Title

Structure-Based Drug Discovery Approaches Against the Novel AntiCancer Therapeutic Target caPCNA

Permalink

https://escholarship.org/uc/item/8pz716cw

Author

Jossart, Jennifer

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Structure-Based Drug Discovery Approaches Against the Novel Anti-Cancer Therapeutic
Target caPCNA

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Biochemistry and Molecular Biology

by

Jennifer N. Jossart

September 2022

Dissertation Committee:
     Dr. Jeff Perry, Chairperson
     Dr. Maurizio Pellecchia
     Dr. Chia-en Chang

The Dissertation of Jennifer N. Jossart is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

I am filled with gratitude for all the people that have helped me throughout my academic career. More than anyone, I would like to thank my graduate advisor and mentor, Dr. Jeff Perry, who has invested so much of his time towards my success. I feel that he saw potential in me that I did not see in myself and his guidance has helped shape me into the researcher I am today. I am indebted to him for the countless hours spent on troubleshooting experiments, editing proposals, and securing funding to ensure I could dedicate my time to this exciting research. Dr. Perry has provided me with countless opportunities to excel as a graduate student and I am so excited to continue this journey as a post-doc working alongside my mentor.

I must also extend my gratitude to my committee members and collaborators who have contributed their time and insight help me grow as a researcher. My committee members over these past five years, Dr. Chia-en Chang, Dr. Sihem Cheloufi, Dr. Sean O'Leary, and Dr. Maurizio Pellecchia, have pushed me to view my experiments from different perspectives and provided new ideas to help guide my studies. My collaborator at UCR, Dr. Pellecchia, provided me the opportunity to contribute to his innovative work in the realm of drug discovery and welcomed me into his lab to learn from both Dr. Baggio and Dr. Udompholkul when I needed help with both NMR and DELFIA techniques. My collaborators at the City of Hope Medical Center in Dr. Malkas and Dr. Hickey's research groups, welcomed me into the fellowship of the ring. I appreciate the feedback they have given me in our weekly meetings and the efforts they have made to make sure that I, as the most junior member of our team, am still part of the conversation and understand each of

iv

the analyses they have mastered. I am very much looking forward to continuing our investigation into the one ring to rule them all and I could not imagine a better group of scientists to work with and learn from.

I consider myself very fortunate to have found a lab that allowed me to pursue my research interests and also establish life-long friendships. It was in this lab I met my best friend, Erica Li, who brightens my days with her infectious laughter. We are kindred spirits and I feel so fortunate to have her in my life. Team Jen will always be dear to me and I am grateful to my undergraduates, Amy, Leandra and Robert for helping me when I was overwhelmed with work and grounding me when I needed to be reminded that I don't live in the lab. I am so proud of you all and I know you will achieve great things. There are so many impressive students that have passed through our lab and I know I will not be able to name them all, but I would like to especially thank Christ for guiding me when I first started in the Perry lab, Desiree for keeping me sane, Jenna for helping me whenever I was in crisis mode, and Mahamaya for always encouraging me when I doubted myself.

Last, but certainly not least I must extend sincerest appreciation to my husband and family. This dissertation would not be possible without my husband's incredible patience as I broke the computer and Microsoft Word countless times. My family supported me when I decided to apply to graduate school and understood when I had to disappear for a while to write this 'book'. My achievements would not have been possible without their love and support.

ACKNOWLEDGEMENT OF PREVIOUSLY PUBLISHED MATERIALS

DEDICATION

This thesis is dedicated to my husband, Juan. You have always been a constant

source of encouragement throughout the highs and lows of both graduate school and life.

Thank you for your love and your light.

This work is also dedicated to my loving parents, Daniel and Maria, for their support

during this process and for providing exceptional examples of what hard work and

determination can achieve in life.

ABSTRACT OF THE DISSERTATION

Structure-Based Drug Discovery Approaches Against the Novel Anti-Cancer Therapeutic
Target caPCNA

by

Jennifer N. Jossart

Doctor of Philosophy, Graduate Program in Biochemistry and Molecular Biology
University of California, Riverside, September 2022
Dr. Jeff Perry, Chairperson

Proliferating cell nuclear antigen (PCNA) acts as a vital scaffold to mediate DNA

metabolism and cell survival mechanisms. The identification by our collaborators at the

City of Hope medical center of a cancer associated isoform of PCNA, caPCNA, provided

a means for the development of anti-cancer therapeutics that are selective to cancer cells.

Our collaborators also developed a selective caPCNA inhibitor, AOH1160, that induced

cancer cell apoptosis at nanomolar concentrations *in cellulo* and in mouse *in vivo* models,

without effecting non-malignant cells. Yet, this AOH1160 molecule was not metabolically

stable, and required further optimization to develop a more drug-like lead. We

hypothesized that structural-based drug design (SBDD) techniques could be used to

characterize the AOH1160 scaffold binding interactions at atomic resolutions and help

guide future caPCNA inhibitor design. Thus far, we confirmed that the city of Hope

AOH1160 series of compounds bind to recombinant PCNA protein. Furthermore, through

our discovery of novel recombinant PCNA crystallization conditions, we have determined

three co-crystal structures of caPCNA:AOH1160 analogs. This includes the recently

developed analog AOH1996, which has superior metabolic stability as compared to the

AOH1160 parent molecule, while maintaining its potency, and it has now entered Phase-1 clinical trials for breast cancer therapy. However, there is a ~90% failure rate of novel chemical entities heading into the clinic, with solid tumors being a particularly challenging area for drug discovery and development. Thus, we are continuing to investigate analogs around this scaffold, as well as aiming to discover and develop PIP-box selective inhibitors that are in different areas of chemical space. Our binding and structural studies have identified a novel PIP-box binding molecule, B05, that has approximately 20 times greater potency *in cellulo* as compared to the AOH1996 scaffold. In addition, we have identified a pool of promising fragment binders through our rapid binding assays to funnel into fragment-based lead design (FBLD) studies that aim to identify advantageous chemistries to be utilize in lead design. A combination of *in silico* docking and with fragment screening led to our discovery of new caPCNA binders with novel chemical scaffolds, which we are exploring further. Also, we have recently identified novel *apo*-PCNA crystallization conditions that diffract to much higher resolutions than previously, which likely allows for the unambiguous characterization of fragment binders of interest by X-ray crystallography. Overall, we provided the first determination that the City of Hope AOH series of compounds are PCNA binders, and our successful structural determinations of AOH1160 analogs, together with our new B05 analog, are currently enabling a new generation of caPCNA inhibitors to be developed with the goal of enhanced clinical utility.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1
PCNA: a critical hub and novel anti-cancer target

Chapter 1: PCNA: a critical hub and novel anti-cancer target

1.1   Abstract

PCNA critically functions as scaffold to organize numerous enzymes for DNA replication, repair, chromatin formation, and cell survival mechanisms. The majority of PCNA protein-protein interactions are mediate through the PCNA-interacting-protein (PIP) box binding site. Recently, a unique isoform of PCNA has been identified to play a critical pathological role in various cancers. Notably, this cancer associated isoform of PCNA (caPCNA) is highly expressed in cancer cells, but not at significant levels in non-malignant cells. In a novel approach to targeting caPCNA function our collaborators first mapped the amino acids residues of the PIP-box cavity to identify key interactions that contributed to caPCNA protein-protein interaction specificity. This study identified the region spanning amino acid residues L126-Y133, which partially delineates the PIP-box binding site, and derived the R9-cc-caPeptide (R9-caPep). This decoy peptide effectively hijacks caPCNA PIP-box interactions and exhibited selective cytotoxicity in cancer cells at IC50s of $10 - 90$ uM, but presented minimal toxicity in non-malignant cells. A subsequent computer aided-drug design (CADD)-based approach to identify small molecule inhibitors to target this L126-Y133 adjacent PIP-box pocket led to the identification of a 'first-in-class' small molecule, AOH1160, that inhibits caPCNA hub activity *in vitro*, *in cellulo* and *in vivo* to drive cancer cell death while not affecting non-malignant cells. However, the AOH1160 scaffold is not metabolically stable and therefore we aim to utilize structural characterization studies through macromolecular crystallography to guide the development of more drug-like lead molecules.

1.2   Key cellular functions of PCNA

Commonly referred to as the 'maestro of the replication fork', proliferating cell nuclear antigen (PCNA) acts as a vital hub to dock and localize enzymes acting in DNA metabolism events and cell survival mechanisms [1-3]. Although initially characterized as an auxiliary factor to conferring processivity to replicative DNA polymerase δ (polδ), further investigation has revealed that PCNA is an essential component of both DNA replication and repair mechanisms, where PCNA acts as a docking scaffold to localize multiple replication and repair protein binding partners to DNA [3, 4]. Additionally, chromatin bound PCNA regulates chromatin organization during S-phase and ensures accurate inheritance of both genetic and epigenetic information passed from parent to daughter cells. The increased expression of PCNA during S-phase has been utilized as a reliable proliferation marker in cancers [5, 6]. Recent studies have discovered that the multi-faceted functions of PCNA extend beyond the nucleus as it has been recently identified that PCNA undergoes translocation to the cytosol. Cytosolic PCNA has been characterized to form binding interactions with six of the ten glycolytic enzymes to potentially link glycolysis and cancer cell proliferation. Further study in mature neutrophils has revealed that PCNA directly binds and sequesters apoptotic procaspases to promote neutrophil survival and aberrant inflammation responses in a variety of inflammatory diseases [7-10].

The essential role of PCNA in eukaryotic DNA replication

High-fidelity replication of chromosomal DNA is dependent on the complex assembly of multiple proteins at the replication fork. Central to this multi-protein complex is the DNA-

sliding clamp, PCNA, which associates replicative enzymes to facilitate timely synthesis of the newly formed daughter strands. The formation of the replication fork occurs in a multi-step mechanism that is first initiated during G1-phase of the cell cycle. To rapidly replicate the vast eukaryotic genome there are multiple encoded sites along the DNA sequence that specify replication starting points that are defined as origins of replication. During G1-phase origin licensing dictates the assembly of the pre-replicative complex (PRC) and is essential to ensure the genetic material is replicated only once per cell cycle. The PRC is composed of the origin of replication complex (ORC), initiation co-factors CDC6 and Cdt1, and the replicative helicase. Origins of replication are bound by the ORC, which stimulates chromatin remodeling to create an accessible, open conformation of the DNA duplex for the recruitment of the remaining PRC components. The initiation factor proteins CDC6 and Cdt1 facilitate interaction between the DNA-bound ORC and inactive helicase, MCM2-7, to complete origin licensing. During the transition into S-phase, PRC assembly is completed with the recruitment of the final helicase component, MCM10. The PRC is then activated via phosphorylation by cyclin-dependent kinases (CDKs) and Cdc7-Dbf4 kinase (DDK) phosphorylate [11, 12]. The now active MCM helicase complex unwinds the DNA duplex to form two bi-directional replication forks within a replication bubble. The single-stranded leading and lagging templates of the replication fork are both stabilized by replication protein A (RPA). The pol α-primase complex binds to RPA, and it synthesizes an RNA primer followed by a short stretch of DNA to prime each of the newly separated strands [13]. The pol α-primase complex is then replaced by the eukaryotic clamp loading factor (RFC), which facilitates the opening and closing of the PCNA ring

around DNA in an ATP-dependent process [14]. RFC-dependent loading occurs once for the leading strand, while the lagging strand requires one RFC-PCNA loading per Okazaki fragment. This loading mechanism always positions the protein-protein interaction (PPI) interface of PCNA facing the direction of replication [15]. Once PCNA is loaded onto the DNA strand high-fidelity polymerases δ (polδ) and ε (polε) are recruited and directly bind PCNA to preform lagging and leading strand synthesis respectively. PCNA confers processivity to these replicative enzymes, by ensuring the polymerase remains tethered to DNA throughout DNA synthesis. The scaffold activity of PCNA enables Okazaki fragment maturation in lagging strand synthesis, by organizing the sequential interactions of polδ, flap endonuclease 1 (FEN1), and DNA Ligase I [16]. On the lagging strand, the 5' to 3' directionality of the primed template results in discontinuous DNA synthesis. As polε extends the primer-template junction to the end of the previous Okazaki fragment, the 5' end of this fragment is then displaced, creating a flap of single-stranded DNA. FEN1 is then recruited to PCNA, and binding stimulates FEN1 endonuclease activity to cleave the flap structure. DNA ligase I is the recruited to seal the nick and complete the joining of Okazaki fragments to synthesize a continuous daughter strand [16].

PCNA facilitates DNA damage tolerance and repair pathways

The critical hub function of PCNA simultaneously coordinates DNA replication and facilitates the resolution of encountered DNA damage. Unresolved DNA damage can result in severe consequences such as degenerative diseases and tumorigenesis. Therefore, as the central component of the replicative machinery PCNA must coordinate the recruitment of enzymes involved in either damage tolerance or repair pathways in a timely manner to

prevent prolonged stalling of the replication fork that can lead to cell death [17]. The three DNA damage tolerance (DDT) mechanisms are translesion synthesis (TLS), template switching (TS) and the homologous recombination (HR) salvage pathway [17]. Upon encountering a DNA lesion, the high-fidelity replicative polymerases polδ and polε are unable to accommodate DNA lesions in their active sites, thereby preventing the progression of DNA replication [18]. Rather than disassemble, the replication machinery will bypass DNA damage and leave a single-stranded DNA (ssDNA) gap behind the replication fork. The exposed ssDNA is bound by RPA, which facilitates the recruitment of TLS pathway enzymes that includes the E3 ligase, Rad18. The recruitment of Rad18 is key to initiating the TLS pathway because this enzyme works in conjunction with the E2 ubiquitin (Ub) conjugating enzyme, UBE2B (Rad 6 in yeast) to mono-ubiquitinate the K164 residue of PCNA. This modification of PCNA triggers the switch between high-fidelity polymerase interaction to TLS polymerase recruitment [19-21]. The TLS pathway utilizes low-fidelity polymerases (polη, ι, κ, ζ, and Rev1) that contain both PCNA and Ub-binding motifs to favor interaction with Ub-PCNA over unmodified PCNA [22]. The TLS polymerases contain flexible active sites to accommodate modified bases so they can replicate across DNA lesions, but they lack proof reading activity to prevent incorrect base incorporation [23]. Therefore, this process effectively rescues replication progression and thus avoids fork stalling-initiated apoptotic signaling. However, this mechanism also circumvents repair and potentially introduces point mutations. These mutations may be resolved by a subsequent major repair pathway, but if left unresolved will lead to mutagenesis [21].

An alternative to error-prone TLS is the error-free TS mechanism, which has yet to be fully characterized in humans, but TS in yeast has been characterized to follow a similar mechanism to HR. Studies in *S. cerevisiae* have revealed a potential model for TS, which is referred to as the 'chicken-foot' model and it is summarized as follows: invasion of the newly synthesized sister DNA duplex by the nascent strand, polδ is 'switched' or transferred to PCNA on the sister strand, error-free replication continues past the lesion, and the produced sister chromatin junction is resolved [24, 25]. Identical to TLS, the first step of TS initiation requires mono-ubiquitinoylation of K164 by the UBE2B/Rad18 complex. As characterized in *S. cerevisiae*, the Rad6/Rad18 complex then recruits an additional E2/E3 complex (Ubc13/Mms2/Rad5) that catalyzes the poly-ubiquitination of K164 of yeast PCNA. In humans, this modification is executed by the Rad5 functional homologues, helicase-like transcription factor (HLTF) and SNF2 histone linker PHD RING-helicase (SHPRH), which work alongside the Ubc13/Mms2 homologs, UBC13/UEV1 [25]. The poly-ubiquitination of PCNA both signals the activation of the TS pathway and inhibits the TLS pathway by blocking the interaction of PCNA with TLS polymerases [21, 22, 26].

The third DDT pathway is the HR salvage pathway, which resolves replication fork stalling, but also introduces potentially lethal chromosomal rearrangements through deleterious recombination events [1, 27]. Therefore, the HR salvage pathway is generally inhibited during S-phase to instead promote damage tolerance by error-free TS in a process regulated by the addition of small-ubiquitin-like modifier protein (SUMO) to PCNA. More specifically, the E2-Ub-conjugating enzyme, Ubc9 and E3-SUMO-ligase SUMOylate the

K164 and K127 residues of PCNA to prevent the recruitment of HR salvage pathway initiation enzymes, Rad51 and Rad52 [1, 21, 28, 29]. Interestingly, the K164 residue also undergoes ubiquitination to initiate the TLS and TS pathways [19]. This indicates potential crosstalk between the SUMO and ubiquitin modifiers, and further investigation in *S. cerevisiae* revealed that Rad18 has SUMO-targeted ubiquitin ligase (STUbL) activity [30, 31]. STUbLs are RING finger E3s that contain a conserved SUMO interaction motif [32, 33]. The attachment of the yeast SUMO homologue to K164 and K127 provides a substrate for Rad18, which can ubiquitinate both the SUMO extension and the protein that the SUMO modification is attached to. The Rad18 STUbL activity explains the switch from SUMO to ubiquitin modification on the same residue of PCNA. The degradation of the SUMO moiety by Rad18 permits the initiation of the alternative DDT pathways depending on Rad6 (TLS) or Rad5 (TS) activity [21, 28, 30]. These findings demonstrate that the post-translational modifications (PTMs) of PCNA that regulate PPIs are critical in preventing unwanted recombination events.

The DNA repair pathways of nucleotide excision repair (NER), base excision repair (BER), and mismatch repair (MMR) each require PCNA to mediate their individual mechanisms and subsequent DNA re-synthesis steps. The NER pathway functions to remove bulky modifications that create distortions in the DNA helix that would otherwise interfere with replication. NER relies on the functions of xeroderma pigmentosum (XP) proteins to both detect and remove DNA damage. This multistep pathway begins with damage recognition by the XPC-RAD23B/XPE complex, which recruits transcription factor II H (TFIIH), XPA, and XPG to the site of damage. The helicase activity of TFIIH

unwinds the DNA surrounding the lesion to allow two endonucleases, XPG and XPF (ERCC1) to make incisions on the 3' and 5' side of the strand containing the lesion. The damaged strand is removed and PCNA facilitates DNA resynthesis, through interaction with replicative polymerases and ligase enzymes. Studies have revealed that XPG endonuclease can bind PCNA to potentially coordinate the DNA excision and synthesis steps of NER [34]. Additionally, PCNA interactions with the CRL4 E3 ligase and histone acetyltransferase p300 are thought to modulate NER progression and termination [35, 36].

The BER pathway removes non-helix-distorting base lesions caused by oxidation, deamination, alkylation, and mis-incorporated uracil [37]. Unique to BER is the utilization of specialized DNA glycosylases, which scan the genome to detect and remove the damaged base. Upon encountering a base lesion, the glycosylase will flip the base into its catalytic site pocket to excise the lesion and produce an abasic site [38]. An apurinic/apyrimidinic endonuclease (APE) cleaves the phosphate backbone 5' to the abasic site. Subsequent DNA resynthesis can then proceed through 'short-patch repair' in which DNA polymerase β (pol β) fills the single nucleotide gap, and the nick is sealed by the XRCC1-DNA ligase III complex. Alternatively, in the 'long-patch repair' pathway, a gap of 2-10 nucleotides is generated. PCNA then orchestrates gap filling and subsequent processing in a manner analogous to lagging strand synthesis by recruiting polδ/ε, FEN1 and DNA Ligase I [37, 39]. Studies have demonstrated PCNA plays a role in BER damage recognition, excision and the concluding DNA ligation, via interaction with various DNA glycosylases, stimulating APE activity, and signaling the recruitment of the XRCC1 cofactor protein [40-42].

MMR resolves single nucleotide misincorporations produced from base-base mismatch or small insertions and deletions generated during DNA replication and recombination, and PCNA functions in the recognition, excision, and resynthesis steps of MMR [43]. The MutSα (Msh2-Msh6) recognition complex detects base-base mismatches and short insertions/deletions, while the less frequently used MutSβ (Msh3-Msh6) complex recognizes longer insertions/deletions. PCNA helps localize the MMR recognition complexes through direct interaction with the Msh3 and Msh6 subunits. Excision of the lesion and surrounding bases from the newly synthesized strand is executed by the exonuclease activity of the MutLα (MHL1-PMS2) complex and exonuclease 1 (Exo1). MutLα activity is stimulated by PCNA interaction with the MHL1 subunit [44]. The DNA segment containing the mismatched base is removed and PCNA recruits polδ to fill the gap. Critical to efficient MMR is the discrimination between the original and newly synthesized strands, which is assigned upon RFC-dependent loading that faces the PCNA PPI site toward the 3'-end of the daughter strand [15, 43].

