylphenyl)ethynyl)trimethylsilane (0.29 g, 0.91 mmol) was dissolved in THF (5 mL) and the solution was cooled to 0 °C. Tetra butyl ammonium fluoride (TBAF) (1.4 mL, 1.4 mmol) in THF (1 M, 2.6 mL) was added dropwise and the reaction mixture was stirred for 10 min. The reaction mixture was diluted with EtOAc (50 mL), extracted with water (4 × 50 mL) and washed with brine (2 × 50 mL). The organic phase was dried over  $\text{MgSO}_4$  and filtered. The solvent was evaporated and the product was used without further purification. Yield 93%. Brown oil.  $R_f = 0.45$  (100% heptane).  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  7.40 (d,  $J = 8.0$ , 2H), 7.13 (d,  $J = 7.8$ , 2H), 3.02 (s, 1H), 2.59 (t,  $J = 7.6$ , 2H), 1.59–1.55 (m, 2H), 1.29–1.26 (m, 14H), 0.88 (t,  $J = 6.8$ , 2H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CDCl}_3$ )  $\delta$  144.22, 132.25, 128.63, 119.41, 105.22, 92.67, 36.12, 32.12, 31.45, 29.82, 29.79, 29.69, 29.55, 29.46, 22.91, 14.34. MS (APPI):  $m/z$  242.2 [ $\text{M}^+$ ].

#### 4.1.37. Benzyl 2,4-bis(benzyloxy)-6-((4-decylphenyl)ethynyl)benzoate

The product was obtained using Sonogashira coupling reaction of 2,4-bis(benzyloxy)-6-(((trifluoromethyl)sulfonyl)oxy)benzoate **10**, and 1-decyl-4-ethynylbenzene **9**. Distilled diethylamine (0.11 mL, 1.05 mmol) and 1-decyl-4-ethynylbenzene **9** (0.18 g, 0.74 mmol) were subsequently added to a solution of the triflate **10** (0.36 g, 0.63 mmol), CuI (9.0 mg, 47  $\mu\text{mol}$ ), and  $\text{PdCl}_2(\text{PPh}_3)_2$  (18 mg, 0.026 mmol) in degassed anhydrous acetonitrile (1.0 mL) under nitrogen atmosphere. The mixture was subjected to microwave irradiation for 35 min at 120 °C (70 W). The reaction mixture was diluted with ether (50 mL) and washed with demi water (2 × 30 mL), brine (2 × 50 mL), dried over  $\text{MgSO}_4$  and filtered. The solvent was evaporated under reduced pressure, and the residue was purified by column chromatography with heptanes/EtOAc 5:1 (v/v) as eluent. Yield 80%. Brown oil.  $R_f = 0.40$  (heptane/EtOAc 1:1).  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  7.43–7.27 (m, 15H), 7.22 (d,  $J = 8.0$ , 2H), 7.12 (d,  $J = 7.9$ , 2H), 6.75 (s, 1H), 6.56 (s, 1H), 5.37 (s, 2H), 5.06 (s, 2H), 5.04 (s, 2H), 2.60 (t,  $J = 7.5$ , 2H), 1.60–1.55 (m, 2H), 1.30–1.26 (m, 14H), 0.88 (t,  $J = 6.6$ , 3H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CDCl}_3$ )  $\delta$  166.91, 160.54, 157.24, 144.06, 136.47, 136.38, 136.09, 131.89, 128.89, 128.76, 128.61, 128.48, 128.44, 128.17, 128.12, 127.75, 127.30, 123.86, 120.06, 119.89, 109.67, 101.97, 93.47, 86.28, 70.80, 70.53, 67.30, 36.15, 32.12, 31.47, 29.81, 29.71, 29.55, 29.46, 22.91, 14.34. MS (APPI):  $m/z$  665.2 [ $\text{M}+\text{H}^+$ ].

#### 4.1.38. 2-(4-Decylphenethyl)-4,6-dihydroxybenzoic acid (20)

Benzyl 2,4-bis(benzyloxy)-6-((4-decylphenyl)ethynyl)benzoate (0.28 g, 0.42 mmol) was dissolved in EtOAc (due to low solubility in MeOH) and added to 10 mol% Pd/C (10%) (45 mg, 42 nmol). The suspension was shaken with 3 atm  $\text{H}_2$ -pressure in a Parr apparatus at 45 °C for 24 h. Subsequently, the mixture was filtered through Celite. The solvent was evaporated under reduced pressure and the residue was purified by column chromatography with heptane/EtOAc 9:1 (v/v) as eluent, followed by EtOAc as eluent to obtain the desired product. Yield 72%. White solid.  $R_f = 0.40$  (heptane/EtOAc 1:1 + acetic acid). The product was more than 95% pure as judged from TLC and NMR.  $^1\text{H}$  NMR (200 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  7.12–7.02 (m, 4H), 6.18 (d,  $J = 2.4$  Hz, 1H); 6.16 (d,  $J = 2.4$  Hz, 1H), 3.18–3.10 (m, 2H), 2.83–2.75 (m, 2H), 2.62–2.48 (m, 2H), 1.60–1.50 (m, 2H), 1.28–1.20 (m, 14H), 0.89 (t,  $J = 6.5$ , 3H).  $^{13}\text{C}$  NMR (50 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  167.82, 164.57, 150.02, 142.23, 141.60, 130.15, 112.86, 106.44, 102.82, 41.15, 40.02, 37.41, 33.93, 33.67, 31.59, 31.47 (2C), 31.31, 31.19, 24.59, 15.30. HRMS:  $m/z$  [ $\text{M}-\text{H}$ ] $^-$  calcd for  $\text{C}_{25}\text{H}_{33}\text{O}_4$  397.23843, found 397.2388.

#### 4.1.39. 2-(Benzo[d]oxazol-2-yl)phenol (26)

Salicylic acid (0.138 g, 1.0 mmol) was mixed with 2-aminophenol (0.11 g, 1.0 mmol) and 3 mL of polyphosphoric acid was added to the mixture. The mixture was heated in oil bath at 180 °C under  $\text{N}_2$  for 6.5 h. The reaction mixture was cooled to room temperature and subsequently poured into 150 mL demi water. The mixture was extracted with ethylacetate (2 × 30 mL). The combined organic layers were washed with demiwater (2 × 40 mL), brine (1 × 40 mL), dried with  $\text{MgSO}_4$ , filtered and concentrated in vacuo. The product was obtained after crystallization or purification using column chromatography with heptane/EtOAc 50:1 (v/v) as eluent. Yield 47%. Colorless to white crystals.  $R_f = 0.69$  (heptane/EtOAc 1:1). The product was more than 95% pure as judged from TLC and NMR.  $^1\text{H}$  NMR (200 MHz, DMSO)  $\delta$  8.04 (dd,  $J = 7.9$ , 1.7, 1H), 7.83–7.89 (m, 2H), 7.53–7.58 (m, 1H), 7.45–7.50 (m, 2H), 7.06–7.16 (m, 2H).  $^{13}\text{C}$  NMR (50 MHz, DMSO) 158.17, 151.40, 149.24, 139.90, 134.41, 128.04, 126.31, 125.77, 120.43, 119.64, 117.63, 111.50, 110.84. HRMS:  $m/z$  [ $\text{M}+\text{H}$ ] $^+$  calcd for  $\text{C}_{13}\text{H}_{10}\text{O}_2\text{N}_1$  212.07061, found 212.07051.

