UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Identification of specific inhibitors for a dual- specificity phosphatase SSH-2

Permalink

https://escholarship.org/uc/item/1tw1z0d3

Author

Mui, Matthew Kwan-Ho

Publication Date

2011

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Identification of Specific Inhibitors for a Dual-Specificity Phosphatase SSH-2

A thesis submitted in partial satisfaction of the requirements for the degree
Master of Science

in

Bioengineering

by

Matthew Kwan-Ho Mui

Committee in charge:

Shu Chien, Chair
Jason Haga
Robert Sah
Gabriele Wienhausen

2011

The Thesis of Matthew Kwan-Ho Mui is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2011

TABLE OF CONTENTS

Signature Page iii
Table of Contents iv
List of Figures vii
List of Tables viii
Acknowledgements ix
Abstract of the Thesis x
I. Introduction 1
1. The DUSP Family 3
1.1 Subcategories of DUSP 5
1.1.1 Slingshot Phosphatases (SSH) 5
1.1.2 Phosphatase of Regenerating Liver (PRL) 5
1.1.3 Cdc14 Phosphatases 5
1.1.4 PTEN-like and Myotubularin Phosphatases 6
1.1.5 Mitogen-Activated Protein Kinase Phosphatases (MKP) 6
1.1.6 Atypical DUSPs 6
1.2 General Mechanism for Dephosphorylation 6
1.3 Screened Proteins 8
1.3.1 SSH-2 8
1.3.2 Cdc14B 8
1.3.3 Cdc25A 9
1.3.4 Cdc25B 9
1.3.5 DUSP18 10
1.3.6 JSP-1 / DUSP 22 10
1.3.7 KAP 10
1.3.8 MKP-3 / DUSP6 11
1.3.9 MKP-4 / DUSP9 11
1.3.10 MKP-5 / DUSP10 12
1.3.11 MKP-6 / DUSP 14 12

1.3.12 MKP-8 / DUSP26 12
1.3.13 MTMR2 13
1.3.14 PAC-1 / DUSP2 13
1.3.15 PTEN 13
1.3.16 PRL-1 14
1.3.17 PRL-3 14
1.3.18 TMDP / DUSP13A 14
1.3.19 VHR / DUSP3 15
1.3.20 VHY / DUSP15 15
1.3.21 VHZ / DUSP 23 16
1.3.22 VH1 / DUSP12 16
1.3.23 VH3 / DUSP5 16
II. Methods 18
1. Virtual Screening 18
1.1 Overview 18
1.2 ZINC Database 19
1.3 Grid Software and Workflow 19
1.4 Clusters and CPUs 21
1.5 DOCK 6.0 21
2. In Vitro Verification 23
2.1 Compound Selection 24
2.1.1 Overview 24
2.1.2 Compound Similarity by Hierarchical Tree 24
2.1.3 Actual Compounds Tested 26
2.2 In Vitro Verification Procedure 26
III. Results 28
1. Virtual Screening 28
1.1 Energy and AMBER Scoring 28
1.2 Consensus List 29
1.3 Identified Specific Inhibitors 30

2. Structure and Similarity of Identified Compounds 33
3. Mutation of Arginine-398 40
4. Compound Verification 44
4.1 ZINC03377116 Treatment 44
4.2 ZINC06601214 Treatment 47
4.3 ZINC04307500 Treatment 49
IV. Discussion 52
1. Virtual Screening 52
1.1 General Trends 52
1.2 Combined Consensus List Implications 56
1.3 Docking Issues 57
1.4 Grid Issues 58
2. In Vitro Verification 59
V. Conclusion 62
VI. Reference 64

LIST OF FIGURES

Figure 1 | Cofilin activation and Inactivation 1
Figure 2 | The DUSP Family..... 4
Figure 3 | General Mechanism for Dephosphorylation..... 7
Figure 4 | DOCK Job Distribution 20
Figure 5 | Hierarchical Tree Example 25
Figure 6 | Hierarchical Tree of the 22 Identified Potential Specific Inhibitors for SSH-2 34
Figure 7 | Identified Potential Inhibitors with Highly Similar Structures..... 35
Figure 8 | ZINC03377116 Docked to SSH-2..... 36
Figure 11 | ZINC06601214 Docked to MKP-5 38
Figure 12 | ZINC03429974 Docked to MKP-5 39
Figure 13 | ZINC03377116 Docked to MKP-5 39
Figure 14 | Arginine-398 Mutation 41
Figure 19 | Cells Treated with ZINC03377116 45
Figure 20 | Average Intensity per Cell under ZINC03377116 Treatment 46
Figure 21 | Cells Treated with ZINC06601214 47
Figure 22 | Average Intensity per Cell under ZINC06601214 Treatment 48
Figure 23 | Cells Treated with ZINC04307500 49
Figure 24 | Average Intensity per Cell under ZINC04307500 Treatment 50
Figure 25 | ZINC04172162 Bound to SSH-2..... 53
Figure 26 | ZINC05373221 and ZINC04307500..... 54
Figure 27 | ZINC05260817..... 55

LIST OF TABLES

Table 1 | Computational clusters. 21
Table 2 | Actual Tested Compounds. 26
Table 3 | Energy Ranking. 28
Table 4 | AMBER Ranking. 29
Table 5 | Consensus Ranking. 29
Table 6 | Combined Consensus List Showing Identified Top Specific Inhibitors. 30
Table 7 | Disparity Scores for Identified Specific Inhibitors. 31
Table 8 | Consensus Ranks for SSH-2 and MKP-5 of Discussed Compounds. 40
Table 9 | Active Site Sequence Alignment. 56

ACKNOWLEDGEMENTS

I would like to acknowledge Dr. Shu Chien for his support as chair of my committee and Dr. Jason Haga, Dr. Robert Sah and Dr. Gabriele Weinhausen for their time as members of the committee. I would like to especially thank Dr. Jason Haga for his continual guidance throughout the years.

I would also like to acknowledge Marshall Levesque, Ph.D. candidate at University of Pennsylvania, for his assistance in computational modeling. Lastly, I would like to thank Dr. Sung Hur for his help in actin quantification, and Phu Nguyen and Brian Tsui of the Vascular Molecular Bioengineering Laboratory for their help in biochemical techniques in the laboratory.

Molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR001081).

ABSTRACT OF THE THESIS

Identification of Specific Inhibitors for a Dual-Specificity Phosphatase SSH-2

by

Matthew Kwan-Ho Mui

Masters of Science in Bioengineering

University of California, San Diego, 2011

Professor Shu Chien, Chair

Slingshot-2 (SSH-2), one of 61 dual specificity phosphatases (DUSPs), has been shown to be associated with the progression of cancer and Alzheimer's disease by other studies. This DUSP activates cofilin through dephosphorylation and serves to regulate several cell functions, including growth and movement. Finding a specific inhibitor for SSH-2 may have profound impact in medicine because of its disease association. This study executed a large-scale virtual screening experiment in an attempt to find specific inhibitors for SSH-2 from a small molecule, chemical database containing over 2 million

compounds, called ZINC. Results from 23 DUSP screenings, specifically SSH-2, Cdc14B, Cdc25A, Cdc25b, DUSP18, JSP-1, KAP, MKP-3, MKP-4, MKP-5, MKP-6, MKP-8, MTMR2, PAC-1,PTEN, PRL-1, PRL-3, TMDP, VHR, VHY, VHZ, VH1 , and VH3, were used to identify specific inhibitors from the large ZINC database. Eleven compounds were identified as potential specific inhibitors for SSH-2 with a disparity mean of at least 8000 and standard deviation less than 7000.

Preliminary testing on the effectiveness of three compounds, ZINC03377116, ZINC04307500, and ZINC05501214, from the list of 11 identified compounds was investigated *in vitro* against HeLa cells. Cells treated with ZINC03377116 and ZINC06601214 showed a decrease in actin filaments. This indicates that these compounds did not inhibit the activity of SSH-2. However, cells treated with ZINC05501214 showed an increase in actin stress fibers at low compound concentrations, suggesting that the activity of SSH-2 was inhibited.

I. INTRODUCTION

Actin is a globular protein that is highly conserved across eukaryotic cells [1]. Through actin polymerization, actin molecules form microfilaments and thin filaments that participate in many cellular activities, including motility, cell division, and signaling. The assembly of these actin filaments is assisted by proteins within the actin-polymerizing factor (ADF) family and the disassembly is specifically regulated by the cofilin family of proteins. Like many other cellular events, this dynamic reorganization of actin filaments is well coordinated and regulated. In particular, the cofilin proteins are regulated through phosphorylation and dephosphorylation, processes by which a phosphate group is either added or removed from the protein, respectively.

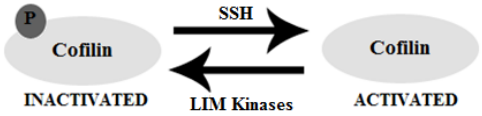


Figure 1 | **Cofilin activation and Inactivation.** Cofilin is activated through dephosporylation by SSH phosphatases and inactivated by phosphorylation by LIM kinases.

Cofilin is known to be activated in part through dephosphorylation at a serine residue by a group of proteins called Slingshot phosphatases (SSH) that belongs to a family of dual-specificity phosphatases (DUSPs) that dephosphorylate targets at both phosphoserine and phospothreonine/tyrosine residues [2]. It is activated through phosphorylation by LIM kinases, as shown in Figure 1. Studies have shown that over-expressed cofilin forms cofilin-actin rods that inhibit transport in the cell. This leads to

an accumulation of amyloid precursor protein and contributes to the pathology of Alzheimer's disease [3]. In addition, fine tuning the balance between phosphorylated and non-phosphorylated form of cofilin is thought to be able to slow down the migratory potential of cancer [4]. Although cofilin has direct association with these two diseases, it also participates in other important cellular processes through actin reorganization. Inhibiting cofilin as a treatment to cancer and Alzheimer's disease is therefore not feasible as it would be lethal to normal cells as well as diseased cells [5]. However, the activity of cofilin can be indirectly modulated by inhibiting the ability of SSH to activate cofilin. Thus, a specific inhibitory compound for SSH may have profound impact in therapeutic medicine [5].

Traditionally, protein inhibitors are found through *in vitro* testing to determine the affinity of each compound to the target. With the current large availability of compounds, the time and costs of testing them one-by-one is enormous and unrealistic to pursue. Through molecular modeling and simulation, a list of high potential inhibitors for a target protein can be generated in a cost- and time-efficient manner. By performing virtual screening experiments over a grid of supercomputers, the computational time can be reduced even further. This study took advantage of grid technologies of the Pacific Rim Applications and Grid Middleware Assembly (PRAGMA) grid for the identification of potential specific inhibitors for SSH-2 through virtual screening experiments.

Although all three SSH homologues, SSH-1, SSH-2 and SSH-3, contribute to the regulation of cofilin, only the three-dimensional crystal structure of SSH-2 has been determined in recent years. This study has discovered a list of specific inhibitors for SSH-2 through virtual screening experiments using the three-dimensional crystal

structures of 23 DUSPs, specifically SSH-2, Cdc14B, Cdc25A, Cdc25B, DUSP18, JSP-1, KAP, MKP-3, MKP-4, MKP-5, MKP-6, MKP-8, MTMR2, PAC-1,PTEN, PRL-1, PRL-3, TMDP, VHR, VHY, VHZ, VH1, and VH3. The other DUSPs were screened in addition to SSH-2, because all DUSPs are involved in one or more cellular processes and only inhibitors specific to SSH-2 will deter the disease progression without causing considerable damage to other cells in the body. Each screening provides the relative binding strength of a list of compounds to the screened target protein. By cross-referencing these results, specific inhibitors for SSH-2 were identified. Ideally, all 61 members of the DUSP family should be screened; however, the crystal structures of all 61 proteins have not yet been determined [6]. Since these virtual screening experiments rely heavily on the crystal structures of these proteins for accurate results, only a subset of the entire DUSP family was screened for this study.

This paper describes the methods of these virtual screening experiments using parallel processing through grid computing. Difficulties and problems with using grid computing are also discussed. Furthermore, the results from each DUSP screening were compared with that of SSH-2 to determine the specificity of each compound. From this, a small set of specific inhibitors for SSH-2 were identified. Several identified inhibitors were then verified through direct application to cancer cells, and these results are also included in this paper.

1. The DUSP Family

Protein tyrosine phosphatases (PTPs) and their counterpart, protein tyrosine kinase (PTKs), regulate a wide range of physiological processes and signaling pathways

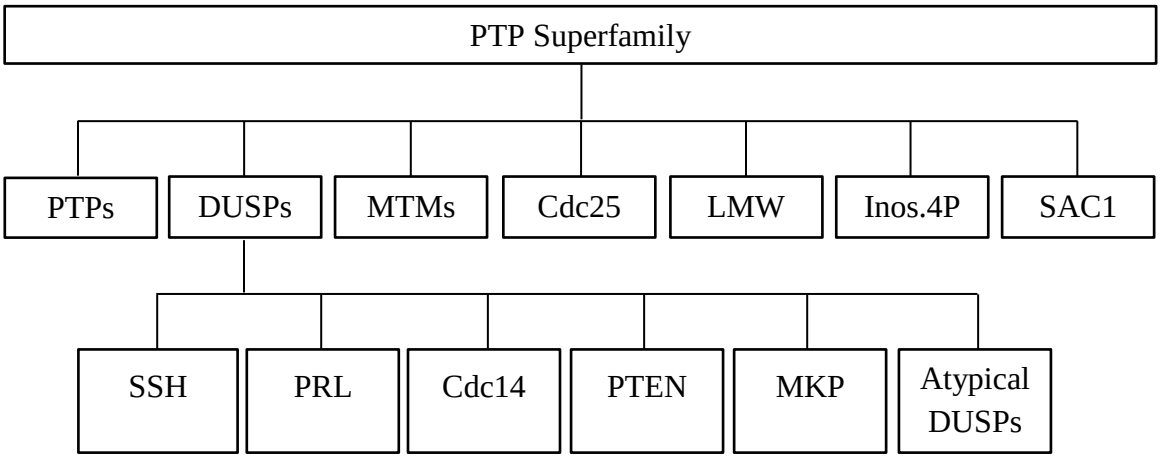


Figure 2 | **The DUSP Family.** The DUSP family is one of seven subcategory of the PTP superfamily. The classification of DUSPs is not standardized, however, proteins within the DUSP family can be subdivided into slingshot (SSH) phosphatases, phosphatase of regenerating liver (PRL), Cdc14 phosphatases, PTEN-like and myotubularin phosphatases, mitogen-activated protein kinase phosphatases (MKPs), or atypical DUSPs [7, 8].

through reversible phosphorylation and dephosphorylation [7]. The PTP superfamily can be divided into seven categories based on structural homology and substrate preference as summarized in Figure 2 [8]. One of the seven categories is the DUSP family, which includes proteins with a highly conserved PTP loop sequence (HCXXXXXR) [7]. The general structure of the sequence within the DUSP family positions the cysteine residue at the base of the phosphate-binding pocket while the arginine residue loops back to the catalytic site to assist with the dephosphorylation process. Although the classification of some DUSPs has not been standardized, members within the family generally fall within six categories: slingshot (SSH) phosphatases, phosphatase of regenerating liver (PRL), Cdc14 phosphatases, mitogen-activated protein kinase phosphatases (MKPs), or atypical DUSPs [7].

1.1 Subcategories of DUSP

1.1.1 Slingshot Phosphatases (SSH)

Slingshot phosphatases are first discovered in *Drosophilia* through mutagenesis studies. The results from these studies show disorganized epidermal morphology because the SSH proteins are involved in the actin depolymerization pathway [9]. Three mammalian SSH proteins have been discovered thus far and they dephosphorylate targets at phosphoserine and phosphothrenine residues.

1.1.2 Phosphatase of Regenerating Liver (PRL)

The exact role of PRL proteins are poorly understood, however, these proteins are overexpressed in various cancers [10]. The three known PRLs, PRL-1, PRL-2 and PRL-3, all share a high sequence homology and contain the signature PTP catalytic domain [11]. Through reversible oxidation, these proteins can be inactivated by forming disulfide bonds within the catalytic pocket.

1.1.3 Cdc14 Phosphatases

The mammalian Cdc14 subfamily of proteins is poorly understood, but they are thought to be involved in control of centrosome and spindle fibers through studies of *S.cerevisiae* and *C.elegans* [12]. There are four Cdc14 phosphatases (KAP, Cdc14A, Cdc14B and PTP9Q22) discovered thus far and they are related to Cdc25 phosphatases of the PTP superfamily.

1.1.4 PTEN-like and Myotubularin Phosphatases

Five PTEN-like and 16 myotubularin phosphatases have been discovered so far. They mainly dephosphorylate D3-phosphyorylated inositol phospholipids. In human cancers, PTEN is frequently mutated, silenced or deleted, leading to uncontrolled proliferation [7].