PCNA function in chromatin organization and epigenetic inheritance

Alongside replication and repair mediation, PCNA must also facilitate the dynamic organization of chromatin states. By localizing transcription activation and silencing enzymes to sites of replication, PCNA helps dictate the transition between chromatin states to enable accessibility of the genomic information to replication/repair enzymes and transcriptional complexes. In addition to regulating the accessibility of the parent strand, PCNA also facilitates the packaging and preservation of the newly formed daughter strands. The condensed, genetically identical DNA molecules generated during S-phase

are referred to as sister chromatids. To ensure the accurate distribution of genetic material the sister chromatids must remain physically linked until anaphase when they are separated prior to cell division [45]. Linking is achieved through the establishment of cohesion molecules immediately following DNA replication, and their accompanying acetylation by Eco1 (Ctf7) protein. Initially demonstrated in yeast, and later in human cells, Eco1 activated sister chromatid cohesion is dependent on direct interaction with PCNA for localization following DNA synthesis [1]. It is essential to preserve the genetic information from the parental DNA and the newly synthesized sister chromatids. This is accomplished through the packaging of DNA into condensed chromatin structures that assemble to form the chromosome superstructures. Chromatin is composed of nucleosome subunits that can be further divided into histone core proteins H3, H4, H2A, and H2B. Prior to packaging, the open conformation of chromatin is compared to 'beads-on-a-string', because the nucleosome subunits are deposited onto the DNA strands and the DNA is wrapped around pairs of histone proteins that make up the nucleosome substructure to condense the total length of DNA [46-48]. This packaging process is partially mediated by chromatin assembly factor 1 (CAF-1), which deposits histones H3 and H4 onto the daughter strands in the initial steps of nucleosome assembly. CAF-1 activity is localized to sites of replication through direct interaction with PCNA, indicating PCNA is essential to couple nucleosome assembly to DNA synthesis. In addition, PCNA-dependent CAF-1 activity has been observed in the final steps of both TLS and NER DNA damage response mechanisms to reassemble chromatin following repair [49, 50].

Continuous nucleosome deposition onto nascent DNA stimulates chromatin condensation, which if left unregulated will lead to aberrant heterochromatin formation. Therefore, it is necessary to disrupt nucleosome assemblies to allow access to the DNA duplex for various DNA repair enzymes, epigenetic modifiers, and transcription factors [51]. This essential function is executed by ATP-dependent chromatin remodeling complexes that interrupt DNA-histone interactions and shift nucleosomes on the DNA duplex. The WICH chromatin remodeling complex is composed of ISWI-type ATPase SNF2H and the Williams Syndrome Transcription Factor (WSTF). Studies utilizing monkey cells have revealed that depletion of WSTF results in heterochromatin formation and inhibition of transcription [52]. Thus far, researchers have determined that the WSTF protein contains four potential PCNA interaction motifs, and it has been hypothesized that the WSTF/PCNA interaction may impart processivity to the WICH complex for chromatin remodeling activity prior to chromatin maturation [52, 53].

The dynamic organization of chromatin dictates the accessibility of the genomic information to replication/repair enzymes and transcriptional complexes. True inheritance not only involves the accurate distribution of the nucleotide sequence, but also the epigenetic information not encoded in the DNA sequence, which defines gene expression patterns and must be passed from the parent chromosome to the sister chromatids [54]. Epigenetic modifications that are applied both to the DNA sequence and the histone core proteins of the nucleosome facilitate chromatin condensation/relaxation and transcriptional activation. The histone modification of lysine acetylation neutralizes the positively charged lysine residue and eliminates the electrostatic interaction between the histone core to

unwind the DNA. Typically, this modification is executed by histone acetyltransferases (HATs) to stimulate chromatin relaxation. HAT p300 is recruited to sites of DNA damage in a PCNA-dependent manner to acetylate the histone core and trigger relaxation of the chromatin structure to enable access to NER components for repair [36]. The deacetylation of histone lysine residues by histone deacetylase enzymes (HDACs) condenses the chromatin structure. PCNA has been characterized to facilitate chromatin remodeling activity by localizing HDAC1, 2 to the to the replication fork during DNA replication in S phase; this localization initiates the formation of the nascent chromatin structure following replication [55].

The chemical modification of adding a methyl group ($-CH_3$) to the C5 position of cytosine bases adjacent a guanine base is referred to as CpG methylation. CpG methylation is an epigenetic modification that does not affect the chromatin assembly, but rather triggers the silencing of gene expression through the recruitment of transcription repression enzymes (e.g., methyl-CpG-binding domain proteins) or physically blocks association of transcriptional activators [41]. The DNA cytosine methyltransferase (DNMT1) enzyme contains a PCNA-interacting domain to directly bind PCNA at sites of replication and maintain methylation patterns on CpG nucleotides. This interaction has been characterized in detail through co-crystallization of DNMT1 peptide with PCNA [56]. The PCNA-dependent function of DNMT1 in early to mid-S phase demonstrates the coupling of DNA replication and maintenance of epigenetic inheritance.

Cell cycle control and PCNA-mediated cell survival

PCNA plays an essential role in both cell cycle regulation and apoptosis signaling through interaction with cell cycle progressor and pro-apoptotic enzymes. In the absence of DNA damage the cell cycle will proceed through the activity of cyclins and cyclin dependent kinases (Cdks) that complex with PCNA, to couple progression of the cell cycle with replication. For example, the ternary complex of PCNA/Cdk2-cyclinA catalyzes phosphorylation of RFC, DNA ligase I, and FEN1 to trigger their dissociation from PCNA and support the transition from S-phase [57]. Upon the detection of DNA damage, the tumor suppressor p53 is upregulated to influence the transcription of genes to facilitate DNA repair or pro-apoptotic signaling. A major target of p53 activity is the potent Cdk inhibitor p21 (WAF1), which forms a p21/PCNA/Cdk2-cyclinA quaternary complex to inhibit cell division. In addition to inhibiting Cdk activity, the p21/PCNA interaction blocks binding with polδ and ε, thus inhibiting DNA replication and inducing cell cycle arrest until the damage has been resolved [58]. Similarly, in response to genotoxic or physiological stress, the growth arrest and DNA damage genes GADD45A (Gadd45), GADD45B (MyD118), and GADD45G (CR6) are stimulated by p53 to signal growth suppression and apoptosis. Studies utilizing yeast and human cells have both detected direct interaction between the GADD45 family proteins and PCNA. This binding interaction results in inhibition of growth suppression function and thereby supports cell survival [59, 60]. Notably, both Gadd45 and MyD118 interactions with PCNA have been demonstrated to support DNA damage response by promoting NER and inhibiting cell cycle progression to prevent development of malignancies [61, 62]. Based on these

findings it has been hypothesized that each of the GADD45 family proteins functions in response to DNA damage by stimulating repair in a PCNA-dependent manner [60].

PCNA activity in the cytoplasm

As demonstrated in the previous sections, the nuclear hub function of PCNA has been well characterized in a variety of DNA synthesis mechanisms. Contrary to previous classifications PCNA functions have expanded beyond the nucleus with the recent identification of cytosolic PCNA involvement in cell growth and survival mechanisms. The nuclear-to-cytosol re-localization of PCNA occurs via a nuclear export sequence that is surface exposed when PCNA is in a monomeric state. Once shuttled out of the nucleus, the PCNA scaffold reassembles the homotrimer and has been characterized to form interactions with proapoptotic enzymes to support cell survival in neutrophils and neuroblastoma cells [63, 64], glycolytic enzymes of cancer cells [7, 8], and has been implicated in cytoskeleton remodeling through its activity in osteoclasts [10].

Neutrophils are a key component of the innate immune system that responds to inflammation and infection to exert antimicrobial effects through the release of proteases and reactive oxygen species (ROS). Neutrophil products can also cause host tissue damage, and therefore it is vital to tightly regulate neutrophil function and clearance [65]. The PCNA scaffold binds and sequesters apoptotic procaspases to inhibit their activation and promote neutrophil survival. This function also prevents the untimely release of toxic neutrophil products upon apoptosis and degradation. Current research is focused on targeting cytoplasmic PCNA interactions to prevent binding of procaspases and promote neutrophil degradation as an approach to resolve inflammation responses in inflammatory

diseases [63, 66, 67]. Evidence of cytoplasmic PCNA regulating apoptotic signaling has also been observed in a Parkinson's disease (PD) study that utilized neuroblastoma cell line (SH-SY5Y) that is commonly used as a PD model. This study revealed that reactive oxygen-dependent S-nitrosylation modifications of cytoplasmic PCNA decreased interaction between PCNA and caspase-9. In the absence of PCNA interaction, caspase-9 is activated and initiates pro-apoptotic signaling in these neuronal cells [64].

Translocation of PCNA to the cytoplasm promotes cell survival in chemotherapeutic resistant acute myeloid leukemia, through direct interactions of PCNA with glycolytic enzymes so as to coordinate glycolytic metabolism and cell proliferation. Analysis of PCNA interaction partners utilizing Far-Western blotting in human cancer cells revealed 14 novel cytosolic binding partners. Of these 14 binding partners, 6 enzymes facilitate steps 4-9 of the glycolysis pathway: fructose-bisphosphate aldolase A (ALDOA), triosephosphate isomerase (TPIS), glyceraldehyde-3-phosphate dehydrogenase (G3P), phosphoglycerate kinase 1 (PGK1), phosphoglycerate mutase 1 (PGAM1), and Alpha-enolase (ENOA). Sequencing of the 14 interaction partners revealed 5 enzymes that contained PCNA interaction motifs: ALDOA, annexin A2 (ANXA2), elongation factor 1A (EF1A), PGK1 and malate dehydrogenase (MDHM) [7, 8]. The binding partners EF1A, MDHM, peptidyl-prolyl cis–trans isomerase A (PPIA), and peroxiredoxin-6 (PRDX6) are upregulated in a variety of malignant tissues, which indicates cytosolic PCNA may play an additional role in regulating cancer cell progression. Subsequent enzymatic studies demonstrated that PCNA stimulated the activity of G3P and ALDOA enzymes [7]. In addition, the direct interaction between annexin A2 (ANXA2) and PCNA garnered a

hypothesis regarding PCNA involvement in regulating cytoskeleton integrity. A 2020 study discovered that cytosolic PCNA facilitates actin-cytoskeleton remodeling during osteoclast differentiation, potentially associating PCNA to osteoclast-related bone diseases, such as osteoporosis and rheumatoid arthritis. However, further characterization is required to elucidate this mechanism [10].

1.2.1 Relating PCNA structure and function

PCNA forms a ring-shaped protein exhibiting pseudo-six-fold symmetry that is achieved through three identical monomers joined in a head-to-tail fashion (**Fig. 1.01A**). The outer surface of the PCNA ring is composed of β-sheets and flexible β-hairpins that partially form sites of protein-protein interactions (PPIs). The internal surface, which frames the central pore of PCNA, is lined with α-helicies rich in positively charged lysine and arginine residues [68, 69]. The 35 Å diameter of this central pore allows for threading of the DNA duplex that is 20 Å in diameter. The DNA duplex is covered in negatively charged phosphate groups that form charged interactions with the internal surface of PCNA (**Fig. 1.01B**) [70]. Each monomer is composed of N- and C-terminal domains homologous to each other, which are linked by a flexible interdomain connector loop (IDCL) which spans amino acid residues 118-135. The PCNA ring is split into two distinct faces. The front face points in the direction of DNA synthesis and it contains both the C-terminus and the IDCL to facilitate docking of various binding partners through protein-protein interactions. The back face contains several extended loops targeted for ubiquitination and SUMOylation modifications (**Fig. 1.01A**) [19]. The PCNA-interacting-protein box (PIP-box) cavity is the key binding site on PCNA that is formed by IDCL amino acids 121-133, β-hairpin amino

acids 41-44, and C-terminal amino acids 254-257 [6]. This hydrophobic cavity displays

plasticity that is achieved through the dynamics of loop residues and facilitates the binding

of over 200 different enzymes involved in DNA metabolism (**Fig. 1.01C**) [4, 71].
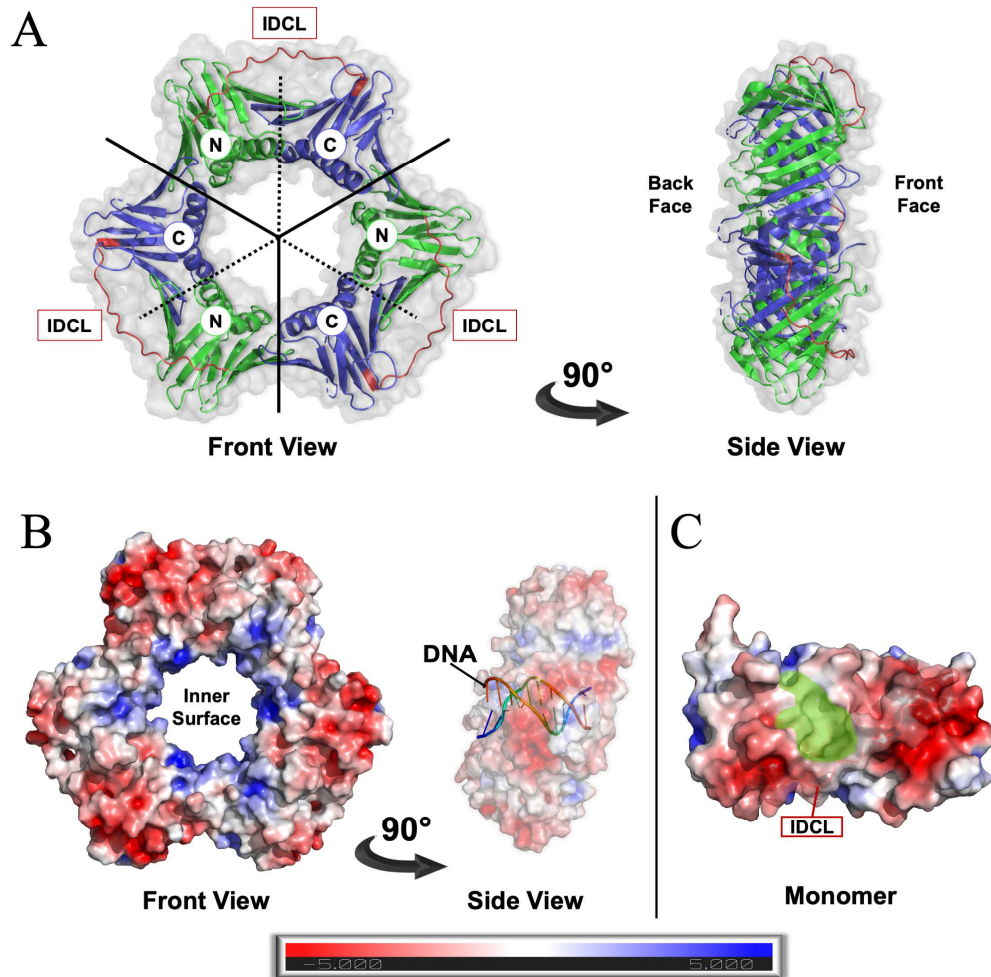
Figure 1.00 | The PCNA homotrimeric ring. (A) Front and 90 ° rotation side view of the PCNA homotrimer assembly in cartoon representation. The front view highlights the head-to-tail interactions of each monomer (solid black line). The N-terminal (green) and C-terminal (blue) domains (dashed line) are linked by the IDCL (red) to make up each monomer. The side view depicts the front and back faces of the ring that specify the direction of replication. (B) Front and 90 ° rotation side view of the PCNA homotrimer in electrostatic surface representation. The Poisson-Boltzmann electrostatic surface potentials are shown in red and blue, corresponding to -5 to +5 kT/e respectively. The side view representation is overlayed with double stranded DNA (rainbow) that threads through the central pore of PCNA. (C) The electrostatic surface representation of a single PCNA monomer. The primary PCNA PPI site adjacent the IDCL is highlighted in green.

The PIP-box cavity is the primary docking site for over 200 proteins that bind PCNA [73]. These PPIs include enzymes involved in DNA replication (RFC, polδ/ε, FEN1, DNA ligase I), repair (polη/ι/κ, XPG), chromatin organization (CAF-1, p300, Ctf7), and cell cycle regulation (p21, Cyclin/CDKs, Gadd45) [72]. PIP-box cavity interacting proteins typically contain a canonical PIP-box motif sequence, Q-x-x-(L/M/I/V)-x-x-(F/Y)-(F/Y), where "x" is any amino acid [73]. The conserved glutamine is bound into a small "Q-pocket" to form hydrogen bonds with the PCNA backbone. This is followed by a helical turn formed by the last five motif residues, whose helix topology contributes to the formation of hydrophobic interactions with the binding site. Two additional PCNA binding motifs have been identified to facilitate interaction with the hydrophobic PIP-box site. The AlkB hydroxylase homolog PCNA interacting motif (APIM), which is composed of a five amino acid consensus sequence, (K/R)-(F/Y/W)-(L/I/V/A)-(L/I/V/A)-(K/R). The APIM motif has been identified in poly(ADP-ribose) family proteins, RNA polymerases, and general transcription factors [75]. The determination of a crystal structure of PCNA in complex with APIM peptide revealed a similar binding mode to the PIP-motif [76]. A third PCNA binding motif, the K-A-box, was identified through random peptide library screening. The consensus sequence, K-A-(A/L/I)-(A/L/Q)-x-x-(L/V), has been detected in tumor suppressors BRCA-1 and BRCA-2 [77], and also the cytosolic proteins ALDOA, ANXA2, EF1A, PGK1, and MDHM [7].

Post-translational modifications regulating PCNA activity

The complex coordination mediated by PCNA raises a key question: what dictates binding partner interaction? Ongoing studies have shown that PCNA protein-protein interactions (PPIs) are primarily controlled by several post-translational modifications (PTMs) to the protein structure,
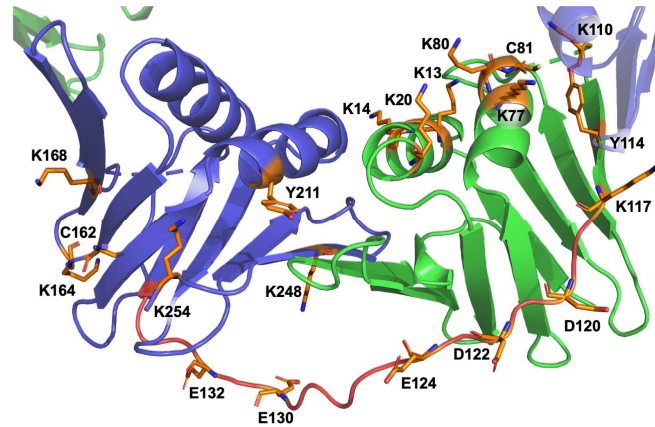
Figure 1.01 | PCNA residues targeted for PTM. Cartoon representation of a PCNA monomer. The residues that have been characterized to undergo PTM are shown in stick representation (*orange*) and labeled.

which directly affect partner binding affinities. As described previously, mono-ubiquitination of K164 triggers the polymerase switch in TLS and further poly-ubiquitination activates the TS mechanism. Adding another level of complexity multiple PTMs have been characterized on the same PCNA homotrimer simultaneously. This is exemplified in the regulation of the HR salvage pathway, which is inhibited in S-phase by SUMOylation of PCNA on K164 or K127 to prevent aberrant HR. Evidence has detected the SUMO modification on K127 alongside a Ub modification on K164, indicating that not every monomer of the homotrimer undergoes the same PTM. In addition, if every PCNA monomer is not equally modified, this proposes a method in which PCNA binds multiple partners simultaneously through each of its monomers [19, 21]. The SUMOylation modification of PCNA has also been implicated in the regulation of sister chromatid recombination [29]. In addition to Ub and SUMO modifications, a few other

PCNA PTMs have been identified to regulate PPIs and overall organization of various DNA metabolism mechanisms. Examples of these modifications include the phosphorylation of Tyr211 by epidermal growth factor receptor (EGFR) to inhibit MMR [78], the acetylation of K20 to promote HR damage repair [79], S-nitrosylation of C81/C162 to regulate the apoptotic pathway in PD [64], and PCNA ISGylation to terminate the TLS DDT pathway [6]. The extensive list of PCNA binding partners and ongoing identification of PTMs offers the potential to correlate PTMs and disease state progression (**Fig. 1.01**).

1.2.2 PCNA and Cancer

Tumor cells express high levels of PCNA to support unrestrained replication and cell proliferation. Therefore, PCNA is considered a proliferation marker in a variety of human tumors, and it is used in conjunction with breast cancer markers (estrogen receptor, progesterone receptor, *Her2/neu*) as a prognostic biomarker [80, 81]. As an essential component to the growth and survival of cancer cells, PCNA also represents an attractive target to develop anti-cancer agents. However, a significant challenge lies in conferring selectivity for malignant cells over normal cells when targeting PCNA in the development of anti-cancer therapeutics. Fortuitously, a cancer associated isoform of PCNA (caPCNA) is ubiquitously and highly expressed in malignant tissues including, breast cancers [82, 83], prostate cancer [84], neuroblastoma [85], and small cell lung cancer [86], but is not significantly expressed in normal cells. This cancer specific isoform was first identified through analysis of the DNA replication machinery components from both malignant and non-malignant breast cells. Two-dimensional SDS-PAGE migration analysis revealed that

malignant cells produced two distinct isoelectric profiles for PCNA; the predominantly expressed caPCNA harbors an acidic isoelectric point (pI) as opposed to the more basic pI of the wildtype (*wt*) PCNA expressed in non-malignant cells [82, 83]. Subsequent studies revealed the expression of caPCNA does not correlate with increased cell proliferation or result from genetic mutations, rather it may arise from alterations to the PTMs typically applied to *wt*PCNA that makes this isoform more accessible to interaction to PCNA binding partners [82]. Thus, this unique isoform of PCNA may offer a novel therapeutic avenue the development of selective anti-cancer therapeutics.