## 4.2. Enzyme activity studies

### 4.2.1. Human 5-LOX inhibition screening UV assay

Enzyme inhibition was measured by the residual enzyme activity after 10 min incubation with the inhibitor at room temperature. The enzyme activity was determined by conversion of lipoxigenase substrate linoleic acid into hydroperoxy-octadecadienoate (HPOD). The conversion rate was followed by UV absorbance of the conjugated diene at 234 nm ( $\epsilon = 25000 \text{ M}^{-1} \text{ cm}^{-1}$ ) over a period of 20 min. The UV absorbance increase over time was used to determine the enzyme activity.

Tris buffer (50 mM) pH 7.5 containing 2 mM EDTA and 2 mM  $\text{CaCl}_2$  was used as an assay buffer for human 5-LOX experiments. The human 5-LOX enzyme was diluted 1:4000 with the assay buffer. The inhibitor (100 mM in DMSO) was diluted with the assay buffer to 1 mM. The substrate, linoleic acid was diluted with EtOH to 20 mM. Subsequently, 1 mL of enzyme solution (1:4000) was mixed with 100  $\mu\text{L}$  ATP (2 mM), 100  $\mu\text{L}$  inhibitor (1 mM) and 790  $\mu\text{L}$  Tris buffer and incubated for 10 min. After that the linoleic acid solution (10  $\mu\text{L}$ , 20 mM) was added to a mixture and the conversion rate of the substrate was measured after 10 s mixing of the enzyme with the substrate. The reaction rate in the absence of the inhibitor was used as positive control. In the positive control experiment, the assay buffer was supplemented with a small amount of DMSO (3  $\mu\text{L}$  of DMSO in 1.5 mL) in order to replace the DMSO inhibitor solution, which was also pre-incubated for 10 min.

### 4.2.2. COX-2 inhibition screening assay

The COX-2 activity was measured spectrophotometrically by measuring the formation of oxidized  $N,N,N,N'$ -tetramethyl-*p*-phenylenediamine (TMPD). Tris buffer (0.1 M) pH 8.0 was used as an assay buffer for COX-2 experiments. The COX-2 enzyme was diluted 1:1000 with the assay buffer. Hematin (7.5 mM in DMSO) was diluted with assay buffer to 330  $\mu\text{M}$ . The inhibitor was diluted to 1.1 mM in DMSO. The substrate, arachidonic acid was diluted by adding 50  $\mu\text{L}$  KOH 0.1 M and 800  $\mu\text{L}$   $\text{H}_2\text{O}$  into 150  $\mu\text{L}$  of a 22 mM arachidonic acid solution in EtOH. The colorimetric substrate TMPD (2.64 mM in  $\text{H}_2\text{O}$ ) was prepared freshly prior to the experiment. The substrate, linoleic acid was diluted with EtOH to 20 mM. Subsequently, in a 96 wells plate, 150  $\mu\text{L}$  assay buffer, 10  $\mu\text{L}$  hematin 330  $\mu\text{M}$ , 10  $\mu\text{L}$  COX-2 enzyme (1:1000) was added. 10  $\mu\text{L}$  inhibitor solution (1.1 mM in DMSO) was added to the inhibitor well and 10  $\mu\text{L}$  of DMSO was added to the positive control wells. The plate was incubated for 5 min at room temperature. 20  $\mu\text{L}$  of freshly prepared TMPD was added followed by the

addition of 20  $\mu\text{L}$  arachidonic acid. Arachidonic acid addition was excluded from the negative control wells. The plate was incubated for another 5 min at room temperature. The absorbance was measured at 550 nm, and the calculations were performed with Excel 2010. The experiments were performed in triplicate, and the presented results were the average of three measurements with the standard deviation.

#### 4.2.3. Michaelis Menten enzyme kinetics

The enzyme kinetics of human 5-LOX were also studied by the formation of the conjugated diene product at 234 nm ( $\epsilon = 25000 \text{ M}^{-1} \text{ cm}^{-1}$ ) using the same experimental setup as for the concentration dependent and the inhibitor screening. Enzyme kinetics were performed against linoleic acid and ATP to get a complete indication about the regulation of human 5-LOX enzyme activity.

In the kinetic study of the human 5-LOX against ATP, the ATP concentration was varied between 12.5–200  $\mu\text{M}$  while the concentration of linoleic acid was fixed (100  $\mu\text{M}$ ). In the kinetic study of the human 5-LOX against the linoleic acid, substrate concentration was varied between 37.5–300  $\mu\text{M}$ . The enzyme activity was measured in the absence or presence of fixed concentrations of activator **23a** (0  $\mu\text{M}$ , 12.5  $\mu\text{M}$ , and 25  $\mu\text{M}$ ) or inhibitor **23d** (0  $\mu\text{M}$ , 25  $\mu\text{M}$ , and 50  $\mu\text{M}$ ). The reaction velocities ( $v$ ), which are the concentration changes over time, were plotted against the substrate concentrations in Michaelis–Menten plots and the  $K_m$  and  $V_{\max}$  and the apparent values ( $K_m^{\text{app}}$  and  $K_m^{\text{app}}$ ) in the presence of the activator were derived. The results from the non-linear curve fitting were in line with the results from Lineweaver–Burk plot.

The slopes and y-intercepts from the Lineweaver–Burk plot of h-5-LOX activation was derived and then re-plotted as  $1/\Delta\text{slope}$  or  $1/\Delta y$  intercept versus  $1/[\text{activator}]$ . From these plots  $\alpha$ ,  $\beta$  and  $K_A$  values were derived as described by Leskovac<sup>44</sup> for non-essential activation. According to this method in the re-plot of  $1/\Delta y$  intercept versus  $1/[\text{activator}]$  the y intercept corresponds to  $\beta (V_{\max}/(\beta - 1))$ , whereas in the re-plot of  $1/\Delta\text{slope}$  versus  $1/[\text{activator}]$  the y intercept correspond to  $\beta \cdot V_{\max}/K_m(\beta - \alpha)$ . The x intersection point of two lines from the plot of  $1/\Delta y$  intercept and  $1/\Delta\text{slope}$  is  $-\beta/\alpha K_A$ , which can be employed to derive the  $K_A$  value if  $\alpha$  and  $\beta$  are known. All the experiments were performed in triplicate and the average triplicate values and their standard deviations are plotted. Calculations were performed with Excel 2010.

### 4.3. Docking studies

#### 4.3.1. Protein and ligand structures

Structures of compounds **23a**, **23b**, **23c**, and **23d** were built using Omega2 from OpenEye<sup>53</sup> We fixed the rotatable bonds between the benzyl ring and the carboxylic acid group on all molecules using AutoDockTools to dock structures that kept a planar geometry for the benzyl ring and carboxylic acid.<sup>54</sup>

#### 4.3.2. Docking and scoring

The receptor structure was prepared for docking using the prepare\_receptor4.py script from AutoDockTools. Compounds were docked using *smina*<sup>55</sup> with the default *vina* scoring function and the argument ‘–num\_modes = 40’. The docked poses were rescored using the ‘–score\_only’ command in *smina* and then reranked. The top ranked poses were used for further analysis.