1.1.5 Mitogen-Activated Protein Kinase Phosphatases (MKP)

This group of DUSPs dephosphorylates mitogen-activated protein kinases (MAPKs) at the phosphothreonine and phosphotyrosine residues. Like other DUSPs, they adopt similar catalytic site structure and the mutation of the catalytic cysteine causes the proteins to become inactive. The regulation of MAPKs leads to many physiological processes, including cellular proliferation, differentiation and apoptosis [13].

1.1.6 Atypical DUSPs

Atypical DUSPs share many features, including catalytic sequence and structure with other DUSPs, especially with the proteins from the MKP subgroup. However, atypical DUSPs are phylogenetically distinct from MKPs; thus, they are categorized in their own separate category [7].

1.2 General Mechanism for Dephosphorylation

Although each DUSP differs slightly in structure and composition, certain residues are conserved across the entire family. The conserved residues, specifically a

cysteine, an arginine and an aspartic acid, from the catalytic pocket are responsible for the overall mechanism of dephosphorylation of DUSPs [14].

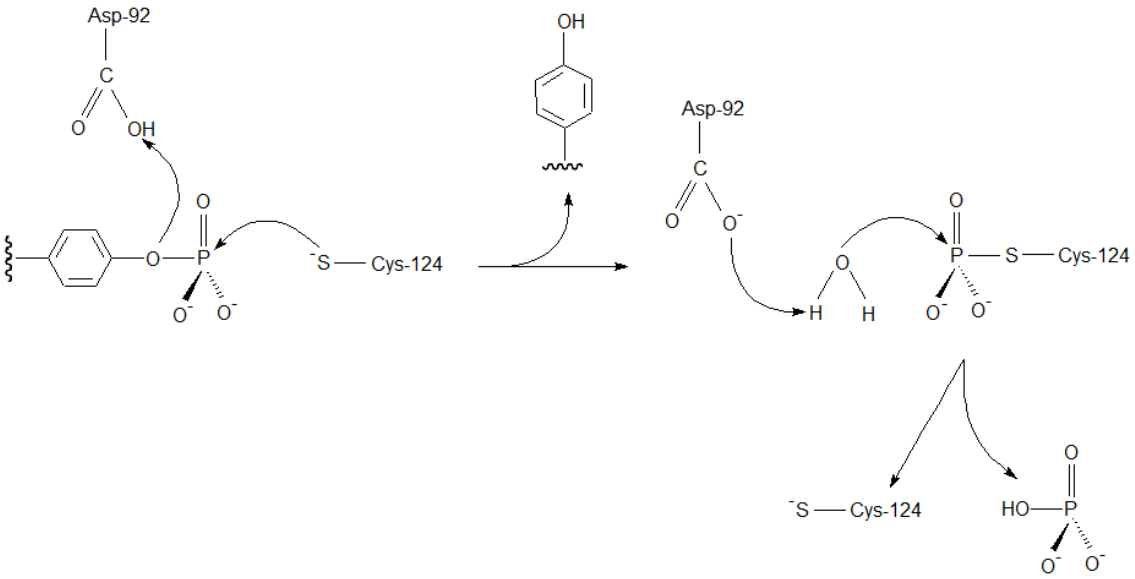


Figure 3 | **General Mechanism for Dephosphorylation.** The catalytic cysteine acts as a nucleophile and accepts the incoming amino-phosphate while the aspartic acid residue enhances the catalytic activity by acting as a general acid. The phosphate is transferred to a water molecule in the next step, and the enzyme is restored [15]. The amino acid numbers refers to the position in the VHR sequence.

The dephosphorylation of a target amino acid follows two general steps as shown in Figure 3. First, the phosphate moiety enters the active site. The cysteine of the signature PTP loop acts as a nucleophile and accepts the phosphate group, forming a phosphocysteine intermediate. Since phosphocysteines are rare in biology, the arginine from the PTP loop stabilizes the transition state by forming three hydrogen bonds with two oxygen atoms from the phosphate group. During this step, a conserved aspartic acid upstream of the PTP sequence further enhances the catalytic activity by acting as a

general acid by donating a proton to the leaving group. Next, the phosphate is transferred to a water molecule in the catalytic site and the enzyme is restored [14, 15].

1.3 Screened Proteins

1.3.1 SSH-2

Of the three SSH proteins, only the crystal structure of SSH-2 has been resolved. SSH-2 (PDB ID: 2NT2) has similar motifs as many other DUSPs, including a five-stranded β-sheet surrounded by four α-helices on one side and two on the other. Structurally, SSH-2 is most similar to VHR in the DUSP family, showing approximately 80% sequence similarity. The catalytic pocket of SSH-2, however, is shallower and wider than that of VHR and is consistent with the size of phosphoserine in cofilin [6].

1.3.2 Cdc14B

The two human isoforms of Cdc14 are poorly understood, however, Cdc14 in *S.cerevisiae* is known to promote inactivation of the mitotic cyclin-dependent kinase at the end of anaphase and Cdc14 of *C.elegans* is recognized to localize key components to the central spindle in anaphase and midbody in telophase [12]. This suggests that the two human isoforms of Cdc14 are regulators of mitotic exit and cytokinesis. The catalytic site of Cdc14B (PDB ID: 1OHE) was solved through x-ray crystallography after trysinolysis of the full protein because the full protein did not readily crystallize. The catalytic site was solved to a 2.5 Å resolution and structure contains two similar domains. Both domains adopt the DUSP-like fold and the catalytic center is defined by the PTP

signature motif (HCXXXXXR). Also, the β-domain of the protein is structurally most similar to PTEN, one of the 23 DUSPs screened in this study [12].

1.3.3 Cdc25A

Since the classification of DUSPs has not been standardized, the Cdc25 group of proteins may be categorized either as a distinct family of the PTP superfamily or listed under the DUSP family [7, 16]. Because Cdc25 contains the conserved catalytic sequence and share the same catalytic mechanism as other DUSPs, they are considered as part of the DUSP family for the purpose of this study [16]. The crystal structure of Cdc25A (PDB ID: 1C25) revealed a five-stranded β-sheet surrounded by three α-helices on one side and two on the other. In comparison to other DUSPs, the active site of Cdc25A is extremely shallow and Phe-432 may be used to recognize phosphotyrosine substrates [16].

1.3.4 Cdc25B

Cdc25B (PDB ID: 1QB0) is known to be overexpressed in a wide variety of cancers and it important in the regulation of cell proliferation [17]. In comparison to Cdc25A, the catalytic site of Cdc25B contains several distinct differences. First, the carbonyl of residue 477 points away from the catalytic pocket. Second, the orientations of Arg436 and Glu431 are also different [18].

1.3.5 DUSP18

DUSP18 (PDB ID: 2ESB) adopts a similar catalytic structure as other DUSPs, but it contains approximately 30 residues in the C-terminus that forms two antiparallel β-strands that exhibit significant interactions with the catalytic site. There is a substantial overlap in structures between DUSP18 and other DUSPs, but major differences exist in the areas around the active site that gives DUSP18 its unique structure and substrate specificity [19].

1.3.6 JSP-1 / DUSP 22

The human JNK stimulatory phosphatase-1 (JSP-1) is genetically identical to two other DUSPs, VHX and VHR. The actual JSP-1(PDB ID: 1WRM) protein consists of 184 amino acids and it is expressed in many different cell types. The physical structure of JSP-1 is similar to VHR and MKP3 with approximately 35-40% sequence homology. The N-terminus follows the motif found in many other DUSPs, but the C-terminus does not [20].

1.3.7 KAP

KAP (PDB ID: 1FPZ) is an important regulator of the cell cycle, because it activates CDK2, one of many proteins in the cyclin-dependent protein kinases (CDKs) family that coordinates the progression of the cell cycle. Like many other DUSPs, KAP has a central twisted five stranded β-sheet surrounded by six α-helices. The marked difference of KAP that is not seen in other DUSPs is an antiparallel β3/β4 hairpin that

projects above the catalytic surface of the molecule. This hairpin loop increases the specificity for CDK2 and is used for forming the CDK2-KAP complex [21].

1.3.8 MKP-3 / DUSP6

The main function of MKP-3 is to inactivate and dephosphorylate mitogen-activated protein kinases (MAPK). A wide-range of cell-surface stimuli can activate MAPKs through signal transduction that leads to cell proliferation, differentiation and cell cycle progression. MKP3 (PDB ID: 1MKP) is highly specific towards two MAPKs, namely ERK1 and ERK2. Structural comparison between VHR and Pyst1 shows that the α2 and α3 helices of VHR are absent in MKP-3. The overall structure of MKP-3 contains six-stranded β-sheet surrounded by four helices on one side and an extra short helix on the other. The catalytic site is formed by a loop known as the phosphate-binding loop (PTP-loop) [22].

1.3.9 MKP-4 / DUSP9

The eighth protein screened in this study is MKP-4 (PDB ID: 2HXP). It belongs to the MKP family as suggested by its name. It has 57% amino acid sequence identity to Pyst1 and 61% identity to Pyst2. Structurally, MKP-4 is very similar to other DUSPs, however, it has an extended active site motif and shows specificity towards ERK family members when expressed in mammalian cells [23].

1.3.10 MKP-5 / DUSP10

Similarly to other MKPs, MKP-5 (PDB ID: 2OUD) dephosphorylate other MAPKs. MKP-5 is thought to interact with JNK, since the activity level of JNK is shown to increase in MKP-5 knock-out mice. Although the catalytic domain is highly conserved across MKPs, the binding domain of MKP5 is unique among other MKPs in that it contains an extra segment at the N terminal. The exact function and mechanism of action of this extra segment remains unknown. Furthermore, the binding domain of MKP5 is similar with other DUSPs, containing a parallel 5-stranded β-sheet surrounded by α-helices on both faces. It is also a close structural homolog to VHR [24].

1.3.11 MKP-6 / DUSP 14

MKP-6 (PDB ID: 2GWP) is known to preferentially dephosphorylate JNK and ERK and act as a negative regulator of CD28. The truncated form of the protein of residues 2-191 was solved at 1.88 A. When compared against DUSP18 and VHR through structural alignments, the catalytic cysteine residue is conserved. The binding pocket of MKP6 is also shallow, approximately 7.0 Å deep, thus allowing either phosphotyrosine or phosphothreonine/serine to enter [25].

1.3.12 MKP-8 / DUSP26

The structural data of MKP-8 (PDB ID: 2E0T) is available on the Protein Databank, however, the actual paper that accompanies the data has yet to be published.

Studies suggest that MKP-8 may act as a tumor suppressor in certain cancers, including neuroblastoma and ovarian cancer [26, 27].

1.3.13 MTMR2

Mutation in MTMR2 (PDB ID: 1M7R) has been discovered to cause neuropathy that is characterized by abnormal myelination of peripheral nerves [28]. The surface of the active site of MTMR2 is strongly electropositive, while the remaining surfaces are mainly electronegative. Furthermore, the catalytic pocket is wider and deeper than VHR, thus contributing to substrate specificity [29].

1.3.14 PAC-1 / DUSP2

PAC-1 (PDB ID: 1M3G) belongs to the MKP group of the DUSP family. The three-dimensional structure of the phosphatase bound protein was solved through nuclear magnetic resonance spectroscopy. The catalytic cysteine was mutated to a serine for the structural analysis process. While the enzymatic activity of the protein is negated, studies have shown that substrate recognition is not affected by the mutation. The NMR-determined overall structure is very similar to MKP3, however, small differences, such as one extra turn in α1 helix and shorter turn in the α4 helix in PAC1 versus MKP3, exist [27].

1.3.15 PTEN

Human cancers often involve the mutation of the tumor suppressor PTEN (PDB

ID: 1D5R). The phosphatase domain of the protein again contains five-stranded β-sheet surrounded by six α-helices. Its overall structure is similar to VHR with 121 Cα atoms overlap when superimposed. PTEN, however, has an 11 amino acid residue insertion in the β2-α1 loop and a 4 residue insertion in the pα5-α6 loop [30].

1.3.16 PRL-1

PRL-1 (PDB ID: 1X24) was first discovered because of its role in liver regeneration. Further studies have shown that its expression level is elevated in certain tumor cells lines. Its crystal structure was solved from N-terminal or C-terminal truncations. The catalytic domain of PRL1 resembles other DUSP and that it is most similar to Cdc14, KAP and PTEN, with decreasing order of similarity [31].

1.3.17 PRL-3

The solution structure of PRL-3 showed that it is highly homologous to VHR, PTEN and KAP. Unlike other DUSPs, PRL3 places the catalytic cysteine and arginine at opposite ends of the catalytic P-loop. This particular arrangement has been shown to be responsible for the substrate specificity. In addition, the catalytic pocket of PRL3 is also extremely shallow and may further contribute to its specificity [32].

1.3.18 TMDP / DUSP13A

TMDP (PDB ID: 2GWO) or testis- and skeletal-muscle-specific dual-specificity phosphatase is shown to be involved in spermatogenesis. The structure is composed of

four molecules (A, B, C, and D) that are arranged in two pairs of dimers, A:B and C:D. The catalytic site has a much flatter topology in comparison to VHR. The entrance to TMDP also has a wider pocket than VHR, thus allowing both phosphotyrosine and phosphothreonine to enter [33].

1.3.19 VHR / DUSP3

In comparison to protein tyrosine phosphatases (PTPs), VHR or DUSP3 (PDB ID: 1VHR) has a shallower catalytic pocket. The shallower pocket allows dephosphorylation at phosphoserine, phosphotyrosine and phosphothreonine residues whereas PTP is only specific to phosphotyrosine. Moreover, the α1-β1 loop is shorter than other PTPs and exposes the positively charged binding pocket that is lined with several arginine residues. These differences may be subtle, but they all contribute to the specificity of the VHR to the target substrates [34].

1.3.20 VHY / DUSP15

VHY is a testes specific DUSP that has been shown to be active in spermatocytes and spermatids and may be a regulator of meiosis. The structure again has a five-stranded β-sheet with five α-helices on one surface and a single α-helix on the other. Its structure is similar to that of VHR and has 132 Cα atoms overlap when they are superimposed on top of each other. The difference between VHR and VHY is that VHR has a deep and narrow active site while VHY has an active site that is flat and wide [35].

1.3.21 VHZ / DUSP 23

X-ray crystallography of VHZ (PDB ID: 2IMG) reveals a 6 Å deep catalytic pocket. The surface of the pocket is composed of hydrophobic and positively charged residues, which play an important role in substrate recognition and stabilization. The overall structure of VHZ is very similar to other DUSPs in that a five-stranded β-sheet is surrounded by five α-helices [36].

1.3.22 VH1 / DUSP12

Although VH1 (PDB ID: 3CM3) is a Vaccinia virus protein, many eukaryotes, including humans encode for DUSPs that are very similar to VH1. In particular, the human genome encodes for at least 38 VH1-like proteins. The topology of VH1 is very similar to other DUSPs, especially of VHZ and VHR. The catalytic pocket of VH1 is approximately 6 Å deep and the residues within the pocket that are mainly responsible for the dephosphorylation process are arginine-116, aspartate-79 and cysteine-110 [37].

1.3.23 VH3 / DUSP5

VH3 or DUSP5 (PDB ID: 2G6Z) belongs to the MKP subfamily and preferentially dephosphorylates extracellular signal-regulated kinases (ERK). VH3 dephosphorylates targets at phosphoserine/threonine and phosphotyrosine residues. Structurally, VH3 adopts a similar fold to many DUSPs, with a central twisted five-stranded β-sheet surrounded by six α-helices. The two DUSPs that VH3 are most structurally to are MKP-3 and VHR. The structural alignment of VH3 and VHR shows

that 136 of 146 Cα atoms were superimposed. In comparison to MKP3, however, the orientation of the backbone amides is significantly different in that they do not point towards the active pocket [38].

II. METHODS

1. Virtual Screening

1.1 Overview

Although molecular modeling and simulation provide an alternative to the drug discovery process, the time it takes to complete an entire screening of a large library of compounds on a single processor may still take years. Using parallel processing and grid technologies, this enormous, computationally-heavy task can be divided into multiple jobs and distributed across a grid of supercomputers. By doing so, the computational time can be cut down significantly; however, the expense of building and maintaining supercomputers creates an obstacle, especially for low-budget projects in academia. Thus, this study used a shared grid of supercomputers among various universities and research institutes around the world called the PRAGMA grid.

Starting with a chemical database, each molecule was screened against a target enzyme to determine their affinity to the target by computationally simulating their interactions. Each compound was first given an energy score based on van der Waals and electrostatic forces. The molecules were screened again and given an Assisted Model Building with Energy Refinement (AMBER) score, which takes into account solvation energies. These molecules were then ranked and compiled into a master list showing possible molecules with high affinity to the target. SSH-2 was first screened against the entire ZINC database of 2,066,906 compounds [39]. Only the top 1% of the library that showed strong association to SSH-2 was retained for further screenings of the other DUSPs. Since the desired specific inhibitor must bind to SSH-2 tightly, eliminating the

compounds that do not bind well to SSH-2 saved a considerable amount of computation time. After each screening, each compound was assigned a score based on the binding strengths to the screened protein. The compounds were then ranked based on their scores into a consensus list, showing in order the best to the worst binding compound. Several specific inhibitors to SSH-2 were identified by cross-referencing the lists to find the difference between the SSH-2 and another DUSP family member's consensus ranking for each individual compound. A large difference in rankings suggested that an inhibitor will bind more favorably to SSH-2 and the most specific inhibitor was selected by finding the compound with the largest ranking difference.

