EVALUATION OF *AUTODOCK VINA* FOR USE IN FRAGMENT-BASED DRUG DISCOVERY

BY

MICHAEL PAUL STARLING

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

In partial fulfillment of the requirements

For the degree of

MASTER OF SCIENCE

in

Chemistry

May, 2013

Nashville, Tennessee

Approved:

Dr. Walter J. Chazin

Dr. Michael P. Stone

To my parents and brother, Robbie, Susan, and Matthew, for always being supportive and encouraging me all along the way.

To my friends who have taken the time to have a drink with me.

To Sandra Ford for everything she has been and done.

ACKNOWLEDGMENTS

First, I would like to thank my advisor, Professer Walter Chazin, who has gone above and beyond to help me succeed in the lab and in life. The amount of gratitude I have for him and what he has done for me cannot be overstated. I must also thank the lab the people that work in his lab. Dr's. Sivaraja Vaithlyngham, Michael Shell, Nicholas George, and Michel Feldkmap have all donated their time directly mentoring me and the projects I have worked on. Although not directly involved in my projects, Dr's. Sarah Soss, Steve Damo, and Chris Brosey all made significant contributions to my development as a scientist and an individual which cannot be ignored. Dr. Markus Voehler for the countless help with the NMR and his opinion on flying toilet paper. Catherine, Kallie, Ewa, Marta, Lian, Norie, Cheryl, Agnieszka, Amanda, Dungeng, Rachel, Gus, and Yan have all made direct impacts in my development and made my time in graduate school a time in my life that I will never forget.

It would be a complete oversight not to include my undergraduate advisor, Dr. Kevin Williams, as much of my understanding and pursuit of science can be traced directly to him. In addition to Dr. Williams, Dr. Pesterfield, Alicia, Dr. Igumenova, Kevin, John, and Goose all have made significant contributions to my love of science and been there when my love was turning to annoyance or hate. I must thank my ex-roommate, Tony, and current roommates, Dennis and Amir, for graciously dealing with me and the battles that come along with graduate school. Although this thesis is in part dedicated them, not enough thanks can be given to my friends and most importantly, my family, who have relentlessly supported me throughout my life and have strived to make sure I am happy and successful.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

Chapter 1

Introduction

1.1 – DNA damage response and repair

Prevention of modifications to our genetic code from spontaneous and environmental modifiers is a constant battle facing all cells. Current estimations suggest that each cell could face up to 10^5 lesions per day from spontaneous errors. The spontaneous errors are the result of mismatching of bases during DNA replication, interconversion of bases by deamination, and modification by alkylation. Additionally, reactive oxygen species (ROS) derived from normal cellular metabolic processes can oxidize DNA bases and create breaks in the DNA backbone. Environmental modifications to the DNA are produced by chemical and physical sources. Examples of these include ultraviolet radiation (UV) and ionization radiation (IR). Ultraviolet radiation from sunlight exposure can induce pyrimidine dimers and 6-4 photoproducts resulting in 10^5 lesions per cell per day[1,2]. Exposure to ionization radiation from cosmic radiation, chemical agents used in cancer chemotherapeutics, and medical treatments using X-ray or radiotherapy can cause DNA bases to be oxidized resulting in the formation of single-strand and double-strand DNA breaks (SSBs and DSBs), as well as interstrand cross-links.

ATR and CHK1 Activation

The major regulators of the DNA-damage response (DDR) are the phosphoinositide 3-kinase (P3IK)-related protein kinases (PIKK's), including the ataxia-telangiectasia mutated (ATM) and RAD3 related kinase (ATR). ATR is a large kinase with a strong preference for phosphorylating serine or threonine residues that follow a glutamine residue. ATR activation occurs every S phase to regulate the firing of replication origins, the repair of damaged replication forks, and to prevent premature mitosis.[3]

Although ATR is activated in response to many different types of DNA damage, there is evidence that a single structure may be responsible for this activation. When a cell undergoes replication stress, stalling or collapsing of the fork, there is an uncoupling of the DNA helicase from the DNA polymerase. This uncoupling result in long stretches of single stranded DNA (ssDNA) that becomes coated with replication protein A (RPA). The RPA coated ssDNA serves as the initial recognition factor for ATR signaling.[4-6]

During DNA repair and replication, stretches of single-strand DNA (ssDNA) are formed as the DNA is unwound. Left unprotected, this ssDNA is susceptible to modifications that could lead to mutations in the genome. Replication protein A is the eukaryotic ssDNA binding protein and by binding to the ssDNA, protects it from modification during repair.[7,8] The RPA-ssDNA complex is required for localization of ATR to sites of damage. The localization occurs through a direct interaction of the RPA-ssDNA complex with ATR-interacting protein (ATRIP). Although ATRIP is a separate protein from ATR, it can be considered an obligate subunit of ATR as the stabilities of ATR and ATRIP are linked, their association is unregulated, and there is no known

difference in phenotypes from the loss of ATRIP or ATR.[3,6] The binding of ATRIP to RPA involves an acidic alpha helix in ATRIP that binds to a basic cleft in the N-terminal domain of the RPA70 subunit (RPA70N).[9]

Although localization of the ATR-ATRIP complex can be attributed to the interaction with the RPA-ssDNA complex, the activation of ATR cannot. Activation of ATR requires the co-localization of the ATR-ATRIP complex with the RAD9-RAD1-HUS1 (9-1-1) complex. The 9-1-1 complex is a heterotrimeric ring-shaped molecule that is homologous in sequence to the replicative sliding clamp proliferating cell nuclear antigen (PCNA). The 9-1-1 complex is loaded onto a stretch of DNA adjacent to the RPA-ssDNA complex, in an ATR-dependent reaction involving the clamp loader RAD17.[10] RPA is also crucial for the loading of the 9-1-1 complex, as it imparts specificity to the loading of the 9-1-1 complex to the 5' primer end. This 5'-junction substrate containing the RPA-ssDNA complex, is the structure that normally results in the activation of ATR as it is the resulting DNA structure when a polymerase stalls and is uncoupled during DNA replication.[11] The Rad 9 protein possesses a 120 amino acid c-terminal "tail" containing at least ten serine and threonine phosphorylation sites, in which one of these sites is phosphorylated by ATR but is not required for CHK1 activation. The other sites are phosphorylated by unknown kinases and at least one other site is essential for the activation of CHK1.[12,13]

In *Saccharomyces cerevisiae* it is reported that the 9-1-1 complex is responsible for directly activating ATR but there is no evidence that this occurs in other organisms. In humans there are two mediator proteins, Topoisomerase Binding Protein 1 (TopBP1) and Claspin, which are responsible for the activation of the CHK1.[13] The 9-1-1 complex

recruits through RAD9 the topoisomerase-binding protein-1 (TOPBP1). TOPBP1 contains multiple BRCA1 C-terminal (BRCT) domains and is required for ATR activation *in vivo*. Recruitment of TOPBP1 requires that the C-terminal tail of RAD9 be phosphorylated at serine 387. Phosphorylation of Ser387, creates a recognition site for the BRCT I and II domains of TOPBP1, which recruits TOPBP1 to ATR. TOPBP1 also contains an ATR activation domain, located between BRCT domains VI and VII, which interacts and activates the ATR-ATRIP complex *in* vitro. The mechanism by which TOPBP1 activates the ATR-ATRIP complex remains poorly understood as the primary binding site for the activation domain of TOPBP1 is located on ATRIP and mutation to this region block activation.[14,15] Claspin is thought to already be present at active replication forks during replication and is phosphorylated in an ATR dependent manner. Once Claspin is modified it binds to CHK1, recruiting CHK1 to sites of RPA-coated ssDNA. This brings CHK1 in close proximity to active ATR, where it can then be phosphorylated by CHK1 rendering it active.[16,17] Recent studies have found that TopBP1 and Claspin are enough to activate CHK1 in the absence of RPA-coated ssDNA *in vitro*.[18]

Additional mediator proteins have been found to be involved in the activation of CHK1 in response to replication stress, including Timeless, Tipin (Timeless-interacting protein) and RHINO (Rad9-Rad1-Hus1 interacting nuclear orphan). Timeless is known to bind both ATR and CHK1 whereas Tipin is known to interact with RPA32 and is required for stable interactions of Claspin with stretches of RPA-coated ssDNA. RHINO was recently identified based on its interactions with 9-1-1 complex. The interaction

maps to the N-terminal portion of RHINO, and depletion of RHINO was shown to lead to a reduction of CHK1 phosphorylation. [19-21]

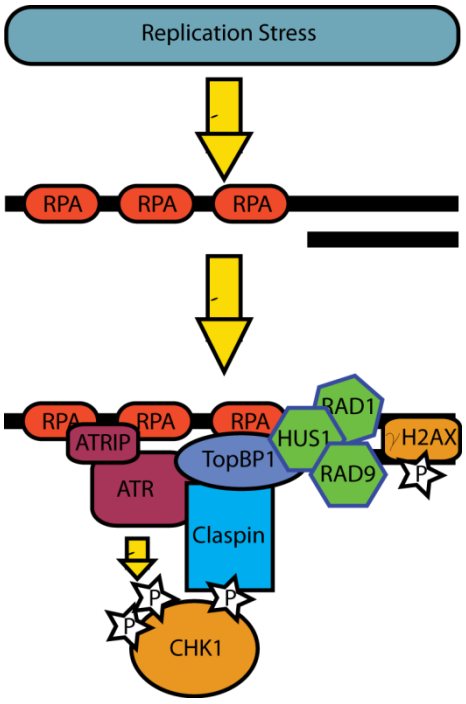


Figure 1. Major players in the CHK1 phosphorylation signaling cascade.

DNA Repair Pathways

To defend our genome against the hundreds of thousands of DNA lesions, our cells have developed multiple repair pathways; each specialized in repairing a different type of damage. Errors in replication that result in the mismatching of bases or insertion/deletion of bases are corrected by mismatch repair (MMR). Sites of dinucleotide repeats are unstable in some human cancers and display a "microsatellite instability" that is the result of defects in MMR. These defects are commonly seen in hereditary nonpolyposis colorectal cancer (HMPCC). The process of MMR can be broken down

into four steps: (1) recognition of the mismatch (2) recruitment of additional MMR factors (3) exonucelase degradation past the mismatch (4) re-synthesis of the excised tract.

Ultraviolet radiation resulting in the formation of 6-4 photoproducts and pyrimidine dimers, as well as bulky lesions like the polycyclic aromatic hydrocarbon benzo[a]pyrene are repaired by nucleotide excision repair (NER). Distortions in duplex DNA are recognized by XPC. The duplex DNA is unwound by two helicases, XPD and XPF, in TFIIH and signals for the formation of the pre-incision complex. Two endonucleases, XPF and XPG, cut the duplex and a 24-32 bp fragment containing the lesion is excised from the DNA. The resulting gap is filled by gap-filling replication machinery.

Ionizing radiation from X-rays, alkylating agents, and reactive oxygen species that induce the formation of abasic sites, 8-oxoguanine bases, and single-strand breaks are repaired through base excision repair (BER). The modified bases are flipped out of the duplex by a DNA glycosylase. The base is then excised from the DNA resulting in an abasic site. Incision of the DNA occurs at the abasic site, allowing for the 5'-baseless sugar to be removed and single base gap filling synthesis. The resulting nick is sealed completing the repair.

Double strand breaks (DSBs) and interstand cross-links arising from X-ray irradiation and anti-tumor agents such as cis-platin and mitomycin C (MMC) are repaired by the two branches of recombinational repair, homologous recombination (HR) and non-homologous end joining (NHEJ). HR is preferred and utilizes sister chromatids as

templates to precisely repair the genomic sequence, while NHEJ is an error-prone approach that links the two ends of a DSB together when HR is not possible.[1,2]

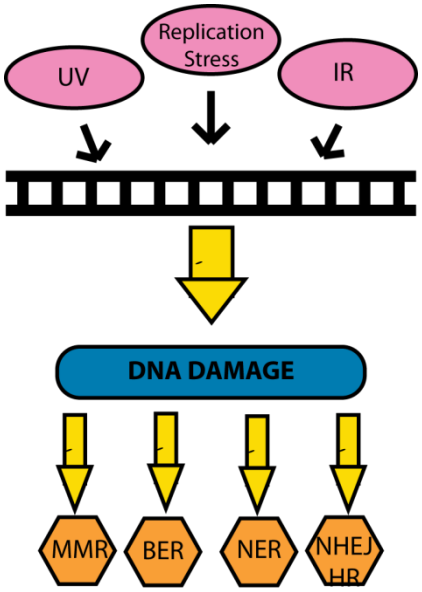


Figure 2. DNA Damage repair pathways specialized in repairing specific types of damage.

1.2 – DDR role in cancer progression

Over time, errors occur during replication and unrepaired damage builds up, which can lead to mutations in the genetic code. These mutations can disrupt the regulation of gene expression, resulting in uncontrolled cell proliferation and cancer. In 2005, two papers were published that brought forth a new model of how the DDR was involved in tumorigenesis in which DNA damage checkpoints become active in the early stages of tumorigenesis, providing a barrier to early tumor progression and directing cells to apoptosis. These studies reported evidence of an activated DDR response, noticeably phosphorylation of histone H2AX, ATM activation, RAD17 phosphorylation, CHK1 phosphorylation, and localization of oncogene binding proteins, from precancerous and cancerous lesions and tumor tissue. The authors proposed a mechanism leading to

genomic instability in early development: aberrant stimulation of cell proliferation leads to replication stress, which can activate the DNA damage checkpoint(s). The activation of the checkpoint(s) induces cell cycle arrest or apoptosis, which functions as a tumor suppressor.[22,23]

The following year two more studies were reported with evidence that the DDR was involved in cellular senescence. Oncogene-induced senescence was found to result from a robust DNA damage response (DDR) to replication stress and double strand breaks. The results, from cells and cancerous tissues, suggested that oncogene expression leads to an increase in DNA replication events, DDR activation, and genome activation. These findings led the authors to suggest a model in which oncogene expression leads to hyper-proliferation and DNA hyper-replication leading to an accumulation of DNA damage and sparking activation of S-phase specific DDR. The DDR proficient cells in turn undergo cellular senescence. Cells with an ineffective DDR undergo unrestrained proliferation fueled by oncogene expression. These reports suggest that senescence, which, like apoptosis, functions as a barrier to tumorigenesis.[24,25]

1.3 – Replication Protein A (RPA)

Replication Protein A (RPA) is a modular, heterotrimeric single-strand DNA (ssDNA) binding protein essential for replication, damage response and repair, which was first identified as an essential factor for replication of simian virus 40 DNA. RPA is composed of three conserved subunits: RPA70, RPA32, and RPA14. The RPA70 subunit is composed of four oligonucleotide/oligosaccharide binding fold (OB-fold) domains, termed 70N, 70A, 70B, and 70C. The DNA binding domains, 70A, 70B, and 70C are

connected by short 10 and 15 residue linkers, while the protein interaction module 70N is connected to 70A by a long (~70 residue) linker. The RPA32 subunit has three domains: the disordered 32N, an OB-fold domain 32D, and a winged-helix domain 32C. RPA32 participates both in protein interactions and binding ssDNA as the 32C domain is the other primary protein interaction domain of RPA, while 32D is involved binding to ssDNA. RPA32N is the site of poly-phosphorylation at seven sites within the first 32 residues, which occurs in response to cell cycle progression or DNA damaging agents. Phosphorylation of RPA70 has also been observed *in* vitro and *in* vivo. RPA70C, RPA32D, and RPA14 form the timerization core of the three RPA subunits. RPA binds DNA in a 5' → 3' direction with decreasing affinity from 70A to 32D.[7,8]

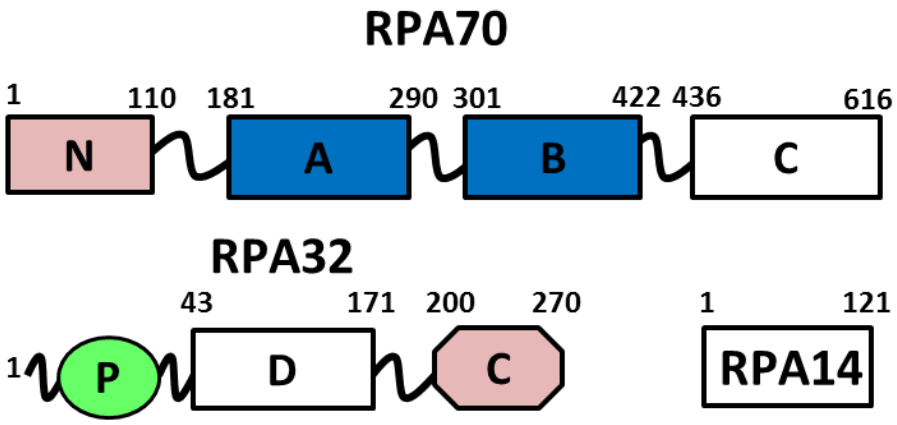


Figure 3. Domain mapping of Replication Protein A. 70N, A, B, C and 32D are all OB fold domains, 32C is unstructured and the n-terminus of RPA32 is unstructured.