### Author contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of

the manuscript. R.W., P.A.M.K. and N.E. synthesized the molecules and performed the biochemical analysis. M.P.B. and C.J.C. performed and analyzed the docking studies and were supported by grant R01GM097082-01 from the National Institutes of Health. H.J.H. and F.J.D. supervised the research. R.W. and F.J.D. wrote the manuscript.

### Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.bmc.2013.10.015>.

### References and notes

- Martel-Pelletier, J.; Lajeunesse, D.; Reboul, P.; Pelletier, J. *Ann. Rheum. Dis.* **2003**, *62*, 501.
- Gillmor, S. A.; Villasenor, A.; Fletterick, R.; Sigal, E.; Browner, M. F. *Nat. Struct. Biol.* **1997**, *4*, 1003.
- Sigal, E. *Am. J. Physiol.* **1991**, *260*, L13.
- Epp, N.; Fürstenberger, G.; Müller, K.; de Juanes, S.; Leitges, M.; Hausser, I.; Thieme, F.; Liebisch, G.; Schmitz, G.; Krieg, P. *J. Cell Biol.* **2007**, *177*, 173.
- Ikei, K. N.; Yeung, J.; Apopa, P. L.; Ceja, J.; Vesci, J.; Holman, T. R.; Holinstat, M. *J. Lipid Res.* **2012**, *53*, 2546.
- Hunter, J. A.; Finkbeiner, W. E.; Nadel, J. A.; Goetzl, E. J.; Holtzman, M. J. *Proc. Natl. Acad. Sci.* **1985**, *82*, 4633.
- Nadel, J. A.; Conrad, D. J.; Ueki, I. F.; Schuster, A.; Sigal, E. *J. Clin. Invest.* **1991**, *87*, 1139.
- Brash, A. R.; Boeglin, W. E.; Chang, M. S. *Proc. Natl. Acad. Sci.* **1997**, *94*, 6148.
- Brown, C. D.; Kilty, I.; Yeaton, M.; Jenkinson, S. *Inflamm. Res.* **2001**, *50*, 321.
- Chanez, P.; Bonnans, C.; Chavis, C.; Vachier, I. *Am. J. Respir. Cell Mol. Biol.* **2002**, *27*, 655.
- Chu, H. W.; Balzar, S.; Westcott, J. Y.; Trudeau, J. B.; Sun, Y.; Conrad, D. J.; Wenzel, S. E. *Clin. Exp. Allergy* **2002**, *32*, 1558.
- Shannon, V. R.; Chanez, P.; Bousquet, J.; Holtzman, M. J. *Am. Rev. Respir. Dis.* **1993**, *147*, 1024.
- Piao, Y.; Du, Y.; Oshima, H.; Jin, J.; Nomura, M.; Yoshimoto, T.; Oshima, M. *Carcinogenesis* **2008**, *29*, 440.
- Rásó, E.; Döme, B.; Somlai, B.; Zacharek, A.; Hagmann, W.; Honn, K. V.; Tímár, J. *Melanoma Res.* **2004**, *14*, 245.
- Gupta, S.; Srivastava, M.; Ahmad, N.; Sakamoto, K.; Bostwick, D. G.; Mukhtar, H. *Cancer* **2001**, *91*, 737.
- Melstrom, L. G.; Bentrem, D. J.; Salabat, M. R.; Kennedy, T. J.; Ding, X.; Strouch, M.; Rao, S. M.; Witt, R. C.; Ternent, C. A.; Talamonti, M. S.; Bell, R. H.; Adrian, T. A. *Clin. Cancer Res.* **2008**, *14*, 6525.
- Nathoo, N.; Prayson, R. A.; Bondar, J.; Vargo, L.; Arrigain, S.; Mascha, E. J.; Suh, J. H.; Barnett, G. H.; Golubic, M. *Neurosurgery* **2006**, *58*, 347.
- Wilborn, J.; Bailie, M.; Coffey, M.; Burdick, M.; Strieter, R.; Peters-Golden, M. *J. Clin. Invest.* **1996**, *97*, 1827.
- Zou, L. Y. *J. Cancer Mol.* **2006**, *2*, 227.
- Serhan, C. N.; Chiang, N.; Van Dyke, T. E. *Nat. Rev. Immunol.* **2008**, *8*, 349.
- Edenius, C.; Haeggström, J.; Lindgren, J. Å. *Biochem. Biophys. Res. Commun.* **1988**, *157*, 801.
- Hashimoto, A.; Hayashi, I.; Murakami, Y.; Sato, Y.; Kitasato, H.; Matsushita, R.; Izuka, N.; Urabe, K.; Itoman, M.; Hirohata, S.; Endo, H. *J. Rheumatol.* **2007**, *34*, 2144.
- Sobrado, M.; Pereira, M. P.; Ballesteros, I.; Hurtado, O.; Fernández-López, D.; Pradillo, J. M.; Caso, J. R.; Vivancos, J.; Nombela, F.; Serena, J.; Lizasoain, I.; Moro, M. A. *J. Neurosci.* **2009**, *29*, 3875.
- Matschinsky, F. M. *Nat. Rev. Drug Disc.* **2009**, *8*, 399.
- Smith, W. L.; DeWitt, D. L.; Garavito, R. M. *Annu. Rev. Biochem.* **2000**, *69*, 145.
- Anderson, G. D.; Hauser, S. D.; McGarity, K. L.; Bremer, M. E.; Isakson, P. C.; Gregory, S. J. *Clin. Invest.* **1996**, *97*, 2672.
- Reuter, B. K.; Asfaha, S.; Buret, A.; Sharkey, K. A.; Wallace, J. L. *J. Clin. Invest.* **1996**, *98*, 2076.
- Hendel, J.; Nielsen, O. H. *Am. J. Gastroenterol.* **1997**, *92*, 1170.
- Kargman, S. L.; O'Neill, G. P.; Vickers, P. J.; Evans, J. F.; Mancini, J. A.; Jothy, S. *Cancer Res.* **1995**, *55*, 2556.
- Cianchi, F.; Cortesini, C.; Bechi, P.; Fantappiè, O.; Messerini, L.; Vannacci, A.; Sardi, I.; Baroni, G.; Boddì, V.; Mazzanti, R.; Masini, E. *Gastroenterology* **2001**, *121*, 1339.
- Wisastra, R.; Ghizzoni, M.; Boltjes, A.; Haisma, H. J.; Dekker, F. J. *Bioorg. Med. Chem.* **2012**, *20*, 5027.
- Brash, A. R. *J. Biol. Chem.* **1999**, *274*, 23679.
- Ghizzoni, M.; Boltjes, A.; Graaf, C. d.; Haisma, H. J.; Dekker, F. J. *Bioorg. Med. Chem.* **2010**, *18*, 5826.
- Ghizzoni, M.; Wu, J.; Gao, T.; Haisma, H. J.; Dekker, F. J.; George Zheng, Y. *Eur. J. Med. Chem.* **2012**, *47*, 337.
- Uchiyama, M.; Ozawa, H.; Takuma, K.; Matsumoto, Y.; Yonehara, M.; Hiroya, K.; Sakamoto, T. *Org. Lett.* **2006**, *8*, 5517.
- Harrowven, D. C.; Woodcock, T.; Howes, P. D. *Angew. Chem., Int. Ed.* **2005**, *44*, 3899.