1.2 ZINC Database

The chemical database used to screen SSH-2 was a "drug-like" subset of the ZINC database containing 2,066,906 compounds [40]. This database was chosen because of its favorable pharmaceutical properties. The compounds were around 150 – 500 daltons with potent hydrogen bonding properties [41]. This study used the results from the SSH-2 screening to further reduce the subset to 20,935 compounds by choosing to screen only the top 1% of the library that demonstrated strong binding to SSH-2. By eliminating the low ranked compounds that had poor binding scores to SSH-2, the total screening time for each target protein was significantly reduced.

1.3 Grid Software and Workflow

This study required an application that facilitated for protein-ligand interaction

simulations over multiple processors. DOCK 6.0, developed by UCSF [42], satisfied these requirements and was chosen for the study because of its versatile screening methods and functionality.

DOCK has a built-in message passing interface (MPI) that enabled parallel computing in a single computing cluster system. The software was implemented on the grid through Opal Operation Provider (Opal OP), a toolkit that wraps scientific applications as Web services. To establish communication between clusters, Jakarta Tomcat was installed, which gives user access to remote clusters allowing jobs to be submitted through a portal with a unique job ID.

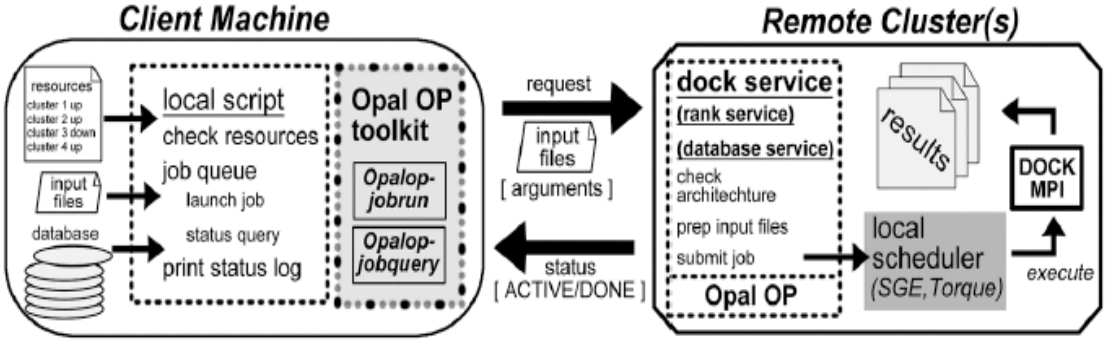


Figure 4 | **DOCK Job Distribution.** DOCK MPI job distribution from master cluster to remote cluster(s) [43].

For the docking experiments, automated scripts were written such that slices of the database were distributed to remote clusters through a scheduler. The workflow of a typical screening is shown in Figure 4. The master cluster performed checks on the availability of remote clusters at regular intervals and distributed jobs accordingly through Opal OP. When the remote cluster received a job, it prepared all the input files and sent each compound to a compute node for docking using the DOCK MPI function.

1.4 Clusters and CPUs

This study utilized a total of 6 clusters located at several different locations, including Japan, Malaysia, Puerto Rico, Switzerland and USA, on the PRAGMA grid. Table 1 summarizes the clusters, locations, and the number of CPUs used for this study. Each cluster contained a different number of CPUs ranging from 20 to 110. This study used approximately 25% of the available CPUs on each cluster to ensure that there would be enough computational resources available for other scientific projects.

Table 1 | **Computational clusters.** Summary of clusters, locations and CPUs used for this study.

Cluster Name	Region	CPUs used
Aurora	Malaysia	10
Café	Japan	10
Komolongma	Puerto Rico	40
Ocikbpra	Switzerland	8
Rocks-52	USA	14
Tea	Japan	20

1.5 DOCK 6.0

The screening process was divided into two phases based on two DOCK scoring methods. The first phase used an energy scoring method. The receptor was first prepared using a molecular visualization software called Chimera [44]. Hydrogen atoms and charges were added as necessary for the docking process. A receptor energy grid was calculated with 0.3 angstroms spacing using the DOCK accessory program GRID. Using this energy grid, energy scores were calculated based on van der Waal and electrostatic forces between receptor and ligand using Equation 1 [45].

$$E = \sum_{i=1}^{lig} \sum_{j=0}^{rec} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + 332.0 \frac{q_i q_j}{D r_{ij}} \right] \qquad \text{(Eq. 1)}$$

Equation 1 sums over ligand atoms *i* and receptors atoms *j*, where A_{ij} and B_{ij} are van der Waals repulsion and attraction terms and r_{ij} is the distance between atoms *i* and *j*. The electrostatic term is represented by $q_i q_j$ and *D*, where q_i and q_j are point charges on atoms *i* and *j* and *D* is the electric function. The number 332.0 is a conversion factor that converts electrostatic energy into kilocalories per mole. Together, these terms give the interaction energy of the ligand-receptor complex [45]. The parameters used for the grid-energy scoring method were 900, 60, 20 and 50 for "max orientations", "pruning cluster cut off", "simplex anchor max", and "simplex grow max" parameters, respectively. Although the accuracy of the docking models will increase with larger parameter values, the computation time for each protein-ligand simulation will also increase. The parameter values used for this study were selected such that the virtual screening experiments would complete in a reasonable amount of time while a high level of accuracy was maintained.

The grid-energy scoring method is fast, but treats the entire ligand and receptor as static structures to find the best orientation of their interaction using the "anchor-and-grow" algorithm [46]. This algorithm uses rigid segments, such as a benzene ring, as structural anchors to find the best orientation that will fit into the protein. The rest of the non-rigid compound, such as a long carbon chain, is then generated or "grown" from the anchor. This regenerated compound is now in the best binding orientation to the catalytic site. The grid-based energy score is calculated after this process based on the Van der Waal forces between the compound and the protein. Once all the compounds were scored using this method first, the entire database was re-ranked back into a master list showing the best to the least binding compound.

Scored compounds from the first phase as well as their best binding orientation to

the receptor were passed on to the second phase, which used the AMBER scoring method. The orientations of the compounds from the previous screening method were required, because AMBER does not reorient them. Also unlike the energy scoring method, AMBER allows the receptor to be flexible and takes into account of solvation energies. The drawback of this method is that each ligand must be prepared separately with another DOCK accessory program, thus requires significant computation time. Some of the specific parameters used for AMBER scoring are "amber score movable distance cutoff", "AMBER score before md minimization cycles", "AMBER score md steps", and "AMBER score after md minimization cycles" with assigned values of 3.0, 250, 2700, and 250 respectively. Similarly to the parameters used for the energy scoring method, larger input values will increase simulation accuracy, but at the same time, increase computation time for these virtual experiments. These numbers were again selected to achieve a high level of model accuracy and to complete each screening in a reasonable amount of time.

2. In Vitro Verification

Although potential specific inhibitors were identified through virtual screening experiments, *in vitro* verification is necessary to confirm that validity of the simulated results. This is because these virtual experiments were based on interactions between compound and the target only. Factors such as the solubility of the compound, the ability for the compound to diffuse through the cell membrane, or cell toxicity were unaccounted

for. The results from the *in vitro* verification experiments will provide the additional data necessary for these compounds to be developed as therapeutic drugs against cancer.

2.1 Compound Selection

2.1.1 Overview

Several factors were considered when choosing compounds for *in vitro* verification. First, compounds were chosen from the list of most potential inhibitors described in previous sections. Second, they were selected based on their molecular structure. Testing compounds with a variety structures will provide insight to the general molecular structure that best inhibits SSH-2. This means that only one of two compounds with similar structures should be tested to avoid redundancy. Compound similarity was determined quantitatively using the Tanimoto coefficient as a similarity measure and plotted in a hierarchical tree as described in the next section. Lastly, due to funding constraints, only a very small subset of compounds was tested *in vitro*.

2.1.2 Compound Similarity by Hierarchical Tree

The degree of similarity between two compounds was determined by the Tanimoto coefficient using a small molecule analytical package called ChemmineR [47]. The Tanimoto coefficient is defined by Equation 2, where *a* and *b* are the number of unique atom pairs between two compounds, while *c* is the number of atom pairs in common.

$$S(A,B) = \frac{c}{a+b+c} \qquad \text{(Eq. 2)}$$

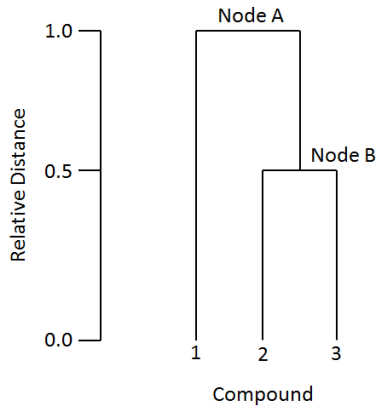


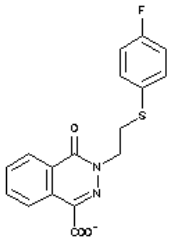
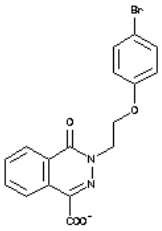
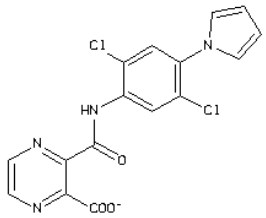
Figure 5 | **Hierarchical Tree Example.** Compound similarity is plotted as a hierarchical tree based on their calculated Tanimoto coefficient. The degree of similarity between compounds is based their nodal distance.

Based on this definition, the Tanimoto coefficient gives the ratio between shared atom pairs versus the total number of atom pairs for the two compounds. Identical compounds will then have a coefficient with the value equal to unity [48].

Through ChemmineR, compound similarity was plotted as a hierarchical tree or a dendrogram based on their calculated Tanimoto coefficient. The compounds were clustered in such a way that highly similar compounds were grouped closely together with the node distance determining the relative similarity of the clustered compounds with the entire database. Figure 5 illustrates an example of a hierarchical tree, where hypothetical compounds 1, 2, and 3 are sorted. The degree of similarity between compound 2 and 3 is given by Node B and the similarity between compound 1 and 2 and compound 1 and 3 are given by Node A. Since the distance of Node B is less than the distance of Node A, compounds 2 and 3 are more similar than they are with compound 1. The top 500 potential inhibitors for SSH-2 were sorted in this same fashion.

2.1.3 Actual Compounds Tested

Table 2 | **Actual Tested Compounds.** The structure, molecular formula, and molecular weight for the tested compounds are listed in this table.

ZINC ID	ZINC06601214	ZINC03377116	ZINC04307500
Structure			
Molecular Formula	$C_{17}H_{13}FN_2O_3S$	$C_{17}H_{13}BrN_2O_4$	$C_{16}H_{10}Cl_2N_4O_3$
Molecular Weight	344.4 g/mol	389.2 g/mol	377.2 g/mol

Three compounds were tested in this study. Their corresponding ZINC IDs were ZINC06601214, ZINC03377116, and ZINC04307500. The structure, molecular formula, and molecular weight are summarized in Table 2. ZINC06601214 and ZINC03377116 were purchased through Enamine while ZINC04307500 was purchased through Maybridge.

2.2 In Vitro Verification Procedure

The *in vitro* verification experiment aimed to provide preliminary results on the effectiveness of several identified inhibitors of SSH-2 through analysis of actin filaments. Cells treated with these compounds were expected to have an increase in actin filament structures, because these compounds were designed to inhibit SSH-2. By inhibiting SSH-2, phosphocofilin would not be dephosphorylated and remains inactive. The depolymerization of actin filaments would then be inhibited, leading to an increase in actin filaments structures.

HeLa cells, cells derived from cervical cancer, were used for the *in vitro* verification portion of the study. Cells were seeded onto 2cm diameter plates and allowed to attach and proliferate for 24 hours. Compounds dissolved in dimethyl sulfoxide (DMSO) were applied the next day at 100µm, 10µm, 1µm, 0.1µm, and 0.01µm concentrations and a DMSO vehicle control was used. The cells were incubated for a period of 24 hours and fixed using a 2% paraformaldehyde solution. Actin filaments were then stained with rhodamine phalloidin and images were acquired with a fluorescent microscope at 40X magnification. The fluorescence intensity of the actin stress fibers was quantified from the entire imaging field. Since the number of cells may differ for each image, the fluorescence intensity was divided by the total number of cells within the acquired image to obtain the average fluorescence intensity per cell. The intensity was then normalized to the intensity of the vehicle control within the experimental set.

III. RESULTS

1. Virtual Screening

1.1 Energy and AMBER Scoring

The degree of binding of each compound to a target protein was determined based on the two scoring methods, grid-based energy score and AMBER score. Each method was initially sorted separately with rank 1 assigned to the compound with the strongest binding association with the screened protein. Because the results are currently under consideration in patent and trade secret applications regarding the use of these data for cancer drug discovery and for brevity, only a small subset of compounds that have already been disclosed in the initial patent are discussed in this paper. Table 3 and 4 lists the grid-based energy rank and AMBER rank, respectively, for three compounds that showed poor binding towards KAP. The binding scores were calculated based on binding free energies, where a more negative score implies stronger binding. All the other screened proteins were ranked using this method.

Table 3 | **Energy Ranking.** Compounds are ranked based on the grid-based energy scoring method, where a negative score implies stronger binding. The energy score and ranking for three compounds of the KAP screening are listed in this table.

ZINC ID	Energy Rank	Energy Score
ZINC04107594	2083	-43.568192
ZINC02655717	3936	-41.538864
ZINC05375291	5323	-40.409500

Table 4 | **AMBER Ranking.** Compounds are ranked based on the calculated AMBER score to generate the AMBER rank. The AMBER score and ranking for three compounds of the KAP screening are listed in this table.

ZINC ID	AMBER Rank	AMBER Score
ZINC02655717	1541	-28.564444
ZINC04107594	3114	-22.781828
ZINC05375291	13546	-9.415719

1.2 Consensus List

The grid and AMBER scores cannot be considered separately as the final ranking of the compounds because the calculated scores by either method were not the absolute binding free energies. The two scores were compiled into a single consensus ranking list for this reason. The consensus rank for a compound was determined by summing the ranked results from the energy and AMBER scores. The compound with the lowest total combined score was assigned as rank 1 and was considered as the highest consensus rank. Each compound was given a consensus rank using this method. The consensus list was compiled based on the consensus rankings as illustrated below in Table 5. This list was used to show the overall ranking and compatibility of each molecule with a specific target protein based on both scoring methods. In Table 5, the three compounds that are listed in the previous two tables are shown. The ranking lists of the other 22 proteins are not shown for brevity, but were generated in identical fashion.

Table 5 | **Consensus Ranking.** Consensus rank for three compounds from the KAP screening.

	Ranking			
ZINC ID	Grid	AMBER	Total	Consensus
ZINC04107594	2083	3114	5197	1032
ZINC02655717	3936	1541	5477	1142
ZINC05375291	5323	13546	18869	8694

Note that approximately 0.5% of the entire database did not receive an AMBER score. This is because they were incompatible with the AMBER preparation process using the script, "prepare_amber.pl." The exact cause of the incompatibility issue remains undetermined; however, spatial overlaps between the protein and the compound may have caused the script to fail as the calculated energy of these overlaps approaches to infinity. Because of this and the very low occurrence of incompatibility, any compounds that did not receive an AMBER score were excluded from further analysis.

1.3 Identified Specific Inhibitors

The consensus lists of all 23 screened proteins were combined into one table to generate the combined consensus list. The combined consensus list allows for cross-referencing compound ranking data for all proteins screened in this study. Only an abridged version of this table showing 11 compounds with the highest specificity for SSH-2 from the virtual screening study is presented in Table 6. In order to interpret this data, disparity or difference between the ranking from the SSH-2 screening and the ranking of all other DUSP screenings was calculated for every compound considered in the study. The disparity ranking for these 11 compounds is listed in Table 7 along with the mean and standard deviation.

Table 6 | **Combined Consensus List Showing Identified Top Specific Inhibitors.** Only 11 compounds are listed due to results being considered as trade secret for the provisional patent on the use of the virtual screening data for cancer drug discovery.

	Consensus Rank							
ZINC ID	**SSH-2**	**KAP**	**MKP3**	**MKP4**	**MKP5**	**VHR**	**VHY**	**VH3**
ZINC05373221	20	19653	17795	5769	15810	19691	14332	5958
ZINC00053046	54	3904	14334	16725	19755	2072	16892	6091

Table 6 | **Combined Consensus List Showing Identified Top Specific Inhibitors,** continued.