DNA processing proteins like RPA are thought to utilize a common set of features to drive the progression of processing by "handing-off" of DNA: modular organization, multiple interaction sites, modest affinity, versatile structural modules, and a competition-based mechanism to promote hand-off. Most, if not all binding partners of

RPA interact through multiple contact points. Although each individual interaction is only of modest affinity, generally high micromolar, the total affinity of the interaction is approximately the product of the individual interactions due to the linkage effect. The use of multiple, modest affinity interactions helps drive the progression of DNA processing as proteins can be shuttled in and out in a sequential order based on their respective affinity. The OB-fold domains of RPA serve as a platform for interactions with both ssDNA and proteins. RPA70N contains an OB fold that does not bind to ssDNA but instead binds a number of proteins including ATRIP, RAD9, RAD17, MRE11, and p53.[9,26] RPA70A and 70B are used primarily for binding to ssDNA, but interactions with SV40 T-antigen helicase, XPA and DNA polymerase alpha-primase have also been mapped to these domains.[27]

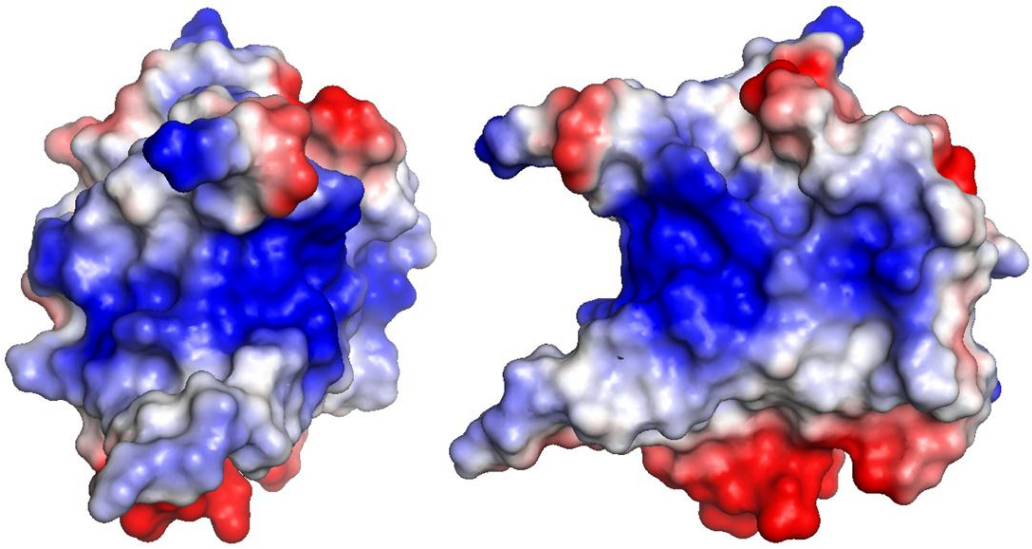


Figure 4. Electrostatic mapping of surface residues of RPA70N. The OB fold contains a large amount of basic residues, referred to as the basic cleft.

The structures of all of the globular domains of RPA and of the multi-domain constructs such as the trimer core have been determined by x-ray crystallography or

NMR, but the knowledge of the organization of the quaternary structure of RPA is scarce.[28-33] Thus our understanding of the global architectural changes that RPA undergoes has been limited. Recently the structure of the DNA binding core, RPA70ABC and 32D, of RPA from *Ustilago maydis* bound to a 32-mer DNA substrate was determined by the Palvetich group using X-ray crystallography. This represented the first view of the quaternary structure of the full ssDNA binding core of RPA engaged on ssDNA. The authors propose a key role for the linker between domains 70B and 70C, which has a central role in a four-way interface between 70B, 70C, and the intervening ssDNA between these domains that are not engaged in the binding pockets. They suggest that the engagement of the BC linker would serve as the key step in transitioning between the three different ssDNA binding modes of RPA.[34] However, no supporting evidence was provided in support of this proposal. Moreover, the crystal structure has an extraordinarily high degree of packing contacts and the solution-based study described below shows that this architecture represents an extreme that cannot be highly populated in solution.

Our lab has completed an in-depth study of the architecture of RPA alone and as it engages ssDNA using a combination of small angle x-ray and neutron scattering (SAXS and SANS) with molecular dynamics simulations. The authors report that: 1) the apo protein is somewhat less extended than anticipated; 2) upon binding to a 10-mer ssDNA substrate, RPA shifts to a more compact arrangement with both 70A and 70B engaging the DNA. At this point the trimer core still retains some flexibility. Upon binding longer ssDNA, which engages 70C and 32D from the trimer core, RPA transitions to a slightly more compact form and the flexibility between RPA70AB and the

trimer core is further reduced. The R_g value for the 30-nt binding mode is only 2 Å smaller than the 10-nt binding, suggesting relatively small global architectural rearrangements when the 70C and 32D in the trimer core are engaged. These results contrast with the widely accepted model in which RPA has three modes of binding ssDNA: there is no evidence for a distinct intermediate DNA binding model.[35] Overall, these studies presented new insights on the remodeling of RPA architecture and how this drives the progression of the DNA processing machinery.

1.4 – Inhibiting the DDR

The key development in understanding DDR was the discovery of its role in creating a barrier to tumorigenesis through apoptosis and senescence. This, along with additional research that has shown initial lesion generation by oncogenes is due to replication stress, has led to efforts to develop small molecule inhibitors (SMIs) of the DDR as a cancer therapeutic. Some success has already been seen with poly ADP-Ribose polymerase (PARP) inhibitors and their involvement in BRCA mutation-associated breast cancer.[36,37] Many groups have also looked into inhibiting key members of the ATR signaling cascade, including ATR, RPA, and CHK1.

ATR signaling inhibitors are expected to sensitize cells to DNA-damaging agents. Thus, cancer cells may be more dependent on replication stress responses than normal cells to complete replication and remain viable. Replication in the presence of DNA damage may lead to mitotic catastrophe, an event in which a cell is destroyed during mitosis because of aberrant chromosome segregation or DNA damage.[9] Current cancer treatments involve chemo- and radio- based therapies that induce DNA damage, resulting

in signaling for repair and/or apoptosis. If this is combined with potential inhibitors of the DDR proteins such as CHK1 or ATR, this could prevent repair of the cancer cells, making them more toxic and ideally signaling them for apoptosis.[38]

Inhibiting ATR

In 1999, caffeine was identified to be an inhibitor of ATR and ATM.[39] Caffeine had previously been known to sensitize cells. Although the concentrations that were determined to inhibit ATR are well above the acceptable concentration for therapeutic use, it serves as good inhibitor for biochemical studies of understanding ATR signaling and CHK1 activation. Recently a group has developed a high throughput cellular assay system to identify ATR inhibitors. Their system used a fusion of the ATR activation domain of TopBP1 with a fragment of the estrogen receptor, $TopBP1^{ER}$ to activate ATR. This screen captures molecule that inactive the ATR response. ATRi only inhibits CHK1 phosphorylation, while DDRi was found to inhibit ATR as well as ATM, CHK2, and DNA-PKcs. DDRi is a previously identified compound as a dual PI3K and mTOR inhibitor with strong anticancer activity.[40]

Inhibiting CHK1

Development of small molecule inhibitors for CHK1 has been under investigation for over a decade. Biochemical studies showed that inhibition of CHK1 results in a rapid and strong phosphorylation of ATR targets in S-phase cells. This is accompanied with an increase in firing of DNA replication origins, massive induction of ssDNA, generation of double strand breaks, resulting in a destabilization of the genome. These observations

suggest that CHK1 mediates control of the initiation of DNA synthesis required for normal progression though S-phase in human cells. Additionally, it is thought that CHK1 inhibition causes DNA breaks and activation of ATR due to lack of CHK1-mediated maintenance of stalled replication forks. CHK1 and ATR are both required to prevent the collapse of these stalled replication forks after treatment with replication inhibitors. ATR is required to prevent fork collapse during normal S-phase and avoid chromosomal break during mitosis. It is possible that CHK1 is required to prevent fork collapse during normal S-phase progression based on the results.[38]

Several CHK1 inhibitors have been tested in pre-clinical and trials as potential cancer therapeutics. Early preclinical trials with UCN-01 were promising but phase 1 clinical trials were unsuccessful as UCN-01 had a small volume of distribution, low systemic clearance, and a prolonged half-life, both due to its high binding affinity to α1-acid glycoprotein (AAG) in human plasma.[41-43] AZD772 is a urea-based inhibitor that is potent for CHK2 but is >10 fold more selective for CHK1, which was optimized from a high-throughput screening hit. It enhances the antitumor activity and abrogates S- and/or G2-phase checkpoints mediated by both antimetabolites and DNA-damaging agents. Results from phase 1 clinical trials have yet to be published.[44-46] PF477736 is a diazepinoindolone-based inhibitor of CHK1 but also significantly inhibits CHK2, although this inhibitor is 100-fold more selective for CHK1. This inhibitor abrogates the induction of cell cycle checkpoints and potentiates the activity of several DNA-damaging agents across a broad spectrum of p53-deficient human cancer preclinical models. This inhibitor was discontinued as of September 2010.[47-49] SCH900776 is an ATP-competitive inhibitor of CHK1. Unlike the previous two inhibitors it does not have affinity for CHK2;

however, it does have activity towards CDK2. This off-targeting effect on CDK2 could reduce the efficacy of this inhibitor as inhibition of CDK2 could induce cell cycle arrest and prevent checkpoint bypass in response to CHK1 inhibition. Phase 1 trials are ongoing with the goal of determining dosing for phase II trials.[50,51]

Inhibiting RPA

Inhibition of RPA has becoming an increasing popular target due to its many roles in DDR and repair. The N-terminal OB-fold domain of the RPA70 subunit (RPA70N) is known to interact with RAD9, RAD17, and ATRIP, all of which are involved in CHK1 activation. Studies from the Turchi group have identified multiple small molecule inhibitors of RPA using an electromobility shift assay (EMSA). They found compounds that inhibit the RPA70AB OB fold domains, which are the initial, high affinity ssDNA binding domains, as well as two compounds that inhibit RPA but not via RPA70AB. Similar to caffeine inhibition of ATR, these inhibitors are very weak (high micromolar) binders that are not potent enough to be used as a therapeutic but have some potential for use in biological assays.[52-54]

Other groups found small molecules that bind to the RPA70N domain, in the basic cleft that is the primary interaction surface with ATRIP and other DDR proteins. One group has identified a compound using an ELISA-like high throughput assay. The compound was selected based on its ability to inhibit the RAD9-RPA70N interaction, a known interaction during ATR signaling. Molecular modeling suggested that it mode of action was by binding to and competing for the RPA70N basic cleft though no direct evidence has been obtained. However, in agreement with the modeling, the compound

was also found to have no effect on the binding of ssDNA, suggesting it does not interact with any of the four domains involved (RPA70A, B, C and RPA32D), leaving RPA70N and RPA32C as prime targets.[55]

Work from a collaboration the Chazin and Fesik laboratories has produced a high-throughput fluorescence polarization assay to identify small molecules that bind to RPA70N. Compounds are selected based on their ability to displace a FITC-labeled ATRIP peptide from the basic cleft of RPA70N. This assay was tested using the compound identified by the ELISA-like technique and a similar binding affinity was obtained. Multiple small molecules have been identified using this fluorescence polarization assay of varying affinity.[56] This thesis presents studies that are part of the effort to identify potent inhibitors of RPA70N protein interactions.

1.5 – Methods to identify small molecule inhibitors of protein-protein interactions

Fragment-based drug discovery (FBDD) has become an increasingly popular route to the development of potent small molecule inhibitors. The approach is based on the identification of multiple low mass, weakly binding fragments that are linked together to generate high overall affinity. FBDD is applicable to high throughput screening, which allows for large fragment libraries typically containing thousands of molecules to be screened in a rather short time frame. Constructing suitable libraries can be a difficult task as fragments for protein-protein interactions tend to be more hydrophobic and heavier than fragments for other applications such as targeting kinases. Since the each fragment could potentially be a scaffold for a future compound, it is important to avoid reactive, toxic, and unstable fragments containing alkylating or acylating groups, as these

can lead to unwanted products. Fragments libraries are typically used in medium or high-throughput screening, which can involve a wide variety of different techniques.[57]

Fluorescence-based thermal shift (TS) detects compounds that increase the unfolding temperature of the target protein by binding to and stabilizing the folded state of the protein. This technique should be seen as an enrichment screening of the target library, as a hit is usually measured a $\Delta T = 1\ ^{o}C$, these results are not always highly reproducible.[57,58]

Fluorescence anisotropy has been used frequently because it can be easily adapted to high throughput screening. A fluorophore is attached to a small molecule tumbling fast in solution. When the small molecule becomes bound to the target biomolecule the tumbling rate slows down dramatically and a difference in anisotropy can be measured. The major challenge in this approach is the need to label each fragment molecule with a fluorophore. It is much more common to use a displacement assay where a fluorescently labeled ligand is pre-bound to the protein and screening is made to find molecules that displace the probe.[56]

Nuclear Magnetic Resonance (NMR) Spectroscopy has been used extensively in identification of small molecules. A wide range of approaches have been developed. Typically in 1D experiments (STD, WaterLOGSY, TINS) protein or ligand signals are monitored, with attenuation of intensity indicating the presence of a binder.[57,59-62] In strategies that use 2-dimensional (2D) NMR like SAR by NMR, target molecules are titrated with ligands and chemical shifts are monitored and perturbations of select shifts indicate the presence of a binder.[63] If assignments are available, 2D experiments can give information on the location of the ligand.