37. Sista, P.; Nguyen, H.; Murphy, J. W.; Hao, J.; Dei, D. K.; Palaniappan, K.; Servello, J.; Kularatne, R. S.; Gnade, B. E.; Xue, B.; Dastoor, P. C.; Biewer, M. C.; Stefan, M. C. *Macromolecules* **2010**, *43*, 8063.
38. Tranchimand, S.; Tron, T.; Gaudin, C.; Iacazio, G. *Synth. Commun.* **2006**, *36*, 587.
39. Back, D. F.; Manzoni de Oliveira, G.; Ballin, M. A.; Corbellini, V. A. *Inorg. Chim. Acta* **2010**, *363*, 807.
40. Gibian, M. J.; Galaway, R. A. *Biochemistry* **1976**, *15*, 4209.
41. Ha, T. J.; Kubo, I. *J. Agric. Food Chem.* **2005**, *53*, 4350.
42. Copeland, R. A.; Williams, J. M.; Giannaras, J.; Nurnberg, S.; Covington, M.; Pinto, D.; Pick, S.; Trzaskos, J. M. *Proc. Natl. Acad. Sci.* **1994**, *91*, 11202.
43. Verhagen, J.; Vliegenthart, J. F.; Boldingh, J. *Chem. Phys. Lipids* **1978**, *22*, 255.
44. Leskovac, V. *Compr. Enzyme Kinet.* **2004**, 111.
45. Gilbert, N. C.; Bartlett, S. G.; Waight, M. T.; Neau, D. B.; Boeglin, W. E.; Brash, A. R.; Newcomer, M. E. *Science* **2011**, *331*, 217.
46. Gilbert, N. C.; Rui, Z.; Neau, D. B.; Waight, M. T.; Bartlett, S. G.; Boeglin, W. E.; Brash, A. R.; Newcomer, M. E. *FASEB J.* **2012**, *26*, 3222.
47. The PYMOL Molecular Graphics System, Version 1.5.0.4. Schödinger LLC.
48. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605.
49. Lee, K. S.; Kim, S. R.; Park, H. S.; Park, S. J.; Min, K. H.; Lee, K. Y.; Jin, S. M.; Lee, Y. C. *J. Allergy Clin. Immunol.* **2007**, *119*, 141.
50. Okamoto, F.; Saeki, K.; Sumimoto, H.; Yamasaki, S.; Yokomizo, T. *J. Biol. Chem.* **2010**, *285*, 41113.
51. Thompson, C.; Cloutier, A.; Bossé, Y.; Poisson, C.; Larivée, P.; McDonald, P. P.; Stankova, J.; Rola-Pleszczynski, M. *J. Biol. Chem.* **2008**, *283*, 1974.
52. Deb, A.; Haque, S. J.; Mogensen, T.; Silverman, R. H.; Williams, B. R. G. *J. Immunol.* **2001**, *166*, 6170.
53. Boström, J.; Greenwood, J. R.; Gottfries, J. *J. Mol. Graph. Model.* **2003**, *21*, 449.
54. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. *J. Comput. Chem.* **2009**, *30*, 2785.
55. Koes, D. R.; Baumgartner, M. P.; Camacho, C. J. *J. Chem. Inf. Model.* **2013**, *53*, 1893.

## APPENDIX C SCRIPT FOR COMPUTING RESIDUE OVERLAPS IN MDS

The following Python script was used to compute the overlaps of the p53 residues as shown in Figure 2.2.

```
#Python for computing the fractional overlap of the p53 residues.
#Part of the Supplemental information for:
#Baumgartner, MP and Camacho, CJ. "The atomistic induced-fit binding
mechanism of p53 to MDM2"

#by Matthew Baumgartner
#1-11-16

import sys
import os
import argparse
import math

#control how much debugging information is printed
DEBUG = 2

def main():

    parser = argparse.ArgumentParser(description = 'Compute the overlap of
md frames to an input structure (ex. smisp)')
    parser.add_argument('-s', dest = 'residues', type = str, help = 'residues
to check overlap of')
    parser.add_argument('-l', dest = 'frame_list', help = 'list of frames')
    parser.add_argument('-ot', dest = 'outtext', help = 'output text file')
    parser.add_argument('-o', dest = 'outfile', help = 'output pdb file')
    parser.add_argument('-start', dest = 'start', default = 0, type = float,
help = 'optional, choose a starting from to start at')
    parser.add_argument('-end', dest = 'end', default = -1, type = float,
help = 'optional, choose a starting from to end at')
    parser.add_argument('--redo', dest = 'redo', action = 'store_const',
const = True, default = False, help = 'If specified, ignore any saved results
and recompute (Default: False)')
```

```

args = parser.parse_args()

resfile = args.residues
frame_list = args.frame_list
outfile = args.outtext
phe_outfile = args.outfile
start = int(args.start)
end = int(args.end)
REDO = args.redo

#https://www.rbvi.ucsf.edu/chimera/docs/UsersGuide/midas/vdwtables.html#all
atom
vdw = {'C' : 1.7,
       'N' : 1.625,
       'O' : 1.48,
       'S' : 1.782,
       'H' : 1.00,
       'P' : 1.871,
       'F' : 1.56,
       'Cl' : 1.735,
       'CL' : 1.735,
       'Br' : 1.978,
       'BR' : 1.978,
       'I' : 2.094}

if not REDO:
    if os.path.exists(outfile):
        print 'Output file:', outfile, 'already exists, non-
recomputing!\nRun with --redo to force recompute'
        sys.exit()

    res_atoms = [ map(float,[ line[30:38], line[38:46], line[46:54]]) for
line in open(resfile).readlines() if (line.startswith('ATOM') or
line.startswith('HETATM')) ]

    #proceed each atom name with the resid to avoid collisions
    res_atom_names = [ line[12:16].strip() + '_' + line[22:26].strip() for
line in open(resfile).readlines() if (line.startswith('ATOM') or
line.startswith('HETATM')) ]

    #Get the atom types from the last column
    res_atom_type = [ line[76:80].strip().strip('-+0123456789') for line in
open(resfile).readlines() if (line.startswith('ATOM') or
line.startswith('HETATM')) ]

    #read in text file that lists the absolute paths of the pdb files of the
frames. I used mdpocket from fpocket to split the simulation and create the
input file
    frames = [ line.strip() for line in open(frame_list).readlines() ]

    #the frames in which an atom is over lapped
    frame_overlaps = { nm : [] for nm in res_atom_names }

```



```

for frame_id, frame in enumerate(frames):

    #check to see if you are inside the range you want to be
    if frame_id < start:
        continue
    if end != -1:
        if frame_id >= end:
            continue

    if frame_id > 0:
        if DEBUG > 2:
            print 'frame:', frame_id, 'frame:', frame

        for line in open(frame).xreadlines():
            if line.startswith('ATOM'):
                atom_name = line[12:16].strip()
                if not atom_name.startswith('H'): #skip hydrogens
                    traj_coor = map(float, [line[30:38], line[38:46],
line[46:54]])

                    for i, res_coor in enumerate(res_atoms):
                        res_name = res_atom_names[i]
                        res_vdw = vdw[res_atom_type[i]] #vdw of the first
letter (element)

                        #distance
                        dist = math.sqrt((res_coor[0] - traj_coor[0])**2
+ (res_coor[1] - traj_coor[1])**2 + (res_coor[2] - traj_coor[2])**2)

                        #if it is closer than the vdw
                        if dist < res_vdw:
                            #get the residue number of the overlapping
atom
                            traj_resid = int(line[22:26].strip())

                            #if you haven't added one already
                            if len(frame_overlaps[res_name]) < frame_id:

frame_overlaps[res_name].append(traj_resid)

#                                break

                            #if you didn't find any overlapping atoms, add a 0 to the list
                            for res_name in res_atom_names:
                                if DEBUG > 2:
                                    print res_name, 'overlap', len:',
len(frame_overlaps[res_name])
                                    if len(frame_overlaps[res_name]) < frame_id:
                                        frame_overlaps[res_name].append(0)

atom_names = []
lists = []