ZINC06601214	56	20578	13533	7226	20183	5615	1491	11708
ZINC03313382	109	21316	12498	7170	9376	7583	-	7050
ZINC03429974	133	16616	12987	5239	13369	7368	-	3354
ZINC03377116	138	15853	13994	6250	9562	659	14086	19544
ZINC00260730	175	7086	-	5869	11210	7130	1041	4148
ZINC04307500	179	19920	8768	6660	14312	17657	16807	129
ZINC06737368	194	8620	12675	3458	6931	13620	1291	14739
ZINC03271868	259	9359	13912	6316	14650	17429	73	20935
ZINC04110856	267	8861	14103	9654	16660	12302	-	2321

Compounds from both tables were sorted according to the SSH-2 consensus rankings because the desired specific inhibitor must bind tightly to SSH-2, thus finding an inhibitor as close to the top of the SSH-2 consensus list as possible is highly desired. To qualify as a specific inhibitor, the compound must also demonstrate weak binding towards other DUSPs. Therefore, the compounds from the SSH-2 consensus list were cross-referenced with the consensus lists of other DUSPs.

Table 7 | **Disparity Scores for Identified Specific Inhibitors.** The disparity was calculated from the difference between consensus ranking of a particular screened DUSP and the consensus ranking of SSH-2.

	Disparity								
ZINC ID	**KAP-SSH2**	**MKP3-SSH2**	**MKP4-SSH2**	**MKP5-SSH2**	**VHR-SSH2**	**VHY-SSH2**	**VH3-SSh2**	**Mean**	**Std Dev**
ZINC05373221	19633	17775	5749	15790	19671	14312	5938	9266	6931
ZINC00053046	3850	14280	16671	19701	2018	16838	6037	8562	6387
ZINC06601214	20522	13477	7170	20127	5559	1435	11652	10634	6729
ZINC03313382	21207	12389	7061	9267	7474	-	6941	8559	4971
ZINC03429974	16483	12854	5106	13236	7235	-	3221	8562	4936
ZINC03377116	15715	13856	6112	9424	521	13948	19406	8782	5678
ZINC00260730	6911	-	5694	11035	6955	866	3973	8336	6576
ZINC04307500	19741	8589	6481	14133	17478	16628	-50	8721	6377
ZINC06737368	8426	12481	3264	6737	13426	1097	14545	8042	5541
ZINC03271868	9100	13653	6057	14391	17170	-186	20676	9016	5915
ZINC04110856	8594	13836	9387	16393	12035	-	2054	8891	5060

The disparity was used to relate the consensus lists of SSH-2 to another DUSP by subtracting the consensus rank of SSH-2 from that of the other screened DUSP. For example, the disparity between SSH-2 and KAP was determined by subtracting the SSH-2 consensus rank from the KAP consensus rank. For example, ZINC05373221 from Table 6, had a relatively low KAP consensus rank of 19653 and a relative high SSH-2 rank of 20. This suggests that the compound binds strongly to SSH-2 and weakly to KAP, since a high rank indicates strong binding and a low rank indicates weak binding. The difference or disparity between the two ranks was calculated to be 19633 in Table 7. Thus, this large positive disparity provided a measure of specificity of a given compound to SSH-2 relative to another DUSP and demonstrates that the compound fits the criteria of a specific inhibitor for SSH-2.

The mean of all disparity rankings is also calculated in Table 7. This relates all the consensus lists and gives a direct indication as to which compound would be most specific to SSH-2. Since large positive disparity indicates good specific inhibitors, large positive mean for a compound was also highly desirable. The standard deviation gave a numerical value to the degree of variability in the binding strength of a compound to the screened proteins. Having a small standard deviation in the disparity would then indicate that the compound of interest has similar binding affinity to all of the DUSPs other than SSH-2. Taken together, both the mean and the standard deviation must be considered in choosing a specific inhibitor.

The 11 listed compounds were chosen based on the above criteria. More specifically, the compounds were chosen from the top 500 binding compounds from the SSH-2 screening that has a disparity mean greater than 8000 and a standard deviation

smaller than 7000. These cutoffs were chosen in order to keep the list of potential specific inhibitors small, while maintaining a relatively large mean and small standard deviation.

2. Structure and Similarity of Identified Compounds

The similarity of the top 500 SSH-2 binding compounds was plotted in a hierarchical tree format based on the Tanimoto coefficient using ChemmineR. ChemmineR is a software package for analyzing and clustering small molecules used for screening data [47]. For brevity, only the top 11 identified potential specific inhibitors for SSH-2 are shown in Figure 6 as a hierarchical tree. This figure shows that there were a variety of structures that may inhibit SSH-2; however, certain structures were more prevalent in this list of 22 compounds, such as the cluster highlighted in red containing ZINC06601214, ZINC03429974, and ZINC04307500.

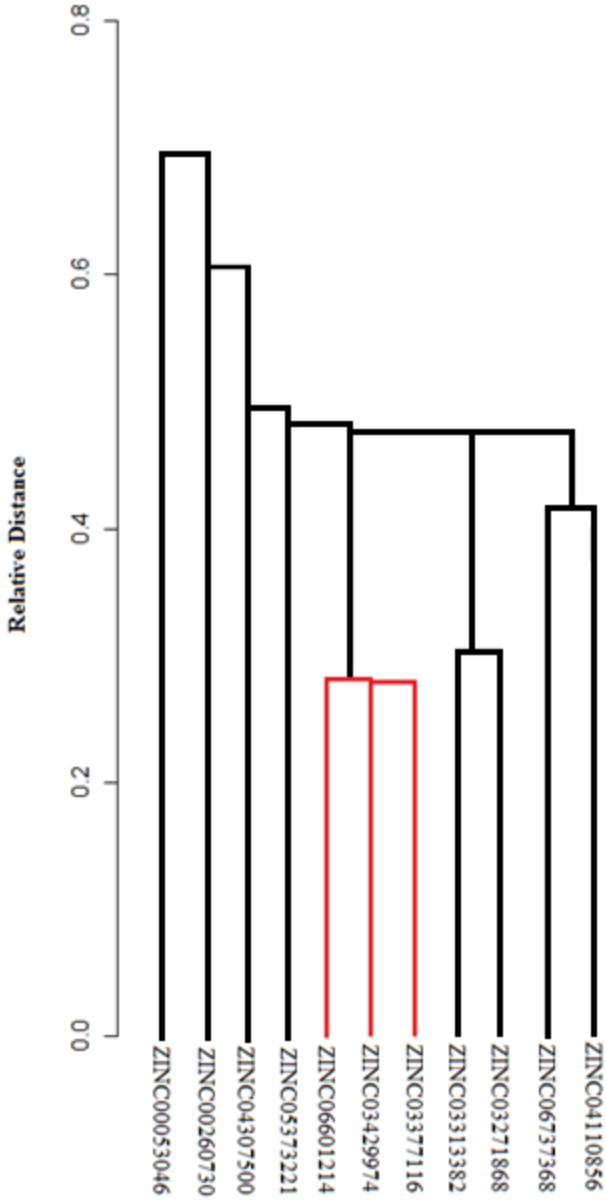


Figure 6 | **Hierarchical Tree of the 22 Identified Potential Specific Inhibitors for SSH-2.** The degree of similarity between the 11 identified potential specific inhibitors for SSH-2 is shown in this figure as a hierarchical tree. The compounds that were verified *in vitro* are highlighted in red. This figure shows that there are a variety of structures that may inhibit SSH-2; however, certain structures were more prevalent, suggesting that certain structures tend to show specificity towards SSH-2.

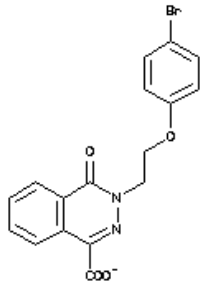
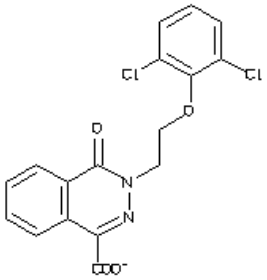
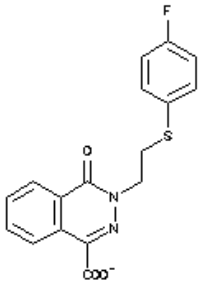
ZINC03377116	ZINC03429974	ZINC06601214
		

Figure 7 | **Identified Potential Inhibitors with Highly Similar Structures.** The molecular structure of the compounds within the highlighted cluster from Figure 6 is shown here. All three compounds contain a carboxylic acid end with highly similar bond connectivity.

The molecular structures of the compounds within the highlighted cluster from the Figure 6 are presented in Figure 7. All three compounds contain a carboxylic acid end with highly similar bond connectivity. The major differences between the compounds were the functional groups on the upper benzene ring and the substitution of a sulfate atom in place of an oxygen atom in ZINC06601214. These highly similar compounds suggest that certain molecular structures provide better binding specificity to SSH-2.

The interactions between these compounds and SSH-2 are visualized in Figures 8, 9, and 10. As illustrated in Figure 8, when ZINC3377116 was docked to the active site of SSH-2, the carboxylic acid clearly fits into the binding pocket and forms multiple hydrogen bonds with nearby residues (highlighted in green). The complex was further stabilized by various interactions, as demonstrated by the hydrogen bonding between the nitrogen atom on the adjacent ring and arginine-398 of SSH-2. Similar interactions were

discovered in the docking of ZINC03429974 and ZINC06601214 to SSH-2 in Figures 9 and 10, respectively.

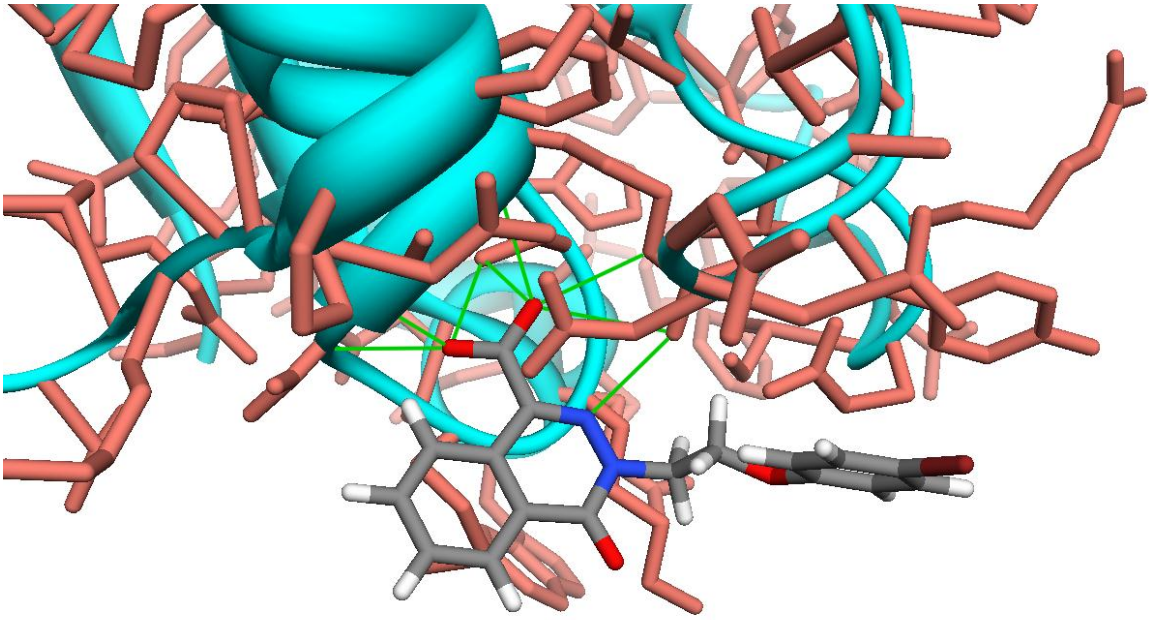


Figure 8 | **ZINC03377116 Docked to SSH-2.** Interactions between the ZINC03377116 and SSH-2 can be seen in this figure. The carboxylic acid clearly fits into the binding pocket and forms multiple hydrogen bonds, highlighted with green lines. Further stability of the enzyme-substrate complex is established by hydrogen bonding of the adjacent nitrogen atom and arginine-398 of SSH-2.

Interestingly, the binding orientations of the compounds to SSH-2 were highly similar, with the only major difference being the position of the rings on the right in these figures. The binding orientation in these figures must be very close to the optimum orientation for compounds that share similar structures, because DOCK 6.0 is designed to find the lowest energy confirmation of the enzyme-substrate complex. The fact that the docking program achieves this through random reorientations and that this orientation appeared repeatedly, suggests that this orientation was especially stable and exhibited tight binding.

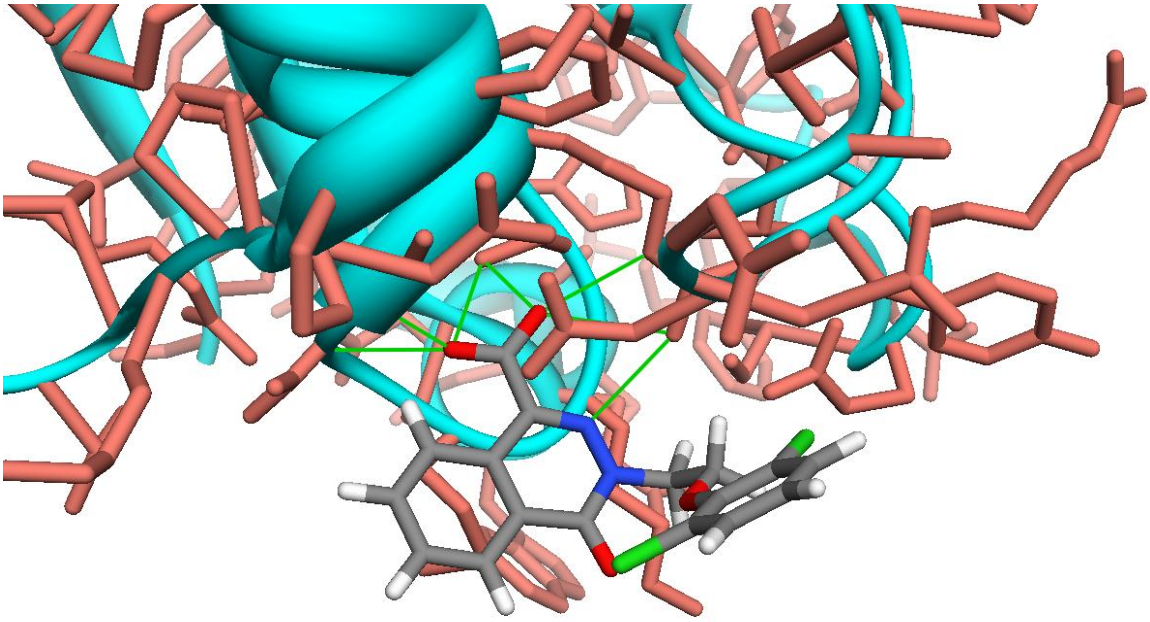


Figure 9 | **ZINC03429974 Docked to SSH-2.** Similar interactions between ZINC03429974 and SSH-2 can be observed. The carboxylic acid again fits into the catalytic pocket and the nitrogen atom on the adjacent ring forms an extra hydrogen bond with arginine-398.

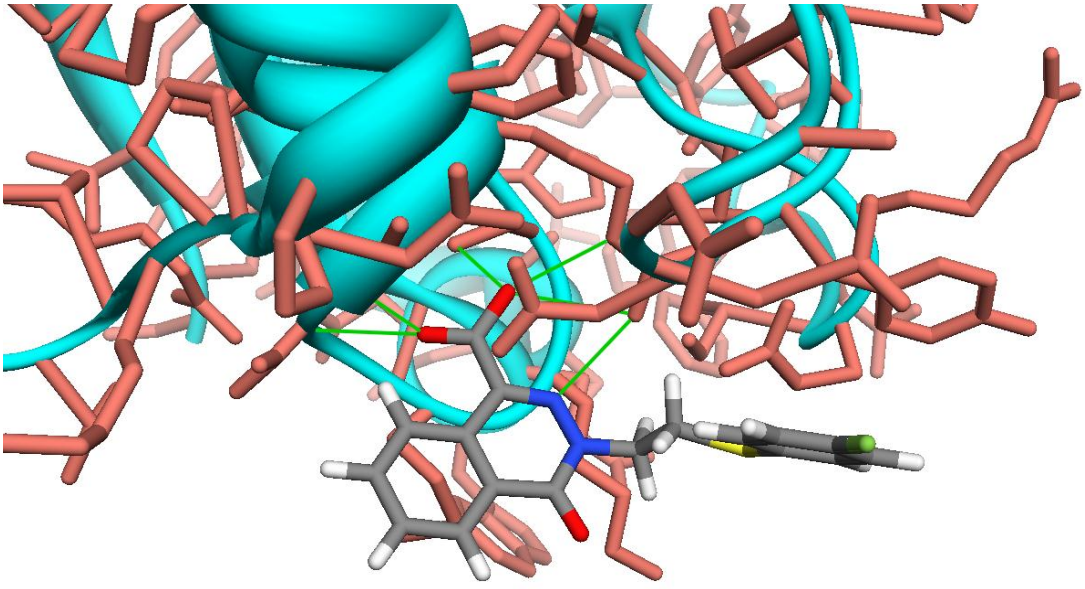


Figure 10 | **ZINC06601214 Docked to SSH-2.** Similar interactions between ZINC06601214 and SSH-2 can be observed. Also, the binding orientations of all three compounds are highly similar, suggesting that this type of compound structure forms stable complexes with SSH-2.