1.2.3   Targeting PCNA functions in cancer

Due to cancer cell dependence on the proliferative activity facilitated by PCNA, multiple research groups have focused their efforts on targeting PCNA function for the development of anti-cancer therapeutics. The clear target to inhibit PCNA function is blocking the PIP-box interaction site that is critical in several PPIs required for accurate DNA synthesis and the more recently characterized cytosolic interactions to promote cell survival and proliferation. However, additional methods to perturb PCNA mediated activities include altering the homotrimeric ring assembly and targeting the signaling pathways leading to PCNA PTMs. To accomplish these inhibitory effects strategies have utilized synthetic peptides, novel small molecule development, small proteins, and DNA aptamers. Here, we present select examples of high-affinity peptide and small molecule-based inhibitors of PCNA and their overall effects on malignancy as reviewed in the literature.

Peptide-based inhibitors of PCNA: p21/PL, ATX-101 and R9-caPeptide

Adopting a peptide-based therapeutic approach is advantageous since peptides derived from bioactive molecules mimic the biological interactions to bestow inherent selectivity for their target. Once the specific peptide sequence required to facilitate interaction has been determined, additional chemical optimization is often required to improve the pharmacokinetic (what the body does to the drug) and pharmacodynamic (what the drug does to the body) profile of these molecules to ensure stability when used in treatments [87]. This peptide-mimetic approach was utilized in the first attempts to inhibit PCNA-mediated functions and was focused on the cyclin dependent kinase (CDK) regulator, p21. The expression of p21 is activated by tumor suppressor protein, p53, in response to DNA damage. p21 binds cyclin-CDK complexes to halt the cell cycle at G1- and S-phase [88]. In addition, via its PIP-motif, p21 binds PCNA and blocks the PCNA interaction with pol δ and FEN-1, ultimately linking cell cycle progression and DNA replication to coordinate DNA repair processes [89]. Focusing on this PIP-motif region in the carboxy-terminus of p21, investigators determined residues 141 - 160 were essential to PCNA interaction. The p21 (141 - 160) 20-mer peptide inhibited *in vitro* SV40 DNA replication and produced an affinity constant of K_d of 87.7 nM for interaction with PCNA, thus capturing the activity of full-length protein [90, 91]. Subsequent efforts to develop peptide-based therapeutics targeting PCNA PIP-box interactions were aimed at minimizing the peptide sequence length and increasing binding affinity. This led investigators to adapt the PIP-motif sequences of two additional PCNA binding partners, Pogo transposase and human DNA ligase I. The hybrid sequence extracted from the PIP-motifs of both binding proteins

resulted in the 16-mer PL-peptide with similar affinity (K_d = 100 nM) to that of p21 (141 - 160) and identical capabilities to inhibit *in vitro* SV40 DNA replication [91, 92]. These examples demonstrate the applicability of developing PIP-motif peptide mimetics for the inhibition of PCNA function.

In a similar approach, peptide mimetics of a secondary PIP-box interacting motif, the APIM, has provided an alternate strategy in targeting PCNA PIP-box interactions. Studies in multiple myeloma therapeutic development have led to the design of the ATX-101 peptide, which contains the APIM consensus sequence and an arginine-rich linker to enable diffusion across the cell membrane. Evidence shows that ATX-101 interaction with PCNA is localized to the cytoplasm where PCNA directly binds procaspases-3, -8, -9, and -10 to reinforce anti-apoptotic signaling. ATX-101 inhibition of PCNA/procaspase sequestering promotes caspase-dependent apoptosis. Treatment of hematological cancer cell lines and healthy lymphocytes revealed that the cancer cell lines were more sensitive to ATX-101 treatment, which induced cancer cell apoptosis in multiple myeloma *in vitro*, *ex vivo* and improved the efficacy of melphalan in mouse *in vivo* models (2 – 6 µM ATX-101 + 1 µM melphalan) [93]. Studies in bladder cancer models have demonstrated that ATX-101-cisplatin combination treatments (8 µM ATX-101 + 10 µM cisplatin) re-sensitize cisplatin resistant cell lines to chemotherapy, which was indicated by an increase of DNA double strand breaks. Based on these observations it is hypothesized that ATX-101 also targets the PIP-box of nuclear PCNA that is involved in mediating DNA damage repair mechanisms, such as translesion synthesis (TLS) that is activated in response to cisplatin-induced

damage [94]. Very recently, studies utilizing glioblastoma cell lines revealed ATX-101 inhibited cell proliferation at IC_{50}s of 4.3 - 10.2 μM, depending on the cell line used [95].

A novel approach to blocking the vital PCNA protein interactions is the design of a PIP-box based peptide rather than PIP-motif derived peptide. More specifically, a PIP-box decoy peptide was derived from the IDCL of caPCNA. Utilizing rabbit polyclonal antibodies, researchers were able to determine the exact amino acid region contributing to caPCNA PPI specificity [83]. This region spanning residues 126-133 of the IDCL was utilized to synthesize caPeptide that was further optimized for cell permeability, ultimately producing a cancer specific mimetic deemed R9-cc-caPeptide (R9-caPep). This peptide exhibits cytotoxicity in breast cancer (IC_{50} = 60 - 90 μM), pancreatic (IC_{50} = 10 - 40 μM), and neuroblastoma cell lines (IC_{50} = 10 - 32 μM), but presented minimal toxicity to non-malignant cell types [85, 96]. Surface plasmon resonance experiments have shown that R9-caPep blocks PPIs between PCNA and PIP-motif proteins FEN-1, DNA ligase 1, and polδ. By mimicking the PIP-box interaction of caPCNA, R9-caPep essentially hijacks vital protein interactions to inhibit replication fork progression and DNA repair *in vitro* and inhibits tumor cell growth in mouse *in vivo* models [85].

Small molecule-based inhibitors: T2AA/T3, PCNA-I1, LRRK2-IN-1 and AOH1160
Small molecule inhibitors have been developed to interrupt the PCNA-mediated protein interactions required for DNA metabolism processes. The small molecule approach is deemed advantageous due to the generally improved pharmacokinetic properties as compared to peptide-based therapeutics. However, an overall challenge of the small molecule drug approach is developing the required selectivity and specificity against the

desired disease target. With the continued advancement of structure-based drug design (SBDD) techniques both *in silico* and *in vitro*, researchers can screen through extensive libraries of potential binders to both identify hits and then progress said hits into lead optimization. By mapping out the relationship between structural characteristics, established protein:ligand interactions and the observed activity researchers can work to improve 'drug-like' properties of these leads to develop novel therapeutics.

To detect small molecule inhibitors of PCNA activity, Punchihewa and co-workers performed *in vitro* high-throughput screens of small molecule libraries. Fluorescence polarization (FP) assays were used to test over 30,000 known therapeutic agents and detected a hit molecule capable of inhibiting PCNA/PL-peptide complex formation at IC_{50} of ~3 μM. This molecule was thyroid hormone triiodothyronine (T3), which clearly lacks an anti-cancer therapeutic application due to its thyroid hormone activity. However, subsequent chemical optimization removed the unwanted activity to produce T2 amino alcohol (T2AA), which was detected to have a slightly greater potency with an IC_{50} of ~1 μM in identical FP studies. Further *in vitro* studies found that T2AA inhibits the high-affinity PCNA/p21 interaction, and it triggers the release of polδ from the PCNA-polδ complex on the chromatin of replicating cells; disruption of the PCNA-polδ complex leads to inhibition of DNA synthesis and induction of replication fork stalling [97]. The introduction of T2AA as a single anti-cancer agent *in cellulo* arrested cells in S phase and promoted apoptosis. The PCNA-T2AA co-crystal structure revealed that T2AA not only binds the PIP-box site, but also forms a second binding interaction near K164. As mentioned previously, the mono-ubiquitination of K164 by the E2/E3 ligase Rad6/Rad18

is required to recruit TLS polymerases and initiate DNA DDT and repair via TLS and the Fanconi anemia pathway [98, 99]. It is hypothesized that the T2AA interaction near K164 and PIP-box binding could block the Rad18 mediated Ub modification and prevent recruitment of TLS polymerases required for DDT and repair. This is further supported by increased DNA damage observed in nucleotide excision repair (NER)-null cells treated with both T2AA and cisplatin that induces damage typically resolved through TLS [98, 100]. Based on these findings, T2AA is an interesting lead for the development of an anti-cancer therapeutic or efficient chemosensitizer to be used alongside existing chemotherapeutics.

A unique approach to inhibiting the proliferative functions of PCNA is targeting its dynamic clamp loading process, which is required for the association of PCNA at the replication fork. The clamp loading process is executed by RFC, which opens and closes the PCNA trimer around DNA in an ATP-dependent process [14]. Through initial *in silico* screens of over 300,000 compounds researchers detected a set of molecules predicted to form interactions with PCNA. These hits were then subjected to a structural similarity search to identify 10 small molecules with optimized chemistries to bind the interfaces of and link the PCNA monomers to greatly increase the stability the trimer. These proposed inhibitors were designated PCNA-Is 1-10. The most potent of these, PCNA-I1 selectively bound the PCNA trimer with K_d values of 0.14 - 0.41 μM. PCNA-I1 is proposed to stabilize the homotrimeric structure by bridging the monomer head-to-tail interaction interfaces. PCNA-I1 binds R146 of one monomer and D86 of the adjacent monomer, while also forming nonpolar interactions with K110 of the adjacent monomer. This small-molecule

mediated interaction repeats at each of the 3 monomer interfaces. Following in cell treatment with PCNA-I1, there was an observed reduction in chromatin-associated PCNA. This reduction is thought to be attributed to the stability of the PCNA-I1/trimer complex, which becomes insensitive to RFC ATP-dependent loading. PCNA-I treatment of a variety of breast, prostate, and melanoma tumor cell lines was found to inhibit DNA replication, trigger cell cycle arrest in S and G2/M phases and decrease cell viability, while having little effect on non-malignant cells [101]. Subsequent structure-activity relationship (SAR) studies focused on the PCNA-I1 scaffold, and these led to the identification of two additional derivatives presenting improved solubility and greater potency when tested in prostate cancer cells lines [80]. Overall, these inhibitors provide a unique approach to modulating PCNA function and anti-cancer therapeutic development.

An intriguing discovery is the antiproliferative activity of the potent inhibitor of leucine-rich repeat kinase 2 (LRRK2) enzyme, which has been implicated in Parkinson's disease progression. Initially designed to inhibit LRRK2, LRRK-IN-1 was confirmed to have selectivity for its target through high-throughput kinome library screening. However, further studies indicated that this molecule also induced off-target effects of cell cycle arrest and apoptosis in cancer cell lines. Large-scale pull-down assays performed in both Jurkat and HeLa cancer cells revealed PCNA was the most significantly targeted protein. This binding interaction was further confirmed by cellular thermal shift assays (CETSA) of LRRK2-IN-1 incubated with PCNA, which resulted in increased protein thermal stability (ΔT_m = 6.7 °C), and therefore indicated binding interaction. Molecular modeling was then utilized to identify the potential binding site of LRRK2-IN-1 on PCNA.

These *in silico* studies revealed a favorable interaction along the IDCL residues G124-I128, near the PIP-box binding site. Co-immunoprecipitation of PCNA with PIP-motif containing replicative polymerases pol δ and pol ε failed following LRRK2-IN-1 treatment, further supporting the PIP-box interaction hypothesis. The anti-proliferative activity of LRRK-IN-1 was tested alongside known PCNA inhibitor, T2AA. LRRK-IN-1 was found to be slightly more potent of an inhibitor with an IC_{50} value of 0.26 μM as compared to T2AA (IC_{50} = 1.63 μM). More detailed characterization of LRRK-IN-1 is required to understand its inhibitory effects against PCNA function and investigate any additional off-target effects. However, this discovery has provided a strong foundation for the development of future leads capable of inhibiting PCNA activity and potential anti-cancer therapeutics [102].

The recently identified AOH1160 molecule is potent inhibitor of caPCNA function. Based on the structural information gathered during the development of R9-caPep, it was determined that the PIP-box binding site partially delineated by IDCL residues L126 - Y133 was essential to caPCNA function in cancer cell survival [83]. Subsequent drug discovery efforts implemented high-throughput virtual screens of over 6-million compounds to detect caPCNA PIP-box selective binders. These initial *in silico* studies revealed a set of 57 predicted binders that were subjected to cell viability assays. The most selective of these, AOH39, contains a scaffold that has not been linked to any previously characterized biological activity. Based on this novel scaffold the potent AOH1160 analogue was developed. Treatment of neuroblastoma, breast cancer, and small cell lung cancer cell lines with AOH1160 induced apoptosis at IC_{50}s of 0.11 - 0.53 μM.

AOH1160 treatment was not significantly toxic to non-malignant cells up to 5 μM, indicating an improved therapeutic index compared to its parent molecule, AOH39. In addition, detailed saturation transfer resonance (STD) nuclear magnetic resonance (NMR) analysis revealed that AOH1160 effectively blocks the well-characterized T3/PIP-box interaction and like T2AA, lacked thyroid hormone activity. Incubation of AOH1160 with neuroblastoma and small cell lung cancer cells resulted in cell cycle arrest and increased apoptosis as compared to treatment in normal cells, which were largely unaffected. In addition, western blot analysis of whole cell extracts of neuroblastoma and non-malignant cells detected increased concentration of caspase-3 and -9 fragments over a 48 hr treatment with AOH1160. The fragmentation of these caspases indicates enzymes activation that is observed in response to pro-apoptotic signaling. The inhibition of homologous recombination-mediated double strand break repair by AOH1160 increased sensitivity to cisplatin treatment in neuroblastoma cells. The AOH1160 molecule was stable in Es1^c/SCID mouse, canine, monkey, and human plasma. Successful oral administration of AOH1160 to Es1^c/SCID modified mice reduced tumor proliferation without introducing significant toxicity. This promising lead moleculeAOH1160 is the first orally available small molecule-based, drug-like lead molecule that selectivity targets malignant cells with clinical potential for selectively targeting PCNA in cancer therapy [86].

1.3 Approach

Multiple groups have targeted this site through peptide, small molecule and other methods in anticancer therapeutic development that have not been covered in detail here. However, a persisting challenge lies in selectively targeting PCNA function in malignant cells

30

without affecting *wt*PCNA that is essential to normal cell survival. The discovery of the caPCNA isoform by our collaborators in the Malkas and Hickey research groups has provided and advantageous approach to develop cancer-selective therapeutics. Their work utilizing virtual screening techniques to identify novel PIP-box selective binders led to the identification of the novel PIP-box selective scaffold, AOH39. Subsequent optimization of this molecule produced the 'first-in-class' small molecule inhibitor of caPCNA activity, AOH1160. Treatment with AOH1160 has been shown to simulate cancer cell apoptosis at nanomolar concentrations in both *in cellulo* and in modified mouse *in vivo* models. Although AOH1160 presented favorable therapeutic properties and desired potency against the caPCNA protein, this molecule is very hydrophobic in nature and found to be prematurely metabolized in liver microsome studies (data not shown). Therefore, the AOH1160 scaffold requires further optimization to improve metabolic stability and increase bioavailability. With the help of our collaborators, our long-term goal is to develop potent and selective caPCNA inhibitors with optimized metabolic stability. We hypothesize that structural-based drug design (SBDD) techniques alongside fragment-based lead discovery methods (FBLD) could be used to characterize the AOH1160 scaffold binding interactions at atomic resolutions and help guide future caPCNA inhibitor design.

1.4 Research Strategy

1.4.1   Fragment-based lead discovery (FBLD)

FBLD focuses on utilizing low molecular weight compounds with simple and diverse chemistries as starting points for lead design. Fragments generally follow the rule of three (RO3), which dictates the following chemical criteria: M.W. $\leq$ 300 Da, number of
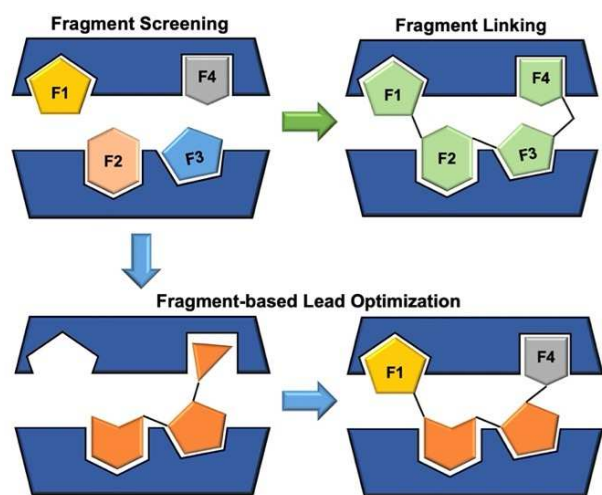
Figure 1.02 | Representation of FBLD compound optimization. The fragment hits detected in screening campaigns can proceed to linking (*green arrow*) to develop target selective small molecule interactors or be utilized to optimize existing lead (*orange*) molecules to enhance target binding interactions (*blue arrows*).

hydrogen-bond donors/acceptors $\leq$ 3, cLogP $\leq$ 3 [103]. The RO3 fragment criteria was derived from Lipinski's rule of five (RO5): M.W. $\leq$ 500 Da, number of hydrogen-bond donors/acceptors $\leq$ 5, cLogP $\leq$ 5, which was developed to provide a chemical guideline for the development of a preclinical lead with good oral bioavailability [104]. Structures following the RO3 are also predicted to be more readily developed into leads that follow the R05 [105]. At a fraction of the size of more complex small molecules, fragments have a greater probability of interacting with binding pockets, clefts, or sites of protein-protein interactions (PPIs). Fragment screening libraries typically composed of $\sim 10^3$ fragments achieve a greater sampling of the chemical space as compared to high-throughput screening (HTS) libraries composed of $\sim 10^6$ small molecules, ultimately making a fragment screen far more cost efficient in comparison [106]. Although initial fragment hits generally bind with low affinities (> 0.1 - 10 mM), due to the small size of the compound, these hits can then be developed into high affinity leads because of their high 'ligand efficiencies' [107]. The ligand efficiency is a measure of the number of fragment/compound atoms that are involved in binding as compared to the overall number of heavy atoms in the structure

[108]. Through fragment screening and identification, multiple hits can be linked to generate selective, high-affinity drug leads. In addition, the identification of fragment binding in the sub-pockets of interaction sites can be used to optimize the chemistries of previously characterized small molecule binders (**Fig. 1.02**) [103]. The Perry lab has a relatively large fragment library of 3,550 diverse fragment compounds with subsets that include 384-drug like cores, high functional diversity, and high Sp^3 characteristics. Notably, two of the sub-libraries are composed of 1,000 brominated and 1,250 covalent fragments. Advantageously, brominated compounds can provide clear identification in crystallographic electron density maps and covalent fragments are designed to target Cys, Ser, His, and Tyr amino acid side chains typically present in PPI sites.

1.4.2   Iterative research cycle

To test our working hypothesis that SBDD techniques could be used to characterize the AOH1160 scaffold binding interactions at atomic resolutions and help guide future caPCNA inhibitor design. (**Fig. 1.03**). The first cycle consists of initial screening assays performed both in *in vitro* thermal shift assays (TSA) and *in silico* molecular docking studies to rapidly detect and rank
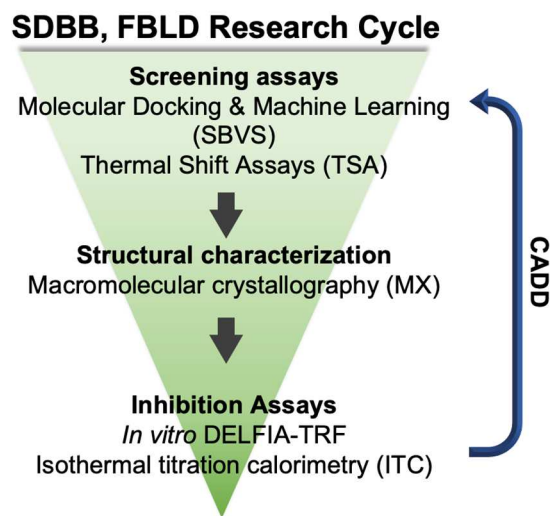

Figure 1.03 | Iterative SBDD, FBLD research cycle. Schematic depicting the overall strategy to generate drug-like, selective caPCNA PIP-box leads.

potential binders. The pool of molecules that will initiate our first research cycle includes

a set of 11 AOH1160 analogs computationally designed by our collaborators. In addition, we have established a recent collaboration with the biotech company, Atomwise, that has developed an innovative artificial intelligence (AI)- driven CADD machine learning technology that was used to detect small molecules predicted to bind the PIP-box binding site of PCNA. Lastly, to investigate novel chemistries for lead generation and optimization we will also screen our in-house library of 3,550 diverse fragments in an FBLD approach. We anticipate that a significant number of promising hits will be detected in our initial screening assays to proceed to the second phase of our research cycle, structural characterization. Utilizing macromolecular crystallography (MX) we will visualize the binding interactions of hits at the atomic level and isolate PIP-box pocket binders. The molecules determined to selectively bind the PIP-box site will then be analyzed in both *in vitro* biochemical affinity assays and *in cellulo* cell viability assays performed by our collaborators. More soluble, high-affinity hit molecules will be analyzed using *in vitro* dissociation-enhanced lanthanide fluorescence immunoassay (DELFIA-TRF) competitive binding assays and isothermal titration calorimetry (ITC). Information gained from these steps is then fed back into the top of the next research cycle. Utilizing computer-aided drug design (CADD) analysis each research cycle will allow us to build upon and further optimize our initial hits and develop high-affinity caPCNA PIP-box inhibitors with improved stability and drug-likeness.