```

```

for res_name, overlapped in frame_overlaps.iteritems():
    if DEBUG > 2:
        print res_name, ':', overlapped
        print len(overlapped)
    atom_names.append(res_name)
    lists.append(overlapped)

fout = open(outfile, 'w')

if start == 0 and end == -1:

    #print the header
    header = 'Frame\t' + '\t'.join(atom_names) + '\n'
    fout.write(header)

    for i in range(len(lists[0])): #
        str_line = str(i)

        for lst in lists:
            str_line += '\t' + str(lst[i])

        fout.write(str_line + '\n')
    fout.close()

    if DEBUG > 2:
        print outfile, 'written'

#write out the residue again but put the perc overlapped in the b factor
field
name_perc = {}
if DEBUG > 2:
    print 'Calculating the percent overlapped'
for i,lst in enumerate(lists):
    non_zero = len([ x for x in lst if x != 0 ])
    perc_overlapped = float(non_zero) / len(lst)
    if DEBUG > 2:
        print atom_names[i], perc_overlapped
    name_perc[atom_names[i]] = perc_overlapped

output = []

for line in open(resfile).readlines():
    if not (line.startswith('ATOM') or line.startswith('HETATM')):
        output.append(line)
    elif line[76:80].strip().strip('-+0123456789').startswith('H'):
        output.append(line)
    else:
        atom_name = line[12:16].strip() + '_' + line[22:26].strip()
        b_fac = ' ' + '%03.2f' % round(name_perc[atom_name],2)

        new_line = line[:60] + b_fac + line[66:]
        output.append(new_line)

```

```
fout = open(phe_outfile, 'w')
fout.write(''.join(output))
fout.close()

print phe_outfile, 'written'

if __name__ == '__main__':
    main()
```

## BIBLIOGRAPHY

- [1] Belleau, B., "Rational drug design: mirage or miracle?", *Can. Med. Assoc. J.*, Vol. 103, 1970, pp. 850-853.
- [2] Cheng, T., Q. Li, Z. Zhou, et al., "Structure-based virtual screening for drug discovery: a problem-centric review", *AAPS J.*, Vol. 14, 2012, pp. 133-141.
- [3] Sukumar, N. and S. Das, "Current trends in virtual high throughput screening using ligand-based and structure-based methods", *Comb Chem High Throughput Screen*, Vol. 14, 2011, pp. 872-888.
- [4] Mason, J. S., A. C. Good and E. J. Martin, "3-D pharmacophores in drug discovery", *Curr. Pharm. Des.*, Vol. 7, 2001, pp. 567-597.
- [5] Bohm, H. J., "The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure", *J. Comput. Aided Mol. Des.*, Vol. 8, 1994, pp. 243-256.
- [6] Erickson, J. and D. Kempf, "Structure-based design of symmetric inhibitors of HIV-1 protease", *Arch Virol Suppl*, Vol. 9, 1994, pp. 19-29.
- [7] Whittle, P. J. and T. L. Blundell, "Protein structure--based drug design", *Annu. Rev. Biophys. Biomol. Struct.*, Vol. 23, 1994, pp. 349-375.
- [8] Lu, H. and P. J. Tonge, "Inhibitors of FabI, an enzyme drug target in the bacterial fatty acid biosynthesis pathway", *Acc. Chem. Res.*, Vol. 41, 2008, pp. 11-20.
- [9] Niinivehmas, S. P., E. Manivannan, S. Rauhamaki, et al., "Identification of estrogen receptor alpha ligands with virtual screening techniques", *J Mol Graph Model*, Vol. 64, 2016, pp. 30-39.

- [10] Lane, D. P., "p53, guardian of the genome", *Nature*, Vol. 358, 1992, pp. 15-16.
- [11] Botuyan, M. V., J. Momand and Y. Chen, "Solution conformation of an essential region of the p53 transactivation domain", *Fold Des.*, Vol. 2, 1997, pp. 331-342.
- [12] Zak, K., A. Pecak, B. Rys, et al., "Mdm2 and MdmX inhibitors for the treatment of cancer: a patent review (2011-present)", *Expert Opin Ther Pat*, Vol. 23, 2013, pp. 425-448.
- [13] Leao, M., C. Pereira, A. Bisio, et al., "Discovery of a new small-molecule inhibitor of p53-MDM2 interaction using a yeast-based approach", *Biochem. Pharmacol.*, Vol. 85, 2013, pp. 1234-1245.
- [14] Wade, M., Y. C. Li and G. M. Wahl, "MDM2, MDMX and p53 in oncogenesis and cancer therapy", *Nat. Rev. Cancer*, Vol. 13, 2013, pp. 83-96.
- [15] Tovar, C., B. Graves, K. Packman, et al., "MDM2 small-molecule antagonist RG7112 activates p53 signaling and regresses human tumors in preclinical cancer models", *Cancer Res.*, Vol. 73, 2013, pp. 2587-2597.
- [16] Schon, O., A. Friedler, M. Bycroft, et al., "Molecular mechanism of the interaction between MDM2 and p53", *J. Mol. Biol.*, Vol. 323, 2002, pp. 491-501.
- [17] Schon, O., A. Friedler, S. Freund, et al., "Binding of p53-derived ligands to MDM2 induces a variety of long range conformational changes", *J. Mol. Biol.*, Vol. 336, 2004, pp. 197-202.
- [18] Bowman, A. L., Z. Nikolovska-Coleska, H. Zhong, et al., "Small molecule inhibitors of the MDM2-p53 interaction discovered by ensemble-based receptor models", *J. Am. Chem. Soc.*, Vol. 129, 2007, pp. 12809-12814.
- [19] ElSawy, K. M., C. S. Verma, T. L. Joseph, et al., "On the interaction mechanisms of a p53 peptide and nutlin with the MDM2 and MDMX proteins: a Brownian dynamics study", *Cell Cycle*, Vol. 12, 2013, pp. 394-404.