X-ray crystallography provides validation of hits and structural binding information in one step. If crystallization conditions are known then a crystallographic approach can be robust; as synchrotron collection and molecular replacement allow for quick, high resolution structures. Direct screening of compounds has been utilized in certain applications where the libraries are limited and high through-put crystallography systems are available.[57,64,65]

Surface plasma resonance (SPR) has also been used for screening. SPR requires the target protein to be covalently linked to gold clusters on the SPR biosensor chip and fragment solutions are passed over the chip. As fragments bind the change in mass is measured in real time. From the time dependent association-dissociation events, binding kinetics and affinities can be calculated.[57,66,67]

Computation-based virtual screening is another widely used approach. It typically involves one of two strategies: (i) produce a structural model of the binding mode of a known hit fragment when no structural data is available; (ii) *in silico* screening. For docking of ligands to be useful for screening, it must not only produce the correct binding mode but must take in to account free energy change of the ligand in solvent binding to the protein.[57,68,69]

In FBDD, after a hit or hits from a fragment are verified, generally the fragment is elaborated to produce analogs that bind tighter, have more drug-like properties, and help selectivity. Elaboration is usually achieved by one or a combination of different techniques. Merging is the incorporation of structural portions of overlapping molecules into a fragment, using structural information of other known fragments, substrates, and ligands in complex with the protein. Linking is the efficient “linking” of two fragments

that are known to bind in non-overlapping sites. A set of fragments that bind at a single site are discovered and "grown" through chemical synthesis to explore further interactions. Typically a common chemical scaffold is found and modifications are grown off of the scaffold.[57]

A paper by Shuker et al, highlights many of the strategies discussed above and introduces the methodology for a technique designed to obtaining structure-activity relationships of small molecules with proteins using NMR. Selected small molecules can then be linked together using synthetic chemistry to obtain high-affinity ligands. This approach, commonly referred to as SAR by NMR (**S**tructure-**A**ctivity **R**elationships by **N**uclear **M**agnetic **R**esonance), allows the user to obtain the structure-activity relationship by monitoring chemical shift perturbations of ^{15}N heteronuclear single quantum coherence spectra (HSQC). The process involves screening against a library for identification of a fragment molecule that binds to a specific site of the protein. This molecule is then optimized to obtain a higher affinity analog. The second fragment molecule is identified by performing the screens with saturating amounts of the first fragment molecule. This helps ensure that the second fragment binds in a different location than the first. Once a second site binder is identified, the molecule is optimized and then is attached to the first fragment molecule by building a carbon linker using synthetic chemistry. These linked fragments should yield a high-affinity compound as the binding affinity of the compound is the product of the two fragment molecules plus a term that accounts for changes in binding affinity due to linkage.[63] The compounds investigated here were generated using the SAR by NMR approach.

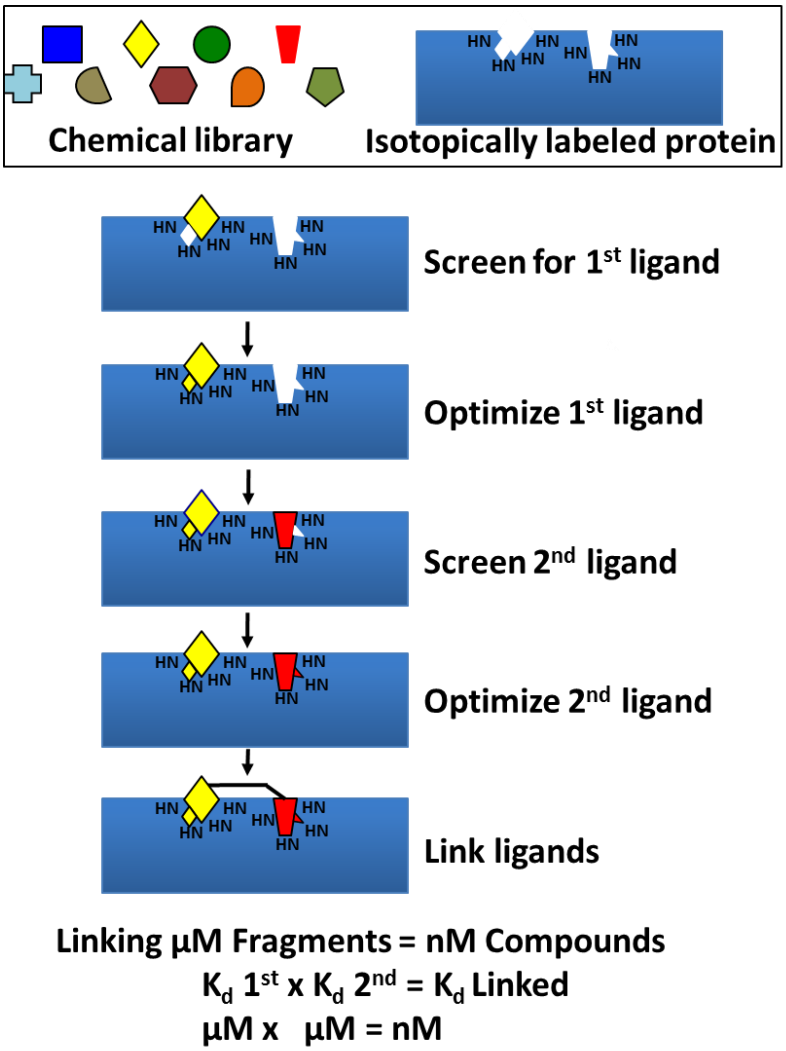


Figure 5. Overview of Structure-activity relationships (SAR) by NMR.

Chapter 2

Evaluation of the Accuracy of *Autodock Vina* for Modeling the Binding to RPA70N of Molecular Fragments from the Vanderbilt SAR-by-NMR Library

2.1 – Introduction and background

Previous work on the project to discover high affinity inhibitors of RPA70N protein interactions involved SAR by NMR screening of a curated 15,000-member fragment library. Two dimensional (2D) ^{15}N-^{1}H heteronuclear signal quantum coherence (HSQC) NMR spectroscopy was used to identify fragments that interact with the basic cleft of RPA70N. Fragments that were determined to interact by monitoring chemical shift perturbations, were then further classified based on whether their primary interaction site was S55 (site 1) or T60 (site 2). Identification of promising fragments were chosen as leads. These leads were optimized, leading to the production of analogs and the synthesis of potent compounds from these analogs. The binding affinities of the compounds were tested using a high throughput fluorescence anisotropy binding assay. Affinities were measured based on the ability of compounds to displace a FITC-labeled ATRIP peptide from the basic cleft of RPA70N.[56] A select number of compounds of interest were chosen for crystallization.

Initial crystallization attempts of wild-type (WT) RPA70N were successful in crystallizing the protein but the crystal packing was not ideal for co-crystallization with ligands as access to the basic cleft is blocked by an interaction with the C-terminus of a symmetry mate in the unit cell. This prompted Dr. Michael Feldkamp to re-engineer RPA70N by introducing charge reversal E-R point mutations at E7 and E100. Both single

mutations and the double mutant were prepared and all resulted in a change in crystal packing. Instead of the basic cleft being occluded, the mutations caused the protein to pack in a way that the basic cleft of symmetry mates forms an open channel. This channel allows for full access to the basic cleft for solvent and ligands. All of the structures discussed here were determined with the E7R mutant. For convenience, the E7R mutation will not be stated each time the structure of RPA70N is mentioned.

In the present study, we were interested is the applicability of *Autodock Vina* as a virtual screening tool.[70] In order to assess the applicability of this approach, six compounds were studied that were first selected from the original SAR by NMR screen and then elaborated. Four of the compounds have been co-crystallized with RPA70N and the other two have no structural data available. For the compounds that were not crystallized, two sets of runs were performed using different RPA70N structures. One was crystal structure of the protein alone (2B29) and the other is the RPA70N structure extracted from the co-crystal structure of the complex with an ATRIP peptide (2B3G).[26] By using a free and bound state of the protein, different binding sites may be revealed and information can be inferred about what residues are involved in binding of the ligand.

2.2 – Experimental methods

Molecular models were generated using Autodock Vina version 1.1.2. A series of comparisons were made for models generated for different exhaustiveness values. The exhaustiveness is a function that controls how "exhaustive" a search the program performs to find a global minimum. Default setting have an exhaustiveness value = 8. Other values tested in this study are 64, 128, 512, and 2048 for select compounds.

Exhaustiveness values between 8 and 64 represent values that would be ideal for screening, as they provide rapid results. Higher exhaustiveness values perform more extensive search but are inefficient time wise for screening. A test was also made to compare free docking to anywhere on the target molecule to guided docking in which a region of the target is specified as containing the binding surface. The results from these studies are presented below for each compound.

To validate that all compounds bound as expected in the basic cleft of RPA70N, NMR chemical shift perturbations induced in the protein were measured for each of the six compounds. Samples of ^{15}N-enriched RPA70N were prepared using standard methods. Solutions of the labeled protein at 130 μM concentration were prepared in a buffer containing, 10 mM Tris pH 7.5, 250 mM NaCl, and 10% $^{2}H_2O$. ^{15}N-^{1}H SO-fast HSQC NMR spectra were acquired for each sample in the absence and presence of the compound, which was added as a 10 μM solution in 13 μL of DMSO. Using the assignments of the RPA70N of ^{15}N and ^{1}H backbone resonances, the location of the binding site was extracted based on the residues whose chemical shifts were perturbed by addition of the compound. The NMR spectral overlays are shown in Figure 6.