Chapter 2
Rapid *in vitro* screening for early-stage drug discovery

Chapter 2: Rapid *in vitro* screening for early-stage drug discovery

2.1   Abstract

The hit identification stage of early drug discovery campaigns can be costly and time consuming. Utilizing *in vitro* high-throughput screening (HTS) assays readily detect hits from large libraries of potential binders. Thermal shift assay (TSA) is a versatile HTS method capable of detecting interactions of high affinity small molecules and weak affinity fragment binders. We hypothesized that TSA could be utilized to both evaluate the druggability of the PCNA target protein, and characterize the binding interactions of novel covalent warheads against the XIAP-BIR3 domain. We have successfully screened our City of Hope CADD-derived AOH1160 analogs, 63 AI-CADD predicted binders and our in-house library of 3,550 drug-like fragments against the PCNA target. We have detected a total of 24 small molecule and 1,183 fragment hits that increased protein thermal stability by 0.5 °C. To focus on our most promising fragment hits we have applied a shift cut-off of ≥ 2 °C, which equates to an overall 6.9% hit rate. Both the small molecule and promising fragment hits have now progressed to structural characterization studies, via macromolecular crystallography, to identify PIP-box selective binders. In addition, TSA screens of synthesized covalent warheads successfully detected the formation of covalent adducts, as indicated by ΔT_m values of > 20 °C. The results from these studies helped guide the design of warheads to promote reactivity and stability of covalent adduct formation between assayed warheads and the XIAP target.

2.2   A rapid primary screening method

Advancements in chemical synthesis have exponentially increased the volume of commercially available compound libraries, such as the Enamine REadily AccessibLe (REAL) Space library, which comprises nearly 29 billion make-on-demand molecules for drug discovery [109]. In addition, proteomic characterization continues to reveal new and exciting disease related targets for therapeutic investigation. Between 2009 and 2018 the median investment in R&D to get a new drug to market was estimated to be $985 million [110]. Therefore, the development of a reliable high-throughput screening (HTS) assay capable of rapidly identifying hit molecules for a variety of disease targets offers a significant advantage to expediting pre-clinical R&D. One such primary screening method is the thermal shift assay (TSA), which detects the thermal stability of a protein target in solution. The binding interaction between protein and ligand will directly influence the protein thermal stability and equate to a shift in the protein melting temperature (T_m), the temperature at which 50% of the protein sample is denatured [111, 112]. The measured fluorescence-based readouts allow researchers to monitor protein unfolding in a direct quantification of protein thermal stability in solution. Therefore, the effects of the addition of a potential ligand binder, as well as varying environmental components such as buffers, additives, and potential ligands, on protein thermal stability can all be measured [113-116]. TSA is a versatile technique that does not require labeling or immobilization procedures and it can therefore be applied to a variety of protein targets. This application is relatively sensitive and so it can be utilized to detect interactions of weak binders (mM) that includes fragments from fragment-based lead discovery (FBLD) studies, as well as high affinity

inhibitors [117]. Both Academic labs and industry-based laboratories such as Astex Pharmaceuticals, Vernalis, Astrazeneca, and Johnson & Johnson Pharmaceuticals employ TSA in both small molecule primary HTS and FBLD campaigns [111, 118-120].

In 1997, Pantoliano and colleagues described a high-density fluorescence-based binding assay that could be performed in 96- or 384-well format [111]. This high-throughput screening format is economical in that the fluorescence-based readouts can be recorded utilizing a commonly available real-time polymerase chain reaction (RT-PCR) instrument, which can apply a controlled temperature gradient. In addition, the requirements for both protein and compound material are relatively small, with average reaction volumes of ≤ 20 µL and typical final protein concentrations ranging from 5 - 10 µM. The average assay could be executed in approximately 1-hour alongside semi-automated data analysis software to obtain thermal shift values from library screens to rapidly analyze a diverse sampling of the chemical space [111].

Proteins exist in a thermodynamic equilibrium of a variety of conformational states. Increasing the temperature of the reaction environment will shift the thermodynamic equilibrium of a folded protein to an unfolded state [121]. In TSA, this shift is quantified by fluorescence readouts where low fluorescence at room temperature indicates a folded state and increasing fluorescence throughout the reaction temperature ramp phase represents unfolding. The fluorescence output over an increasing temperature range produces a hyperbolic melt curve. The midpoint of excitation is extrapolated as the T_m, or the temperature at which 50% of the protein sample is folded and 50% is unfolded [112]. Methods for T_m determination such as non-linear Boltzmann fitting and higher order

polynomial equations have been reported. However, a straightforward alternative is the application of a first-order derivative to the relative fluorescence units (RFU) as a function of temperature ($-d$RFU$/d$T). The T_m is then the lowest point of the derivative plot and provides a simple T_m read-out [122].

The additional assay components, reaction buffer, additives, ligands, and fluorescent dye will influence the produced melt curve and protein T_m extrapolation. TSA optimization begins with buffer screens to identify a reaction buffer that promotes protein stability and the production of an unambiguous melt curve. Determining a stabilizing buffer condition is also advantageous for crystallization studies. As demonstrated in a 2006 study, Ericsson and colleagues performed TSA screens of varying buffering conditions such as pH, ionic strength, precipitants and additives against a panel of 25 *E. coli* proteins to identify stabilizing conditions. The conditions that resulted in a more significant protein T_m were then translated to crystallization experiments, and 10 of the 25 proteins yielded increased crystal formation [115].

Detecting protein unfolding

Visualization of protein thermal denaturation is dependent on detectable fluorescence output. The fluorescence readings can be produced from either intrinsic or extrinsic sources. Intrinsic fluorescence refers to the fluorescence produced from within the protein sample. Two examples of intrinsic fluorescence are green fluorescence protein (GFP) tags [123] or tryptophan fluorescence detection [124]. Both provide increased fluorescence output upon temperature induce denaturation, however both present potential limitations such as a small molecule binding to GFP rather than the target protein and so creating a

false positive shift [113, 123]. Tracking tryptophan fluorescence at 330 nm and 350 nm falls under a subcategory of TSA methodology that is termed nano-differential scanning fluorimetry (nanoDSF). This technique requires significant tryptophan residues in the target protein of interest and considerable investment in specialized equipment to obtain accurate readings [124, 125].

The extrinsic fluorescence approach capitalizes on the use of environmentally sensitive dyes. These environmentally sensitive dyes have low quantum yields in solvents with high dielectric constants and their fluorescence is quenched in aqueous environments. Upon protein denaturation, dye association with hydrophobic regions of the protein that are exposed during unfolding increases the quantum yield or produced fluorescence [126]. Environmentally sensitive dyes utilized in TSA include 8-Anilinonaphthalene-1-sulfonic acid (bis-ANS), Nile Red, and SYPRO® orange. However, more recent studies have favored SYPRO® orange due to its signal-to-noise ratio and fluorescence signal (λ_{ex} 470 nm, λ_{em} 570 nm) that is compatible with FRET filters found in RT-PCR instruments [112, 127]. Environmentally sensitive dyes aid the universal applicability of TSA by avoiding fluorescent labeling of the protein target of interest. However, the occurrence of false positive or negative shifts can arise from compound interaction with SYPRO® orange dye [113]. Therefore, preliminary screens in absence of protein aid in determining dye-ligand interaction and eliminate false positives.

Induced thermal stability

TSA operates on the principle of ligand-induced conformational stabilization. Preferential ligand binding to the folded protein state will increase the population of the folded protein

conformation and increase the difference in free energy between the two states $\Delta G_{unfolded-folded}$. The overall effect will enhance the protein thermal stability in solution and increase the T_m [111, 112]. To extrapolate ligand-induced T_m shifts, it is necessary to maintain a negative control or reference sample for *apo*-protein T_m, which contains an identical volume of the ligand vehicle that is typically 1 μL 100% dimethyl sulfoxide (DMSO). A positive shift in melting temperature induced by compound-protein interaction as compared to an *apo*-protein reference is indicative of binding interaction ($T_{m\ +ligand} - T_{m\ apo} = \Delta T_m$) [113].

TSA offers a platform to rapidly detect the low molecular weight binders, such as fragments. Additionally, the overall number of fragments hits versus the total number of fragments assayed through TSA provides a hit rate that indicates the 'druggability' of the target protein. Typically, hit rates are calculated from shifts of ≥ 0.5 °C and fall within the given criteria: 1% hit rate = poor drug target, 2-3% = good drug target, > 3% hit rate = highly druggable [128]. Therefore, fragment-based TSA screens provide a rapid, low-cost approach to determine whether a particular drug target is appropriate for subsequent drug discovery efforts.

Fragments have weak binding affinities and consequently require high concentrations (mM) to observe a significant shift in T_m. The relatively high concentration of fragments can give rise to issues such as fragment solubility and aggregation [129]. Moreover, fragments may bind at more than one site and may not bind only to the folded state. Multiple binding events may generate falsely large increases in shifts [130]. In the event ligand binding interaction preferentially targets the unfolded state or induces unfolding of

the protein target a negative shift will be detected. This indicates the protein is destabilized in solution. Typical ΔT_m values can range from 0.5 - 2 °C, depending on the ramp-rate (e.g., 0.5 °C/min) used to heat the reaction mixture and dictate fluorescence measurement periods [122]. However, these shifts provide narrow margins to discriminate hits from non-hits. Therefore, it is advantageous to utilize previously characterized binders as positive controls when available and confirm ligand-induced protein stability with follow-up dose-dependent TSA studies [130, 131].

Notably, shift values obtained utilizing RT-PCR based TSA do not directly correlate to ligand affinities, because TSA measurements do not provide ligand binding enthalpies and heat capacities. For a given protein, ligands with different scaffolds and equivalent affinities can induce varying shift values since there is no distinction between enthalpically or entropically driven interactions [127]. Excluded from TSA studies are pan-assay interference compounds (PAINS) molecules, which have been characterized to react non-specifically with a variety of biological targets, induce aggregation, and quench fluorescence to ultimately produce false positives in HTS assays such as TSA [132, 133].

In summary, the development of high-throughput TSA screening techniques has provided a significant advantage in detecting large libraries of potential binders against a variety of disease targets. By converting the traditional DSC technique to a miniaturized fluorescence-based binding assay a variety of reaction conditions can be assayed against a protein target, while utilizing minimal material. In addition, the use of environmentally sensitive dyes, such as SYPRO® Orange, produces fluorescence readings that are measurable through a commonly utilized RT-PCR machine and avoids the need for tedious

sample preparation methods, such as immobilization or labeling. By monitoring the effect

to protein thermal stability through TSA, researchers detect increased protein stabilization

induced by the addition of buffering conditions, additives, or ligands. This information can

be applied in subsequent structural characterization, which requires a stabilizing buffer

condition to promote protein folding, as demonstrated by Ericsson and coworkers [115].

The increase in protein melting temperature in the presence of ligand indicates ligand-

protein interaction through increased protein stability and the produced shifts (ΔT_m) can be

analyzed to readily detect hits. Advantageously, low molecular weight ligands, such as

fragment interactions can also be detected through TSA. Although non-covalent binders

produce potentially narrow shifts of 0.5 – 2.0 °C, the use of previously characterized

positive controls and follow-up titration TSA further validate the pool of hit compounds.

It is advantageous to pair this technique with more detailed binding assays to characterize

hit molecules in subsequent drug development. Throughout the drug design and

development stages, TSA can also be used to detect and compare the binding interaction

of candidate molecules. Overall, the development of TSA has created a relatively universal

binding assay that can be used to screen hundreds to thousands of reaction conditions in a

single day to significantly reduce cost and time of screening campaigns and provide a rapid

platform for subsequent validation of lead molecule binding interaction.

2.2.2   TSA screening to detect potential PIP-box binders

Our collaborators in the Malkas and Hickey labs at the City of Hope Medical Center

conducted an *in-silico* screening of over 7-million molecules to identify predicted PIP-box

selective binders. From these screens, a novel first-in-class chemical scaffold, referred to

AOH39 was subsequently optimized to the nanomolar nM affinity *in-cellulo* inhibitor, AOH1160 [86]. Structure-activity relationship (SAR) studies guided the development of an additional 11 compounds for a total of 12 City of Hope CADD-based small molecules. In an additional collaboration with Atomwise, the AtomNet™ AI-CADD technology executed an *in-silico* screening of 10-million small molecules against the PCNA PIP-box cavity. The AI-driven CADD screens produced a library of 63 predicted PIP-box pocket binders.

In addition to the aforementioned small molecule libraries, TSA screens have been performed against our in-house library of 3,550 drug-like fragments. This library is divided into subsets with individual specificities. The brominated/covalent library subset is composed of molecules known to form covalent interactions with binding site amino acid residues (Cysteine, Lysine, Serine, Histidine, and Threonine), brominated fragments and brominated-covalent molecules. The presence of a covalent interaction between protein binding site and fragment molecule provides a starting point, or anchor that can be built-out to design a high-affinity selective lead. Brominated fragments are advantageous in subsequent MX studies since bromine has a significant number of electrons as compared to the elements of the peptide chain, and would be readily detectable in electron density maps. The comprehensive library is composed of "building block" fragments that have characterized synthesis pathways for hit-to-lead development. Lastly, the diverse fragment library subset is composed of drug-like cores that are derived from previously characterized therapeutics. Each library is pre-screened to exclude PAINS molecules and provide a diverse sampling of the chemical space.

2.2.3   Characterizing covalent warheads in XIAP-BIR3 model

Working in collaboration with the Pellecchia research group at UCR, we utilized TSA to characterize a set of peptide-based covalent warheads designed inhibit protein-protein interactions (PPIs). The regulatory PPI between inhibitor of apoptosis proteins (IAPs) and second mitochondria-derived activator of caspases (SMACs) was utilized as a model system for the design of these peptide-based inhibitors. In normal cells, the IAP, X-linked inhibitor of apoptosis (XIAP) functions to suppress cell death, potentially by binding pro-apoptotic enzymes caspases 3, 7, and 9 [134]. Simultaneously, SMACs are sequestered in the mitochondria to prevent their activation. In response to pro-apoptotic stimuli, SMACs are released into the cytosol and expose a previously buried N-terminal peptide, Ala-Val-Pro-Phe (AVPF) [135]. The AVPF peptide region forms a PPI with the BIR3 domain of XIAP, and upon binding triggers the release of caspases to promote programmed cell death [136].

The development of SMAC mimetics through structure-based studies to improve the drug-like characteristics of the AVPF peptide has led to the development of potent XIAP inhibitors and potential therapeutic agents such as, GDC-0152 (Genentech Inc.) [137] and clinical candidates LCL161 (Novartis) [138] and AT-406 [139]. In a novel approach, our collaborators introduced aryl-sulfonyl fluoride or aryl-fluorosulfate warheads to improve the potency and selectivity inhibitors developed from the AVPF peptide sequence. These agents were designed to target the PPI surface residue of the XIAP-BIR3 domain, Lys311. Furthermore, to assess the potential for covalent adduct formation with additional surface residues, our collaborators also prepared BIR3 mutants

K311Y, K311H, K311S, and K311T alongside an alanine mutant (K311A) to act as a negative control for the detection of covalent adduct formation. [140, 141]. Here, we present the TSA results of the covalent warheads interaction with the XIAP-BIR3 domain, as published in the Journal of Medicinal Chemistry in 2019 [140, 141].

2.3   Results

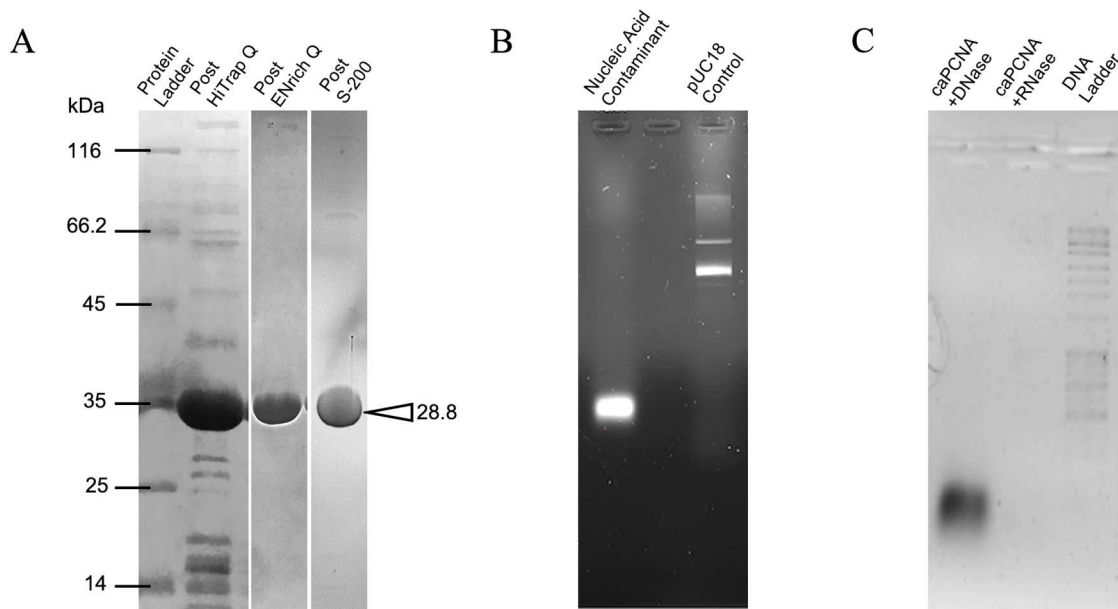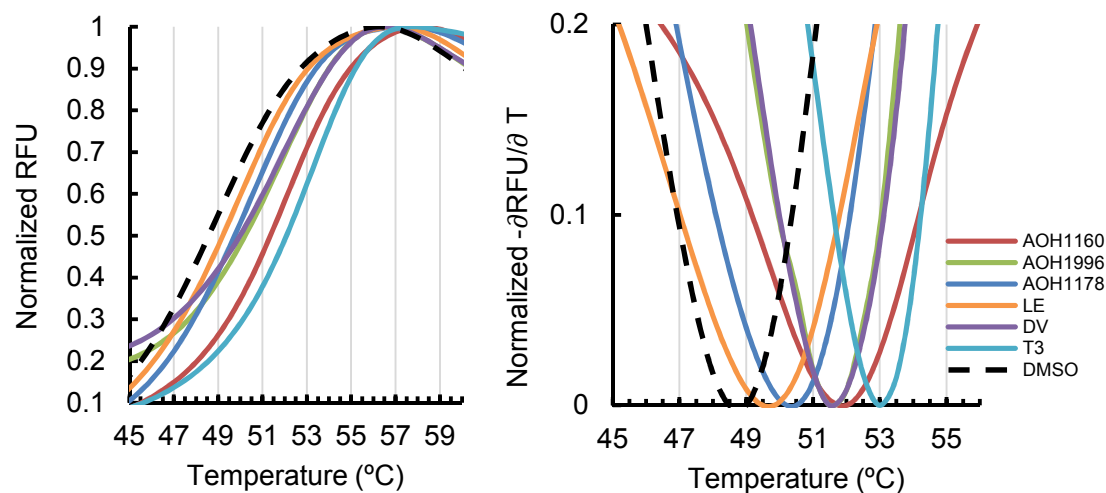### 2.3.1   PCNA TSA screening to detect potential PIP-box binders

Figure 2.00 | Purification of caPCNA. (A) Elution samples following stages of FPLC purification scheme visualized on 12% SDS-PAGE. Protein marker used was Unstained Protein MW Marker (Pierce™). (B) Confirmation of nucleic acid contaminant following Enrich Q elution as compared to pUC18 plasmid control sample as judged by UV illumination of 1% agarose gel. (C) Result of 30-minute incubation of caPCNA Enrich Q elution with 5 mg/mL DNase or RNase run alongside 1 Kb Plus DNA Ladder (Invitrogen™).

Expression and purification of recombinant hPCNA- To screen for potential binders to the caPCNA protein we first needed to produce purified recombinant protein. Since *E. coli* lacks the post translational modification machinery, this expression system was utilized to produce PCNA protein samples similar to the acidic caPCNA isoform that lacks PTMs. A pre-existing construct of human PCNA (hPCNA) gene in pET 22b vector was transformed into the *E. coli* expression cell line Rosetta 2 DE3 (R2DE3). Successful transformation was followed by large scale expression for subsequent purification. Using fast protein liquid chromatography (FPLC) caPCNA was isolated utilizing a series of ion exchange and size

exclusion chromatography (SEC) (**Fig. 2.00A**). As visualized on a 1% agarose gel, the post-SEC protein sample was determined to contain nucleic acid contamination (**Fig. 2.00B**). Two aliquots of this contaminated sample were treated with 5mg/mL DNase or RNase at 25 °C for 30 minutes. Following RNase treatment, the nucleic acid contaminant had been removed (**Fig. 2.00C**) and concluded to be residual RNA carried through the existing purification scheme. Therefore, future purification protocols included an RNase incubation step following high-resolution ion exchange and the RNase was removed with SEC to produce isolated PCNA protein at approximately 16 mg/ liter of E coli. expression. Purified protein was then tested at varying concentrations and buffering conditions in the presence of 1 μL 100% DMSO control to determine optimal TSA reaction parameters to produce an unambiguous melt curve (**Supp. Table 2.00**). TSA studies were carried out with final in-well concentrations of 9 μM PCNA, 20x SYPRO® orange, and 9% DMSO diluted with reaction buffer 1x PBS pH 7.4 (1.8 mM KH_2PO_4, 10 mM Na_2HPO_4, 2.7 mM KCl, and 137 mM NaCl, pH 7.4). All caPCNA TSA studies were preformed alongside previously characterized PCNA inhibitor triiodothyronine (T3) as a positive control and 100% DMSO as a negative control in triplicate.