- [20] Sanchez, M. C., J. G. Renshaw, G. Davies, et al., "MDM4 binds ligands via a mechanism in which disordered regions become structured", *FEBS Lett.*, Vol. 584, 2010, pp. 3035-3041.
- [21] Chen, J., J. Wang, W. Zhu, et al., "A computational analysis of binding modes and conformation changes of MDM2 induced by p53 and inhibitor bindings", *J. Comput. Aided Mol. Des.*, Vol. 27, 2013, pp. 965-974.
- [22] Freedman, D. A., L. Wu and A. J. Levine, "Functions of the MDM2 oncoprotein", *Cell. Mol. Life Sci.*, Vol. 55, 1999, pp. 96-107.
- [23] Chen, J., X. Wu, J. Lin, et al., "mdm-2 inhibits the G1 arrest and apoptosis functions of the p53 tumor suppressor protein", *Mol. Cell. Biol.*, Vol. 16, 1996, pp. 2445-2452.
- [24] Bista, M., S. Wolf, K. Khoury, et al., "Transient protein states in designing inhibitors of the MDM2-p53 interaction", *Structure*, Vol. 21, 2013, pp. 2143-2151.
- [25] Miyazaki, M., H. Naito, Y. Sugimoto, et al., "Lead optimization of novel p53-MDM2 interaction inhibitors possessing dihydroimidazothiazole scaffold", *Bioorg. Med. Chem. Lett.*, Vol. 23, 2013, pp. 728-732.
- [26] Popowicz, G. M., A. Czarna, S. Wolf, et al., "Structures of low molecular weight inhibitors bound to MDMX and MDM2 reveal new approaches for p53-MDMX/MDM2 antagonist drug discovery", *Cell Cycle*, Vol. 9, 2010, pp. 1104-1111.
- [27] Rew, Y., D. Sun, F. Gonzalez-Lopez De Turiso, et al., "Structure-based design of novel inhibitors of the MDM2-p53 interaction", *J. Med. Chem.*, Vol. 55, 2012, pp. 4936-4954.
- [28] Rajamani, D., S. Thiel, S. Vajda, et al., "Anchor residues in protein-protein interactions", *Proc Natl Acad Sci U S A*, Vol. 101, 2004, pp. 11287-11292.
- [29] Koes, D., K. Khoury, Y. Huang, et al., "Enabling large-scale design, synthesis and validation of small molecule protein-protein antagonists", *PLoS One*, Vol. 7, 2012, pp. e32839.

- [30] Uhrinova, S., D. Uhrin, H. Powers, et al., "Structure of free MDM2 N-terminal domain reveals conformational adjustments that accompany p53-binding", *J. Mol. Biol.*, Vol. 350, 2005, pp. 587-598.
- [31] Joseph, T. L., A. Madhumalar, C. J. Brown, et al., "Differential binding of p53 and nutlin to MDM2 and MDMX: computational studies", *Cell Cycle*, Vol. 9, 2010, pp. 1167-1181.
- [32] Kussie, P. H., S. Gorina, V. Marechal, et al., "Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain", *Science*, Vol. 274, 1996, pp. 948-953.
- [33] Gotz, A. W., M. J. Williamson, D. Xu, et al., "Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born", *J Chem Theory Comput*, Vol. 8, 2012, pp. 1542-1555.
- [34] Case, D. A. , V. Babin, J. T. Berryman, et al. "AMBER 14", University of California, San Francisco,
- [35] Wang, J., R. M. Wolf, J. W. Caldwell, et al., "Development and testing of a general amber force field", *J. Comput. Chem.*, Vol. 25, 2004, pp. 1157-1174.
- [36] Wang, J., W. Wang, P. A. Kollman, et al., "Automatic atom type and bond type perception in molecular mechanical calculations", *J Mol Graph Model*, Vol. 25, 2006, pp. 247-260.
- [37] Schrödinger, LLC., "The PyMOL Molecular Graphics System, Version 1.7.4.0".
- [38] Humphrey, W., A. Dalke and K. Schulten, "VMD: visual molecular dynamics", *J Mol Graph*, Vol. 14, 1996, pp. 33-38, 27-38.
- [39] Romo, T. D. and A. Grossfield, "LOOS: an extensible platform for the structural analysis of simulations", *Conf Proc IEEE Eng Med Biol Soc*, Vol. 2009, 2009, pp. 2332-2335.
- [40] Pettersen, E. F., T. D. Goddard, C. C. Huang, et al., "UCSF Chimera--a visualization system for exploratory research and analysis", *J. Comput. Chem.*, Vol. 25, 2004, pp. 1605-1612.

- [41] Levine, A. J., A. Chang, D. Dittmer, et al., "The p53 tumor suppressor gene", *J Lab Clin Med*, Vol. 123, 1994, pp. 817-823.
- [42] Vassilev, L. T., B. T. Vu, B. Graves, et al., "In vivo activation of the p53 pathway by small-molecule antagonists of MDM2", *Science*, Vol. 303, 2004, pp. 844-848.
- [43] Grasberger, B. L., T. Lu, C. Schubert, et al., "Discovery and cocrystal structure of benzodiazepinedione HDM2 antagonists that activate p53 in cells", *J. Med. Chem.*, Vol. 48, 2005, pp. 909-912.
- [44] Michelsen, K., J. B. Jordan, J. Lewis, et al., "Ordering of the N-terminus of human MDM2 by small molecule inhibitors", *J. Am. Chem. Soc.*, Vol. 134, 2012, pp. 17059-17067.
- [45] Allen, J. G., M. P. Bourbeau, G. E. Wohlhieter, et al., "Discovery and optimization of chromenotriazolopyrimidines as potent inhibitors of the mouse double minute 2-tumor protein 53 protein-protein interaction", *J. Med. Chem.*, Vol. 52, 2009, pp. 7044-7053.
- [46] Wolf, S., Y. Huang, G. M. Popowicz, et al., "Ugi Multicomponent Reaction Derived p53-Mdm2 Antagonists", To be published., 2013.
- [47] Huang, Y., S. Wolf, D. Koes, et al., "Exhaustive fluorine scanning toward potent p53-Mdm2 antagonists", *ChemMedChem*, Vol. 7, 2012, pp. 49-52.
- [48] Furet, P., P. Chene, A. De Pover, et al., "The central valine concept provides an entry in a new class of non peptide inhibitors of the p53-MDM2 interaction", *Bioorg. Med. Chem. Lett.*, Vol. 22, 2012, pp. 3498-3502.
- [49] Michelsen, K., J. B. Jordan, J. Lewis, et al., "Ordering of the N-terminus of human MDM2 by small molecule inhibitors", *J Am Chem Soc*, Vol. 134, 2012, pp. 17059-17067.



- [50] Anil, B., C. Riedinger, J. A. Endicott, et al., "The structure of an MDM2-Nutlin-3a complex solved by the use of a validated MDM2 surface-entropy reduction mutant", *Acta Crystallogr D Biol Crystallogr*, Vol. 69, 2013, pp. 1358-1366.
- [51] Gonzalez-Lopez de Turiso, F., D. Sun, Y. Rew, et al., "Rational design and binding mode duality of MDM2-p53 inhibitors", *J. Med. Chem.*, Vol. 56, 2013, pp. 4053-4070.
- [52] Sun, D., Z. Li, Y. Rew, et al., "Discovery of AMG 232, a potent, selective, and orally bioavailable MDM2-p53 inhibitor in clinical development", *J. Med. Chem.*, Vol. 57, 2014, pp. 1454-1472.
- [53] Gonzalez, A. Z., J. Eksterowicz, M. D. Bartberger, et al., "Selective and potent morpholinone inhibitors of the MDM2-p53 protein-protein interaction", *J. Med. Chem.*, Vol. 57, 2014, pp. 2472-2488.
- [54] Gonzalez, A. Z., Z. Li, H. P. Beck, et al., "Novel inhibitors of the MDM2-p53 interaction featuring hydrogen bond acceptors as carboxylic acid isosteres", *J. Med. Chem.*, Vol. 57, 2014, pp. 2963-2988.
- [55] Vaupel, A., G. Bold, A. De Pover, et al., "Tetra-substituted imidazoles as a new class of inhibitors of the p53-MDM2 interaction", *Bioorg. Med. Chem. Lett.*, Vol. 24, 2014, pp. 2110-2114.
- [56] Bajorath, J., "Integration of virtual and high-throughput screening", *Nat. Rev. Drug Discov.*, Vol. 1, 2002, pp. 882-894.
- [57] Koes, D. R. and C. J. Camacho, "ZINCPharmer: pharmacophore search of the ZINC database", *Nucleic Acids Res.*, Vol. 40, 2012, pp. W409-414.