In comparison, the binding orientations of these three compounds to another DUSP other than SSH-2 are generally different. In Figures 11, 12, and 13, the binding orientations of the same three compounds, ZINC06601214, ZINC03429974 and ZINC03377116, to MKP-5 demonstrated markedly dissimilar orientations.

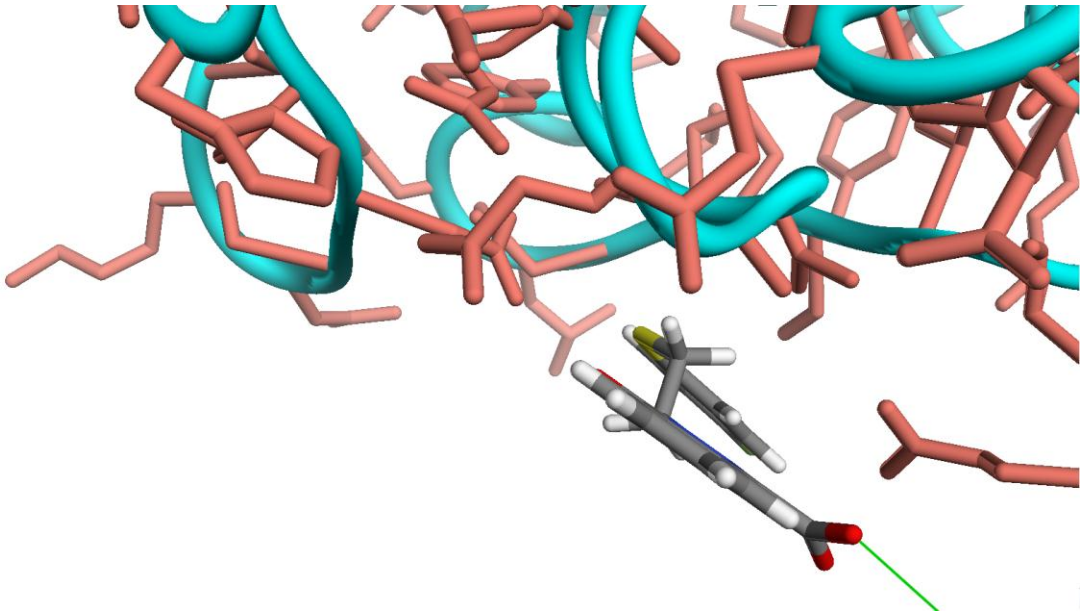


Figure 11 | **ZINC06601214 Docked to MKP-5.** The carboxylic acid is facing a different direction in comparison to the other two docked compounds to MKP-5. Note the lack of hydrogen bonds (green lines) between the compound and MKP-5.

In Figure 11, the carboxylic acid of ZINC06601214 was facing in the opposite direction as the other two compounds in Figures 12 and 13. In these two figures, the carboxylic acids of ZINC03429974 and ZINC03377116 were shifted and the orientations of the attached rings were clearly different. These differences appear because there are multiple orientations that will produce approximately equal binding affinities. Thus, unlike SSH-2, this type of compound structure does not have a binding orientation that produced a high affinity interaction to MKP-5.

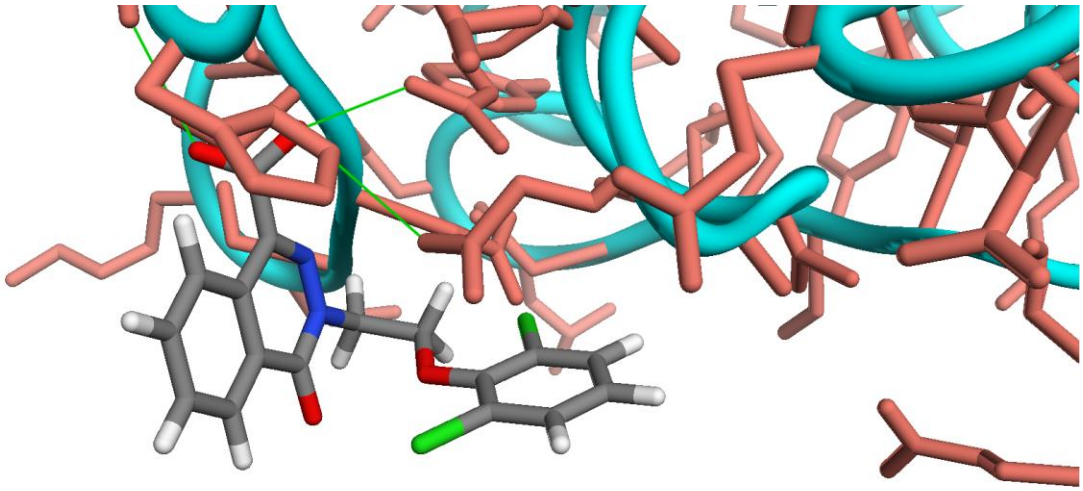


Figure 12 | **ZINC03429974 Docked to MKP-5.** The carboxylic acid is located within the catalytic site, however, the number of hydrogen bonds is considerably less and no hydrogen bonds are observed between the nitrogen atom and MKP-5.

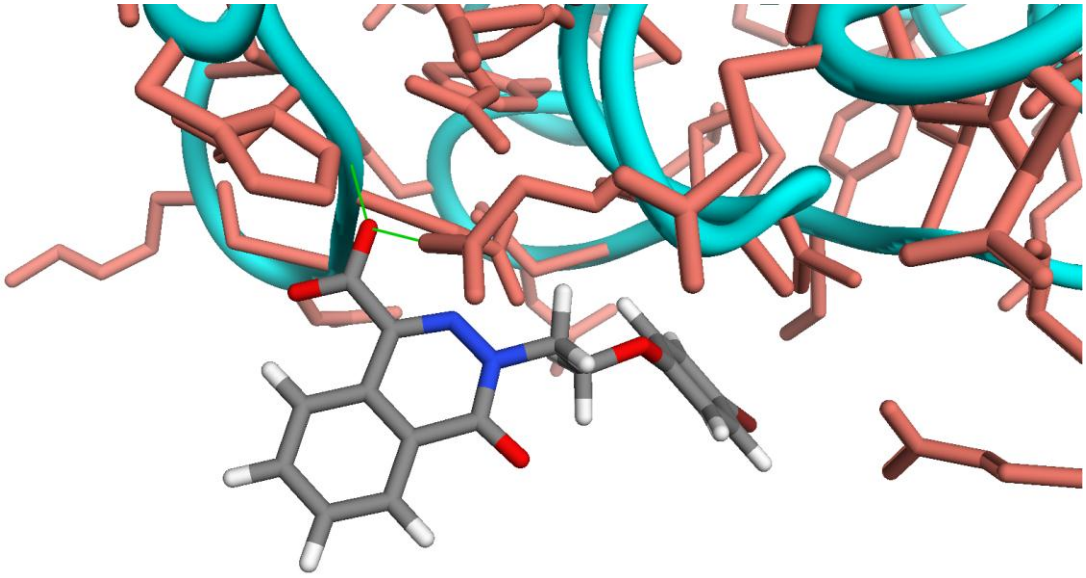


Figure 13 | **ZINC03377116 Docked to MKP-5.** The carboxylic acid is shifted to a different location and again, very few hydrogen bonds exist between the compound and MKP-5. No hydrogen bonds are again observed between the nitro gen atom and MKP-5.

Table 8 | **Consensus Ranks for SSH-2 and MKP-5 of Discussed Compounds.** The general binding orientations suggest stability of the compound structure to the target. This is in agreement to the consensus rankings.

	Consensus Rank	
ZINC ID	**SSH-2**	**MKP-5**
ZINC06601214	56	20183
ZINC03429974	133	13369
ZINC03377116	138	9562

The general orientation of a particular compound structure suggests stability to a protein target. Compound structures with repeated general orientation would then lead to high binding affinities and high consensus rankings. On the other hand, compound structures that do not have a general orientation would most likely have low binding affinities and low consensus rankings. This is generally supported as shown in Table 8, where ZINC06601214, ZINC03429974, and ZINC03377116 have very low consensus rankings for SSH-2 and high rankings for MKP-5.

3. Mutation of Arginine-398

Since ZINC06601214, ZINC03429974, and ZINC03377116 all demonstrated hydrogen bonding between the nitrogen atom and arginine-398 of wild-type SSH-2 (SSH-2-WT), mutation of this arginine to lysine (SSH-2-R398K) should disrupt the binding of these compounds. The general binding orientation of these compounds should vary more after the mutation because of the loss in stability of the complex.

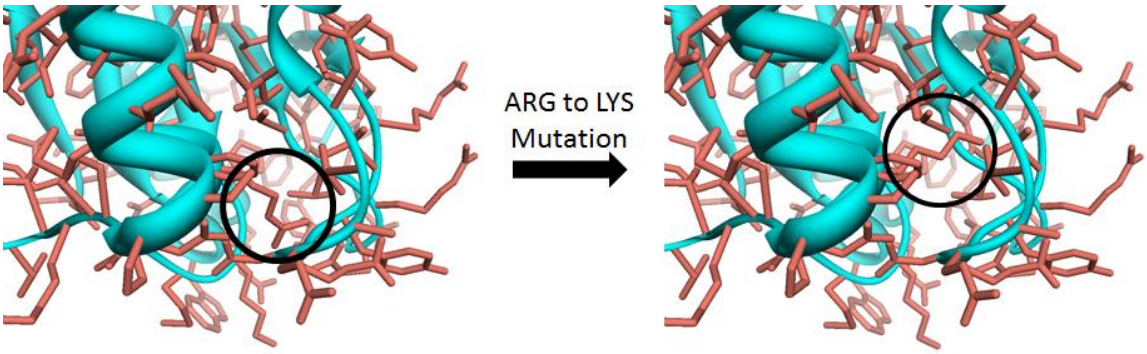


Figure 14 | **Arginine-398 Mutation.** Arginine-398 of SSH-2 was mutated into a lysine residue using Chimera. The positioning of the side chain is greatly shifted as a result.

The arginine was mutated to a lysine residue using Chimera. The model with the highest confirmation probability was selected. The side chain of the mutated residues greatly shifted in position, as shown in Figure 14. All three compounds were unable to form hydrogen bonds between the nitrogen atom and SSH-2-R398K as shown in Figures 15, 16 and 17. A significant number of hydrogen bonding still exists between the oxygen atoms of the carboxylic acid and SSH-2-R398K. Furthermore, the general binding orientation of the carboxylic acid between the three compounds remained highly similar to SSH-2-WT even in the presence of the mutation.

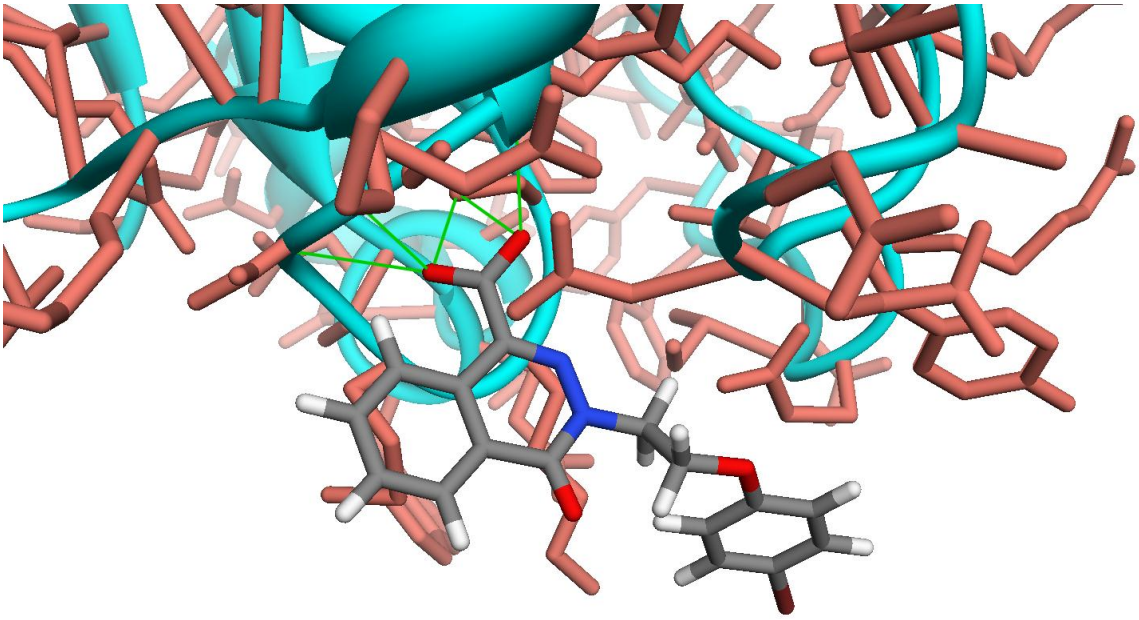


Figure 15 | **ZINC03377116 Docked to Arginine-398 Mutated SSH-2.** No hydrogen bonding is formed between the nitrogen atom and SSH-2 after the mutation. Multiple hydrogen bonds are still formed between the carboxylic acid and SSH-2.

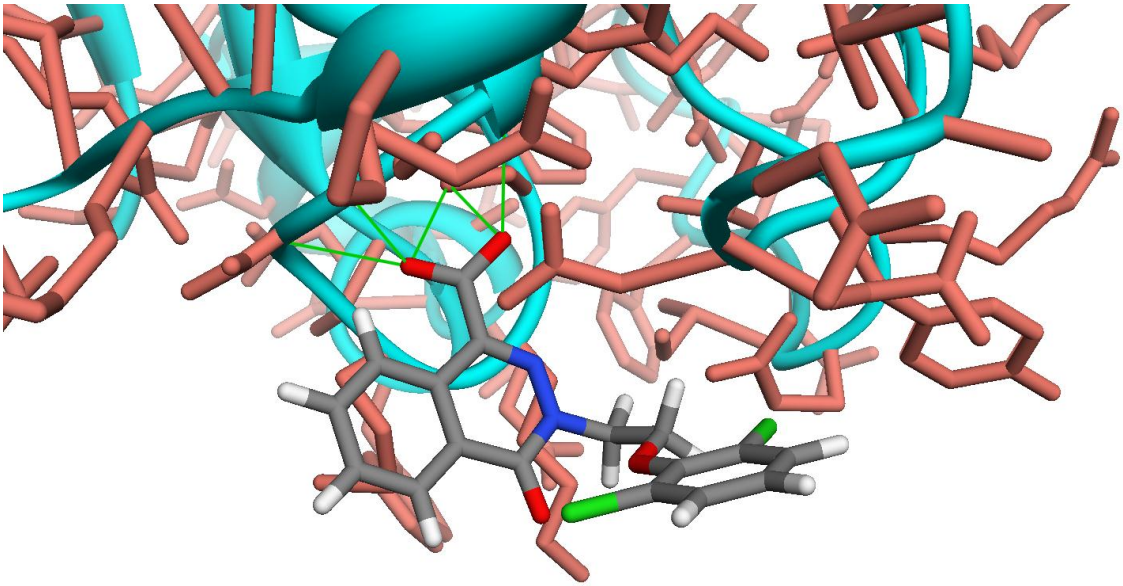


Figure 16 | **ZINC03429974 Docked to Arginine-398 Mutated SSH-2.** Again, no hydrogen bonding is formed between the nitrogen atom and SSH-2 after the mutation. The major difference in the binding orientation of these three compounds is the rings located on the right.

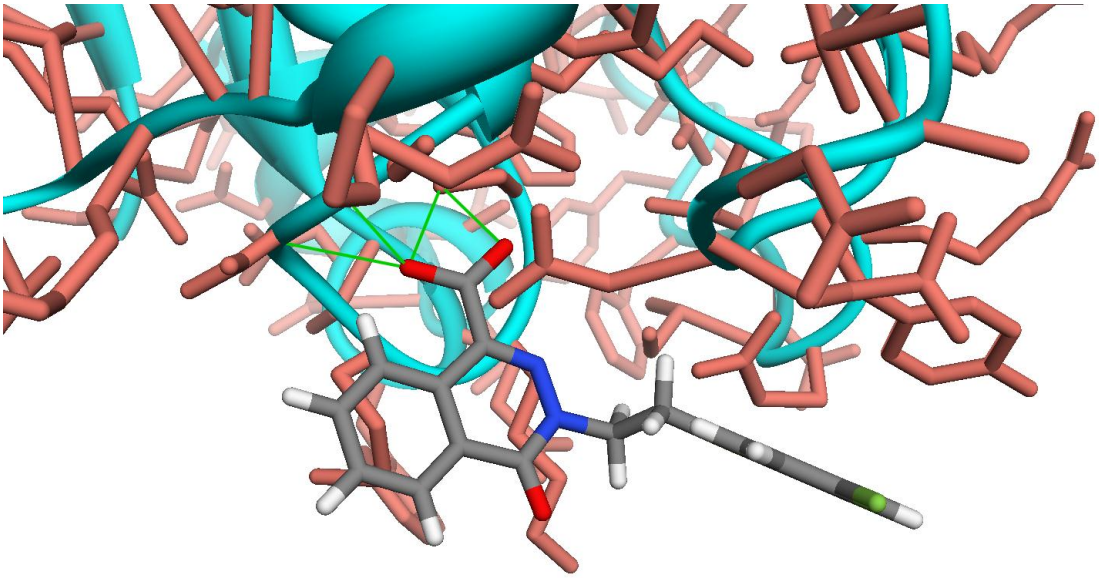


Figure 17 | **ZINC06601214 Docked to Arginine-398 Mutated SSH-2.** The carboxylic acid is still docked to the active site of the mutated SSH-2, with no hydrogen bonds between the nitrogen atom and SSH-2R398K.