Figure 2.01 | City of Hope CADD-molecule TSA. Normalized melt curve and inverse first derivative plots of PCNA melts in the presence of 1 mM hit. Results are an average of assays preformed in triplicate. The DMSO-only control is plotted in black dashed line.

Table 2.00	COH CADD-based Small Molecule ΔTm												
Name	T3	AOH1160	AOH1996	AOH1178	PCNA2	PCNA3	AOH1160e	DV	LV	DE	LE	DK	LK
ΔT_m (°C)	3	2	1.5	1	-5	-0.5	SD	1.5	0	-1.5	0.5	-0.5	-1
cLogP	3.3	4.5	4.2	3.9	5.0	4.7	4.7	5.9	5.9	4.6	4.6	5.9	5.9
cLogD	3.3	4.5	4.2	3.9	5.0	4.7	4.7	5.9	5.9	1.5	1.5	2.4	2.4

City of Hope CADD-based Molecules- A total of 12 City of Hope CADD-based small molecules were analyzed at a final concentration of 1 mM. Positive shift values were obtained for 5 of the 12 compounds, AOH1160, AOH1996, and AOH1178, AOH1160_DV (DV), and AOH1160_LE (LE). The compounds AOH1160_DE (DE), AOH1160_DK (DK), AOH1160_LK (DK), PCNA3, and PCNA2 produced negative shifts that are indicative of protein destabilization. The remaining two compounds produced no shift (AOH1160_LV) or saturated the detector (AOH1160e), which prevented accurate Tm determination (**Fig. 2.01, Table 2.00).**
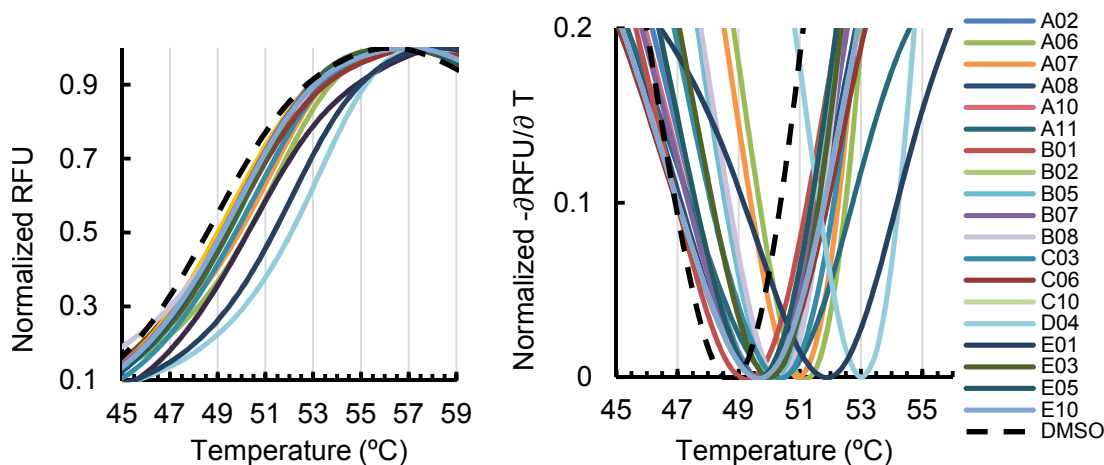
Table 2.01	AI-CADD ΔTm- Positive Shifters						
Name	ΔT_m (°C)	cLogP	cLogD	Name	ΔT_m (°C)	cLogP	cLogD
A02	0.5	2.31	0.07	B08	0.5	2.82	1.77
A06	1.0	2.60	2.02	C03	1.0	3.09	0.72
A07	1.0	2.04	2.04	C06	0.5	3.25	2.20
A08	0.5	-0.16	-1.28	C10	0.5	1.47	1.51
A10	1.5	4.26	1.90	D04	2.5	5.38	5.90
A11	1.5	1.69	-0.92	E01	2.0	4.30	3.78
B01	0.5	2.18	0.92	E03	0.5	1.39	0.69
B02	0.5	3.66	2.24	E05	0.5	2.81	1.23
B05	0.5	4.63	4.63	E10	0.5	0.95	1.07
B07	0.5	4.39	2.10				

Figure 2.02 | Atomwise AI-CADD library TSA hits. Normalized melt curve and inverse first derivative plots of PCNA in the presence of 1 mM of AI-CAD hit. Results are an average of assays preformed in triplicate. The DMSO-only control is plotted in black dashed line.

Atomwise Artificial Intelligence (AI)-CADD Molecules- A total of 63 predicted PIP-box cavity binders that were supplied at 20 mM in 100% DMSO and diluted to a final concentration of 1 mM for TSA screening. TSA detected 19 of the AI-CADD identified molecules produced a ΔTm of 0.5 °C or greater (**Fig.2.02, Table 2.01**).

Figure 2.03 | Fragment Library Screens- Top Shifters. Normalized melt curve and inverse first derivative plots of PCNA in the presence of 20 mM of fragment library subsets (A) comprehensive, (B) brominated, (C) combination brominated and covalent, and (D) diverse fragments. The DMSO-only control is plotted in black dashed line.
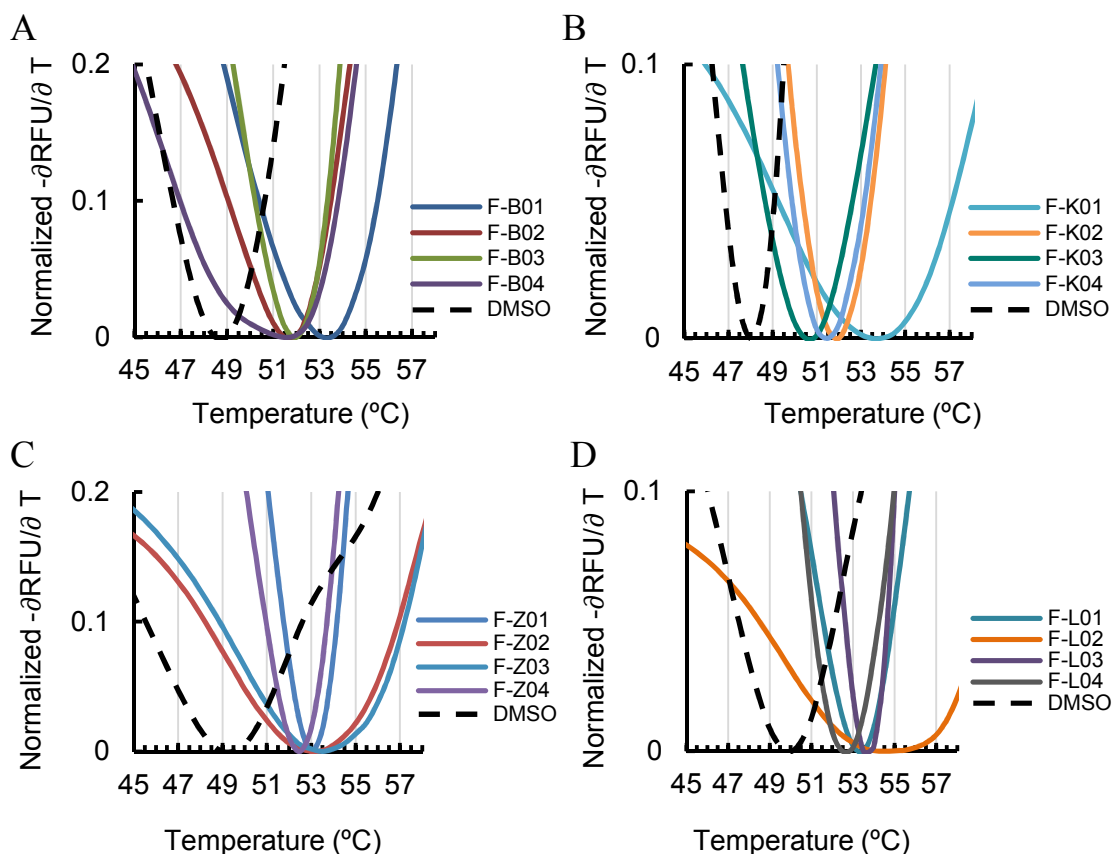
Table 2.02 | Fragment Libraries ΔT_m - Top Shifters

Library	ID	ΔT_{m} (°C)	Library	ID	ΔT_{m} (°C)
Comprehensive	F-B01	5.00	Brominated/Covalent	F-L01	3.50
Comprehensive	F-B02	3.00	Brominated/Covalent	F-L02	4.50
Comprehensive	F-B03	3.50	Brominated/Covalent	F-L03	4.00
Comprehensive	F-B04	3.00	Brominated/Covalent	F-L04	3.50
Brominated	F-K01	6.00	Diverse	F-Z01	4.50
Brominated	F-K02	4.00	Diverse	F-Z02	4.00
Brominated	F-K03	3.00	Diverse	F-Z03	4.00
Brominated	F-K04	3.50	Diverse	F-Z04	3.50

Drug-like fragment library screening- In addition to screening CADD-based small molecules, we have screened our entire fragment library comprising 3,550 drug-like fragments divided into library subsets of: comprehensive, brominated, combination brominated and covalent, and diverse libraries. These screening assays were performed at final in-well concentrations of 9 μM PCNA, 20x SYPRO® orange, and 10 mM fragment. Initial screens detected a total of 1,183 hits producing a positive shift of 0.5 °C or greater across all libraries tested (Comprehensive, 89; Brominated, 157; Brominated/covalent, 693; Diverse, 244). The top 4 most significant shifters from each library subset are plotted in **Figure 2.03A-D** with produced shifts listed in **Table 2.02**.

2.3.2 Characterizing covalent warheads in XIAP-BIR3 model system

We were provided two sets of XIAP inhibitors: previously characterized/published non-covalent inhibitors and novel agents designed to form covalent adducts with the XIAP-BIR3 domain surface residue 311, within the PPI site. These inhibitors were developed from the AVPF peptide sequence and contain either aryl-sulfonyl fluoride or aryl-fluorosulfate electrophile inserts at the "X" position (**Fig. 2.04B**). The TSA reaction parameters used to obtain an unambiguous melt curve were final in-well concentrations of 5 µM XIAP-BIR3wt/mut, 10 µM compound, 20x SYPRO® orange, and 5% DMSO. To also assess reactivity of the inhibitor covalent interaction, each compound was analyzed following two incubation parameters of 2 hours at 25 °C or 6 hours at 37 °C.

A

2h, 25 °C Incubation

6h, 37 °C Incubation

B
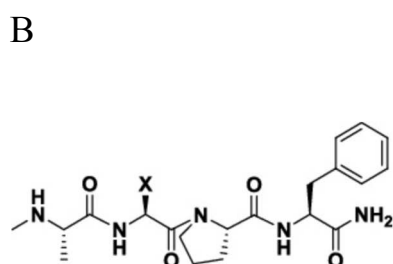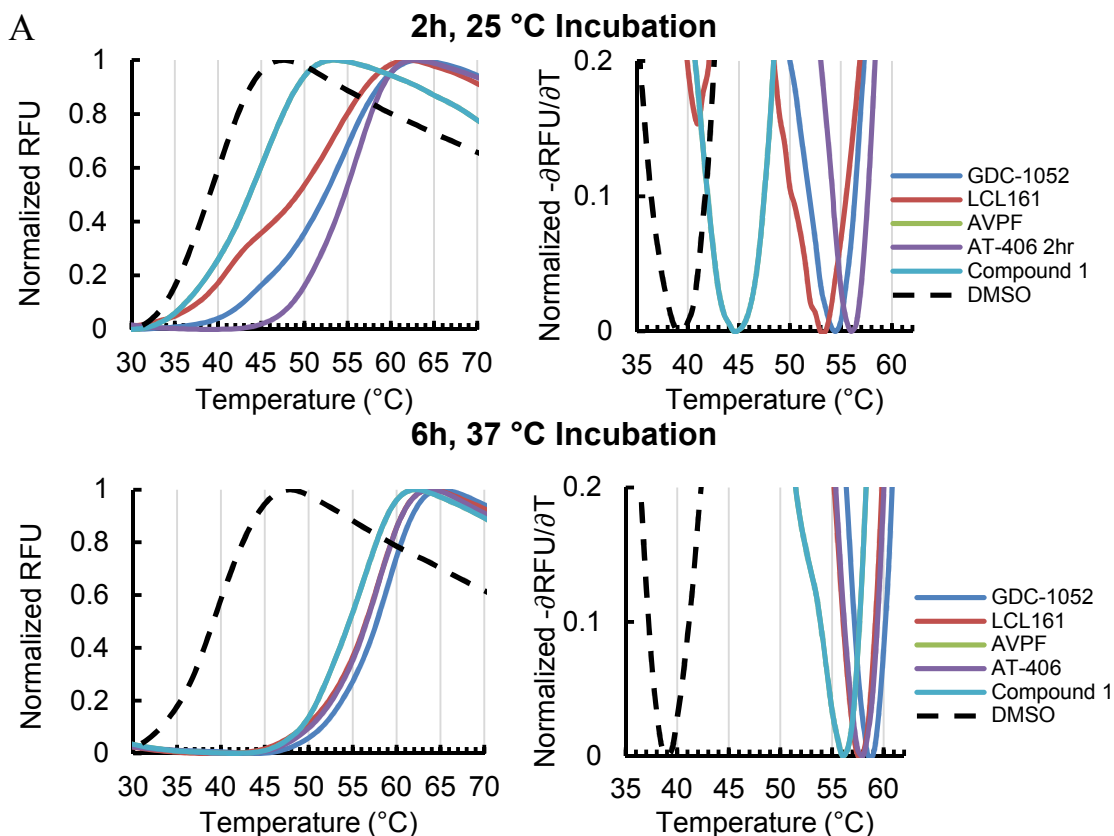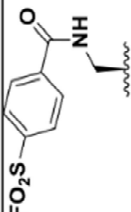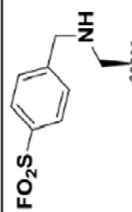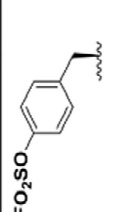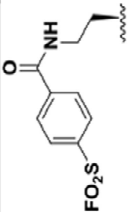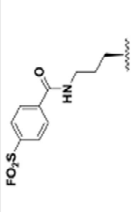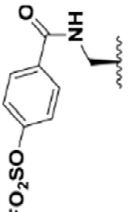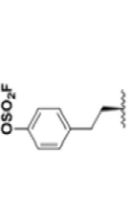
Table 2.03 | Non-covalent inhibitor ΔTm			
		BIR3wt ΔT_m (°C)	
Compound	Inhibitor/ Electrophile	2h 25 °C	6h 37□
AVPF		5.5	17.0
LCL161		14.0	18.5
GDC-0152		15.0	19.5
AT-407		17.0	19.0
1		13.5	17.0

Figure 2.04 | Non-covalent XIAP inhibitors. (A) Normalized melt curve and inverse first derivative plots of 2 h at 25 °C and 6 h at 37 °C incubations. The DMSO-only control is plotted in black dashed line. (B) AVPF peptide backbone with electrophile modification site labeled "X".

54

Covalent vs. non-covalent interaction- To establish a criterion for discriminating between non-covalent vs. covalent adduct formation from extrapolated ΔT_m values, a set of known non-covalent inhibitors were analyzed. These non-covalent inhibitors included the AVPF peptide, LCL161 (Novartis), GDC-0152 (Genentech), AT-406, and 'Compound 1' that had been previously developed by the Pellecchia research group. These non-covalent positive controls did not produce ΔT_m values greater than 20 °C following either incubation parameter, indicating the absence of covalent interaction (**Fig. 2.04A, Table 2.03**). Therefore, ΔT_m values greater than 20 °C were indicative of covalent adduct formation.

Targeting Lys, Tyr, His, Ser, and Thr for covalent adduct formation- An initial set of 7 aryl-sulfonyl fluoride and aryl-fluorosulfate warheads were tested against the wtBIR3 and the following single point mutants of BIR3: K311A, K311Y, K311H, K311T, and K311S. The extrapolated ΔT_m values are summarized in **Table 2.04** as extrapolated from their TSA plots (**Appendix A**). Based on the established criteria, compounds 2 and 3 readily formed covalent interaction with the wtBIR3 as well as Tyr and His mutants following the 2 h and 6 h incubations. Compound 7 also formed covalent interactions but only with the K311H mutant. Additional covalent adduct formation was observed for Compounds 4, 7 and 8 with either the wtBIR3, K311Y or K311H, but were less reactive and could only be detected following the 6 h incubation. For each compound incubation that produced a significant ΔT_m of greater than 20 °C, the K311A mutant comparison produced a ΔT_m less than 20 °C, thus supporting a selective covalent interaction between compound and corresponding BIR3 construct. None of the inhibitors were found to form covalent adducts with either the Ser or Thr mutants.

Table 2.04 | Covalent Inhibitors Targeting Lys, Tyr, His, Thr and Ser Residues

Compound	Agent/ Electrophile	BIR3wt		BIR3 K311A		BIR3 K311Y		BIR3 K311H		BIR3 K311T		BIR3 K311S	
		2hr RT	37°C 6hr	2hr RT	37°C 6hr	2hr RT	37°C 6hr	2hr RT	37°C 6hr	2hr RT	37°C 6hr	2hr RT	37°C 6hr
2	FO$_2$S—C(=O)NH~	31.0	37.0	1.5	8.5	37.5	34.5	35.5	28.5	4.0	6.5	5.0	11.0
3	FO$_2$S—CH$_2$NH~	35.5	35.0	5.0	7.5	36	33	31	28.5	5.0	6.0	5.0	7.5
4	FO$_2$SO—CH$_2$~	8.5	33.0	11.0	13.5	8.5	35.5	11	12.5	9.5	11.5	10.5	14.5
5	FO$_2$S—C(=O)NH~	11.5	16.0	10.5	11.5	10.5	11.5	8.5	8.5	8.5	9.5	6.5	13.5
6	FO$_2$S—C(=O)HN~	11.5	15.0	10.5	11.5	10.5	11.5	11.5	12	7.5	9.5	8.0	13.5
7	FO$_2$SO—C(=O)NH~	4.5	31.0	8.0	8.5	5.5	32	36.5	38	5.5	6.5	8.0	13.5
8	OSO$_2$F— ~	11.5	28.5	11.0	12.5	10.5	29.5	11	29.5	8.5	10	10.5	10.5

56

2h, 25 °C Incubation

6h, 37 °C Incubation
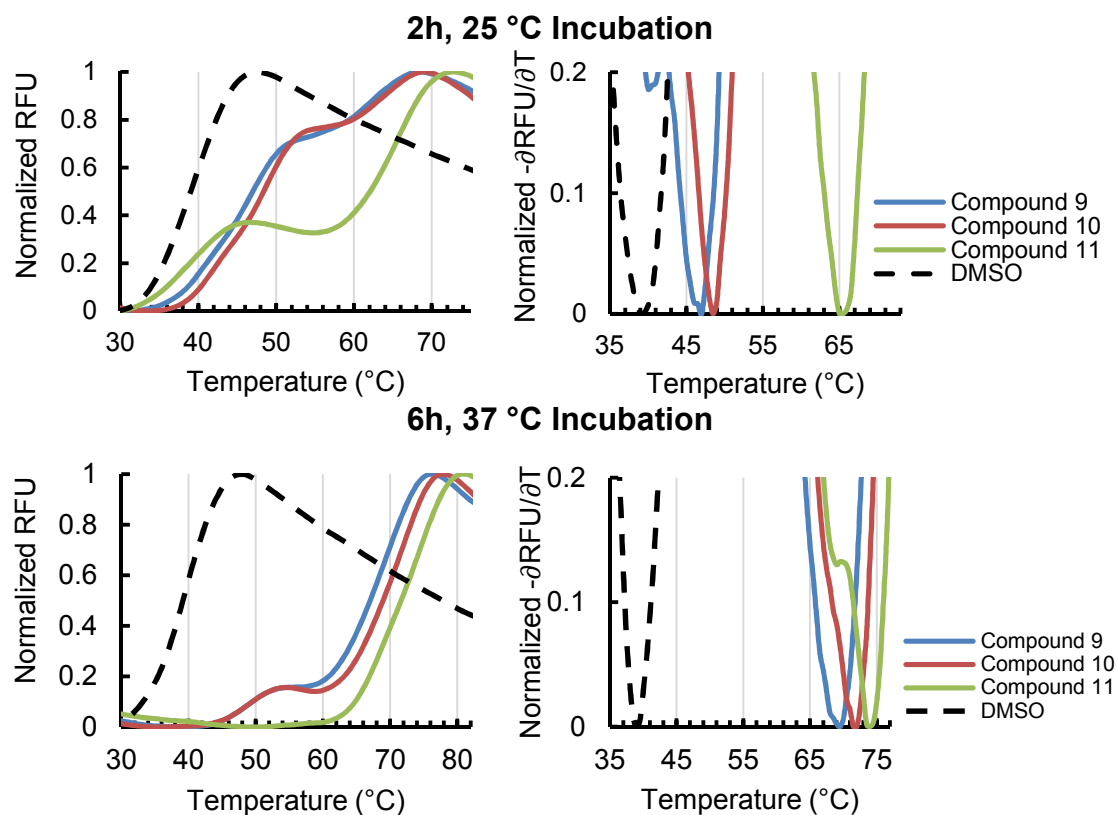
Table 2.05 | Covalent inhibitors 9, 10 and 11 ΔT_m			
Compound	Inhibitor/ Electrophile	BIR3wt	
		2h 25☐	6h 37☐
9		7.5	31.0
10		9.5	33.5
11		35	34.5

Figure 2.05 | Covalent inhibitor TSA, Compounds 9 – 11. (A) Normalized melt curve and first derivative plots following 2 h at 25 °C and 6 h at 37 °C incubations. The DMSO-only control is plotted in black dashed-line.