- [58] Koes, D. R. and C. J. Camacho, "PocketQuery: protein-protein interaction inhibitor starting points from protein-protein interaction structure", *Nucleic Acids Res.*, Vol. 40, 2012, pp. W387-392.
- [59] McGaughey, G. B., R. P. Sheridan, C. I. Bayly, et al., "Comparison of topological, shape, and docking methods in virtual screening", *J Chem Inf Model*, Vol. 47, 2007, pp. 1504-1519.
- [60] Jorgensen, W. L., "The many roles of computation in drug discovery", *Science*, Vol. 303, 2004, pp. 1813-1818.
- [61] Heikamp, K. and J. Bajorath, "The future of virtual compound screening", *Chemical biology & drug design*, Vol. 81, 2013, pp. 33-40.
- [62] Kitchen, D. B., H. Decornez, J. R. Furr, et al., "Docking and scoring in virtual screening for drug discovery: methods and applications", *Nat. Rev. Drug Discov.*, Vol. 3, 2004, pp. 935-949.
- [63] Chang, D. T., T. J. Yao, C. Y. Fan, et al., "AH-DB: collecting protein structure pairs before and after binding", *Nucleic Acids Res.*, Vol. 40, 2012, pp. D472-478.
- [64] Word, J. M., S. C. Lovell, J. S. Richardson, et al., "Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation", *J. Mol. Biol.*, Vol. 285, 1999, pp. 1735-1747.
- [65] Systems, Daylight Chemical Information, "Daylight Theory Manual v4.9", 2011.
- [66] O'Boyle, N. M., M. Banck, C. A. James, et al., "Open Babel: An open chemical toolbox", *Journal of cheminformatics*, Vol. 3, 2011, pp. 33.
- [67] Ross, G. A., G. M. Morris and P. C. Biggin, "Rapid and accurate prediction and scoring of water molecules in protein binding sites", *PLoS One*, Vol. 7, 2012, pp. e32036.

- [68] Stams, T., Y. Chen, P. A. Boriack-Sjodin, et al., "Structures of murine carbonic anhydrase IV and human carbonic anhydrase II complexed with brinzolamide: molecular basis of isozyme-drug discrimination", *Protein Sci.*, Vol. 7, 1998, pp. 556-563.
- [69] Cox, J. D., J. A. Hunt, K. M. Compher, et al., "Structural influence of hydrophobic core residues on metal binding and specificity in carbonic anhydrase II", *Biochemistry*, Vol. 39, 2000, pp. 13687-13694.
- [70] Yang, F. and G. N. Phillips, Jr., "Crystal structures of CO-, deoxy- and met-myoglobins at various pH values", *J. Mol. Biol.*, Vol. 256, 1996, pp. 762-774.
- [71] Goldbeck, R. A., M. L. Pillsbury, R. A. Jensen, et al., "Optical detection of disordered water within a protein cavity", *J. Am. Chem. Soc.*, Vol. 131, 2009, pp. 12265-12272.
- [72] Gibson, A. E., C. E. Arris, J. Bentley, et al., "Probing the ATP ribose-binding domain of cyclin-dependent kinases 1 and 2 with O(6)-substituted guanine derivatives", *J. Med. Chem.*, Vol. 45, 2002, pp. 3381-3393.
- [73] Schulze-Gahmen, U., H. L. De Bondt and S. H. Kim, "High-resolution crystal structures of human cyclin-dependent kinase 2 with and without ATP: bound waters and natural ligand as guides for inhibitor design", *J. Med. Chem.*, Vol. 39, 1996, pp. 4540-4546.
- [74] Yde, C. W., I. Ermakova, O. G. Issinger, et al., "Inclining the purine base binding plane in protein kinase CK2 by exchanging the flanking side-chains generates a preference for ATP as a cosubstrate", *J. Mol. Biol.*, Vol. 347, 2005, pp. 399-414.
- [75] Battistutta, R., E. De Moliner, S. Sarno, et al., "Structural features underlying selective inhibition of protein kinase CK2 by ATP site-directed tetrabromo-2-benzotriazole", *Protein Sci.*, Vol. 10, 2001, pp. 2200-2206.

- [76] Skrzypczak-Jankun, E., K. Zhou and J. Jankun, "Inhibition of lipoxygenase by (-)-epigallocatechin gallate: X-ray analysis at 2.1 Å reveals degradation of EGCG and shows soybean LOX-3 complex with EGC instead", *Int. J. Mol. Med.*, Vol. 12, 2003, pp. 415-420.
- [77] Skrzypczak-Jankun, E., L. M. Amzel, B. A. Kroa, et al., "Structure of soybean lipoxygenase L3 and a comparison with its L1 isoenzyme", *Proteins*, Vol. 29, 1997, pp. 15-31.
- [78] McLean, K. J., P. Lafite, C. Levy, et al., "The Structure of Mycobacterium tuberculosis CYP125: molecular basis for cholesterol binding in a P450 needed for host infection", *J. Biol. Chem.*, Vol. 284, 2009, pp. 35524-35533.
- [79] Hakansson, K. and A. Liljas, "The structure of a complex between carbonic anhydrase II and a new inhibitor, trifluoromethane sulphonamide", *FEBS Lett.*, Vol. 350, 1994, pp. 319-322.
- [80] Ippolito, J. A. and D. W. Christianson, "Structure of an engineered His3Cys zinc binding site in human carbonic anhydrase II", *Biochemistry*, Vol. 32, 1993, pp. 9901-9905.
- [81] Wei, B. Q., W. A. Baase, L. H. Weaver, et al., "A model binding site for testing scoring functions in molecular docking", *J. Mol. Biol.*, Vol. 322, 2002, pp. 339-355.
- [82] Blaber, M., W. A. Baase, N. Gassner, et al., "Alanine scanning mutagenesis of the alpha-helix 115-123 of phage T4 lysozyme: effects on structure, stability and the binding of solvent", *J. Mol. Biol.*, Vol. 246, 1995, pp. 317-330.
- [83] Koes, D. R. and C. J. Camacho, "Pharmer: efficient and exact pharmacophore search", *J. Chem Inf Model*, Vol. 51, 2011, pp. 1307-1314.
- [84] Macarron, R., M. N. Banks, D. Bojanic, et al., "Impact of high-throughput screening in biomedical research", *Nat. Rev. Drug Discov.*, Vol. 10, 2011, pp. 188-195.
- [85] Keseru, G. M. and G. M. Makara, "The influence of lead discovery strategies on the properties of drug candidates", *Nat. Rev. Drug Discov.*, Vol. 8, 2009, pp. 203-212.