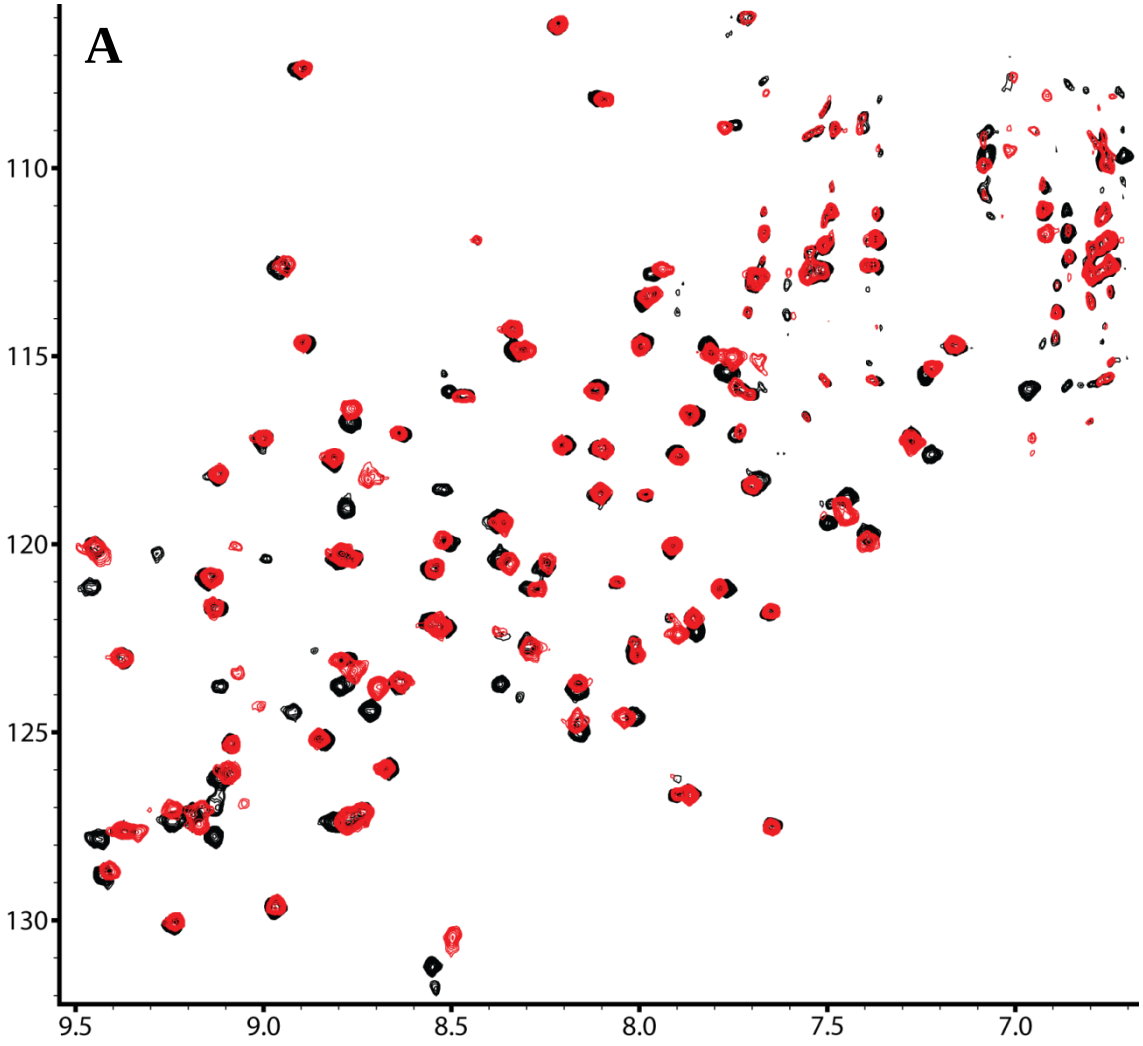
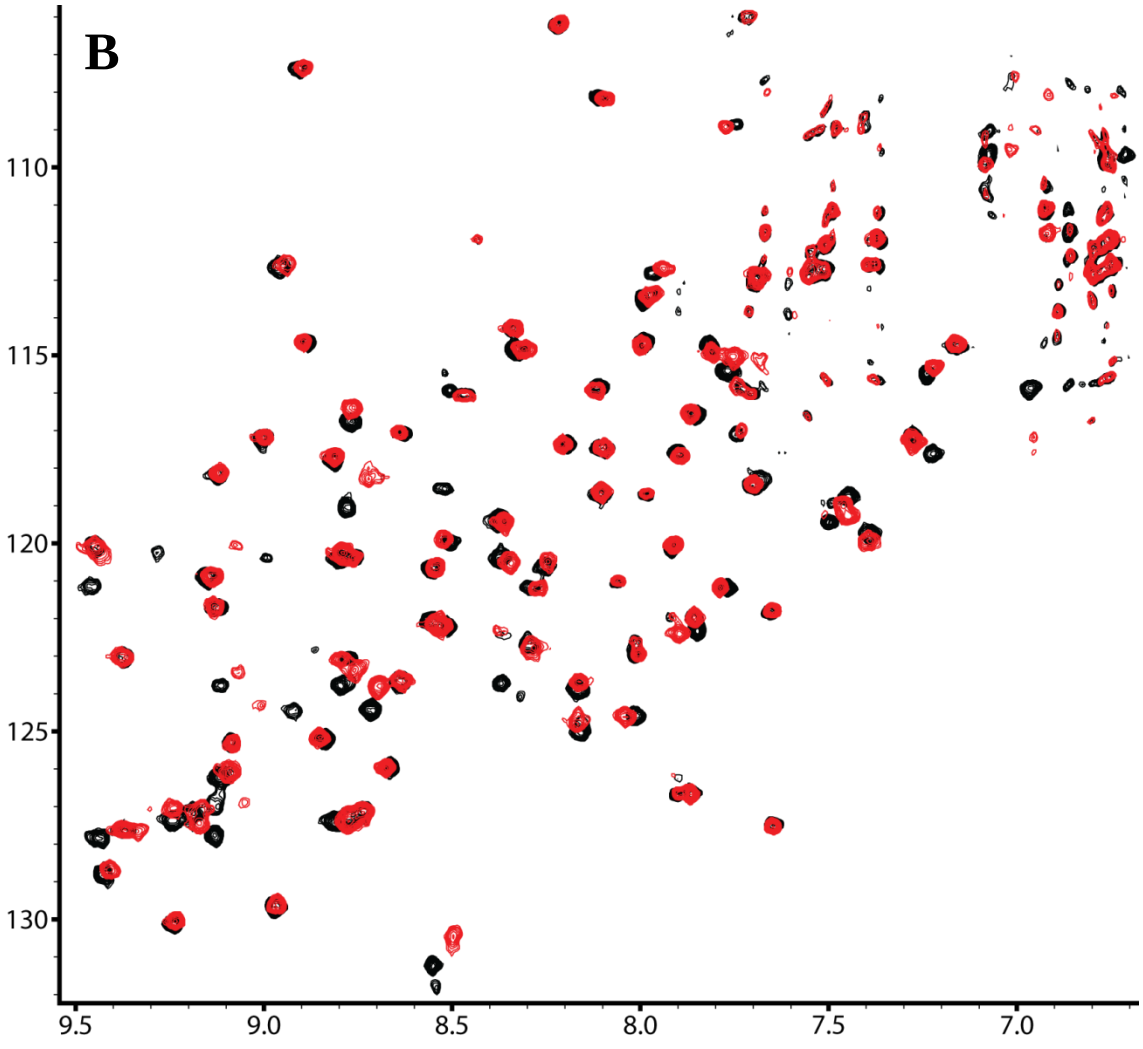
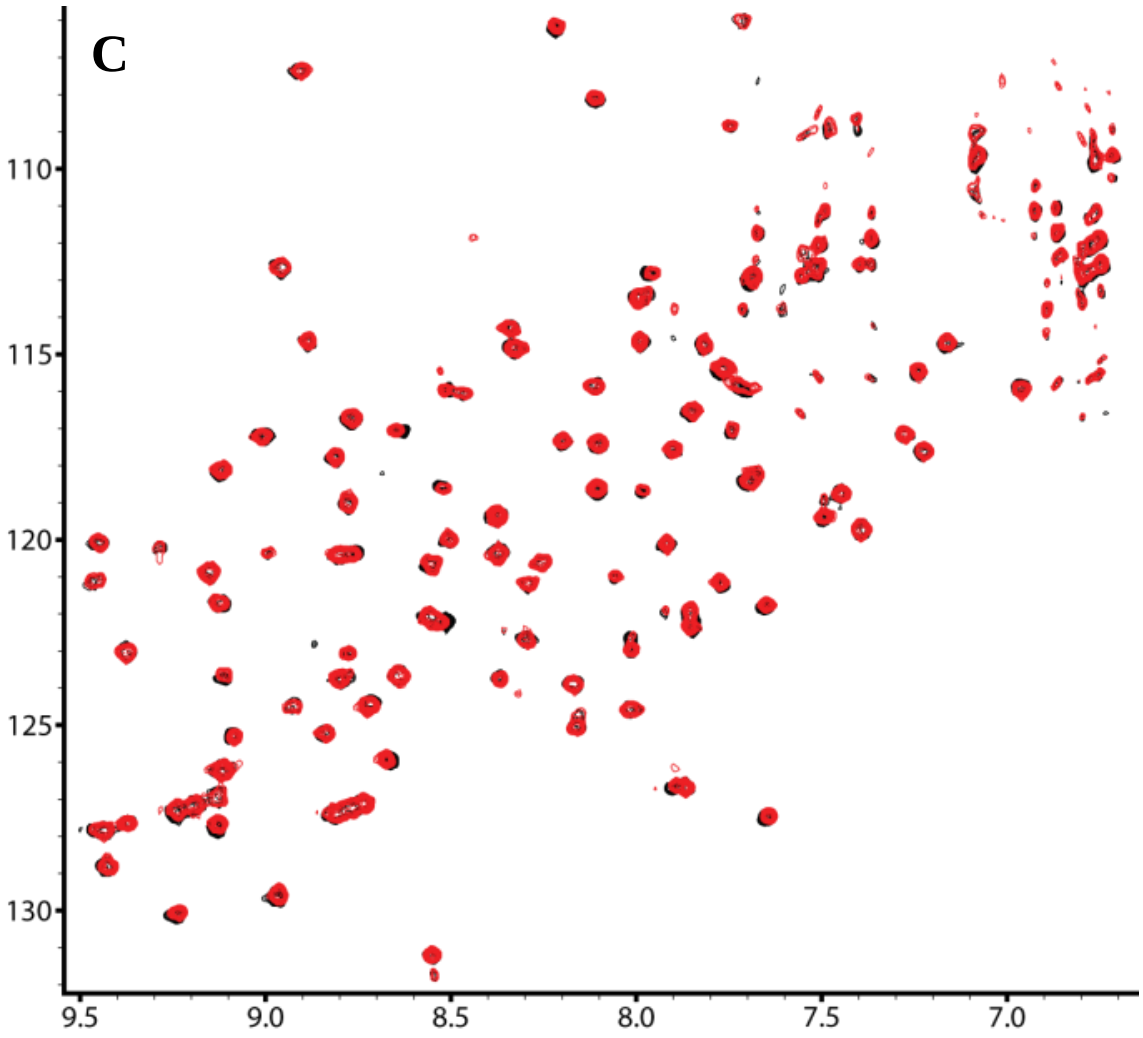
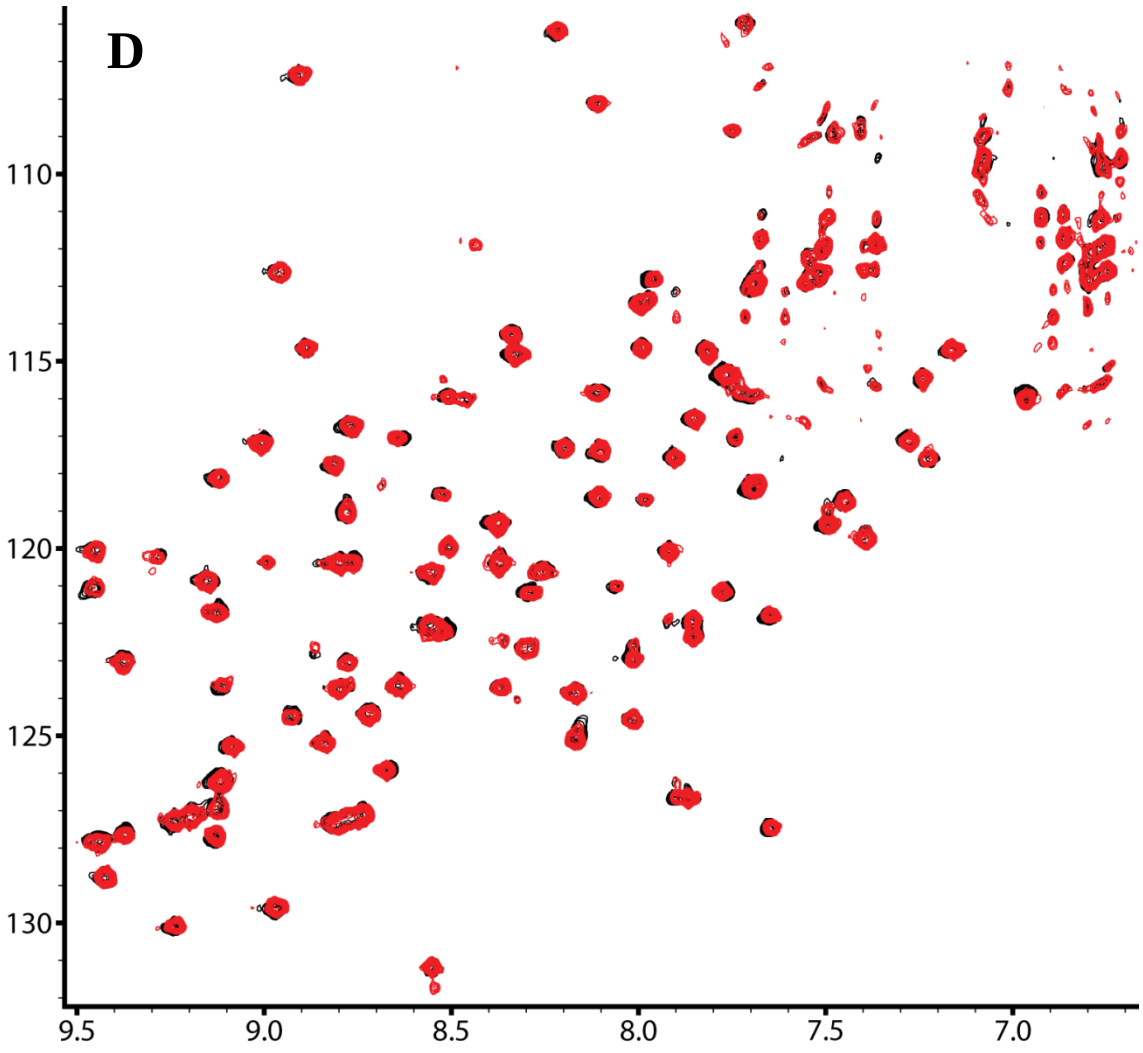
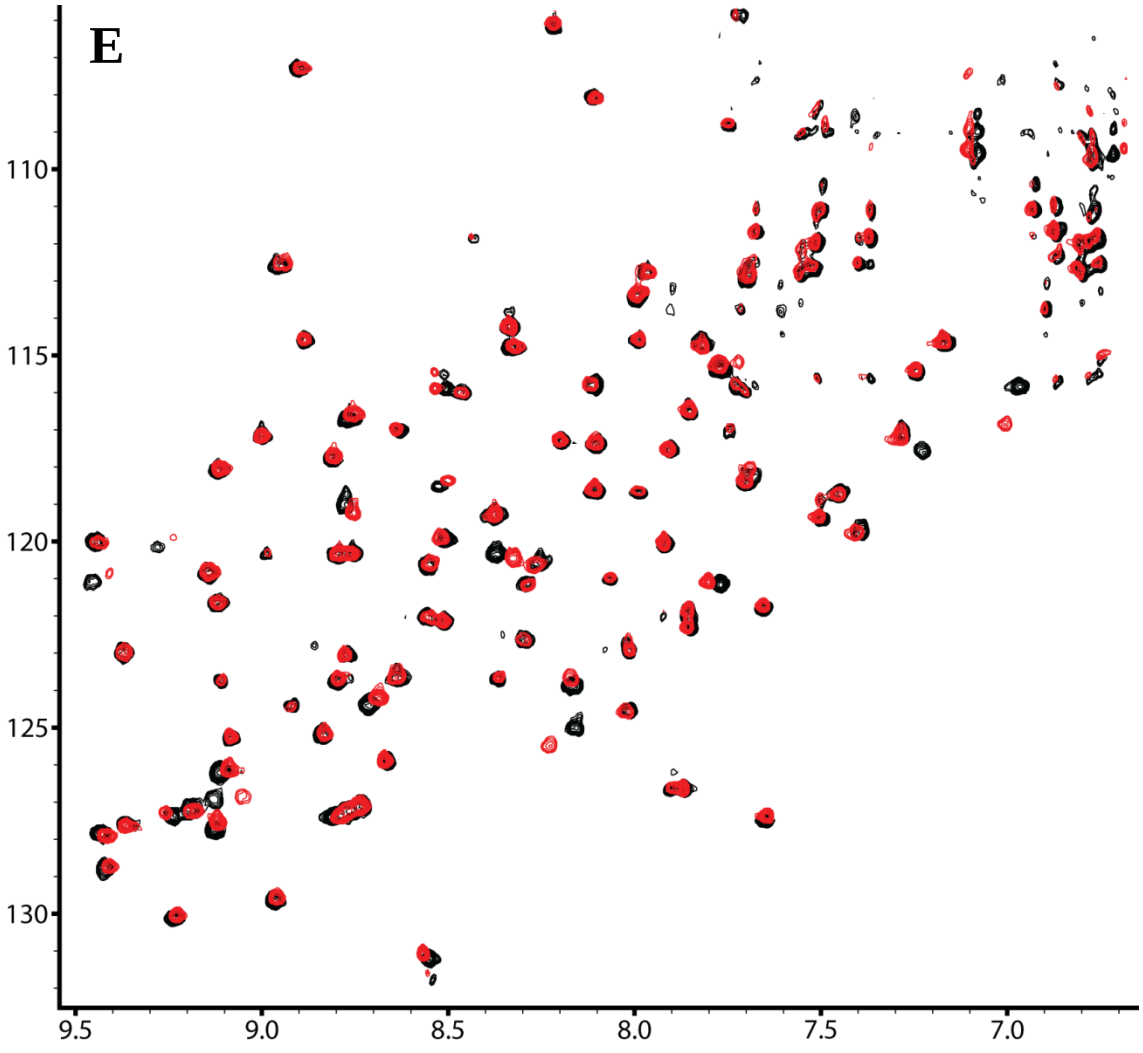