Aryl-sulfonyl fluorides and aryl-fluoro sulfates with improved drug-likeness. Based on the *in vitro* studies performed by our collaborators and by us, Compounds 2, 3, and 4 were determined effective at forming covalent adducts with Lys-, Tyr-, and His- residues with ΔT_m values greater than 20 °C. Based on the structures of Compounds 3 and 4 the subsequent Compounds 9, 10, and 11 were developed to have greater plasma stability, cell permeability and activity while maintaining reactivity. The TSA analysis of Compounds 9, 10, and 11 in the presence of wtBIR3 construct revealed that Compound 11 was the most reactive of these 3 inhibitors (**Fig. 2.05, Table 2.05**) with detected ΔT_m values of 35 °C and 34.5 °C following 2 h and 6 h incubations, respectfully.

2h, 25 °C Incubation

6h, 37 °C Incubation
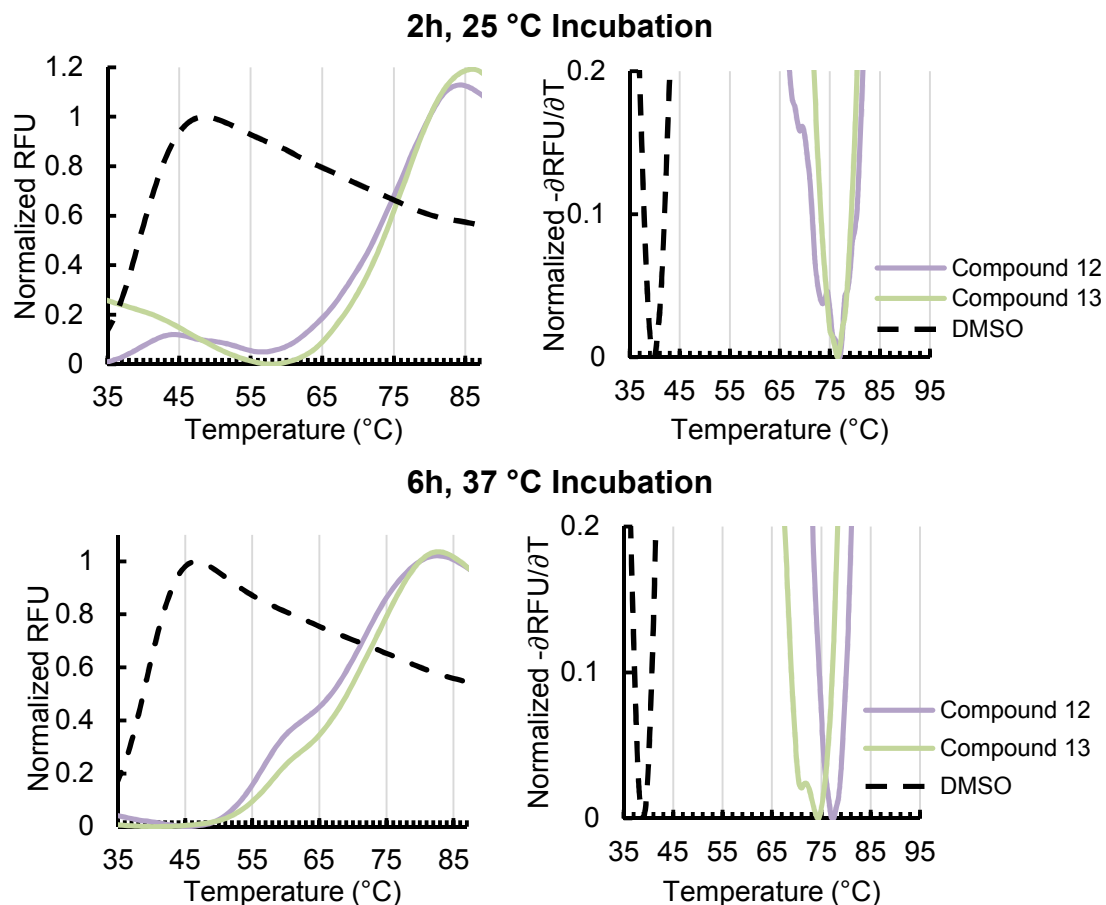
Table 2.06	Optimally reactive covalent inhibitors of IAP TSA			
Compound	Inhibitor/ Electrophile	BIR3wt ΔT_m (°C)		
		2h 25 °C	6h 37□	
12		37.5	38.0	
13		37.5	35.5	

Figure 2.06 | Covalent inhibitor TSA, Compounds 12 and 13. Normalized melt curve and first derivative plots following 2 h at 25 °C and 6 h at 37 °C incubations. The DMSO-only control is plotted in black dashed-line.

Reactive Aryl-fluoro sulfates with improved cellular efficacy- From data gathered on the structure-activity relationship (SAR) of the initial set of 10 Compounds, our collaborators followed with a second study focused on the aryl-fluorosulfate electrophile and produced Compounds 12 and 13 (**Table 2.06**). The TSA data gathered for both compounds in the presence of the *wt*BIR3 construct indicated desired reactivity with ΔT_m values of 37.5 °C following a 2 h incubation (**Fig. 2.06, Table 2.06**).

All XIAP-BIR3 TSA data have been adapted with permission from:

Gambini, L., Baggio, C., Udompholkul, P., **Jossart, J**., Salem, A. F., Perry, J. J. P., & Pellecchia, M. (2019). Covalent Inhibitors of Protein-Protein Interactions Targeting Lysine, Tyrosine, or Histidine Residues. *Journal of Medicinal Chemistry*, *62*(11), 5616–5627.https://doi.org/10.1021/acs.jmedchem.9b00561.

Baggio, C., Udompholkul, P., Gambini, L., Salem, A. F., **Jossart, J**., Perry, J. J. P., & Pellecchia, M. (2019). Aryl-fluorosulfate-based Lysine Covalent Pan-Inhibitors of Apoptosis Protein (IAP) Antagonists with Cellular Efficacy. *Journal of Medicinal Chemistry*. https://doi.org/10.1021/acs.jmedchem.9b01108

2.4 Discussion

Utilizing TSA, we have tested a total of 75 CADD identified small molecules and 3,550 drug-like fragments against PCNA. The City of Hope CADD-based molecules targeting caPCNA are hydrophobic in nature as indicated by their significant partition coefficient (cLogP) values, which predict the solubility of a molecule in an organic solvent vs. an aqueous solution (**Table 2.00**). The ideal cLogP of a drug-like molecule will fall between 2 - 3.2, although as many as 70% of drugs in pre-clinical development are poorly soluble [142]. Although hydrophobic, the 9% in-well DMSO concentration in TSA is high enough to support solubility of the AOH molecule derivatives to obtain shift values. Of the 12 analogs tested, 5 (AOH1160, AOH1996, AOH1178, DV, and LE) produced positive ΔT_m values of 0.5 °C or greater, indicating increased protein stability and binding interaction. The remaining 7 analogs either resulted in a negative ΔT_m values, no shift, or detector saturation. The 5 detected negative shifters (DE, DK, LK, PCNA3, and PCNA2) destabilized PCNA upon binding interaction (**Fig. 2.01, Table 2.00**). Although an interaction has been detected, binding-induced destabilization could potentially hinder subsequent characterization via macromolecular crystallography (MX) because this technique requires stabilization of the protein target to form a well-ordered structure. However, the binding interaction between negative shifters and PCNA could be further investigated through more detailed binding assays.

Notably, a set of more soluble derivatives DE, LE, DK, and LK were designed to ionize in an aqueous environment. These analogs contain the parent AOH1160 scaffold and either glutamate (E) or lysine (K) extensions. The increased hydrophilicity of these

analogs is demonstrated by the cLogD values (**Table 2.00**) which accounts for ionization, unlike the cLogP. The cLogD of these analogs falls within the ideal range of 0 - 3. Therefore, these analogs could be more readily utilized in structural studies that are performed in aqueous environments. Altogether, the chemistries of both positive and negative shifters will be taken into consideration to guide future analog design.

Of the 63 AI-CADD predicted PIP-box binders a total of 19 compounds produced positive ΔT_m values of 0.5 °C or greater (**Fig 2.02**). This equates to a 30% positive hit rate for the entire library of AI-CADD molecules tested. As indicated by the cLogP and cLogD values (**Table 2.01**) this pool of molecules has greater hydrophilicity, as compared to the City of Hope CADD-molecules. This increased solubility is especially advantageous for structural characterization and each of these hits has advanced to MX screens to identify a working crystallization condition to characterize inhibitor binding interaction.

Screening of our extensive library of 3,550 drug-like fragments against PCNA has revealed a substantial number of hits. Across all libraries tested a total of 1,183 fragments produced a positive ΔT_m values of 0.5 °C or greater. Due to the significant number of hits, we have applied a higher shift cut-off of 2 °C or greater, for initial focus. This equates to a total of 245 hits and a 6.9% hit rate. Based on the hit rate criteria described by Chilton and coworkers, this hit rate is very significant and is indicative of a highly druggable target. Notably, PCNA exists as a trimer in solution, providing three potential binding sites, which may account for the significant number of hits detected. The 245 hits have now progressed to titration-based TSA studies for hit validation and MX for binding interaction characterization.

In addition to screening potential binders to PCNA, TSA was also used to characterize/validate aryl-sulfonyl fluorides and aryl-fluorosulfate warheads designed to form covalent interaction with PPI site residues such as lysine, histidine, and tyrosine. Each TSA reaction was carried out following one of two protein/compound incubation parameters: 2 h at 25°C or 6 h at 37 °C. To establish a criterion for covalent vs non-covalent interaction, first a set of known non-covalent binders were tested (**Fig. 2.04**). Each of these molecules produced a ΔT_m less than 20 °C following either incubation condition (**Table 2.03**). Therefore, it was determined that a detected ΔT_m of greater than 20 °C would be indicative of covalent interaction. Next, XIAP-BIR3 mutations to the binding site residue K311 were prepared and tested against a set of 7 covalent warheads. The results of these assays are summarized in **Table 2.04**. From these studies, we detected stable covalent adduct formation between Compounds 2, 3, 4, 7, and 8 and one or more of the BIR3 mutants. Orthogonal characterization assays performed by our collaborators determined that Compounds 2, 3, and 4 were the most effective at forming covalent adducts against tyrosine, histidine, and lysine residues. Thereby elucidating chemical warheads capable of targeting binding site residues outside the typical cystine target.

Follow-up studies lead by our collaborators focused once again on targeting the binding site lysine and utilized the structures of Compounds 3 and 4 from previous studies as well as the subsequently developed Compounds 9, 10, and 11 (**Fig. 2.05, Table 2.05**) Ultimately, further development was required to produce a the structures of Compounds 3 and 4 guided the design of subsequent Compounds 9, 10, and 11 that were developed to improve upon the reactivity of the covalent adduct formation while maintaining compound

stability. Compound 11 was determined to be the most reactive and influenced the design of Compounds 12 and 13 (**Fig. 2.06, Table 2.06**). Both compounds were reactive in forming covalent adducts with the K311 residue in the PPI site of XIAP-BIR3 following both incubation conditions and produced significant ΔT_m values greater than 35 °C. Ultimately, TSA characterization of all synthesized covalent warheads helped detect the most reactive molecules against the surface lysine and elucidated reactive chemistries effective in targeting additional surface residues such as tyrosine and histidine.

2.5   Methods

2.5.1   Expression and purification of recombinant hPCNA

Expression- The hPCNA gene in pET 22b vector was transformed into the *E. coli* expression cell line (R2DE3). Small scale expression tests were performed by previous lab members and generated a glycerol stock from a previously uninduced sample. An overnight (O/N) 50 mL culture of luria broth (LB) was inoculated utilizing this glycerol stock. The O/N culture was secondarily inoculated into 6 L of LB supplemented with ampicillin and incubated at 37 °C and 225 rpm (New Brunswick Scientific Innova©43R) for approximately 1 hour until the $OD_{600nm} = 0.3$. The temperature was then lowered to 16 °C and induced with 0.4 mM isopropyl ß-D-1-thiogalactopyranoside (IPTG) at $OD_{600nm} = 0.6$ and set to shake O/N for 15 hours. The cells were harvested by centrifugation at 4 °C and 5,000 rpm for 30 minutes (Sorvall LYNX 4000 Superspeed Centrifuge with a Fiberlite^{TM} F9-6 x 1000 LEX Fixed Angle Rotor). The cell pellets were collected and stored at -80 ºC.

Purification- Pelleted cells were resuspended with lysis buffer (25 mM Tris-HCl pH 8.5, 150 mM NaCl, 10% glycerol, 1 mM β-ME) at a ratio of 3 mL buffer/ 1 g pellet. The resuspended cells were sonicated (Q-Sonica Q125, ¼ inch probe) on ice at 70% amplitude for 6 minutes, with pulsing for 10 seconds on and 10 seconds off. The lysate was then centrifuged at 4 °C and 30,000 xg for 30 minutes (Sorvall LYNX 4000 Superspeed Centrifuge with a Fiberlite^{TM} F21-8 x 50y Fixed Angle Rotor). The supernatant was carefully separated and loaded onto a 5 mL HiTrap Q FF (Cytiva) and eluted against a gradient of 1 M NaCl using an NGC Quest 10 Plus Liquid Chromatography System (Bio-Rad). Fractions confirmed by SDS-PAGE to contain hPCNA were pooled and manually

diluted with No-salt Buffer (25 mM Tris-HCl pH 8.5, 10% glycerol, 1 mM β-ME) to a final concentration of 50 mM NaCl. The diluted sample was loaded onto an ENrich™Q 10 x 100 (BioRad) for high-resolution anion exchange. The elution fractions containing hPCNA were pooled and concentrated to 5 mL and injected onto a HiLoad® 16/600 Superdex® 200 pg gel filtration column (GE Life Sciences) and eluted with SEC-200 Buffer (25 mM HEPES-NaOH pH 7.4, 100 mM NaCl, 1 mM β-ME). Samples containing purified hPCNA were concentrated to 10 mg/mL ($\varepsilon = 0.5075$) and flash frozen in liquid nitrogen to be stored at -80 ºC.

2.5.2   PCNA TSA screening to detect potential PIP-box binders

Compound Screening- To screen potential binders against hPCNA, 1 μL of compound in 100% DMSO was aliquoted into a white, 0.2 mL semi-skirted 96-well tray (AB-0900/W; Thermo Scientific) and centrifuged for 1 min at 1500 rpm in a PlateFuge MicroPlate MicroCentrifuge (Benchmark Scientific). Then 19 μL of master mix containing 9 μM hPCNA, 20x SYPRO® Orange Dye, in 1x PBS pH 7.4 was added to each well and mixed by pipetting. The 96-well plate was then sealed with optically clear Microseal® 'B' Seals (MSB1001; Bio-Rad) and loaded into a CFX Connect™ Real-Time PCR Detection System (Bio-Rad). The reaction plates were heated from 25 °C to 95 °C at increments of 0.5 °C every 30 s, taking fluorescence scans at excitation/emission ranges of 470/570 nm utilizing the FRET filter. Both the City of Hope CADD-based and Atomwise AI-CADD identified molecules were tested at a final concentration of 1 mM in triplicate format. The drug-like fragment libraries were tested at a final concentration of 10 mM and a single reaction was performed for each fragment to screen for hits. Each experiment was also repeated in the

absence of protein to test potential interaction between compound and SYPRO® Orange Dye. Compounds determined to interact and produce fluorescence were noted and removed from the pool of hits.

2.5.3   Characterizing covalent warheads in XIAP-BIR3 model system

Protein/inhibitor incubations- the XIAP-BIR3 constructs and inhibitors were pre-incubated following one of two parameters, 37 °C for 6 h or 25 °C for 2 h. The 6 h incubations were executed by Dr. Carlo Baggio from the Pellecchia research group and delivered for TSA analysis. The 2h incubations were executed in the Perry lab utilizing provided protein and inhibitor stock solutions.

Characterizing covalent inhibitors- post-incubation, of XIAP-BIR3/inhibitor complexes were prepared in a similar manner to the steps described in the previous section (**1.5.2**). Final concentrations of 5 μM BIR3, 10 μM inhibitor, and 20x SYPRO were obtained following dilution with reaction buffer (50 mM Tris pH 8.0, 150 mM NaCl, 50 μM Zinc acetate) to a final volume of 20 μL. The reaction plates were heated from 10 °C to 95 °C at increments of 0.5 °C every 30 s. Fluorescence intensity was measured utilizing the FRET filter within the excitation/emission ranges 470/570 nm.

2.6   Supplemental Information

Reaction Buffer	Tm (°C)	
50 mM sodium citrate pH 4.0, 150 mM NaCl	27.0/33.0	27.0/33.5
50 mM sodium citrate pH 4.5, 150mM NaCl	SD	SD
50 mM sodium citrate pH 5.0, 150mM NaCl	SD	SD
50 mM sodium citrate pH 5.5, 150mM NaCl	55.5	56.0
50 mM sodium citrate pH 6.0, 150mM NaCl	57.0	56.5
50 mM sodium citrate pH 6.5, 150mM NaCl	56.0	56.0
50 mM sodium citrate pH 7.0, 150mM NaCl	54.5	54.5
50 mM MES pH 5.5, 150mM NaCl	27.5/49	27.5/49
50 mM MES pH 6.0, 150mM NaCl	52	53
50 mM MES pH 6.5, 150mM NaCl	55.0	54.5
50 mM PIPES pH 6.0, 150mM NaCl	54.0	54.0
50 mM PIPES pH 6.5, 150mM NaCl	55.5	55.0
50 mM PIPES pH 7.0, 150mM NaCl	54.5	54.0
50 mM PIPES pH 7.5, 150mM NaCl	52.5	52.5
50 mM Bis-Tris pH 6.0, 150mM NaCl	52.0	51.5
50 mM Bis-Tris pH 6.5, 150mM NaCl	54.0	54.0
50 mM Bis-Tris pH 7.0, 150mM NaCl	54.5	54.0
50 mM sodium phosphate pH 6.0, 150mM NaCl	56.0	56.0
50 mM sodium phosphate pH 6.5, 150mM NaCl	55.5	55.5
50 mM sodium phosphate pH 7.0, 150mM NaCl	53.5	53.5
50 mM sodium phosphate pH 7.5, 150mM NaCl	52.0	52.0
50 mM sodium phosphate pH 8.0, 150mM NaCl	50.0	50.5
50 mM HEPES pH 7.0, 150mM NaCl	54.5	54.5
50 mM HEPES pH 7.5, 150mM NaCl	52.0	53.0
50 mM HEPES pH 8.0, 150mM NaCl	51.0	51.5
50 mM Tris pH 7.0, 150mM NaCl	54.5	55
50 mM Tris pH 7.5, 150mM NaCl	54.5	54.5
50 mM Tris pH 8.0, 150mM NaCl	53.0	52.5
50 mM Tris pH 8.5, 150mM NaCl	50.0	49.5/57.5
50 mM Tris pH 9.0, 150mM NaCl	47/55.5	47/55
50 mM CHES pH 8.5, 150mM NaCl	50.5	50.0
50 mM CHES pH 9.0, 150mM NaCl	45.5/54.5	46.5/55
50 mM CHES pH 9.5, 150mM NaCl	42/52	42/52
50 mM CHES pH 10.0, 150mM NaCl	36.5/48	37/48
1x PBS pH 7.5, 150 mM NaCl, 27 mM KCl	51.5	52.0
1x PBS pH 7.4, 137 mM NaCl, 27 mM KCl	49.0	49.0

Supplementary Table 2.00 | PCNA TSA buffer optimization screen. The reaction buffers highlighted in orange resulted in double melt curve peaks. The reaction buffers highlighted in red resulted in detector saturation. The buffer highlighted in green was utilized in PCNA TSA studies

Chapter 3
Computer-aided techniques for drug discovery and development

Chapter 3: Computer-aided techniques for drug discovery and development

3.1   Abstract

Computer-aided drug discovery (CADD) methods provide a cost-effective and rapid approach to predict binding interactions and pharmacokinetic properties of compounds to help guide the development of drug-like lead molecules. It was hypothesized that CADD-based screening studies would allow us to detect PIP-box selective binders with drug-like chemistries to analyze in subsequent *in vitro* studies. Both our City of Hope and Atomwise collaborators have performed virtual screenings of a combined 13 million drug-like small molecules against the PCNA PIP-box binding site to produce a total of 75 CADD-derived small molecule binders. We have also implemented a SAR-by-catalog approach to leverage the fragment hit chemistries detected in our initial TSA screens. From a small test subset of brominated fragment hits were have identified 150 small molecule superstructures that through docking analysis were shown to retain the original fragment:PIP-box cavity binding interaction. Here, we have presented a list of top 20 small molecule *in silico* binders that produced binding energies of $\geq$ -7.95 kcal/mol and > 30% clustering. Two of these molecules produced significant shifts > 2 °C, indicating binding interaction to PCNA, and have progressed to crystallization studies. In a separate collaborative project, CADD-based studies were utilized to investigate the pro-apoptotic mechanism of oxyresveratrol in cancer cells. Utilizing molecular docking we detected a predicted binding energy of -7.77 kcal/mol and 100% clustering of oxyresveratrol to the catalytic site of caspase-3 indicating potential inhibitory interaction that was further supported by our collaborator's studies in cancer cells that found caspase-3 was not activated following oxyresveratrol treatment.