In order to better understand the differences in orientation of the rings on the right side of the compounds, the position of the carboxylic acid in all three compounds was super imposed, as shown in Figure 18. The binding orientation of the three compounds to SSH-2-WT is shown in Panel A and to SSH-2-R398K in Panel B. This revealed that the compounds were shifted to a greater degree when docked to SSH-2-R398K in comparison to SSH-2-WT. This indicates that the hydrogen bonding between the nitrogen atom and arginine-398 of SSH-2 does indeed stabilize the binding of the compounds. Furthermore, there are small shifts in the carboxylic acid position, further supporting the fact that the nitrogen atom and arginine-398 of SSH-2 stabilize the interactions.

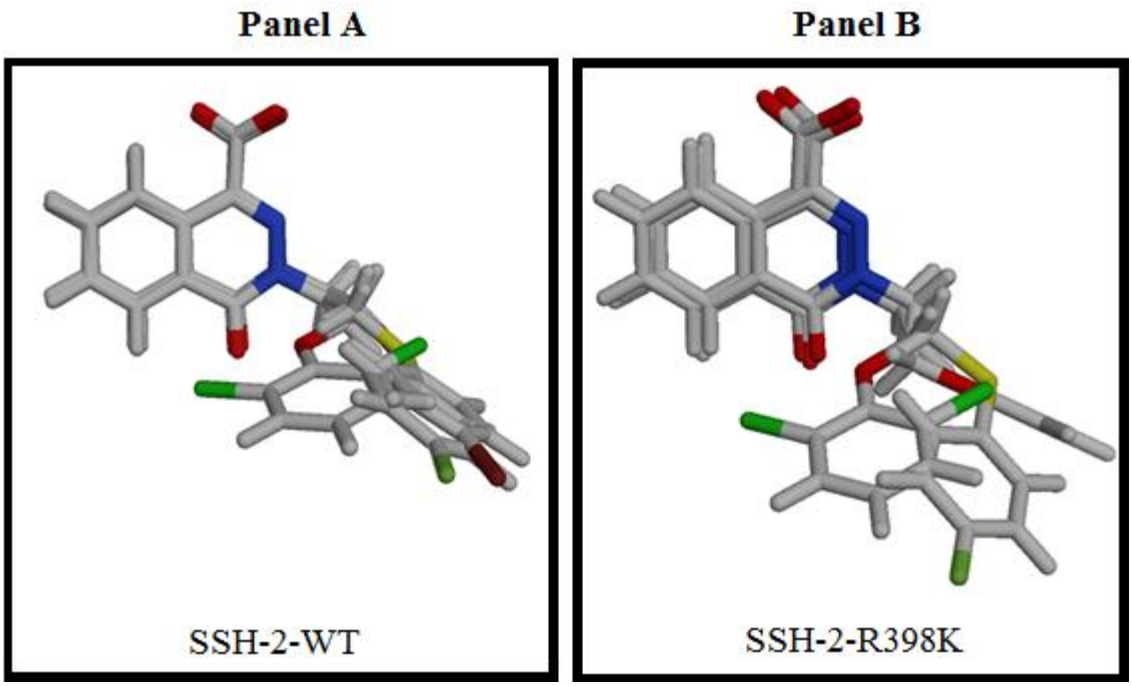


Figure 18 | **Compounds Superimposed.** The docking orientations of three compounds to (A) SSH-2-WT and (B) SSH-2-R398K superimposed. This figure revealed that the compounds were shifted to a greater degree when docked to the mutated model.

4. Compound Verification

4.1 ZINC03377116 Treatment

Cells treated with ZINC03377116 did not show an increase in actin filaments at the applied concentrations. Figure 19 shows the stained actin filaments at 100 µM, 10 µM, 1 µM, 0.1 µM, and 0.01 µM applied compound concentrations. In this figure, linear actin filament structures were seen at all applied concentrations. The actin fluorescence intensity does not seem to have changed significantly at high compound concentrations; however, at 0.1 µM and 0.01 µM compound concentrations, the intensity of actin filaments appeared to have decreased.

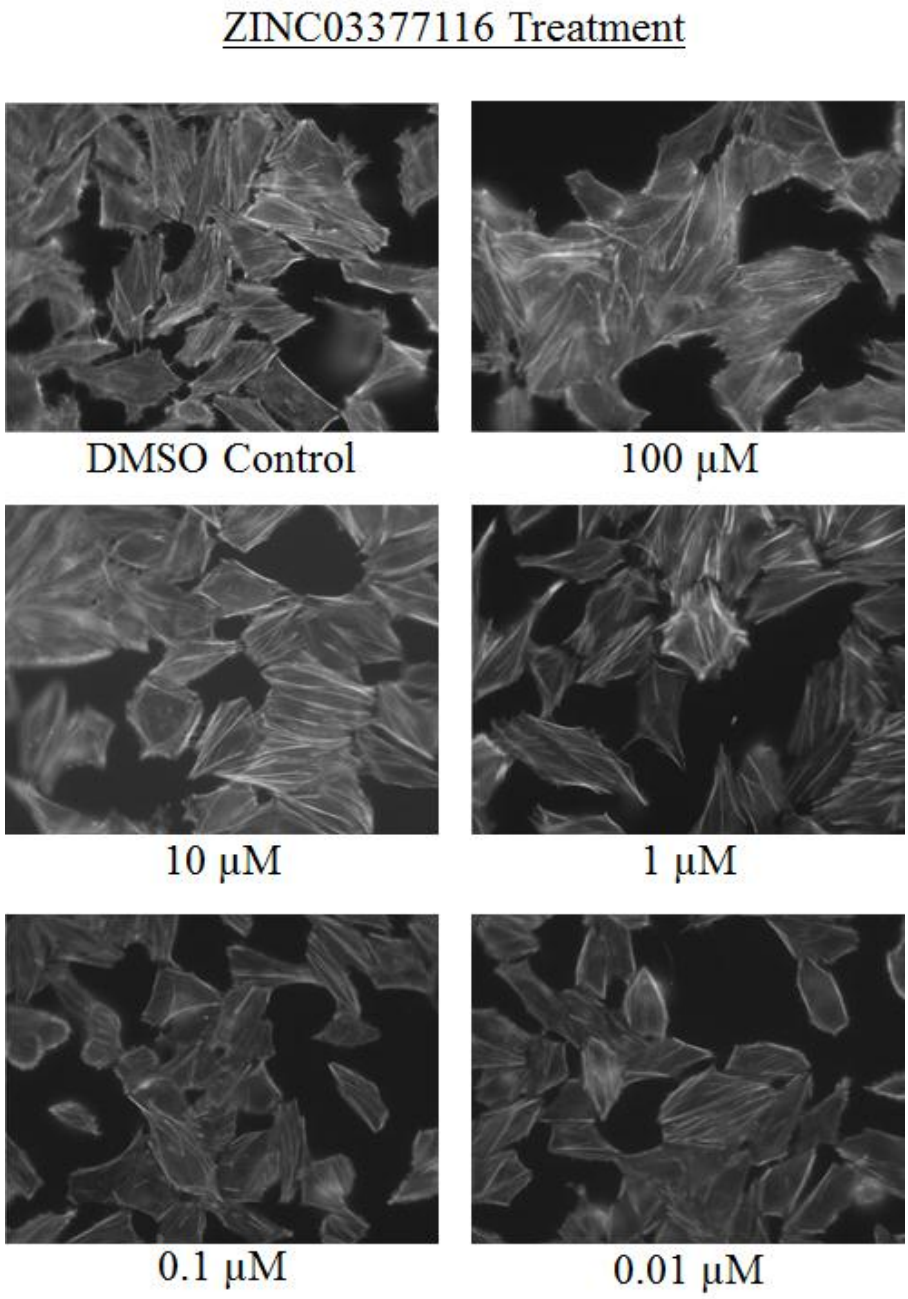


Figure 19 | **Cells Treated with ZINC03377116.** HeLa cells treated with ZINC03377116 showed no significant loss in actin filaments at high concentrations, but showed a decrease in fluorescence intensity at lower concentrations.

Quantification of the fluorescence intensity, shown in Figure 20, confirms this observation. There was a significant decrease in fluorescence intensities under 0.01 µM, 0.1 µM and 1 µM compound concentrations, but remained unchanged when treated at 10 µM and 100 µM concentrations.

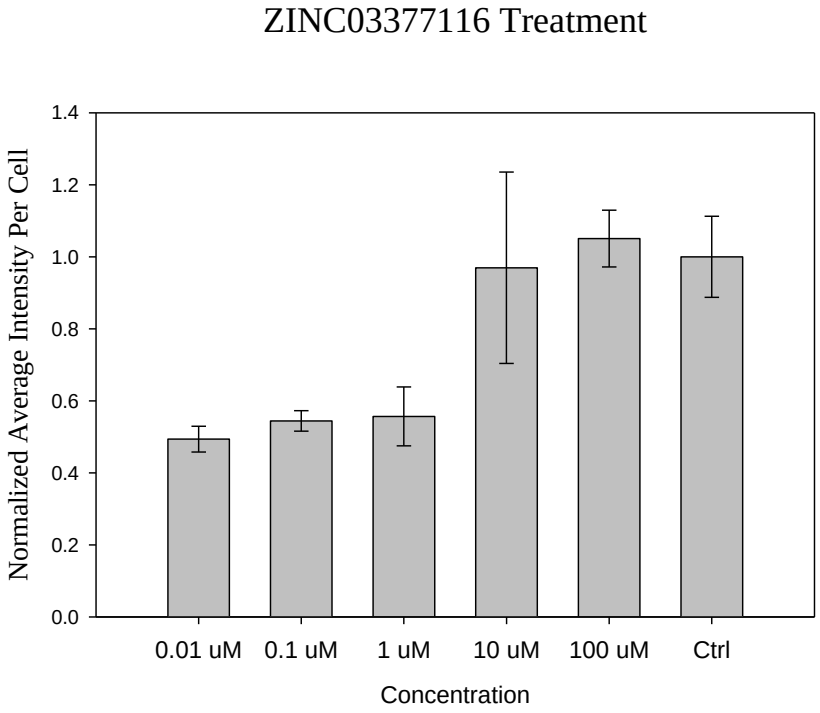


Figure 20 | **Average Intensity per Cell under ZINC03377116 Treatment.** Quantification of the actin filaments fluoresce intensity shows a decrease in actin filaments when cells are treated with ZINC03377116 at 0.01 µM, 0.1 µM and 1 µM. However, at 10 µM and 100 µM compound concentrations, the actin filaments remained unchanged.

This data, showing a decrease in actin filament fluorescence intensities, was opposite of what was expected and suggests that SSH-2 was not inhibited by ZINC03377116. At 10 µM and 100 µM concentrations, the actin fluorescence intensities were the same as control suggesting that the compound did not bind to SSH-2. The reason for the decrease in fluorescence intensities at 0.01 µM, 0.1 µM, and 1 µM is

unknown at this time. One possible explanation is that the compound had a significant nonspecific binding to proteins that regulate the actin polymerization pathway leading to a decrease in actin filament formation.

4.2 ZINC06601214 Treatment

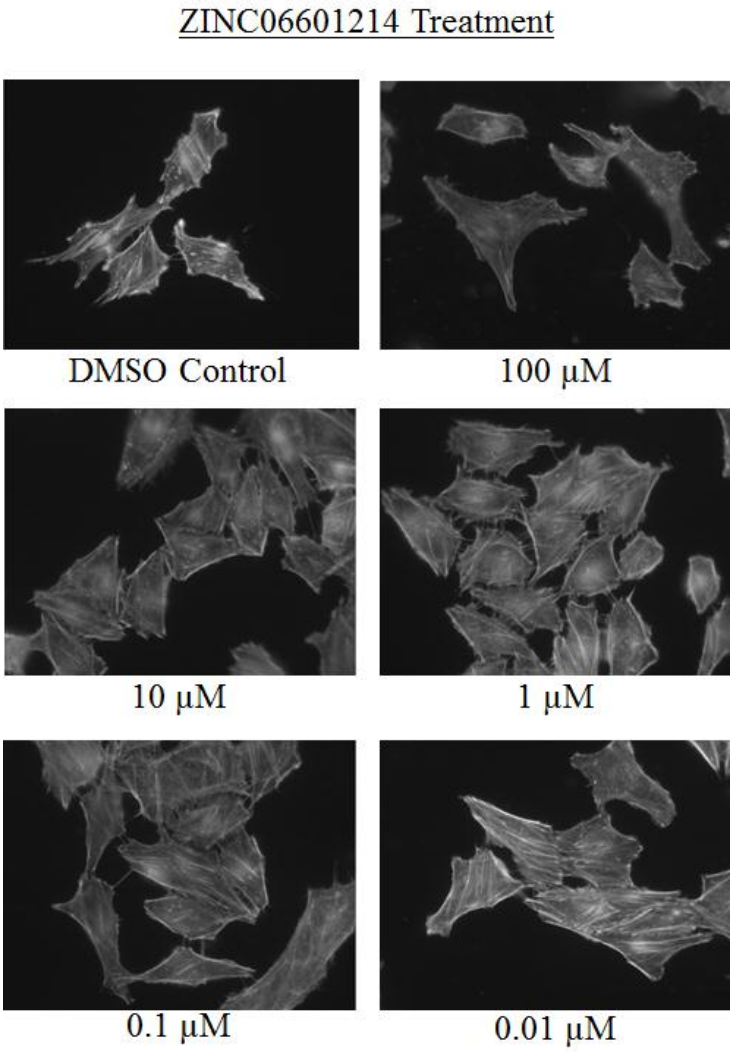


Figure 21 | **Cells Treated with ZINC06601214.** A decrease in actin stress fibers was observed in HeLa cells treated with ZINC06601214 across all tested concentrations.

A decrease in actin stress fibers was observed in HeLa Cells treated with ZINC06601214 across all tested concentrations. Figure 21 shows the stained actin filaments at 100 µM, 10 µM, 1 µM, 0.1 µM, and 0.01 µM applied compound concentrations and DMSO control. In this figure, actin filaments can easily be seen in the control treated with DMSO. When the cells were treated with ZINC06601214 at 100 µM, 10 µM, 1 µM and 0.1 µM concentrations, actin fibers were either not observed or the actin intensity was much less prominent than that of control. At 0.01 µM compound concentration, the actin filaments reappear at a higher intensity than the other treated samples.

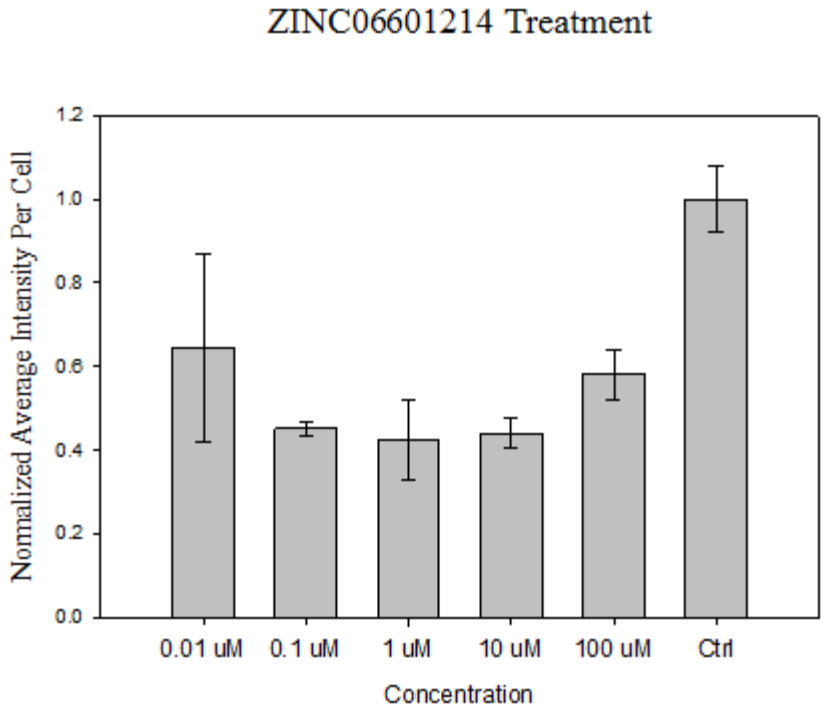


Figure 22 | **Average Intensity per Cell under ZINC06601214 Treatment.** Quantification of the actin filaments fluoresce intensity show a decrease in actin filament intensities in all applied compound concentrations.