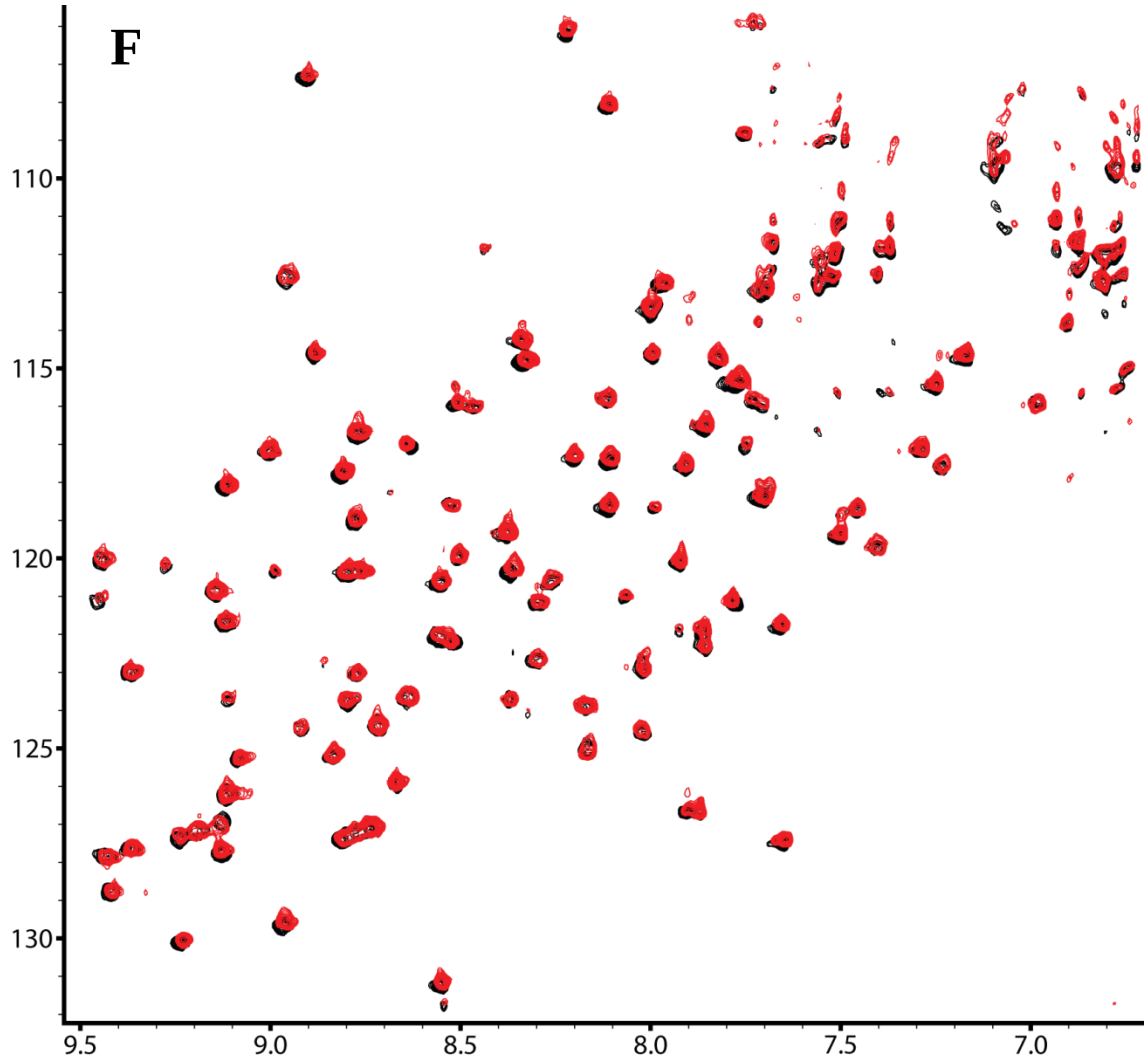
Figure 6. ^{15}N-^{1}H HSQC overlays of compounds used in this study with black as the free protein and red as the final titration point. (a-VU0467976, b-VU0468049, c-cVU0085636, d-VU0466242, e-VU0469701, f-VU0100560)


B
110
115
120
125
130
9.5
9.0
8.5
8.0
7.5
7.0


C
110
115
120
125
130
9.5
9.0
8.5
8.0
7.5
7.0


D
110
115
120
125
130
9.5
9.0
8.5
8.0
7.5
7.0


E
110
115
120
125
130
9.5
9.0
8.5
8.0
7.5
7.0



2.3 – Results

VU0467976

Guided dock experiments resulted in placement of the phenyl furan 2-carboxylic acid portion of the compound in the correct orientation using the default settings (exhaustiveness = 8). Increasing the exhaustiveness to 64 generated a model that closely mimics the position in the crystal structure as seen in Figure 7. Further increasing the exhaustiveness resulted in similar models as seen previously. The lowest energy model generated had an affinity = -10.6 kcal/mol.

Allowing the program to search the entire molecule yielded results similar to what was seen when the search space was restricted to the basic cleft. Default settings were able to place the phenyl furan 2-carboxylic acid in the "correct" location. Increasing the exhaustiveness gave similar results as the program was able to orient the molecule in the same space as seen in the crystal structure, also shown in Figure 7. The lowest energy model had an affinity = -10.6 kcal/mol. This value is the same as what was seen for the guided model. Overlaying both generated lowest energy structures onto the crystal structure, good agreement of the placement and orientation of the compound is observed. NMR titrations show perturbations of residues in expected localized areas.

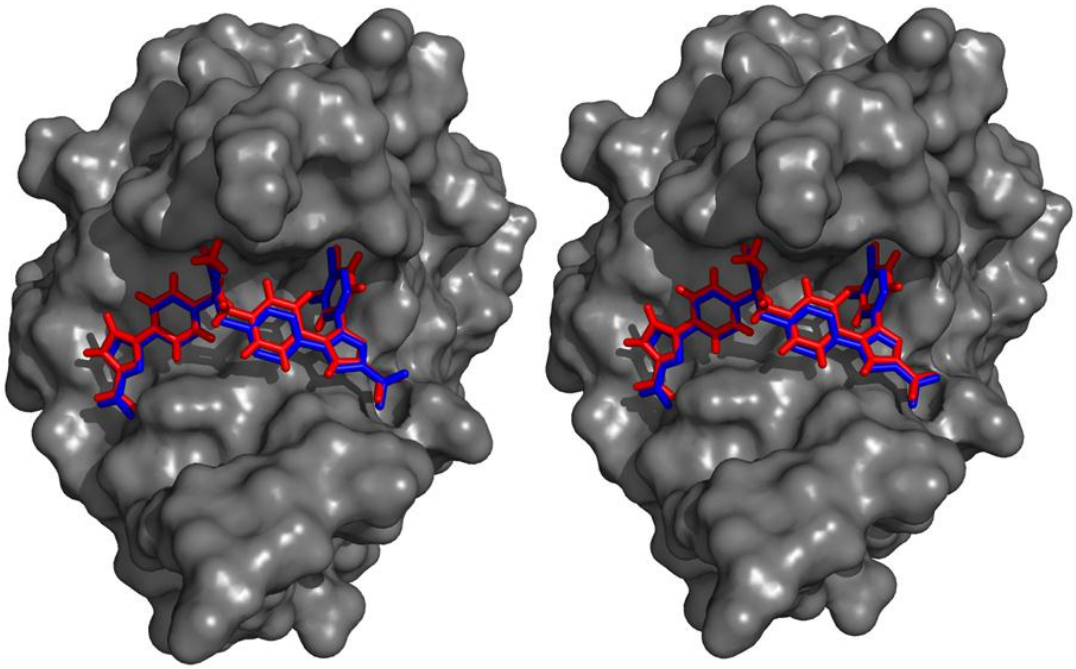


Figure 7. Lowest energy models from guided (left, blue) and unguided (right, blue) overlaid on the crystal structure (red).

VU0468049

For this compound, guided docking with default settings were sufficient to correctly orient the phenyl furan 2-carboxylic acid portion of the compound at site 2. Increasing the exhaustiveness to 128 generated a low energy model (affinity = -8.4 kcal/mol) that recapitulated the crystal structure as seen in Figure 8. NMR titration experiments found a subset of chemical shift perturbations that have were mapped onto the surface of the protein. These perturbations are localized to the basic cleft where the compound was bound in the crystal structure.

Unguided docking generated a model of equal predicted affinity as the guided model, but the site found was outside of the basic cleft. The compound was docked to a patch on the backside of RPA70N shown in Figure 8. The carboxylic acid off of the phenyl furan constituent is stabilized by positive charge from L104 and the amine groups are stabilized by negative charge from E98 and E100. No compounds were observed in the basic cleft in the unguided experiments even though model from the guided experiments had a predicted affinity equal or weaker than all complexes generated in the unguided experiments.

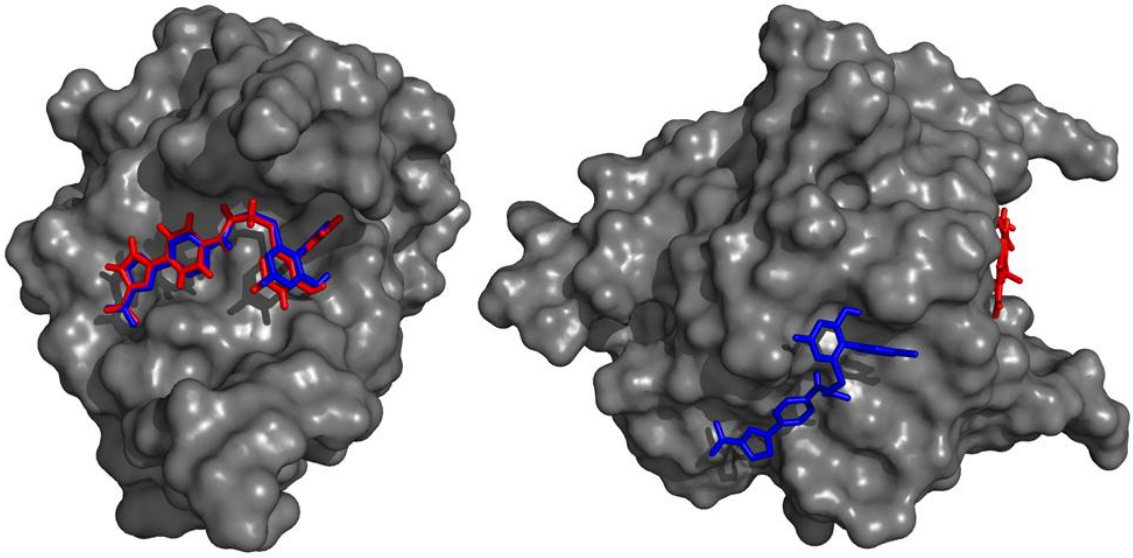


Figure 8. Lowest energy models from guided (left, blue) and unguided (right, blue) of VU0468049 overlaid on the crystal structure (red).

VU0085636

For this compound, guided docking with default settings were sufficient to model the compound in the basic cleft but unable to align the model with the crystal structure. Increasing the exhaustiveness did not generate a lower energy model, all of the lowest energy models from each different exhaustiveness value had an affinity = -7.7 kcal/mol. Additionally none of these models correctly aligned with the structure showing; variation in the linkers between each of the phenyl rings was observed as shown in Figure 9.

Unguided docking did not reproduce what was seen in the guided experiments. Unguided models favored a slightly basic and hydrophobic pocket formed on the back side of RPA involving the –SH of the GSH tag, resides 3-5 of the N-terminus, V76, L102, and K103. The lowest energy model (affinity = -8.7 kcal/mol) derived from unguided searches was positioned in this pocket on the backside of RPA and was 1 kcal/mol lower than the guided model. This model is shown in Figure 9.

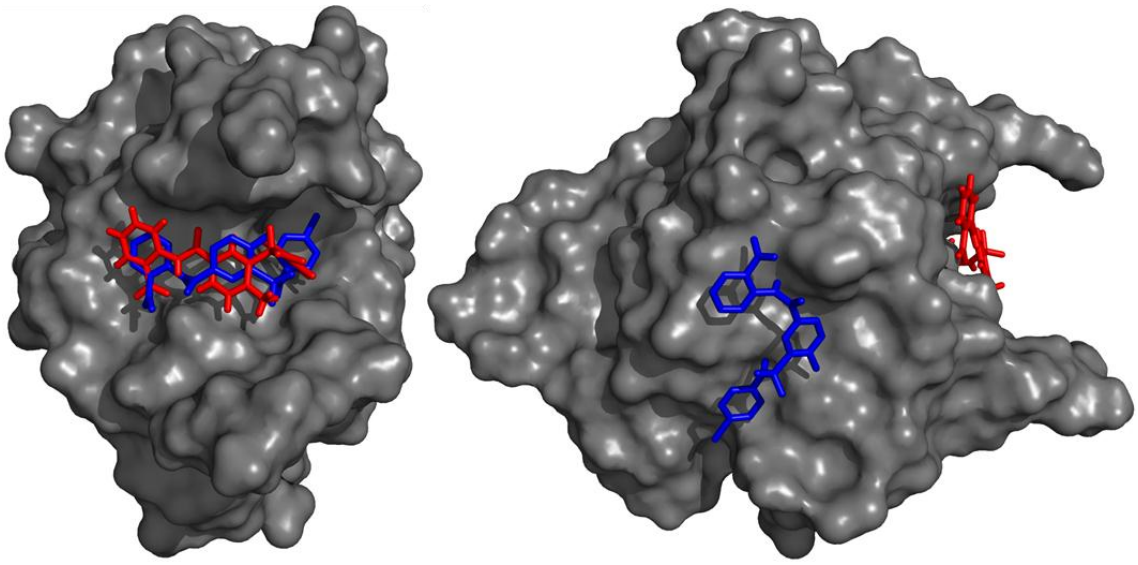


Figure 9. Lowest energy models from guided (left, blue) and unguided (right, blue) of VU0085636 overlaid on the crystal structure (red).

VU0466242

For this compound, guided docking with default settings generated models with the compound at site 2. Increasing the exhaustiveness to 128 generated a model with similar orientation but the sulfur adopts a different conformation. A low energy model (affinity = -6.7 kcal/mol) with good agreement with the crystal structure was generated with an exhaustiveness = 2048, shown in Figure 10. Values this high allow for much greater sampling of the space and lower energy models at the cost of time.

Unguided experiments did not yield any models with the compound bound to the basic cleft (Figure 10). The majority of the unguided models were on the backside of the protein in small pockets. These pockets are composed of mainly hydrophobic residues with one or two charged residues to bind the carboxylic acid or amine groups of the compound. The lowest energy model (affinity = -6.5 kcal/mol) docked to the backside of S55 away from the basic cleft.