3.2   Computer-Aided Drug Design (CADD)

The development of high-throughput screening (HTS) techniques to identify lead-like molecules for therapeutic development has reduced the amount of time to discover and design drug-like leads in pre-clinical R&D efforts [143]. This reduced time benefits academic based labs, but the synthesis of these small molecule libraries for *in vitro* screening studies can be costly. The exploration of FBDD approaches has afforded researchers the ability to assay a diverse chemical space through fragment libraries a fraction of the size (10^3) of small molecule HTS libraries (10^6) [144]. Although this approach is more economical in comparison, transitioning from fragment hit-to-lead design can be a challenging pursuit of trial and error. Computer-aided drug discovery (CADD) methods provide a cost-effective approach to identify hits of both small molecules and fragments as well as guide optimization of said hits into drug-leads [145]. CADD-based methods were first described in the 1980s and continued development of docking algorithms, increased computing power and structural characterization of disease target structures have led to an exponential increase in development of potent therapeutics *in silico*. One of the first therapeutics originating from CADD-based investigation is the potent EGFR inhibitor and chemotherapeutic, gefitinib, which was discovered via *in silico* screens in 1992 [146].

The overarching CADD approach is broken down into two classifications: virtual screening and computational design. Both classes of CADD study can be applied in structure-based and ligand-based approaches to develop therapeutics [143]. Modern computational advancements have enabled high-speed virtual screening of potential

71

binders. Programs available on the Schrodinger platform like shape screening and Glide docking as well as the AutoDock Vina suite developed at the Scripps Research Institute are capable of evaluating each compound in a matter of seconds [147-149]. Subsequent ranking of these ligands based on predicted binding affinity and interaction allows researchers to readily identify hits and thereby avoid the cost of synthesized small molecule screening libraries. More recently, virtual screening methods have been successfully applied to FBLD campaigns to guide fragment growth and identify advantageous chemistries for lead design [150].

The computational design aspect of CADD simulates and predicts the physiochemical and pharmacokinetic properties of leads as molecules are optimized for increased selectivity towards a protein target [151]. It is important to monitor these properties since approximately 40 - 60% of later stage clinical trials fail due to lack of optimized pharmacokinetic properties [152]. The physicochemical properties include, but are not limited to solubility (LogP), hydrogen bonding acceptors, donors, polar surface area and rotatable bonds. Each of the physiochemical properties will contribute to the drug's pharmacokinetics, or absorption, distribution, metabolism, and excretion (ADME) in the human body [153]. In addition to *in silico* derivation of molecular properties the CADD virtual screens provide chemical information to develop pharmacophore and structure-activity relationship (SAR) models. The pharmacophore model describes the steric and electronic features that contribute to molecular interactions. This model can be developed for either the ligand/lead molecule or the target structure. The SAR model describes the relation between a ligand's chemical structure and the observed biological activity. The

72

quantitative SAR (QSAR) model provides a mathematical relationship between the determined SAR to statistically compare the alterations in chemical characteristics of a ligand to its potency. However, unlike SAR, QSAR models can be developed independent of target structural information. The derivation of these models stems from an iterative process of computational analysis and *in vitro* characterization. Creating a model to describe ideal protein:ligand interactions can help optimize screening predictions and guide lead design [154, 155]. By first evaluating lead libraries through either structure-based or ligand-based approaches, a clear criterion is set to prioritize ligands that are more likely to form a complex over molecules that have poor interaction potential. This prevents future time and resources wasted on *in vitro and in vivo* characterization of inactive ligands [143]. Here, we will overview the two virtual screening approaches and provide examples demonstrating how *in silico* techniques can help streamline pre-clinical R&D.

Structure-based Virtual Screening (SBVS)

Also referred to as target-based virtual screening, SBVS is focused on identifying ligands that form favorable binding interaction with a target protein. Predicted ligand binding poses and their corresponding binding affinity are all computationally calculated [143]. The top-ranked ligands with the greatest calculated binding affinity will theoretically have greater activity against the target *in vitro* and *in vivo*. The application of this virtual screening method is dependent on the availability of a 3D protein structure to utilize as a target receptor. Target structures can be accessed from either online databases (RCSB-PDB) or determined through homology modeling programs [156, 157]. Additionally, small molecule libraries such as those discussed in the previous chapter, Enamine-REAL and

others (eMolecules, ZINC, etc) provide structure files for candidate ligands [109, 158]. Therefore, researchers do not need to produce target proteins or synthesize potential inhibitors to analyze interaction potential and complex formations. The most predominant technique to preform SBVS is molecular docking.

In the 1980s Kuntz and his colleagues were the first to describe molecular docking by developing a set of algorithms to generate and evaluate the binding geometries of candidate ligands against a target structure [159]. Advancements in algorithm derivation paired with increased computational power has led to the creation of a variety of molecular docking programs including the publicly available Autodock4 and commercially available Glide (Schrodinger) docking suites [148, 160]. Molecular docking accuracy is dependent on both the search algorithm (SA) and scoring function (SF) [143]. The SA generates possible ligand conformations, or binding poses, at the target binding site to detect potential interactions. Typically, SAs utilized in screening campaigns will also introduce computed ligand flexibility to the limits of defined chemical geometry restraints to prevent the generation of a chemically impossible complexes [143]. Generally, 'flexible' SAs will fall into one of four categories: shape matching, systematic, stochastic, and deterministic [161].

Shape matching is the simplest SA and is based on the shape complementarity between the generated ligand pose and the target receptor binding site. By applying three degrees of freedom each to both the translational and rotational manipulations of the ligand many potential binding poses are generated [161]. Systematic search algorithms progressively increase the degrees of freedom within ligand rotatable bonds to explore all possible binding poses. Depending on the number of rotatable bonds the number of

generated poses will exponentially increase and require more time and computational power, therefore investigators will typically cap the number generated poses to focused on a set number of initial predictions [143, 162]. The stochastic search algorithm introduces random changes in the degrees of freedom for the rotatable bonds and ligand rotation/translation in space to generate many possible conformations. Each new pose is either accepted or rejected based on its calculated probability of favorable interaction with the receptor [161]. The deterministic algorithm examines the potential movement of the preceding pose to generate the next. In order to be an acceptable conformation, the next conformation must be less than or equal to in energy to the preceding pose. Therefore, it is common that generated poses fall into a local minima, thus preventing more global exploration of potential ligand conformations [143]. A virtual simulation that utilizes energy minimization calculations is molecular dynamics. This technique enables movement of both the target protein and docked ligand atoms to simulate the binding interactions that may take place in solution. However, this method is generally not used for initial virtual screening campaigns since it requires significant computing power and time for each individual simulation [163].

The success of a molecular docking screen is dependent on the performance of the applied SF. The SF estimates the non-covalent intermolecular interaction potential between ligand and target [143, 164]. The three major applications of SF calculations are to determine the most energetically favorable ligand binding pose, predict the binding affinity between ligand and target in lead optimization, and rank ligands docked in virtual screening campaigns [165, 166]. Therefore, an accurate SF is essential to distinguish active ligands

from inactive ones. Recent studies have shown that SFs can be classified as one of the following: physics-based, empirical, knowledge-based and machine learning-based [165, 167]. The physics-based SF, force field function, directly calculates the binding energy by adding the van der Waals and electrostatic interactions between ligand and receptor. By pairing the classical force field functions with solvent models and quantum mechanics principles, the entropy-based interactions (solvation effects, covalent bonding, charge transfers, etc..) can accounted for in physics-based SF calculations [165, 168-170]. Empirical SFs estimate binding affinities through the summation of energetic contributors (hydrogen bonds, hydrophobics, steric clashes, etc…) to ligand:receptor interactions [165]. Training sets of ligand:protein complexes with known binding affinities are used to determine empirical weight constants to estimate the value of each contributing energy factor. As compared to physics-based functions empirical SFs required significantly less computing power and are frequently implemented in virtual screening campaigns [149, 171]. Knowledge-based SFs first assign ligand:receptor atom pairings to determine interaction potentials based off energy potentials extracted from known ligand:protein complexes. The calculated binding affinity is the summation of these extrapolated energy potentials [172]. Unlike these first three SFs that utilize linear regression method the final SF calculation implements non-linear machine-learning methods such as random forest, neural network, and deep-learning [165]. Machine learning SFs are implemented in one of two approaches: improve/rescore the predictions of linear SFs or 'learn' interaction properties (pharmacophore, protein/ligand geometries, QSAR, etc…) from known complexes to predict binding affinities of novel ligands [165, 173, 174]. Although, machine

learning is not typically implemented in initial virtual screening campaigns since it is a comparatively more convoluted technique than docking and requires training sets to develop more accurate predictions and significant computing power [173]. A specific example of a machine learning method will be covered in a later section. More recently, an approach referred to as consensus docking has been described in the literature as a means to validate docking results and involves utilizing two or more SFs to analyze top-ranked ligands and rule-out false positives [175, 176].

Additional docking validation methods involve evaluating the docking simulation quality prior to library screening. These methods include the root-mean square deviation (RMSD), receiver operating characteristic (ROC/AUC-ROC) and enrichment factors (EF) [177, 178]. A very popular validation method, RMSD, evaluates a docking simulation's ability to reproduce the binding conformation of a known ligand as characterized utilizing structural biology techniques [177, 179]. The RMSD threshold between published binding conformation and predicted binding pose is typically 2.0 Å, although more complex ligands will have increased thresholds to up to 2.5 Å [179, 180]. The ROC curve and area under the curve (AUC) are used in conjunction to estimate the probability of returning a known active as opposed to a decoy molecule from a docking screen [181]. The ROC curve plots the distribution of true and false outputs and the AUC-ROC is the probability of returning a true hit. The AUC-ROC values will range from 0 to 1, with 1 indicating excellent ability to distinguish active molecules from inactive ones [143, 182]. Lastly, EFs calculate the number of known actives detected as compared to the total fraction of molecules tested.

Typically, EFs are estimated for a set percentage of a total database (e.g., EF20%= number of active molecules detected from screening 20% of the database) [143, 178].

Ligand-based virtual screening (LBVS)

In the event structural information of the biological target is absent or lacking, an alternative method to identifying potential binders is LBVS. This approach is dependent on the availability of known bioactive molecules and their corresponding affinities for the target of interest. LBVS techniques operate on the 'similarity property principle', which assumes that similar chemical structures will have similar physiochemical properties and bioactivity [183]. Therefore, the chemical information derived from known bioactive molecules can be used to compare their similarity to library ligands [182]. Ligand similarity descriptors vary in dimensionality (D), where the complexity of chemical information increases with increased dimensionality (1D < 2D < 3D) [184, 185]. Generally, 1D descriptors capture physiochemical properties such as, LogP, number of atoms/bonds, and molecular weight [184]. The 2D descriptors include structure fragments and atoms connections [182, 185]. Lastly, the 3D descriptors factor in molecular volume, spatial assembly, and solvent-accessible surface area [185]. In general, these search procedures will utilize either 2D or 3D descriptors to identify chemically similar ligands. Selecting the descriptor to be used in screening is critical and dependent on the characteristics of available bioactive molecules as well as the properties deemed most vital to promote ligand interaction. The 3D descriptor approach is considered advantageous since novel scaffolds of similar shape can share biological activity while divergent in 1D and 2D properties [186, 187]. This consideration of ligand stereochemistry typically produces a larger pool of

potential binders, but also requires greater computational power and time for calculation. Therefore, by comparison, 2D approaches are considered a practical alternative for initial screening campaigns [187]. Accurate LBVS predictions require both precise similarity search method and reliable scoring calculation.

Fingerprinting-based similarity search method is used most frequently because it increases computational speed and condenses chemical information [187, 188]. This method fragments ligand structures and the resulting substructures are assigned 'fingerprints', or identifiers in the form of binary strings [185, 188]. Standard substructure analysis is classified as 2D-fingerprinting, and the addition of stereochemical information to the binary strings generates 3D-fingerprints [187]. Fingerprinting chemical characterization is not only useful in screening campaigns, but can also gather information to describe both 2D and 3D pharmacophore and QSAR models of bioactive molecules [142, 184, 189]. In addition to fingerprinting, there are countless methods utilizing thousands of 2D and 3D descriptor combinations that have been developed for LBVS [190, 191]. The most common scoring method is the Tanimoto coefficient/index, because it is a simple direct calculation to determine the ratio of similar elements vs. the total described elements [189]. This method is often paired with the fingerprinting approach to assign fingerprints with values ranging from 0 to 1 as a measure of similarity. The Tanimoto coefficient is the summation of scores and a score of 1 indicates exact identity [192]. The accuracy of Tanimoto indexing is dependent on identical ligand size for similarity scoring. The Tanimoto indexes of ligands larger than the known bioactive will score more favorably and the opposite is true for comparatively smaller ligands. Therefore, the Tanimoto

function has a potential for bias based on ligand size and this must be accounted for in LBVS campaigns [182]. The implemented LBVS experiment can also be evaluated utilizing the AUC-ROC validation approach previously described.

Fragment-based Virtual Screening (FBVS)

Typically, virtual screening techniques would not be utilized to screen low molecular weight ligands such as fragments. This is because fragments tend to bind promiscuously and as described in the SBVS section there are already challenges in accurately determining binding modes of more complex small molecule drug-leads [143, 193]. However, multiple groups have successfully implemented FBVS and CADD-based approaches in FBLD campaigns. This includes the recently developed potent inhibitors against β-lactamase, glycogen synthase kinase-3 β and MTH1 enzyme, which are all derived from fragment hits that were optimized to more complex high affinity binders via *in silico* methods [6, 194, 195]. Advantageously, fragment libraries provide a diverse sampling of the chemical space and offer good starting points to generate high affinity binders [196, 197]. However, following fragment screening campaigns, it may be difficult to determine next steps to initiate development from fragment to lead.

Studies have shown that computational analysis can be utilized to score fragments and prioritize them for subsequent lead synthesis [150]. Since fragments are low affinity and low molecular weight binders they will not score as well in docking simulations as a more complex small molecule, but size-normalized parameters such as Ligand Efficiency (LE), Fit Quality (FQ), and Binding Efficient Index (BEI) can be utilized to evaluate and prioritize fragment hits [198]. Docking screen predictions of fragments hits to monitor both

predicted fragment binding location and frequency target sites can reveal potential hotspots for binding. These hotspots are binding sites that contain residues that contribute to binding free energy [199]. Due to fragments small size and high ligand efficiency, they can anchor into sub pockets and provide starting points for lead development [103, 200]. Both applications help prioritize those fragment hits that target interaction sites for further investigation.

Advancing fragments to *de novo* synthesis involves growing, linking, or merging [201]. Information gathered from either target-based pharmacophore or SAR models can help guide researchers as they grow fragments into compounds containing chemistries to form favorable interactions with the target binding site [196]. If more than one fragment binds in adjacent pockets these can be linked to design a selective small molecule [202, 203]. The compounds produced from either growing or linking can be redocked to predict binding affinities and guide subsequent optimization. By first scoring novel leads and monitoring their calculated ADME *in silico*, researchers can avoid costly synthesis of potentially poor drug-leads [204]. A fast and cheap alternative in fragment hit to lead development is SAR by catalog. This approach takes advantage of databases of commercially available small molecules, or catalogs (ZINC, eMolecules, MCULE, etc…) to preform substructure and similarity searches based off the initial fragment hit [205]. These automated database searches identify small molecule superstructures that contain the fragment substructure, and therefore are more likely to maintain the fragment binding mode while improving binding energy [206]. Success of this approach is demonstrated in the heat shock protein 90 (HSP90) inhibitor studies led by Drysdale and Brough. The

resorcinol fragment identified from initial screening studies was subjected to a substructure catalog search of 2.7 million commercially available compounds [207]. This led to the identification of a pre-clinical hit that produced an $IC_{50} < 500$ nM against HSP90 [208]. Similarity searches, like the ones described in the LBVS section, can also be used in follow up analog generation to identify similar molecules to test via molecular docking. The combination of substructure and similarity search in SAR-by-catalog studies led to successful identification of a potent MTH1 enzyme inhibitor that produced an *in cellulo* IC_{50} of 170 nM, a nearly 500-fold improvement from the initial fragment hit (79 μM) [209]. Ultimately, SAR by catalog enables rapid development in the fragment-to-lead pipeline for subsequent high affinity analog development [196].

The advancements in CADD-based methods over the past 40 years has significantly reduced the time and cost of pre-clinical therapeutic discovery and development. Virtual platforms provide researchers the ability to screen millions of molecules in a fraction of the time of *in vitro* HTS. The discussed *in silico* approaches do not require the costly synthesis of small molecule libraries or production of the target molecule to estimate druggability or initiation of lead design. In addition, recent studies have demonstrated the successful application of CADD-based FBLD to grow fragments into larger small molecules, increasing lead:target interaction and overall binding affinity. Although there is difficulty capturing the dynamics of target and ligand interactions in virtual screening campaigns, the use of scoring functions and simulation validation calculations, readily generated in most CADD-based software helps guide researchers to design a simulation that produces more accurate predictions. Advantageously, the computational calculation of

pharmacokinetic characteristics of initial small molecule hits and developing analogs increases the likelihood of designing a more drug-like lead. Here, we present the SBVS approaches implemented in two separate collaborative efforts to identify selective, small molecule binders to the PIP-box binding site of PCNA. In addition, we have investigated FBLD through a SAR-by-catalog approach to leverage our TSA detected fragment binders. This is in hopes of identifying a fast hit-to-lead avenue to generate high-affinity small molecule PIP-box cavity binders that retain fragment anchoring interactions. Lastly, in an SBVS investigation separate from the PCNA story, molecular docking was utilized to elucidate a novel inhibitory interaction against caspase-3 enzyme that promotes cancer cell apoptosis.

3.2.1 Virtual Screening against the PCNA PIP-box binding site

City of Hope CADD-based molecules

Our collaborators in the Malkas and Hickey groups identified the novel AOH-scaffold via *in silico* virtual screening studies, as detailed in their 2018 publication to the Journal of Cancer Research [86]. Their approach was to first compile a large pool of synthetically available compounds from the Albany Molecular Research Institute and other unspecified compound databases to generate a virtual library of 6.8 million small molecules. This collection of molecules was then subjected to filters based on the criteria specified by Lipinski's rule of five. A total of 3 million compounds were deemed drug-like and proceeded to ligand-based pharmacophore screens in MOE that detected 8,000 hits. These LBVS hits then advanced to SBVS. Docking studies were performed utilizing the PCNA structure extracted from the published complex structure of PCNA bound to PIP-box

protein, FEN-1 (PDB:1UL1), in Glide (Schrodinger, LLC). The predicted binding affinities of SBVS hit molecules were validated/re-scored through energy minimization studies and confirmed 57 total hits. Hits were acquired and tested in cell toxicity studies to identify the parent AOH39 scaffold that was characterized to induce cell apoptosis in malignant cells at micromolar concentrations. The CADD-based optimization to improve AOH39 potency was performed by medicinal chemists working on the project, and led to the development of nanomolar inhibitor, AOH1160.

Atomwise Artificial Intelligence (AI)-driven CADD molecules

Through their Artificial Intelligence Molecular Screens (AIMS) award program, the Atomwise company collaborated with us to identify novel active compounds against the PIP-box binding site of PCNA. SBVS was performed against the PCNA structure extracted from the PCNA:T3 complex (PDB:3VKX) utilizing their machine learning platform, AtomNet®. The AtomNet® platform is the first deep convolution neural network to utilize target structure information to contribute to ligand binding affinity predictions. As opposed to standard LBVS and SBVS, this developed neural network combines information gathered from both the ligand and protein target structures to perform screening studies. In LBVS information is derived from known bioactives and used to intentionally identify similar ligand structures. However, this approach can also be limiting and prevent the identification of novel ligands that form similar and/or additional interactions with the target structure. Therefore, by including the detailed atomic arrangement of the target structure the AtomNet® neural network can elucidate potential interactions that are left unfulfilled by the screened compounds to enhance binding interaction predictions. A

virtual library of millions of ligands was first subjected to a series of filters to remove PAINS molecules, promiscuous binders, while retaining molecules with optimized pharmacokinetic properties [210]. Decoy analyses were used to validate the prepared simulation and achieved an average AUC score of > 0.75. The validated simulation then screened 10-million compounds against the PIP-box binding site of PCNA to detect a total of 63 predicted binders for *in vitro* study [211].