In Figure 22, the intensity of the actin filaments in the cells for each treatment is plotted. The actin filament intensities of the treated cells were dramatically less than control, implying that ZINC06601214 did not inhibit SSH-2. The decrease in the actin intensity may be attributed to the nonspecific binding of the compound to other targets in the cell, most likely proteins that facilitate actin polymerization and filament formation. This is somewhat consistent with the results from ZINC03377116 treatment, especially given the similarity in chemical structures between ZINC03377116 and ZINC06601214.

4.3 ZINC04307500 Treatment

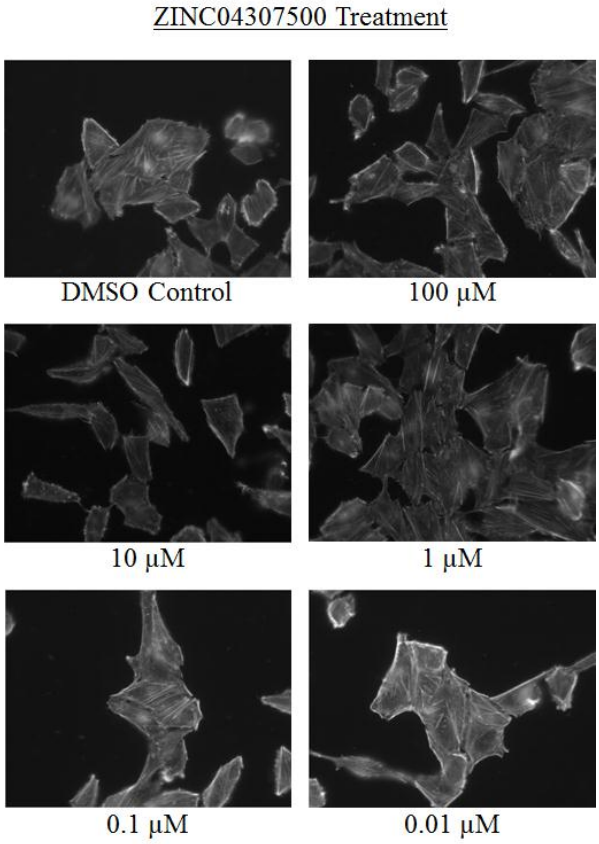


Figure 23 | **Cells Treated with ZINC04307500.** HeLa cells treated with ZINC04307500 showed an increase in actin intensity at 0.01 µM.

When cells were treated with ZINC04307500, the actin filament intensities were decreased at 1 µM, 10 µM and 100 µM concentrations in comparison to the control. At 0.1 µM, the intensity of actin was at about the same level as that of control. At even lower concentrations of 0.01 µM, the actin filaments were clearly much brighter than control. In Figure 23, the actual image of the stained actin filaments of 100 µM, 10 µM, 1 µM, 0.1 µM, and 0.01 µM compound treatment and DMSO control are shown.

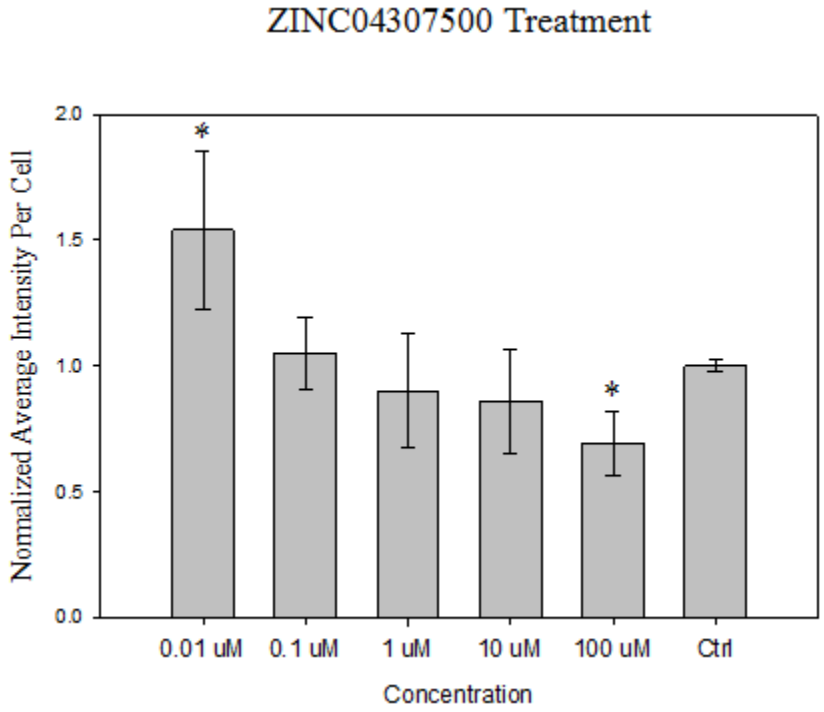


Figure 24 | **Average Intensity per Cell under ZINC04307500 Treatment.**
Quantification of the actin filaments fluorescence intensity shows an increase in intensity at 0.01 µM concentration. The intensity begins to decrease as concentration is increased. At 100 µM concentration, the intensity is significantly lower than that of control.
* Denotes statistically significant changes in fluorescence intensities between sample and control.

The quantification of actin intensities, shown in Figure 24, confirms these observations. In this graph, the actin filament intensities of the samples treated with 0.01 µM of ZINC04307500 were significantly higher than control with a t-score of 2.99 and a

p-value of 0.04. This suggests that SSH-2 was inhibited at this concentration because the inhibition of SSH-2 should result in a decrease in actin depolymerization, however, as the applied compound concentration increased, the observed actin filament intensities decreased. This observation may be explained by the prozone effect. This phenomenon is observed when the concentration of ligands is saturated, leading to lower ligand efficiency due to either interactions between ligands or positive cooperative binding to target protein [49, 50]. In this respect, ZINC04307500 may successfully and specifically bind to the active site of SSH-2 and inhibiting its action at or below 0.01 μM concentrations, however, as concentration is increased, the ligand-receptor system is slowly being overloaded, where ZINC04307500 no longer binds well to SSH-2. SSH-2 would then remain active and resulting in unchanged level of actin filament fluorescence intensity at 0.1 μM, 1 μM and 10 μM concentrations. At 100 μM concentration, the actin intensity is significantly lower than that of control with a t-score of 4.10 and a p-value of 0.01. This may be because at such high concentration, the compound begins to bind nonspecifically. As compounds bind to other proteins, off-target effects may result in a decrease in actin polymerization. A combination of the prozone effect and a decrease in actin polymerization due to nonspecific binding would then lead to a decrease in actin filament at 100 μM or even higher concentrations.

IV. DISCUSSION

1. Virtual Screening

1.1 General Trends

By analyzing the overall consensus ranking list, general trends regarding the compound structures and their ability to specifically inhibit SSH-2 can be inferred. As discussed in the results section, a specific inhibitor for SSH-2 should have a low consensus rank for SSH-2 and high consensus rank for the other DUSPs as well as large disparity mean and large standard deviation. The 11 identified potential specific inhibitors listed in Table 7 all follow this criterion. The structures of these compounds all have a few distinct features. First, all 11 compounds have a carboxylic acid group. Second, most of the carboxylic acids are attached to a ring structure.

A possible reason that a carboxylic acid is present in every identified compound is because it resembles a phosphate group in overall size and shape. Since phophoserine and phosphothreonine are substrates for SSH-2, compounds that are similar to these phosphorylated amino acids should fit well into the binding pocket of SSH-2. In this respect, the carboxylic acid would enter the catalytic pocket and the attached ring would act as an anchor in providing extra support and stability to the enzyme-substrate complex through various interactions. The described binding interactions of the identified potential inhibitors and SSH-2 can be seen in Figure 25 as well as Figures 8, 9 and 10. In Figure 25, ZINC04172162, the compound with the highest disparity mean of 11117, is

bound to the active site of SSH-2. The carboxylic acid and the attached ring are located inside the pocket.

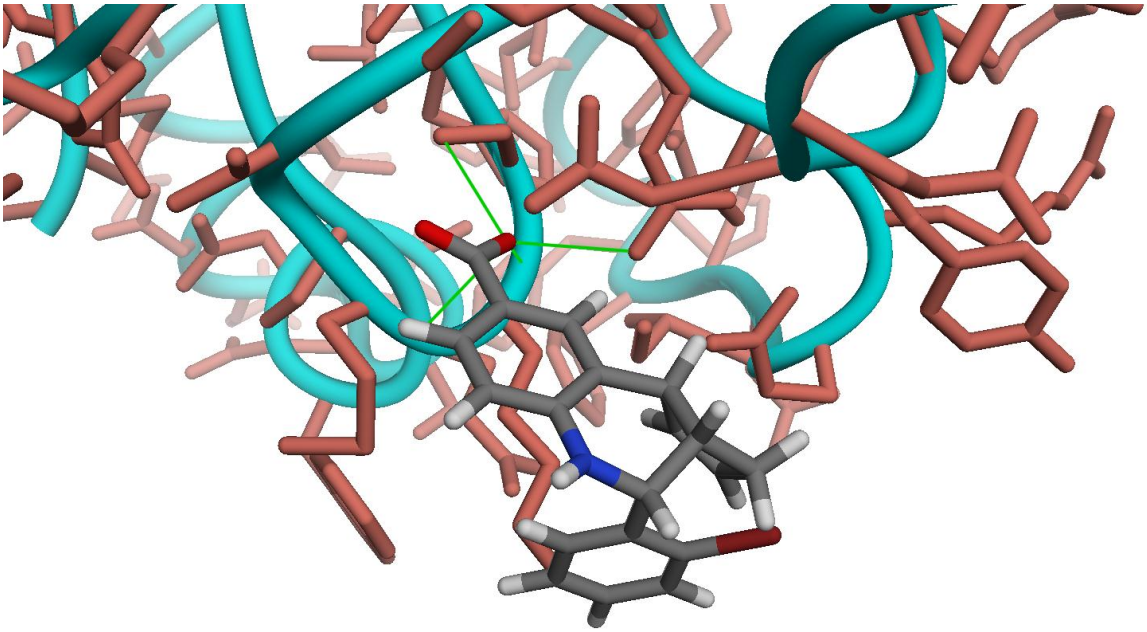
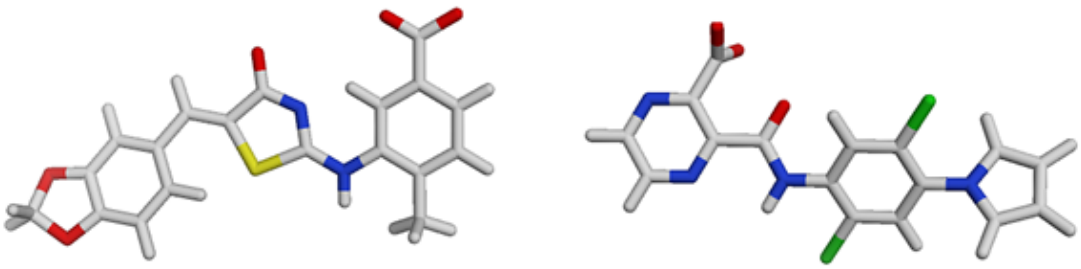


Figure 25 | **ZINC04172162 Bound to SSH-2.** ZINC04172162, the compound with the highest disparity mean, is bound to SSH-2. A possible explanation that this compound binds well to SSH-2 is because the carboxylic acid and the attached ring resemble the overall size and shape of a phosphotyrosine residue.

Furthermore, the compounds that show high specificity towards SSH-2 are generally larger and more rigid with multiple ring structures. Of the 11 identified compounds, 8 compounds contain three or more rings and all 11 contain at least 2 rings. ZINC05373221 and ZINC04307500, shown in Figure 26, are two examples from the list of identified compounds that show this trend. Both of these compound structures contain a carboxylic acid end and several nitrogen atoms.



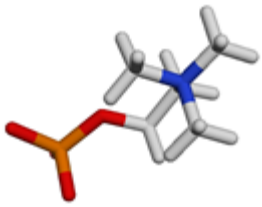
ZINC05373221 ZINC04307500

Figure 26 | **ZINC05373221 and ZINC04307500.** Compounds that show high specificity towards SSH-2 are generally larger and more rigid with multiple ring structures. ZINC05373221 and ZINC04307500 are two such examples from the list of identified potential inhibitors of SSH-2.

On the other hand, compounds that do not show specificity towards SSH-2 tend to be smaller with fewer rings and more linear in shape. For example, one of the compounds with the least specificity is ZINC05260817. It is ranked 1, 3, 7, 11, 35, 48, and 79 on the consensus list of SSH-2, VHY, VHZ, KAP, CDC14b, MKP4, and VH3, respectively. This compound, shown in Figure 27, is much smaller in comparison. The carboxylic end that is present in the identified compounds is replaced by a phosphate group, which is consistent with the fact that DUSPs are designed to accommodate phosphorylated amino acids.

By analyzing the structural differences between these compounds, additional patterns regarding compound specificity can be inferred. These differences suggest that bulky compounds tend to bind more strongly to SSH-2 while smaller compounds tend to bind well to most DUSPs indiscriminately. These results are consistent with the fact that SSH-2 has a wider catalytic pocket [5], that only allows large molecules to enter the catalytic pocket of SSH-2. These large compounds can then interact with the walls of the

SSH-2 catalytic pocket and create strong association. Like large molecules, small molecules can also enter the catalytic pocket of SSH-2; however, their size allows them to enter the catalytic pockets of other DUSPs. This may be the reason why they bind indiscriminately and are ranked high on most DUSPs consensus list.



ZINC05260817

Figure 27 | **ZINC05260817.** This compound is ranked very high across consensus list of several DUSPs and show little specificity towards SSH-2. The size and shape is much smaller and less rigid in comparison to compounds that are specific towards SSH-2.

There are some exceptions to the trend that small compounds binding well to DUSPs. For example, ZINC05260817 bind well to most DUSPs, but bind poorly to MKP-5 and MKP-8. Its consensus rankings for these two proteins are 5380 and 8420, respectively. The active site sequences for proteins that ZINC06260817 bind well to (SSH-2, VHY, VHZ) and proteins that ZINC06260817 do not bind well to (MKP-5, MKP-8) are aligned. The polar amino acids are highlighted in yellow in this table. The amino acids in the aligned sequences are mostly nonpolar. The combined hydrophobicity index, calculated from adding the index of each amino acid in a sequence, ranges from -3.5 to 4.6. Thus, this table shows no significant patterns in hydrophobicity that could explain difference in binding specificity. This indicates that the spatial location of the

amino acids around the active site sequence as well as the size and shape of the catalytic pocket should be considered when determining specificity and interactions of compounds with DUSPs.

Table 9 | **Active Site Sequence Alignment.** The active site sequences of proteins that that ZINC06260817 bind well to (SSH-2, VHY, VHZ) and proteins that ZINC06260817 do not bind well to (MKP-5, MKP-8) are aligned. Polar amino acids are highlighted in yellow. The consensus rankings for ZINC06260817 to the respective proteins are shown. This table shows no significant patterns in hydrophobicity that could explain difference in binding specificity.

DUSP	Consensus Rank	Sequence
SSH-2	1	HCKMGVSR
VHY	3	HCFAGISR
VHZ	7	HCALGFGR
MKP-5	5380	HCQAGVSR
MKP-8	8420	HCAVGVSR

1.2 Combined Consensus List Implications

Although the combined consensus list was used to identify specific inhibitors for SSH-2, it can be used to identify specific inhibitors for other proteins screened in this study. Each column in the combined consensus list ranks compounds by its binding affinity to the screened protein. The combined list was sorted by the SSH-2 column for this study in order to find a specific inhibitor. Similarly, the list can be sorted by a different column to identify inhibitors for a different protein. The inhibitors must again be ranked high for the protein of interest and low for all other proteins in order to be considered a specific inhibitor. Since DUSPs are involved in normal cellular processes as

well as disease states such as inflammation, diabetes and cancer, the ability to identify inhibitors for different proteins allows for drug discovery in several different diseases [7].

Furthermore, the consensus list can be used to identify compounds that are less specific in order to inhibit multiple cellular proteins and pathways. For example, both DUSP1 and DUSP4 are overexpressed in lung cancer [51]. A compound that will inhibit both proteins may be more effective in cancer treatment than a compound that targets only one protein. Instead of having a high consensus ranking for only one protein, this less specific compound must rank high for the proteins of interest and low for all other proteins. The flexibility of this combined consensus list provides the practicality of identifying various inhibitors and drug discovery to multiple pathological states.