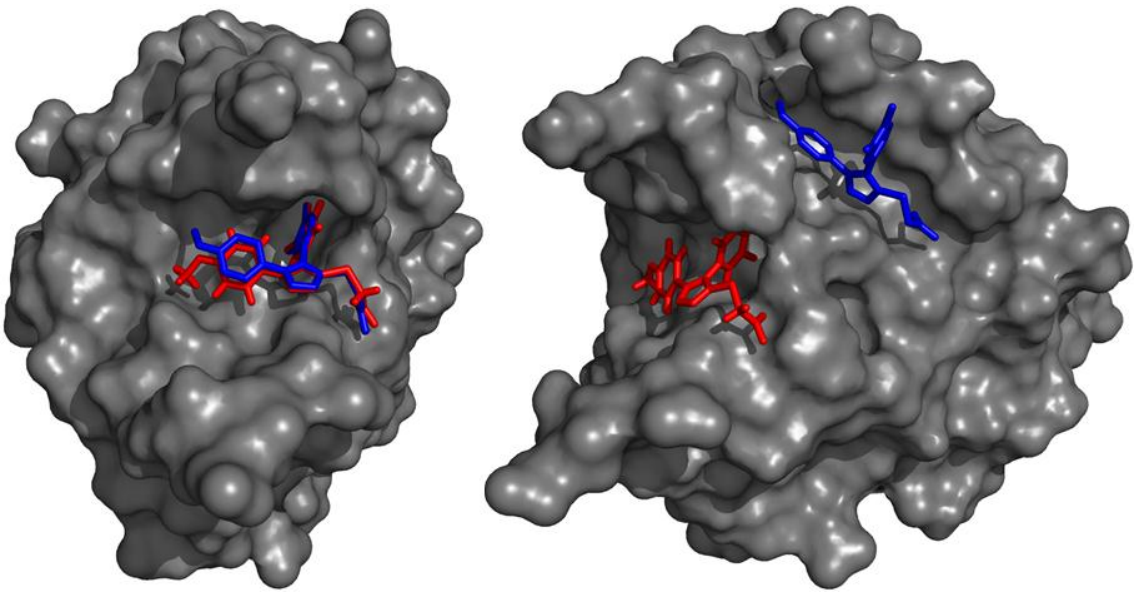


Figure 10. Lowest energy models of VU0466242 from guided (left, blue) and unguided (right, blue) experiments overlaid on the crystal structure (red).

VU0469701

This compound was docked to two different structures of RPA70N (PDB id: 2B29 and 2B3G) because no structural data exists for the complex of this compound with RPA70N. Guided docking using the structure of RPA70N alone (2B29) identified a surface on the basic cleft that was favored for most models, including the lowest energy model (Figure 11). Residues I33, S38, P39, P40, R41, R43, and M57 form a basic and hydrophobic surface on loop 12(L12). Docking to this surface using the default parameters produced a low energy model and exhaustiveness did not further refine this model.

Unguided docking produced a low energy model that binds in a hydrophobic and basic pocket behind the basic cleft. Residues T52, L53, S54, S55, R91, Y118, N119, and E120 form the pocket and are shown in Figure 11. The unguided model had an affinity = -6.0 kcal/mol compared to -5.0 kcal/mol for the guided model.

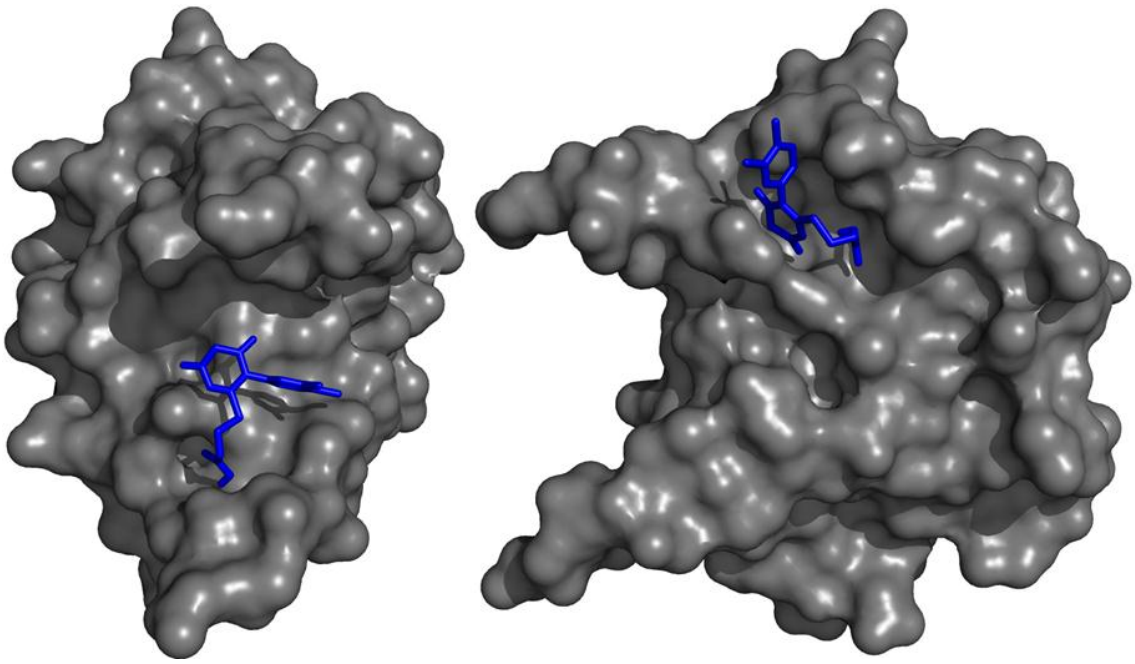


Figure 11. Lowest energy model of VU0469701 from guided (left) and unguided (right) experiments with pdb 2B29.

Guided docking using the structure of RPA70N from its complex with an ATRIP peptide (2B3G) provided a slightly lower energy model than with docking to the free RPA70N domain, -5.2 kcal/mol versus -5.0 kcal/mol, respectively. The compound was positioned in a small hydrophobic and basic pocket involving residues R43, S55, M57, L87, R91, and V93 (Figure 12). Use of the default settings did not identify the lowest energy model but as the exhaustiveness of the search was increased, the same location was targeted in all successive runs. Rearrangements in the orientation of the compound account for the changes seen in the affinity for each run.

Unguided docking using structure 2B3G yielded models that were similar to the models produced with 2B29 (Figure 12). The compound was positioned in a basic and hydrophobic pocket formed by residues R43, L45, T52, L53, S54, S55, R91, and E120. Docking to 2B3G produced a model that was of higher energy than the 2B29 model, e.g.

when exhaustiveness is set to 512, values of -5.7 kcal/mol and -6kcal/mol, respectively, were produced.

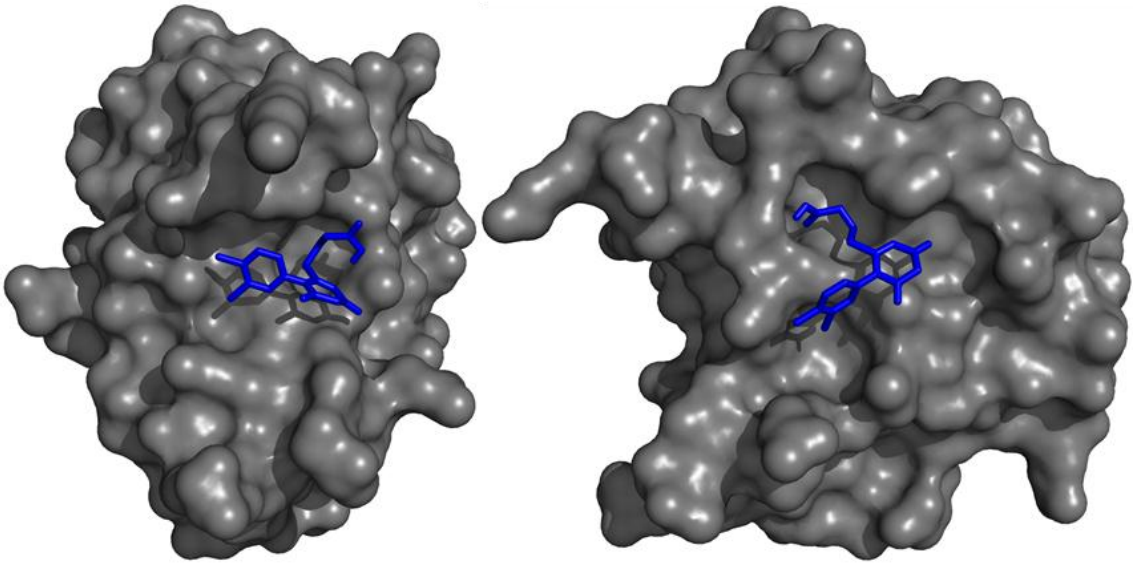


Figure 12. Lowest energy model of VU0469701 from guided (left) and unguided (right) experiments with pdb 2B3G.

VU0100560

This compound was also docked onto two different structures of RPA70N (PDB id: 2B29 and 2B3G) because no co-crystal structure of this complex with RPA70N has been determined. Guided docking using the structure of RPA70N alone (2B29) with Wednesday default settings place the compound on a hydrophobic and basic surface on L12. This surface is comprised of residues I33, T35, S38, R41, and R43 and is highlighted in cyan in Figure 13. Increasing the exhaustiveness did not yield a lower energy model as the affinity remained -5.5 kcal/mol.

Unguided docking experiments did not generate any models with the compound bound to the basic cleft. Using default parameters, models with lower energy than the guided docking model were obtained. The lowest energy model had an affinity = -6.4

kcal/mol and was docked to a hydrophobic pocket comprised of residues L62, P64, L65, and L99 (Figure 13).

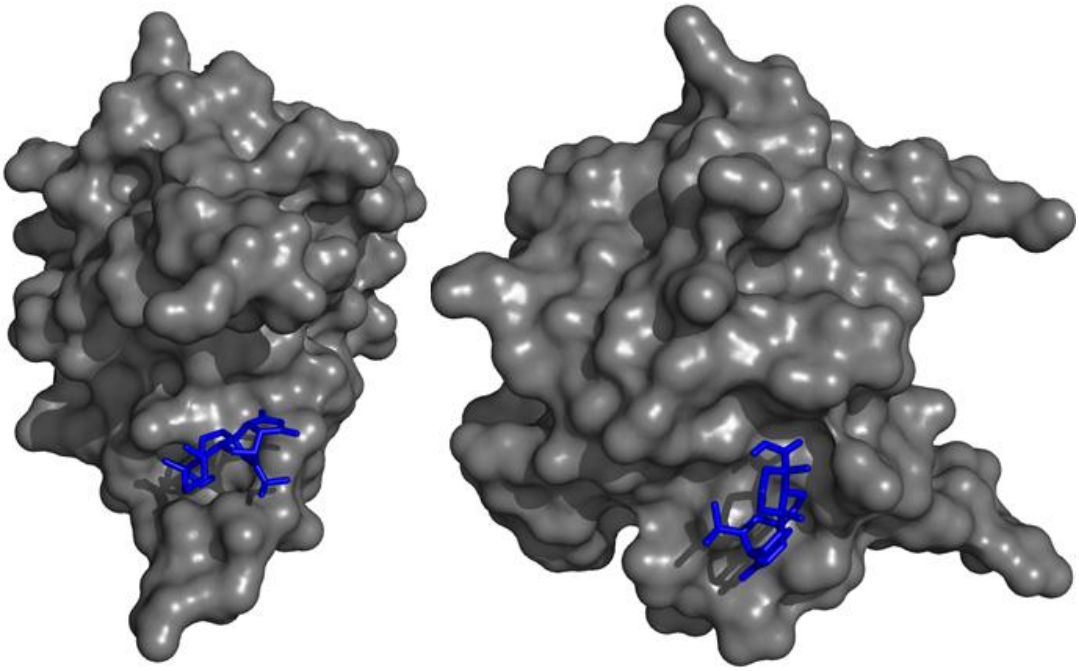


Figure 13. Lowest energy model of VU0100560 from guided (left) and unguided (right) experiments with pdb 2B29.

Guided docking experiments using the structure of RPA70N in complex with an ATRIP peptide (2B3G) yielded lower energy models than the docking experiments performed using the structure of free RPA70N, -6.2 kcal/mol versus -5.5 kcal/mol, respectively. With the default settings, the compound was positioned in a basic and hydrophobic pocket formed by residues R43, S55, M57, N85, L87, R91, and V93 (Figure 14). Increasing the exhaustiveness function from 8 to 512 resulted in a slightly lower energy model, -5.9 kcal/mol versus -6.2 kcal/mol.

Unguided docking experiments generated models with the compound positioned in a patch of hydrophobic residues including I30, V66, E69, L71, S72, and S73 (Figure 14). This model had an affinity = -6.5 kcal/mol. Additionally, two models were generated with the compound docked to the basic pocket as seen in the guided search, but these had

higher energies than the best model with the ligand positioned outside the RPA70N basic cleft.

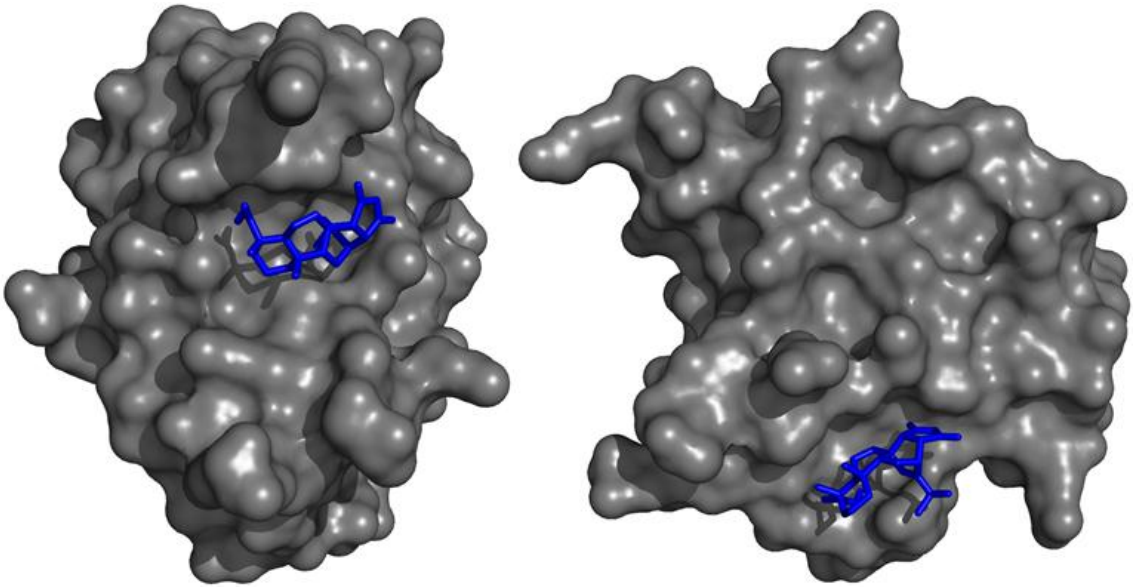


Figure 14. Lowest energy model of VU0100560 from guided (left) and unguided (right) experiments with pdb 2B3G.

2.4 – Summary of findings

Guided docking experiments were successful in recapitulating the position and orientation for the compounds that have previously been crystallized. All compounds with crystal structures of their complex with RPA70N overlay nicely with the position observed in the crystal structure except for VU0085636. The lowest energy model of VU0085636 provided the correct relative positions of the phenyl rings but did not match the orientation seen in the crystal structure. A general trend in the results was that the accuracy of the prediction correlated with the affinity of the compounds. Moreover, the compounds with highest affinity (lower K_d values) required less exhaustive searches to recreate the structure observed in the crystal structure. VU0467976 was the tightest binder tested with a K_d of 8 μM. It also had the lowest energy model from all the

compounds tested, -10.6 kcal/mol. VU0468049 was the next tightest binder and produced the second lowest energy model, -8.4 kcal/mol. These two compounds were both successful in generating models that mimic the respective crystal structures in less exhaustive searches (64 and 128 respectively). VU0466242 a smaller compound and weaker binder (K_D=164 μM) produced a higher energy "best" model, -6.7 kcal/mol and required a more exhaustive search (2048) to obtain that model. This correlation between the affinity of the compound and relative values of the energies reported is interesting, and presumably arises because the structure of the protein does not vary and the energy is dominated by the number of stabilizing contacts made by the compound.