- [86] Alkhalfioui, F., T. Magnin and R. Wagner, "From purified GPCRs to drug discovery: the promise of protein-based methodologies", *Curr. Opin. Pharmacol.*, Vol. 9, 2009, pp. 629-635.
- [87] Irwin, J. J., T. Sterling, M. M. Mysinger, et al., "ZINC: a free tool to discover chemistry for biology", *J Chem Inf Model*, Vol. 52, 2012, pp. 1757-1768.
- [88] Trott, O. and A. J. Olson, "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading", *J. Comput. Chem.*, Vol. 31, 2010, pp. 455-461.
- [89] Ewing, T. J., S. Makino, A. G. Skillman, et al., "DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases", *J. Comput. Aided Mol. Des.*, Vol. 15, 2001, pp. 411-428.
- [90] Kramer, B., M. Rarey and T. Lengauer, "Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking", *Proteins*, Vol. 37, 1999, pp. 228-241.
- [91] Wang, R., Y. Lu and S. Wang, "Comparative evaluation of 11 scoring functions for molecular docking", *J. Med. Chem.*, Vol. 46, 2003, pp. 2287-2303.
- [92] Baumgartner, M. P. and C. J. Camacho, "Choosing the Optimal Rigid Receptor for Docking and Scoring in the CSAR 2013/2014 Experiment", *J Chem Inf Model*, 2015.
- [93] Damm-Ganamet, K. L., R. D. Smith, J. B. Dunbar, Jr., et al., "CSAR benchmark exercise 2011-2012: evaluation of results from docking and relative ranking of blinded congeneric series", *J Chem Inf Model*, Vol. 53, 2013, pp. 1853-1870.
- [94] Dunbar, J. B., Jr., R. D. Smith, K. L. Damm-Ganamet, et al., "CSAR data set release 2012: ligands, affinities, complexes, and docking decoys", *J Chem Inf Model*, Vol. 53, 2013, pp. 1842-1852.

- [95] Fourches, D., E. Muratov, F. Ding, et al., "Predicting binding affinity of CSAR ligands using both structure-based and ligand-based approaches", *J Chem Inf Model*, Vol. 53, 2013, pp. 1915-1922.
- [96] Fourches, D., R. Politi and A. Tropsha, "Target-specific native/decoy pose classifier improves the accuracy of ligand ranking in the CSAR 2013 benchmark", *J Chem Inf Model*, Vol. 55, 2015, pp. 63-71.
- [97] Grudin, S., P. Popov, E. Neveu, et al., "Predicting Binding Poses and Affinities in the CSAR 2013-2014 Docking Exercises Using the Knowledge-Based Convex-PL Potential", *J Chem Inf Model*, 2015.
- [98] Hsieh, J. H., S. Yin, S. Liu, et al., "Combined application of cheminformatics- and physical force field-based scoring functions improves binding affinity prediction for CSAR data sets", *J Chem Inf Model*, Vol. 51, 2011, pp. 2027-2035.
- [99] Huang, S. Y. and X. Zou, "Scoring and lessons learned with the CSAR benchmark using an improved iterative knowledge-based scoring function", *J Chem Inf Model*, Vol. 51, 2011, pp. 2097-2106.
- [100] Koes, D. R., M. P. Baumgartner and C. J. Camacho, "Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise", *J Chem Inf Model*, Vol. 53, 2013, pp. 1893-1904.
- [101] Kramer, C. and P. Gedeck, "Three descriptor model sets a high standard for the CSAR-NRC HiQ benchmark", *J Chem Inf Model*, Vol. 51, 2011, pp. 2139-2145.
- [102] Martiny, V. Y., F. Martz, E. Selwa, et al., "Blind Pose Prediction, Scoring, and Affinity Ranking of the CSAR 2014 Dataset", *J Chem Inf Model*, 2015.

- [103] Nedumpully-Govindan, P., D. B. Jemec and F. Ding, "CSAR Benchmark of Flexible MedusaDock in Affinity Prediction and Nativelike Binding Pose Selection", *J Chem Inf Model*, 2015.
- [104] Novikov, F. N., A. A. Zeifman, O. V. Stroganov, et al., "CSAR scoring challenge reveals the need for new concepts in estimating protein-ligand binding affinity", *J Chem Inf Model*, Vol. 51, 2011, pp. 2090-2096.
- [105] Prathipati, P. and K. Mizuguchi, "Integration of Ligand and Structure Based Approaches for CSAR-2014", *J Chem Inf Model*, 2015.
- [106] Smith, R. D., K. L. Damm-Ganamet, J. B. Dunbar, Jr., et al., "CSAR Benchmark Exercise 2013: Evaluation of Results from a Combined Computational Protein Design, Docking, and Scoring/Ranking Challenge", *J Chem Inf Model*, 2015.
- [107] Smith, R. D., J. B. Dunbar, Jr., P. M. Ung, et al., "CSAR benchmark exercise of 2010: combined evaluation across all submitted scoring functions", *J Chem Inf Model*, Vol. 51, 2011, pp. 2115-2131.
- [108] Sulea, T., Q. Cui and E. O. Purisima, "Solvated interaction energy (SIE) for scoring protein-ligand binding affinities. 2. Benchmark in the CSAR-2010 scoring exercise", *J Chem Inf Model*, Vol. 51, 2011, pp. 2066-2081.
- [109] Sulimov, A. V., D. C. Kutov, I. V. Oferkin, et al., "Application of the docking program SOL for CSAR benchmark", *J Chem Inf Model*, Vol. 53, 2013, pp. 1946-1956.
- [110] Yan, C., S. Z. Grinter, B. R. Merideth, et al., "Iterative Knowledge-Based Scoring Functions Derived from Rigid and Flexible Decoy Structures: Evaluation with the 2013 and 2014 CSAR Benchmarks", *J Chem Inf Model*, 2015.

- [111] Zhu, X., W. H. Shin, H. Kim, et al., "Combined Approach of Patch-Surfer and PL-PatchSurfer for Protein-Ligand Binding Prediction in CSAR 2013 and 2014", *J Chem Inf Model*, 2015.
- [112] Scientific, Andrew Dalke, "Chemfp v. 1.1", 2013.
- [113] Kroon, E., J. O. Schulze, E. Suss, et al., "Discovery of a Potent Allosteric Kinase Modulator by Combining Computational and Synthetic Methods", *Angew. Chem. Int. Ed. Engl.*, Vol. 54, 2015, pp. 13933-13936.
- [114] Rahaman, M. M., F. G. Reinders, D. Koes, et al., "Structure Guided Chemical Modifications of Propylthiouracil Reveal Novel Small Molecule Inhibitors of Cytochrome b5 Reductase 3 That Increase Nitric Oxide Bioavailability", *J. Biol. Chem.*, Vol. 290, 2015, pp. 16861-16872.
- [115] Moroco, J. A., M. P. Baumgartner, H. L. Rust, et al., "A Discovery Strategy for Selective Inhibitors of c-Src in Complex with the Focal Adhesion Kinase SH3/SH2-binding Region", *Chemical biology & drug design*, Vol. 86, 2015, pp. 144-155.
- [116] Grover, P., H. Shi, M. Baumgartner, et al., "Fluorescence Polarization Screening Assays for Small Molecule Allosteric Modulators of ABL Kinase Function", *PLoS One*, Vol. 10, 2015, pp. e0133590.
- [117] Wisastra, R., P. A. Kok, N. Eleftheriadis, et al., "Discovery of a novel activator of 5-lipoxygenase from an anacardic acid derived compound collection", *Bioorg. Med. Chem.*, Vol. 21, 2013, pp. 7763-7778.