1.3 Docking Issues

Overall, DOCK 6.0 completed the intended task without many problems. There were minor issues associated with DOCK accessory programs, such as AMBER preparation as discussed earlier. Several compounds were incompatible with AMBER such that the preparation script, "prepare_amber.pl," failed causing the entire data slice to be dropped from the docking process. The preparation problem may be due to spatial overlaps between the protein and the compound, but the exact cause of the problem is unknown. This problem was easily resolved by removing the failed compound from the data slice manually and resubmitting the job to a cluster.

1.4 Grid Issues

The PRAGMA grid has been particularly busy this year with large number of projects being conducted. Processors on certain clusters, especially Rocks-52, showed high work load. This caused longer queue time and increased the overall experimental completion time.

Another issue was the accumulation of java zombie processes on Rocks-52. Jobs were distributed to other clusters for computation using "slice_distribute.pl" script on the master cluster. When Rocks-52 was set as the master, more and more of these "java <defunct>" processes were created as the distribution script runs. An over-accumulation of java processes consumes the entire process quota of the user and prevents further access by the user. These processes cannot be deleted by the "kill -9" function and must be stopped by the cluster administrator. There was speculation that the cluster's java version (jdk1.5.0_05) was incompatible with the distribution script, however, switching java versions did not eliminate the problem. When the master cluster was switched from Rocks-52 to Tea and the problem was resolved, suggesting that the zombie processes originate from the cluster setup itself.

Since the docking experiments in this study were performed through parallel processing, a large amount of grid resources was required. This became a problem when crashed jobs failed to release semaphores, a system resource, from the processors. Further jobs were then unable to be completed and caused even more resources to be sequestered. Several commands were then necessary to be executed on the clusters to release the held semaphores. These problems increased the screening time, but did not impact the resulting data and specific inhibitors for the dual-specificity phosphatase SSH-

2 were still identified.

2. In Vitro Verification

Through computational simulations, 22 compounds were identified as potential inhibitors for SSH-2, however, these results were obtained through simulations under ideal conditions. Compounds are placed close to the active site during the *in silico* experiments without competition of other substrates or interactions with other cellular processes, which is not the case inside the cell. Furthermore, compound concentration, solubility, and diffusivity are not taken into in the virtual experiments. Testing these compounds *in vitro* was then necessary to provide confirmation of the simulated results and demonstrate the effect of the identified compounds on cells.

Of the three compounds tested, only ZINC04307500 appeared to be able to inhibit SSH-2. In particular, cells treated with this compound showed an increase in actin filament intensity at 0.01 μM concentration. The cells showed either an unchanged or decreased amount of actin filaments when treated with higher concentrations. This observation may be explained by the prozone effect, where a saturated ligand concentrations leads to a decrease in lower ligand efficiency as discussed in the results section.

When cells were treated with ZINC03377116 and ZINC06601214, a decrease in actin filaments is generally observed. This decrease may be attributed to the nonspecific binding of these compounds. They may inhibit key proteins that are involved in the actin polymerization process and, thus, decreasing the amount of actin filaments being formed.

Another reason that these compounds may not have successfully inhibited SSH-2 is because of errors in the algorithm of DOCK 6.0 or the parameters used for the virtual experiments. Since SSH-2 was crystallized through a cysteine to serine mutation [6], the residue was mutated back for the purpose of the *in silico* experiments. Although the model with the highest confirmation probability was selected, the three-dimensional structure of the wild type SSH-2 may deviate from the selected model and may introduce some error in the virtual screening experiments. Furthermore, grid-scoring takes into account van der Waal and electrostatic forces, which represent an estimate of the enthalpy of the enzyme-substrate system. The calculation of entropy is not accounted for in this simulation, hence, the free energy of the system, which ultimately dictates the binding of the compound to the complex, is estimated from the enthalpy alone [45].

Since the virtual experiments are conducted under the assumption of ideal conditions, the compounds may not have interacted with the protein as predicted. First, the compounds may not pass readily through the cell membrane because of their size and structural properties. These compounds often contain polar or ionic groups such as the carboxylic acid end that is present in every identified potential inhibitor. Even if these compounds pass through the cell membrane, they may have been actively pumped back out by membrane transporters. This is a known problem for chemotherapy against cancer as multi-drug resistant proteins have been discovered to remove antitumor drugs from cells. An example is the multiple drug resistance-associated protein 1 (Mrp1), which has been found to be overexpressed in glioblastoma [52].

A second possibility is that the compounds did in fact inhibit SSH-2 but the effect is compensated by other cellular processes. Since the SSH group of proteins has three

isoforms, inhibiting only SSH-2 may not be enough to disrupt the structures of actin filaments. The cells may be able to continue to dephosphorylate cofilin through SSH-1 and SSH-3, thus no change in actin stress fibers is detected. It is important to note that the three dimensional structures of SSH-1 and SSH-3 have not been resolved by x-ray crystallography and could not be screened in this study.

Lastly, the binding affinity of these compounds to SSH-2 may be much lower than that of cofilin. These compounds would then be poor competitive inhibitors of SSH-2, because cofilin would easily displace these compounds from SSH-2. Furthermore, these two compounds may bind nonspecifically to other proteins not screened in this study. This would then decrease the effective concentration and, thus, unable to inhibit SSH-2 successfully.

V. CONCLUSION

This study was a large-scale virtual screening experiment to identify specific inhibitors for SSH-2. Studies have shown that inhibiting SSH-2 may decrease cancer metastasis [4]. Since SSH-2 belongs to the DUSP family that shares highly similar spatial structures, a specific inhibitor has not yet been identified. Using grid resources and parallel processing, 20,935 compounds from the ZINC database were docked to all 23 DUSPs with known crystal structures. Since the accuracy of DOCK depends heavily on the actual protein three dimensional structures, the remaining 38 DUSPs were not screened. It would be interesting, however, to model the structures of these DUSPs using a protein folding software, such as MODELLER [53], and screen them against the ZINC library used in this study. The results from these simulations may further narrow down the list of potential inhibitors identified. Furthermore, more specific trends regarding the structural and chemical properties of the identified compounds may be inferred.

Through mutational analysis, this study concluded that arginine-398 of SSH-2 contributes to the binding stability of several compounds through hydrogen bonding. By using the compound docking orientations in combination with mutations to other SSH-2 residues, detailed interactions between SSH-2 and the bound compounds can be determined. Knowing the residues that are responsible for compound stability may be crucial for drug designs. Compounds that exhibit even higher binding affinities to SSH-2 may be synthesized based on the studied interactions, thus increasing the efficacy of these inhibitors, which is desirable in their application as cancer treatments.

The preliminary results of the *in vitro* verification of ZINC04307500 seem to be promising as the data supports the notion that SSH-2 was inhibited. Further testing, such as cancer cell migration studies, is required to fully assess the potential of the use of ZINC04307500 as a cancer treatment. If this compound can successfully and specifically inhibit SSH-2, it may have tremendous therapeutic benefits in a wide number of diseases.

VI. REFERENCE

1. Sheterline, P., J. Clayton, and J. Sparrow, *Actin.* Protein profile, 1995. **2**(1): p. 1-103.

2. Huang, T.Y., C. DerMardirossian, and G.M. Bokoch, *Cofilin phosphatases and regulation of actin dynamics.* Current opinion in cell biology, 2006. **18**(1): p. 26-31.

3. Maloney, M.T. and J.R. Bamburg, *Cofilin-mediated neurodegeneration in Alzheimer's disease and other amyloidopathies.* Molecular neurobiology, 2007. **35**(1): p. 21-44.

4. Scott, R.W. and M.F. Olson, *LIM kinases: function, regulation and association with human disease.* Journal of molecular medicine, 2007. **85**(6): p. 555-68.

5. Wang, W., R. Eddy, and J. Condeelis, *The cofilin pathway in breast cancer invasion and metastasis.* Nature reviews. Cancer, 2007. **7**(6): p. 429-40.

6. Jung, S.K., et al., *Crystal structure of human slingshot phosphatase 2.* Proteins, 2007. **68**(1): p. 408-12.

7. Patterson, K.I., et al., *Dual-specificity phosphatases: critical regulators with diverse cellular targets.* The Biochemical journal, 2009. **418**(3): p. 475-89.

8. Jeffrey, K.L., et al., *Targeting dual-specificity phosphatases: manipulating MAP kinase signalling and immune responses.* Nature reviews. Drug discovery, 2007. **6**(5): p. 391-403.

9. Niwa, R., et al., *Control of actin reorganization by Slingshot, a family of phosphatases that dephosphorylate ADF/cofilin.* Cell, 2002. **108**(2): p. 233-46.

10. Bessette, D.C., D. Qiu, and C.J. Pallen, *PRL PTPs: mediators and markers of cancer progression.* Cancer metastasis reviews, 2008. **27**(2): p. 231-52.

11. Stephens, B.J., et al., *PRL phosphatases as potential molecular targets in cancer.* Molecular cancer therapeutics, 2005. **4**(11): p. 1653-61.

12. Gray, C.H., et al., *The structure of the cell cycle protein Cdc14 reveals a proline-directed protein phosphatase.* The EMBO journal, 2003. **22**(14): p. 3524-35.

13. Denu, J.M., et al., *Form and function in protein dephosphorylation.* Cell, 1996. **87**(3): p. 361-4.

14. Fauman, E.B. and M.A. Saper, *Structure and function of the protein tyrosine phosphatases.* Trends in biochemical sciences, 1996. **21**(11): p. 413-7.

15. Denu, J.M. and J.E. Dixon, *A catalytic mechanism for the dual-specific phosphatases.* Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(13): p. 5910-4.

16. Bakan, A., et al., *Toward a molecular understanding of the interaction of dual specificity phosphatases with substrates: insights from structure-based modeling and high throughput screening.* Current medicinal chemistry, 2008. **15**(25): p. 2536-44.

17. Kristjansdottir, K. and J. Rudolph, *Cdc25 phosphatases and cancer.* Chemistry & biology, 2004. **11**(8): p. 1043-51.

18. Reynolds, R.A., et al., *Crystal structure of the catalytic subunit of Cdc25B required for G2/M phase transition of the cell cycle.* Journal of molecular biology, 1999. **293**(3): p. 559-68.

19. Jeong, D.G., et al., *Structure of human DSP18, a member of the dual-specificity protein tyrosine phosphatase family.* Acta crystallographica. Section D, Biological crystallography, 2006. **62**(Pt 6): p. 582-8.

20. Yokota, T., et al., *Crystal structure of human dual specificity phosphatase, JNK stimulatory phosphatase-1, at 1.5 A resolution.* Proteins, 2007. **66**(2): p. 272-8.

21. Song, H., et al., *Phosphoprotein-protein interactions revealed by the crystal structure of kinase-associated phosphatase in complex with phosphoCDK2.* Molecular cell, 2001. **7**(3): p. 615-26.

22. Stewart, A.E., et al., *Crystal structure of the MAPK phosphatase Pyst1 catalytic domain and implications for regulated activation.* Nature structural biology, 1999. **6**(2): p. 174-81.

23. Muda, M., et al., *Molecular cloning and functional characterization of a novel mitogen-activated protein kinase phosphatase, MKP-4.* The Journal of biological chemistry, 1997. **272**(8): p. 5141-51.

24. Tao, X. and L. Tong, *Crystal structure of the MAP kinase binding domain and the catalytic domain of human MKP5.* Protein science : a publication of the Protein Society, 2007. **16**(5): p. 880-6.

25. Rosengren, K.J., et al., *Structural and functional characterization of the conserved salt bridge in mammalian paneth cell alpha-defensins: solution*

structures of mouse CRYPTDIN-4 and (E15D)-CRYPTDIN-4. The Journal of biological chemistry, 2006. **281**(38): p. 28068-78.

26. Vasudevan, S.A., et al., *MKP-8, a novel MAPK phosphatase that inhibits p38 kinase.* Biochemical and biophysical research communications, 2005. **330**(2): p. 511-8.

27. Patterson, K.I., et al., *DUSP26 negatively affects the proliferation of epithelial cells, an effect not mediated by dephosphorylation of MAPKs.* Biochimica et biophysica acta, 2010. **1803**(9): p. 1003-12.

28. Bolino, A., et al., *Charcot-Marie-Tooth type 4B is caused by mutations in the gene encoding myotubularin-related protein-2.* Nature genetics, 2000. **25**(1): p. 17-9.

29. Begley, M.J., et al., *Crystal structure of a phosphoinositide phosphatase, MTMR2: insights into myotubular myopathy and Charcot-Marie-Tooth syndrome.* Molecular cell, 2003. **12**(6): p. 1391-402.

30. Lee, J.O., et al., *Crystal structure of the PTEN tumor suppressor: implications for its phosphoinositide phosphatase activity and membrane association.* Cell, 1999. **99**(3): p. 323-34.

31. Sun, J.P., et al., *Structure and biochemical properties of PRL-1, a phosphatase implicated in cell growth, differentiation, and tumor invasion.* Biochemistry, 2005. **44**(36): p. 12009-21.

32. Kozlov, G., et al., *Structural insights into molecular function of the metastasis-associated phosphatase PRL-3.* The Journal of biological chemistry, 2004. **279**(12): p. 11882-9.

33. Kim, S.J., et al., *Crystal structure of human TMDP, a testis-specific dual specificity protein phosphatase: implications for substrate specificity.* Proteins, 2007. **66**(1): p. 239-45.

34. Yuvaniyama, J., et al., *Crystal structure of the dual specificity protein phosphatase VHR.* Science, 1996. **272**(5266): p. 1328-31.

35. Yoon, T.S., et al., *Crystal structure of the catalytic domain of human VHY, a dual-specificity protein phosphatase.* Proteins, 2005. **61**(3): p. 694-7.

36. Agarwal, R., S.K. Burley, and S. Swaminathan, *Structure of human dual specificity protein phosphatase 23, VHZ, enzyme-substrate/product complex.* The Journal of biological chemistry, 2008. **283**(14): p. 8946-53.

37. Koksal, A.C., J.D. Nardozzi, and G. Cingolani, *Dimeric quaternary structure of the prototypical dual specificity phosphatase VH1.* The Journal of biological chemistry, 2009. **284**(15): p. 10129-37.

38. Jeong, D.G., et al., *Crystal structure of the catalytic domain of human DUSP5, a dual specificity MAP kinase protein phosphatase.* Proteins, 2007. **66**(1): p. 253-8.

39. Pham, P.D.L., M. J.; Ichikawa, K.; Date, S.; Haga, J. H.;, *Identification of a Specific Inhibitor for the Dual-Specificity Enzyme SSH-2 via Docking Experiments on the Grid.* Fourth IEEE International Conference on eScience, 2008: p. 547-554.

40. Irwin, J.J. and B.K. Shoichet, *ZINC--a free database of commercially available compounds for virtual screening.* Journal of chemical information and modeling, 2005. **45**(1): p. 177-82.

41. Lipinski, C.A., *Drug-like properties and the causes of poor solubility and poor permeability.* Journal of pharmacological and toxicological methods, 2000. **44**(1): p. 235-49.

42. *UCSF DOCK*. Available from: http://dock.compbio.ucsf.edu/.

43. Levesque, M.J.I., K.; Date, S.; Haga, J. H., *Bringing Fleixibility to Virtual Screening for Enzymatic Inhibitors on the Grid.* 9th IEEE/ACM International Conference on Grid Computing, 2008.

44. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* Journal of Computational chemistry, 2004. **25**(13): p. 1605-12.

45. Meng, E.C.S., B. K.; Kuntz, I. D., *Automated Docking with Grid-Based Energy Evaluation.* Journal of Computational chemistry, 2004. **13**(4): p. 505-524.

46. Moustakas, D.T., et al., *Development and validation of a modular, extensible docking program: DOCK 5.* Journal of computer-aided molecular design, 2006. **20**(10-11): p. 601-19.

47. Cao, Y., et al., *ChemmineR: a compound mining framework for R.* Bioinformatics, 2008. **24**(15): p. 1733-4.

48. Holliday, J.D., et al., *Analysis and display of the size dependence of chemical similarity coefficients.* Journal of chemical information and computer sciences, 2003. **43**(3): p. 819-28.

49. Barbarakis, M.S., et al., *Observation of "hook effects" in the inhibition and dose-response curves of biotin assays based on the interaction of biotinylated glucose oxidase with (strept)avidin.* Analytical chemistry, 1993. **65**(4): p. 457-60.

50. Kim, J.-H.Y., J-Y;, *Protein Adsorption on Polymer Particles*, in *Encylopedia of Surface and Colloid Science*2002. p. 4373-4381.

51. Britson, J.S., et al., *Deregulation of DUSP activity in EGFR-mutant lung cancer cell lines contributes to sustained ERK1/2 signaling.* Biochemical and biophysical research communications, 2009. **390**(3): p. 849-54.

52. Peignan, L., et al., *Combined Use of Anticancer Drugs and an Inhibitor of Multiple Drug Resistance-Associated Protein-1 Increases Sensitivity and Decreases Survival of Glioblastoma Multiforme Cells In Vitro.* Neurochemical research, 2011.

53. Eswar, N., et al., *Comparative protein structure modeling using MODELLER.* Current protocols in protein science / editorial board, John E. Coligan ... [et al.], 2007. **Chapter 2**: p. Unit 2 9.