Unguided docking experiments allowed for the entire surface of the protein to be sampled as opposed to be optimized to only sample the basic cleft of RPA70N, which serves as the binding site for target proteins. However, these unguided experiments were largely unsuccessful in regenerating the position of the compound in the crystal structure. In many cases, the pockets in which the compound was positioned was away from the basic cleft, i.e. lower energy models were generated with the compound docked outside versus in the basic cleft. Additionally, these experiments take a substantially longer time to run, as expected, since a higher exhaustiveness value is needed to adequately sample all of the space. The problem is made even more challenging when a smaller compound or fragment is investigated, as its small size allows for more binding sites to be sampled. Given the higher computational cost and uncertainty in the result, spending larger amounts of time modeling does seem to be advantageous. This is particularly true in the light of the availability of very rapid screening methods. The development of NMR pulse sequences allowing for collection of complete HSQC spectra in 15 minutes for well-

behaved proteins at reasonably low concentrations is especially valuable in this context because it would allow at least a direct albeit coarse mapping of the ligand binding site. Overall, unguided docking does not appear to currently lend itself to be a valuable approach for screening.

Chapter 3

Conclusion and Future Directions

3.1 – Evaluation of overall performance of *Autodock Vina*

This study evaluated the ability of *Autodock Vina* to serve as a tool in the development of small molecule inhibitors of RPA70N. Given the lack of success in recapitulating the crystal structure and the need for long, exhaustive unguided searches to adequately sample the space proved to be inefficient and unable to adequately predict the correct binding site. On the other hand, with some knowledge of the binding site on the target, the program did a remarkably good job at predicting the binding site. Additionally, we found a correlation between the accuracy of the prediction and the size of the compound: larger compounds like VU0467976 worked better. We attribute this observation to the fact that as the molecule gets larger the number of potential binding sites is reduced compared to small fragments like VU0466242. An inverse correlation between exhaustiveness and the K_d was observed for the tight binders in this group. These compounds required less exhaustive searches and produced lower energy models.

Overall, my results suggest that *Autodock Vina* is not an efficient tool for screening large libraries of fragments but has the potential to be useful for rational drug design. Analogs of lead compounds can quickly be designed and prepared for docking as compared to synthesizing all of the compounds. Importantly, large number of analogs can be designed and efficiently modeled in one day. Given the correlations seen in this study, relative binding affinity between the compound and protein can be assessed from the

calculated energies. In particular, the lower the energy is, the tighter the binding. This will allow for the synthesis and testing of only those compounds that are highly likely to bind more tightly to the target. Should *Autodock Vina* program prove generally applicable, it would enhance the overall efficiency of the compound optimization process.

3.2 – Future directions

In efforts to continue evaluating the role of virtual screening and molecular docking in drug discovery, additional steps could be implemented for further evaluation of *Autodock Vina*. All of the experiments that were performed in this study kept the protein as a rigid body and only allowed certain bonds in the compounds to rotate in the docking. *Autodock Vina* allows for residues to set as flexible. Allowing residues in the basic cleft to retain flexibility should allow for better models. RPA70N was identified as having two discrete ligand binding sites in the basic cleft, centered on S55 and T60. Additionally, the side chains of residues R41, R43, and M57 were consistently identified as key contributors to the binding of the six compounds tested here. Allowing flexibility of these key residues alone may be a good compromise to obtain additional information about the targeted binding site without having as high of a computational cost as allowing flexibility to all side chains.

If structural information is available on the target binding site, it would be worth evaluating the use of the program as a virtual screening tool. Fragments would be screened in a defined target area and could be separated based on their binding location. By performing this virtual screening, a chemist could see orientations and positions for

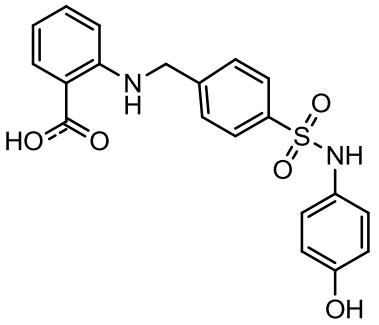
each fragment at a targeted binding site. Distances can be measured and the chemist can rationally design a linker to connect the fragments based on the set of models generated by the program. This approach would be beneficial in situations where complexes do not readily crystallize and no additional structural data is available.

Additional testing should be performed to see if the correlation between tighter binding of the protein-ligand complexes and lower energies in *Autodock Vina* models. This study examined compounds with affinities between 8 μM and 170 μM. If this trend was to continue in to the nanomolar range, this approach would be an ideal method to predict which compounds and future analogs are worth pursuing synthetically. Ideally, this particular step would come later in the process, well after screening. However, as we have noted above, the use of computational evaluation of binding of ligands to a target protein can be useful at multiple stages of the compound optimization process. With the continual development of *Autodock Vina* and many other computational approaches, the future of drug discovery will very likely see an increase in overall efficiency as a result of these developments.

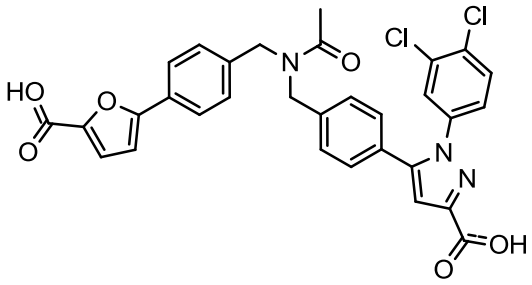
Appendix A

Compounds used in study

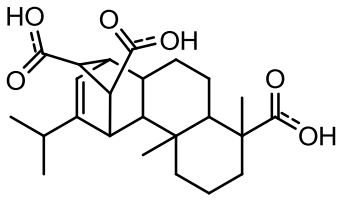
VU0085636



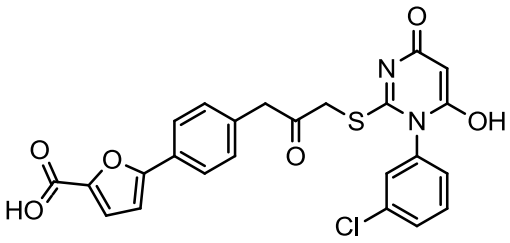
VU0467976



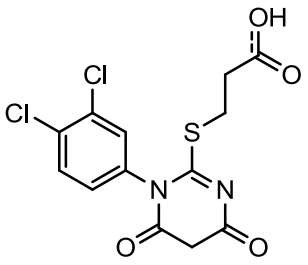
VU0100560



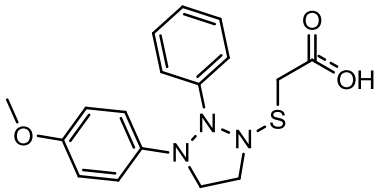
VU0468049



VU0469701



VU0466242



Appendix B

Energy values for all models

VU0467976 Model Position	Guided Ex = 8 Affinity (kcal/mol)	VU0467976 Model Position	Unguided Ex = 8 Affinity (kcal/mol)
1	-9.1	1	-8.1
2	-8.9	2	-7.8
3	-8.5	3	-7.5
4	-8.4	4	-7.4
5	-8.3	5	-7.4
6	-8.2	6	-7.3
7	-8.1	7	-7.3
8	-8.0	8	-7.3
9	-7.9	9	-7.2

VU0467976 Model Position	Guided Ex = 64 Affinity (kcal/mol)	VU0467976 Model Position	Unguided Ex = 64 Affinity (kcal/mol)
1	-10.6	1	-10.4
2	-9.3	2	-9.8
3	-9.3	3	-9.3
4	-9.3	4	-9.2
5	-9.3	5	-9.1
6	-9.0	6	-9.1
7	-9.0	7	-9.1
8	-8.8	8	-8.7
9	-8.8	9	-8.7

VU0467976 Model Position	Guided Ex = 128 Affinity (kcal/mol)	VU0467976 Model Position	Unguided Ex = 128 Affinity (kcal/mol)
1	-10.4	1	-10.6
2	-9.4	2	-9.7
3	-9.2	3	-9.4
4	-9.2	4	-9.1
5	-9.1	5	-9.1
6	-9.1	6	-8.7
7	-9.0	7	-8.6
8	-8.8	8	-8.6
9	-8.7	9	-8.5

VU0468049 Model Position	Guided Ex = 8 Affinity (kcal/mol)	VU0468049 Model Position	Unguided Ex = 8 Affinity (kcal/mol)
1	-7.5	1	-6.8
2	-7.3	2	-6.7
3	-7.2	3	-6.5
4	-7.1	4	-6.5
5	-7.0	5	-6.5
6	-7.0	6	-6.4
7	-6.8	7	-6.4
8	-6.6	8	-6.2
9	-6.5	9	-6.0

VU0468049 Model Position	Guided Ex = 64 Affinity (kcal/mol)	VU0468049 Model Position	Unguided Ex = 64 Affinity (kcal/mol)
1	-7.9	1	-8.4
2	-7.8	2	-8.3
3	-7.8	3	-8.2
4	-7.8	4	-8.1
5	-7.7	5	-8.1
6	-7.7	6	-8.1
7	-7.6	7	-7.9
8	-7.4	8	-7.8
9	-7.4	9	-7.8

VU0468049 Model Position	Guided Ex = 128 Affinity (kcal/mol)	VU0468049 Model Position	Unguided Ex = 128 Affinity (kcal/mol)
1	-8.4	1	-8.3
2	-8.1	2	-8.2
3	-8.0	3	-8.2
4	-8.0	4	-8.2
5	-8.0	5	-8.1
6	-7.9	6	-8.0
7	-7.8	7	-8.0
8	-7.8	8	-7.9
9	-7.8	9	-7.9

VU0085636 Model Position	Guided Ex = 8 Affinity (kcal/mol)	VU0085636 Model Position	Unguided Ex = 8 Affinity (kcal/mol)
1	-7.7	1	-8.4
2	-7.6	2	-8.3
3	-7.6	3	-8.3
4	-7.3	4	-8.1
5	-7.3	5	-8.1
6	-7.3	6	-8.0
7	-7.1	7	-8.0

8	-7.1	8	-8.0
9	-7.0	9	-8.0

VU0085636 Model Position	Guided Ex = 64 Affinity (kcal/mol)	VU0085636 Model Position	Unguided Ex = 64 Affinity (kcal/mol)
1	-7.6	1	-8.7
2	-7.6	2	-8.4
3	-7.5	3	-8.4
4	-7.5	4	-8.3
5	-7.3	5	-8.3
6	-7.3	6	-8.3
7	-7.3	7	-8.2
8	-7.2	8	-8.2
9	-7.2	9	-8.2

VU0085636 Model Position	Guided Ex = 128 Affinity (kcal/mol)	VU0085636 Model Position	Unguided Ex = 128 Affinity (kcal/mol)
1	-7.6	1	-8.7
2	-7.6	2	-8.5
3	-7.5	3	-8.4
4	-7.5	4	-8.4
5	-7.4	5	-8.3
6	-7.3	6	-8.3
7	-7.3	7	-8.2
8	-7.3	8	-8.2
9	-7.3	9	-8.2

VU0085636 Model Position	Guided Ex = 512 Affinity (kcal/mol)	VU0085636 Model Position	Unguided Ex = 512 Affinity (kcal/mol)
1	-7.7	1	-8.7
2	-7.6	2	-8.5
3	-7.5	3	-8.5
4	-7.5	4	-8.4
5	-7.4	5	-8.3
6	-7.3	6	-8.3
7	-7.3	7	-8.2
8	-7.3	8	-8.2
9	-7.3	9	-8.2

VU0466242 Model Position	Guided Ex = 8 Affinity (kcal/mol)	VU0466242 Model Position	Unguided Ex = 8 Affinity (kcal/mol)
1	-5.4	1	-5.6
2	-5.2	2	-5.5
3	-5.2	3	-5.4
4	-5.2	4	-5.3

5	-5.0	5	-5.3
6	-4.9	6	-5.3
7	-4.8	7	-5.1
8	-4.8	8	-5.1
9	-4.7	9	-5.1

VU0466242 Model Position	Guided Ex = 64 Affinity (kcal/mol)	VU0466242 Model Position	Unguided Ex = 64 Affinity (kcal/mol)
1	-5.8	1	-6.2
2	-5.8	2	-5.9
3	-5.7	3	-5.8
4	-5.7	4	-5.7
5	-5.6	5	-5.7
6	-5.6	6	-5.7
7	-5.6	7	-5.7
8	-5.5	8	-5.7
9	-5.4	9	-5.7

VU0466242 Model Position	Guided Ex = 128 Affinity (kcal/mol)	VU0466242 Model Position	Unguided Ex = 128 Affinity (kcal/mol)
1	-6.2	1	-6.2
2	-6.1	2	-5.6
3	-6.0	3	-5.6
4	-5.9	4	-5.6
5	-5.7	5	-5.5
6	-5.7	6	-5.5
7	-5.7	7	-5.5
8	-5.7	8	-5.5
9	-5.7	9	-5.5

VU0466242 Model Position	Guided Ex = 512	VU0466242 Model Position	Unguided Ex = 512 Affinity (kcal/mol)
1	-5.8	1	-6.5
2	-5.8	2	-6.4
3	-5.7	3	-6.0
4	-5.7	4	-6.0
5	-5.7	5	-5.9
6	-5.6	6	-5.9
7	-5.6	7	-5.9
8	-5.6	8	-5.8
9	-5.5	9	-5.8

VU0466242 Model Position	Guided Ex = 2048 Affinity (kcal/mol)	VU0466242 Model Position	Unguided Ex = 2048 Affinity (kcal/mol)
1	-6.7	1	-6.6

2	-6.4	2	-6.5
3	-6.3	3	-6.4
4	-6.2	4	-6.3
5	-6.2	5	-6.3
6	-6.1	6	-6.1
7	-6.1	7	-6.1
8	-6.1	8	-6.1
9	-6.1	9	-6.0

VU0469701 - 2B29 Model Position	Guided Ex = 8 Affinity (kcal/mol)	VU0469701 - 2B29 Model Position	Unguided Ex = 8 Affinity kcal/mol)
1	-4.9	1	-6.0
2	-4.7	2	-5.6
3	-4.6	3	-5.4
4	-4.6	4	-5.4
5	-4.5	5	-5.1
6	-4.5	6	-5.1
7	-4.5	7	-5.0
8	-4.5	8	-4.9
9	-4.3	9	-4.9

VU0469701 - 2B29 Model Position	Guided Ex = 64 Affinity (kcal/mol)	VU0469701 - 2B29 Model Position	Unguided Ex = 64 Affinity (kcal/mol)
1	-5.0	1	-6.0
2	-4.8	2	-5.8
3	-4.7	3	-5.5
4	-4.7	4	-5.4
5	-4.6	5	-5.3
6	-4.6	6	-5.2
7	-4.6	7	-5.2
8	-4.5	8	-5.1
9	-4.5	9	-5.1

VU0469701 - 2B29 Model Position	Guided Ex = 128 Affinity (kcal/mol)	VU0469701 - 2B29 Model Position	Unguided Ex = 128 Affinity (kcal/mol)
1	-5.0	1	-6.0
2	-4.8	2	-5.8
3	-4.7	3	-5.6
4	-4.7	4	-5.5
5	-4.7	5	-5.4
6	-4.6	6	-5.4
7	-4.6	7	-5.4
8	-4.6	8	-5.3
9	-4.6	9	-5.2

VU0469701 - 2B29 Model Position	Guided Ex = 512 Affinity (kcal/mol)	VU0469701 - 2B29 Model Position	Unguided Ex = 512 Affinity (kcal/mol)
1	-5.0	1	-6.0
2	-4.8	2	-5.7
3	-4.8	3	-5.5
4	-4.7	4	-5.5
5	-4.7	5	-5.5
6	-4.7	6	-5.4
7	-4.7	7	-5.3
8	-4.6	8	-5.2
9	-4.6	9	-5.2

VU0469701 - 2B3G Model Position	Guided Ex = 8 Affinity (kcal/mol)	VU0469701 - 2B3G Model Position	Unguided Ex = 8 Affinity (kcal/mol)
1	-5.2	1	-5.4
2	-4.7	2	-5.4
3	-4.5	3	-5.4
4	-4.5	4	-5.3
5	-4.5	5	-5.1
6	-4.4	6	-4.9
7	-4.4	7	-4.9
8	-4.2	8	-4.8
9	-4.1	9	-4.8

VU0469701 - 2B3G Model Position	Guided Ex = 64 Affinity (kcal/mol)	VU0469701 - 2B3G Model Position	Unguided Ex = 64 Affinity (kcal/mol)
1	-5.2	1	-5.7
2	-4.9	2	-5.6
3	-4.8	3	-5.6
4	-4.6	4	-5.6
5	-4.6	5	-5.5
6	-4.6	6	-5.4
7	-4.6	7	-5.4
8	-4.6	8	-5.4
9	-4.5	9	-5.3

VU0469701 - 2B3G Model Position	Guided Ex = 128 Affinity (kcal/mol)	VU0469701 - 2B3G Model Position	Unguided Ex = 128 Affinity (kcal/mol)
1	-5.3	1	-5.7
2	-4.7	2	-5.7
3	-4.7	3	-5.6
4	-4.6	4	-5.6
5	-4.6	5	-5.6
6	-4.6	6	-5.5
7	-4.6	7	-5.5

8	-4.5	8	-5.4
9	-4.5	9	-5.4

VU0469701 - 2B3G Model Position	Guided Ex = 512 Affinity (kcal/mol)	VU0469701 - 2B3G Model Position	Unguided Ex = 512 Affinity (kcal/mol)
1	-5.2	1	-5.7
2	-4.9	2	-5.7
3	-4.8	3	-5.6
4	-4.7	4	-5.6
5	-4.6	5	-5.6
6	-4.6	6	-5.5
7	-4.6	7	-5.5
8	-4.6	8	-5.5
9	-4.6	9	-5.5

VU0100560 - 2B29 Model Position	Guided Ex = 8 Affinity (kcal/mol)	VU0100560 - 2B29 Model Position	Unguided Ex = 8 Affinity (kcal/mol)
1	-5.5	1	-6.4
2	-5.4	2	-6.2
3	-5.3	3	-6.0
4	-5.1	4	-6.0
5	-5.1	5	-6.0
6	-5	6	-5.9
7	-5	7	-5.9
8	-4.8	8	-5.9
9	-4.8	9	-5.8

VU0100560 - 2B29 Model Position	Guided Ex = 64 Affinity (kcal/mol)	VU0100560 - 2B29 Model Position	Unguided Ex = 64 Affinity (kcal/mol)
1	-5.5	1	-6.4
2	-5.4	2	-6.2
3	-5.3	3	-6.0
4	-5.3	4	-6.0
5	-5.2	5	-6.0
6	-5.1	6	-6.0
7	-5.1	7	-6.0
8	-5.1	8	-5.9
9	-5	9	-5.9

VU0100560 - 2B29 Model Position	Guided Ex = 128 Affinity (kcal/mol)	VU0100560 - 2B29 Model Position	Unguided Ex = 128 Affinity (kcal/mol)
1	-5.5	1	-6.4
2	-5.4	2	-6.2
3	-5.3	3	-6.1
4	-5.3	4	-6.1

5	-5.2	5	-6.0
6	-5.1	6	-6.0
7	-5.1	7	-6.0
8	-5.1	8	-5.9
9	-5	9	-5.9

VU0100560 - 2B29 Model Position	Guided Ex = 512 Affinity (kcal/mol)	VU0100560 - 2B29 Model Position	Unguided Ex = 512 Affinity (kcal/mol)
1	-5.5	1	-6.4
2	-5.4	2	-6.2
3	-5.3	3	-6.1
4	-5.3	4	-6.1
5	-5.2	5	-6.0
6	-5.1	6	-6.0
7	-5.1	7	-6.0
8	-5.1	8	-5.9
9	-5.1	9	-5.9

VU0100560 - 2B3G Model Position	Guided Ex = 8 Affinity (kcal/mol)	VU0100560 - 2B3G Model Position	Unguided Ex = 8 Affinity (kcal/mol)
1	-5.9	1	-6.5
2	-5.9	2	-6.2
3	-5.7	3	-6.2
4	-5.6	4	-6.2
5	-5.5	5	-6.2
6	-5.3	6	-6.0
7	-5.2	7	-5.8
8	-5.1	8	-5.7
9	-5.1	9	-5.7

VU0100560 - 2B3G Model Position	Guided Ex = 64 Affinity (kcal/mol)	VU0100560 - 2B3G Model Position	Unguided Ex = 64 Affinity (kcal/mol)
1	-5.9	1	-6.5
2	-5.9	2	-6.2
3	-5.8	3	-6.2
4	-5.6	4	-6.2
5	-5.6	5	-6.2
6	-5.6	6	-6.0
7	-5.5	7	-6.0
8	-5.5	8	-6.0
9	-5.4	9	-5.9

VU0100560 - 2B3G Model Position	Guided Ex = 128 Affinity (kcal/mol)	VU0100560 - 2B3G Model Position	Unguided Ex = 128 Affinity (kcal/mol)
1	-6.2	1	-6.5
2	-6.1	2	-6.3

3	-6.1	3	-6.2
4	-5.8	4	-6.2
5	-5.8	5	-6.2
6	-5.8	6	-6.0
7	-5.7	7	-6.0
8	-5.7	8	-6.0
9	-5.5	9	-5.9

VU0100560 - 2B3G Model Position	Guided Ex = 512 Affinity (kcal/mol)	VU0100560 - 2B3G Model Position	Unguided Ex = 512 Affinity (kcal/mol)
1	-6.2	1	-6.5
2	-6.1	2	-6.2
3	-6.1	3	-6.2
4	-5.8	4	-6.2
5	-5.8	5	-6.2
6	-5.8	6	-6.0
7	-5.8	7	-6.0
8	-5.7	8	-6.0
9	-5.7	9	-6.0

References Cited

1. Sancar, A., et al. *Annu Rev Biochem.* **2004**, *73*, 39-85.
2. Ciccia, A. and Elledge, SJ. *Molecular Cell*. **2010**, *40*, 179-204.
3. Cimprich, KA. and Cortez, D. *Mol. Cell. Bio.* **2008**, *9*, 616-627.
4. Namiki, Y, and Zou, L. *PNAS.* **2006**, *103*, 580-585.
5. Choi, JH., et al. *PNAS.* **2010**, *107*, 13660-13665.
6. Lindsey-Boltz, LA., et al. *J. Biol. Chem.* **2012**, *287*, 36123-36131
7. Wold, MS. *Annu. Rev. Biochem.* **1997**, *66*, 61-92.
8. Fanning, E., Klimovich, V., and Nager, AR. *Nucleic Acids Research*. **2006**, *34*, 4126-4137.
9. Xu, Xin., et al. *Mol. Cell. Biol.* **2008**, *28*, 7345-7353.
10. Zou, L., Liu, D., and Elledge, SJ. *PNAS.* **2003**, *100*, 13827-13832.
11. Wu, X., Shell, SM., and Zou, Y. *Oncogene*. **2005**, *24*, 4728-4735.
12. Sohn, SY. And Cho, Y. *J. Mol. Biol.* **2009**, *390*, 490-502.
13. Parrilla-Castellar, ER, Arlander, SJ, and Karnitz, L. *DNA Repair (Amst).* **2004**, *8-9*, 1009-1014.
14. Delacroix, S., et al. *Genes Dev*. **2007**, *21*, 1472-1477.
15. Kumagai, A., et al. *Cell*. **2006**, *124*, 943-955
16. Jeong, SY., et al. *J. Biol. Chem.* **2003**, *278*, 46782-46788.
17. Kumagai, A. and Dunphy, WG. *Nat. Cell. Biol.* **2003**, *5*, 161-165.
18. Lindsey-Boltz, LA. and Sancar, A. *J. Biol. Chem.* **2011**, *286*, 19229-36
19. Cotta-Ramusino, C., et al. *Science*. **2011**, *332*, 1313-1317.
20. Kemp, MG., et al. *J. Biol. Chem.* **2010**, *285*, 16562-16571.

21. Kondratov, RV. and Antoch, MP. *Trends Cell Biol*. **2007**, *17*, 311-317.

22. Gorgoulis, VG., et al. *Nature*. **2005**, *434*, 907-912.

23. Bartokova, J., et al. *Nature*. **2005**, *434*, 864-870.

24. Bartkova, J., et al. *Nature*. **2006**, *444*, 633-637.

25. Di Micco, R., et al. *Nature*. **2006**, *444*, 638-642.

26. Bochkareva, E., et al. *PNAS*. **2005**, *102*, 15412-15417.

27. Sttauffer, ME. and Chazin, WJ. *J. Biol. Chem.* **2004**, *279*, 30915-30918.

28. Bochkarev, A., et al. *Nature*. **1997**, *385*, 176-181.

29. Bochkarev, A., et al. *The EMBO Journal*. **1999**, *16*, 4498-4504.

30. Jacobs, DM. et al. *J Biomol NMR*. **1999**, *14*, 321-331.

31. Mer, G., et al. *Cell*. **2000**, *103*, 449-456.

32. Bochkareva, E., et al. *The EMBO Journal*. **2001**, *20*, 612-618.

33. Bochkareva, E., et al. *The EMBO Journal*. **2002**, *21*, 1855-1863

34. Fan , J. and Pavletich, NP. *Genes Dev*. **2012**, *26*, 2337-2347.

35. Brosey, CA., et al. *Nucleic Acids Research*. **2013**, *XX*, in press.

36. Ratnam., K., and Low, AJ. *Clin. Cancer. Res*. **2007**, *13*, 1383-1388.

37. Rouleau, M., et al. *Nat. Rev. Cancer.* **2010**, *10*, 293-301

38. Toledo, LI., Murga, M. and Fernandez-Capetillo, O. *Molecular Oncology*. **2011**, *5*, 368-373

39. Sarkaria, JN., et al. *Cancer Res*. **1999**, *59*, 4375-4382.

40. Toledo, LI., et al. *Nat. Struct. Mol. Biol*. **2011**, *18*, 721-727.

41. Sausville, EA., et al. *J.Clin. Oncol*.**2001**, *19*, 2319-2333.

42. Perez, RP., et al. *Clin. Cancer Res.* **2006**, *12*, 7079-7085.

43. Fracasso, PM., et al. *Cancer Chemother. Pharmacol*.**2011**, *67*, 1225-1237.

44. Zabludoff, SD., et al. *Mol. Cancer Ther.* **2008**, *7*, 2955-2966.

45. Mitchell, JB., et al. *Clin. Cancer Res.* **2010**, *16*, 2076-2084.

46. Morgan, MA., et al. *Cancer Res.* **2010**, *70*, 4972-4981.

47. Blasina, A., et al. *Mol. Cancer Ther.* **2008**, *7*, 2394-2404.

48. Zhang, C., et al. *Clin. Cancer Res.* **2009**, *15*, 4630-4640.

49. Brega, N., et al. *J. Clin. Oncol.* **2010**, *28*, 3062.

50. Daud, A., et al. *J. Clin. Oncol*. **2010**, 28, 3064.

51. Ma, CX., Janetka, JW., and Piwnica-Worms, H. *Trends in Molecular Medicine*. **2011**, *17*, 88-96.

52. Syljuasen, RG., et al. *Mol. Cell. Biol.* **2005**, *25*, 3553-3562.

53. Anciano Granadillo, VJ., et al. *Journal of Nucleic Acids*. **2010**, *10*, 304035.

54. Neher, TM., et al. *ACS Chem Biol*. **2010**, 10, 953-65.

55. Glanzer, JG., Liu, S., and Oakley, GG. *Bioorg. Med. Chem.* **2011**, *8*, 2589-2595.

56. Souza-Fagundes, EM., et al. *Analytical Biochemisty.* **2012**, *421*, 742-749.

57. Scott, DE., et al. *Biochemistry*. **2012**, *51*, 4990-5003.

58. Kranz, JK., and Schalk-Hihi. *Methods Enzymol*. **2011**, *493*, 277-298.

59. Vanwetswinkel, S., et al. *Chemistry & Biology*. **2005**, *12*, 207-216.

60. Mayer, M., and Meyer, B. *Angew. Chem., Int, Ed.* **1999**, *38*, 1784-1788.

61. Dalvit, C., et al. *J. Biomol. NMR*. **2000**, *18*, 65-68.

62. Klages, J., Coles, M., and Kessler, H. *Analyst*. **2007**, *132*, 693-705

63. Shuker, SB., et al. *Science*. **1996**, *274*, 1531-1534.

64. Blundell, TL., Jhoti, H., and Abell, C. *Nat. Rev. Drug Discovery*. **2002**, *1*, 45-54.

65. Drinkwater, N., et al. *Biochem. J.* **2010**, *431*, 51-61.

66. Navratilova, L., and Hopkins, AL. *ACS Med. Chem. Lett*. **2010**, *1*, 44-48.

67. Congreve, M., et al. *Methods Enzymol*. **2011**, *493*, 115-136.

68. Verdonk, ML., et al. *J. Med. Chem.* **2011**, *54*, 5422-5431.

69. Favia, AD., et al. *J. Chem. Inf. Model*. **2011**, *51*, 2882-2896.

70. Trott, O., and Olson, AJ. *J. Comput. Chem.* **2010**, *31*, 455-461.