3.2.2 SAR by catalog for FBLD against the PIP-box binding site

Our initial *in vitro* TSA fragment screening studies detected an over 1,183 positive shifters producing a ΔT_m of ≥ 0.5 °C. To focus on the most promising hits we have applied a more significant shift cut-off of ≥ 2.0 °C to advance 245 fragments to initial structural characterization via crystallography. However, we also wanted to explore our fragment hits utilizing *in silico* FBLD methods. An undergraduate mentee in our research group, Amy Domae, executed extensive molecular docking studies to first prioritize hits within our brominated fragment subset and then perform SAR by catalog screens against fragment substructures with the aim to identify high affinity small molecules to bind the PIP-box binding site. Amy was able to screen and rank over 100 small molecules identified from similarity and substructure searches in the eMolecules and ZINC compound databases. These SAR by catalog efforts led to a list of top 20 predicted PIP-box binders that scored high predicted binding affinities with significant binding pose agreement (clustering). These promising leads are currently being investigated in both *in vitro* TSA and MX experiments to validate docking predictions.

85

3.2.3 Molecular docking analysis of oxyresveratrol binding to caspase-3

In addition to CADD-based studies against the PCNA protein, molecular docking studies were utilized to characterize the binding interactions of oxyresveratrol to the proapoptotic enzyme, caspase-3. Initial investigation into the anti-cancer effects of oxyresveratrol were performed by Dr. Bipin Nair's research group at Amrita University in India. Their studies observed oxyresveratrol-induced apoptosis in human melanoma, neuroblastoma, and hormone dependent breast cancer cell lines [212-214]. Treatment with oxyresveratrol resulted in increased generation of reactive oxygen species (ROS), poly(ADP-ribose) polymerase (PARP) cleavage and the down regulation of several cancer biomarkers: matrix metalloprotease enzymes (MMP) -2 and -9, epidermal growth factor receptor (EGFR), vascular endothelial growth factor (VEGF), and cyclooxygenase 2 (COX-2) enzyme [215]. Notably, across all cell lines tested our collaborators did not detect caspase-3 fragments. Caspase-3 activation results in cleavage of this enzyme to promote pro-apoptotic signaling, therefore these findings indicate that apoptosis was induced in a caspase-independent manner. Based on these observed anti-cancer characteristics the Nair group performed follow-up studies in the triple-negative breast cancer cell line, MDA-MB-231. Treatment of cells with oxyresveratrol induced cell death in a dose-dependent manner. In addition, immunofluorescence assays in primary cells and 3D tumor models visualized the nuclear translocation of apoptosis-inducing factor (AIF) in response to increased concentration of ROS. Translocation of AIF protein stimulated chromatin condensation, DNA fragmentation and caspase-independent cell death. The culmination of these findings

triggered investigation into the underlying mechanisms of oxyresveratrol-induced, caspase-independent cancer cell apoptosis [216].

3.3 Results

3.3.1 SAR by catalog for FBLD against the PIP-box binding site
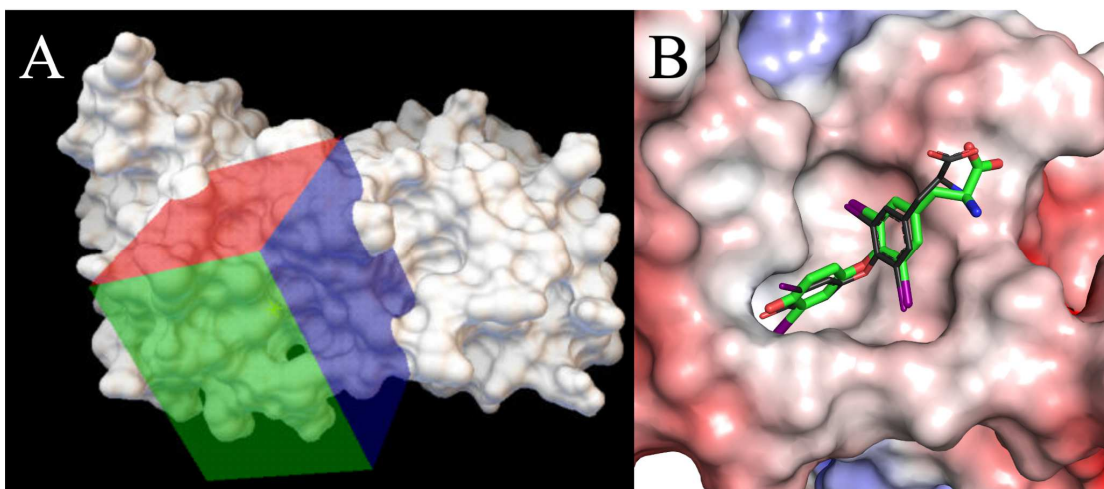
Figure 3.00 | Known bioactive docking validation. (A) Surface representation of the PCNA monomer (white) extracted from the PDB:3VKX co-crystal structure displayed in Autodock Tools, Autodock4.2 GUI. The docking site boundaries are indicated by the displayed grid box. (B) The binding site pocket is depicted with Poisson-Boltzmann electrostatic surface potentials shown in red and blue, corresponding to -5 to +5 kT/e respectively. The co-crystal structure T3 binding mode (*black*) overlayed with the re-docked T3 molecule (*green*).

Validating the PCNA docking simulation- The PCNA monomer structure extracted from the PCNA:T3 co-crystal structure was utilized for molecular docking studies in Autodock4. The AutoDock4 suite was utilized to obtain clustering information that indicates agreement between predicted binding poses. The 3D-docking boundaries are specified by the prepared grid-box (**Fig. 3.00A**) that captures the amino acid residues of the PIP-box binding pocket characterized to interact with known bioactive, T3 and PIP-motif binders. This includes the IDCL residues L126-Y133, β-hairpin amino acids 41-44, and C-terminal amino acids 250-254. To validate the docking strategy and prepared grid-box, the T3 molecule was re-

docked into the PIP-box binding pocket and compared to the co-crystal structure binding mode (**Fig. 3.00B**). The predicted binding energy of T3 was -7.55 kcal/mol with 100% clustering and suitable agreement to the crystal structure.

FBLD via SAR-by-catalog- The SAR-by-catalog approach was implemented to discover more complex, high affinity binders that retain the starting fragment binding mode. An initial pool of 58 brominated fragment hits detected to increased PCNA thermal stability by 1 °C or greater in TSA screens. All 58 fragments were docked to the PIP-box binding pocket and ranked based on predicted binding energy and clustering. Fragments with predicted binding energies less than or equal to -5.3 kcal/mol and clustered with greater than 70% pose agreement were characterized as hits. A total of 9

Figure 3.01 | FBLD SAR-by-catalog workflow. Diagram representation of SAR-by-catalog approach utilizing molecular docking to identify predicted high affinity small molecule binders.

fragment hits then advanced to automated substructure searches in both the ZINC (> 750 million) and eMolecules (26.4 million) compound databases. Four of the nine fragments

did not return any small molecule superstructures. The fifth fragment hit returned over 100 small molecule superstructures. Two of these small molecules were also subjected to similarity searches in the databases to return 11 molecules with Tanimoto similarity scores of 50% or greater. Between the remaining four fragment hits, a total of 39 superstructures were detected. A combined total of 150 small molecule superstructures were once again docked against the PIP-box binding site. The hit criteria for small molecule binding interactions were developed following the analysis of the collective docking results. Presented, are the top 20 hits that produced predicted binding energies of less than or equal to -7.95 kcal/mol and clustering greater than 30%. Four of these top 20 molecules were identified from similarity searches, while the remaining 16 were produced from substructure searches. The predicted binding energies and clustering scores for each of the top 20 hits are summarized in **Table 3.00** and depicted in **Supp. Fig. 3.00**.

Table 3.00 | SAR-by-catalog Top 20 Hits				
Small Molecule Hit	Highest Binding Energy Orientation		Highest Clustering Orientation	
	Binding Energy (kcal/mol)	Clustering	Binding Energy (kcal/mol)	Clustering
FBSM01	-9.06	73	-9.06	73
FBSM02	-9.87	7	-8.18	37
FBSM03	-8.92	13	-8.63	77
FBSM04	-8.9	17	-8.67	36
FBSM05	-8.66	94	-8.66	94
FBSM06	-8.58	60	-8.58	60
FBSM07	-8.58	29	-7.19	30
FBSM08	-8.56	48	-8.56	48
FBSM09	-8.5	85	-8.5	85
FBSM10	-8.38	89	-8.38	89
FBSM11	-8.3	6	-7.73	43
FBSM12	-8.22	88	-8.22	88
FBSM13	-8.2	76	-8.2	76
FBSM14	-8.13	66	-8.13	66
FBSM15	-8.1	56	-8.1	56
FBSM16	-8.1	31	-7.74	58
FBSM17	-8.08	10	-7.76	46
FBSM18	-8.06	76	-8.06	76
FBSM19	-7.98	50	-7.98	50
FBSM20	-7.95	99	-7.95	99

[FBSM06]	ΔT$_m$ (°C)
100 μM	**2.50**
50 μM	**1.50**
25 μM	**1.00**
12.5 μM	**0.50**
0 (DMSO)	**0**

[FBSM09]	ΔT$_m$ (°C)
100 μM	**-1.00**
50 μM	**-2.5/+3.0**
25 μM	**3.00**
12.5 μM	**1.50**
0 (DMSO)	**0**
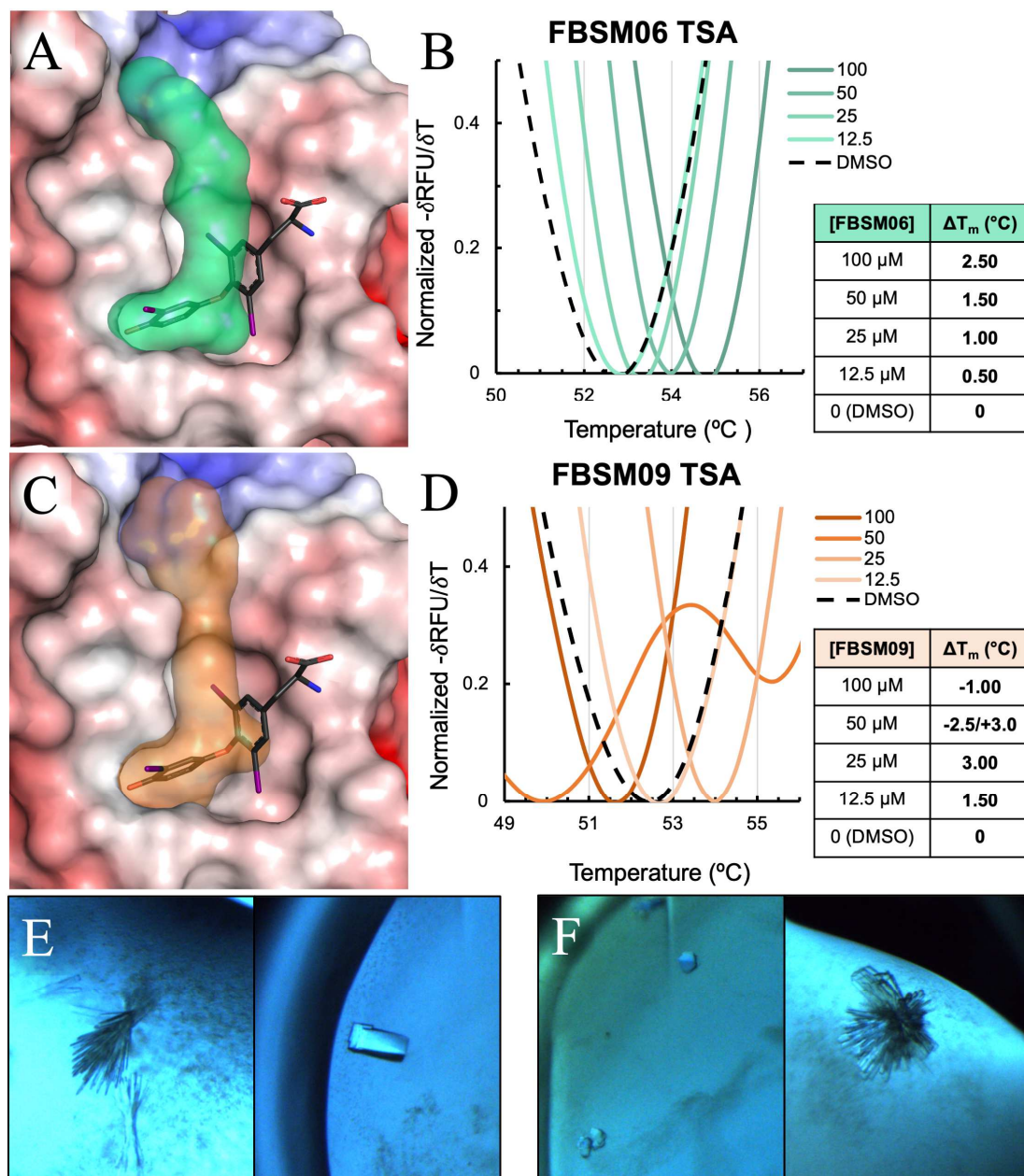
Figure 3.03 | SAR-by-catalog hit TSA and crystallography. (A) Lowest predicted binding energy and greatest clustering pose of FBSM06 (surface, green). (B) Normalized derivative plot of FBSM06 titration TSA. (C) Lowest predicted binding energy and greatest clustering pose of FBSM06 (surface, orange). (D) Normalized derivative plot of FBSM06 titration TSA. Binding pose images are overlayed with known bioactive, T3 (black). (E) Two sparse matrix hits for PCNA:FBSM06 co-crystals. (F) Two sparse matrix screening hits of PCNA:FBSM09 co-crystals. The PCNA Poisson-Boltzmann electrostatic surface potentials are shown in red and blue, corresponding to -5 to +5 kT/e respectively. The TSA plots and shifts are representative of triplicate assay results.

SAR-by-catalog hits, TSA analysis- Three SAR-by-catalog hits, FBSM06, FBSM09, and FBSM10 were readily available for purchase and acquired for *in vitro* analysis. Follow-up TSA studies screened these compounds at varying concentrations starting at 100 μM over 2-fold serial dilutions to 0.78 μM, following the assay set-up described in chapter 2. The FBSM10 molecule fluctuated between +/- 0.5 °C, for each triplicate of all concentrations assayed, and thus prevented accurate T_m extrapolation (data not shown). The two remaining molecules did induce thermal shifts starting at the 12.5 μM final compound concentrations (**Fig. 3.03B and D**). FBSM06 increased PCNA thermal stability by 0.5 °C at 12.5 μM and 2.5 °C at 100 μM final compound concentrations. The FBSM09 compound significantly increased PCNA melting temperature at 12.5, 25 and 50 μM concentrations. Notably, the addition of 25 μM FBSM09 significantly increased the thermal stability of PCNA by 3.0 °C. However, at the 50 μM two separate derivative peaks, indicating the formation of two separate PCNA complexes formed. The first peak indicates compound induced destabilization with a shift of -2.5 °C. The second peak indicates compound induced stabilization with a shift of 3.0 °C, similar to the reaction observed at the 25 μM concentration.

SAR-by-catalog hits, crystallography- All three acquired hits have proceeded to crystallization studies. We have tested both our existing PCNA crystallization conditions and performed sparse matrix screens of over 200 diverse crystallization conditions. Both positive shifters, FBSM06 and FBSM09, have produced two sparse matrix screening hits each (**Fig. 3.03E and F**) that are currently undergoing optimization.

3.3.2 Molecular docking analysis of oxyresveratrol binding to caspase-3
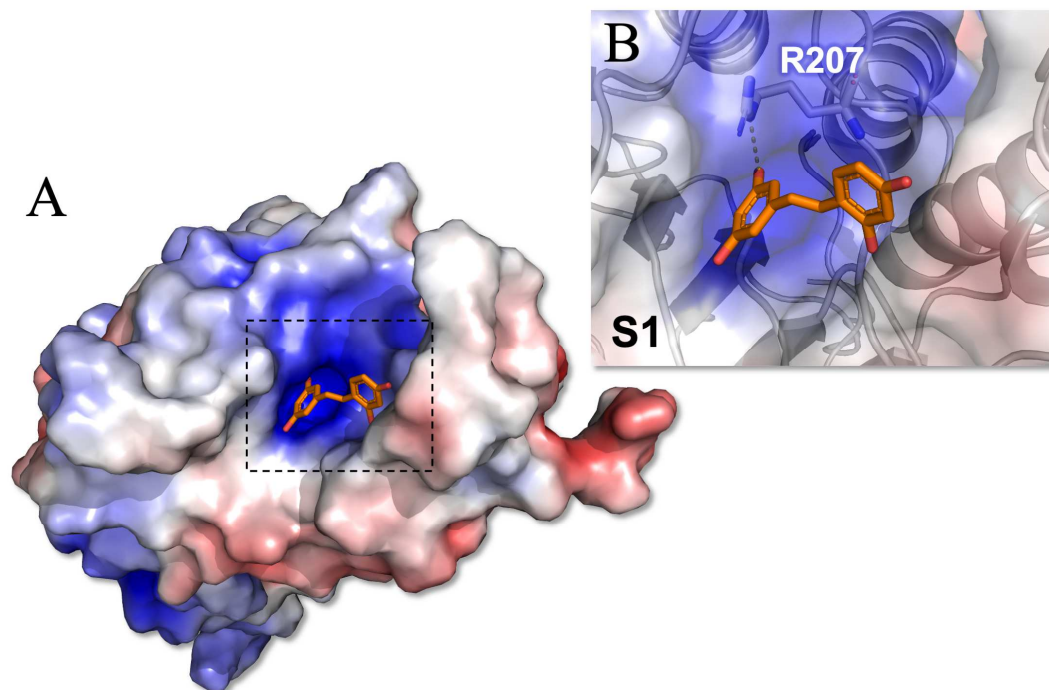
Figure 3.04 | Docking of oxyresveratrol against caspase-3. (A) The predicted highest energy, largest clustering binding pose of oxyresveratrol (orange sticks) to caspase-3. The caspase-3 Poisson-Boltzmann electrostatic surface potentials are shown in red and blue, corresponding to -5 to +5 kT/e respectively. (B) Oxyresveratrol hydrogen bonding to the R207 residue of the S1 catalytic pocket of caspase-3.

Oxyresveratrol predicted binding mode to caspase-3- To characterize potential binding interactions between oxyresveratrol and caspase-3 we first performed all-over docking against the capase-3 structure extracted from the PDB:2H5I structure in Autodock4. The largest clustering of predicted binding modes for oxyresveratrol was 44% to the caspase-3 catalytic site. The caspase-3 catalytic site is composed of surface pockets S1-4 that form the biological substrate binding groove. The 3D-docking boundaries of the grid box were then focused on the catalytic site and the predicted binding energy was -7.99 kcal/mol with 100% clustering in the S1 pocket (**Fig. 3.04A**). Oxyresveratrol also forms a hydrogen bonding interaction with the R207 residue of this pocket (**Fig. 3.04B**).

3.4 Discussion

3.4.1 FBLD via SAR-by-catalog with PCNA fragment hits

We have utilized molecular docking to leverage the fragment substructure hits determined to form binding interaction with PCNA from our TSA screens of over 3,550 drug-like fragments. To test our designed SAR-by-catalog approach, our undergraduate mentee was assigned a pool of brominated fragment hits selected due to the significant electron density of bromine atoms. The bromine electron density signal can be readily identified in the electron density maps generated from crystallization studies and allow us to easily detect bound compound in the crystal structure. The shift cut-off of $\geq 1°C$ was applied to result in a more practical pool of hits, 58 total, for initial investigation. Our initial SAR-by-catalog studies implemented both substructure and similarity search against a combined virtual database of over 775 million small molecules, to produce a total of 150 small molecule hits. Based on docking analysis these small molecule superstructures have retained the anchoring interaction formed by the originating fragment hit. The criteria applied to generate our pool of 150 small molecule hits specified a predicted binding energy greater than -7.5 kcal/mol, which is generally considered a hit in AutoDock4 molecular docking screens. In addition, the binding modes of each of these hits produced notable agreement across the 100 GA runs (> 30%). From this significant pool of hit molecules we have presented a list of the top 20 molecules that produced the combined lowest predicted binding energy and greatest clustering (**Table 3.00**). Three of these top 20 SAR-by-catalog hits were readily available for purchase and acquired for *in vitro* analysis. One of these hits, FBSM06, produced a predicted binding energy of -8.58 kcal/mol and 60% clustering

(**Table 3.00**). The titration-based TSA started at an in-well compound concentration of 100 μM that was subjected to 2-fold serial dilutions to 0.78 μM. We observed increasing thermal stability of PCNA with increasing compound concentration starting at 12.5 μM FBSM06, and the most significant shift was registered at 2.5 °C at 100 μM (**Fig. 3.03B**). The FBSM09 hit produced a predicted binding energy of -8.5 kcal/mol and 85% clustering (**Table 3.00**). Similar to the FBSM06 TSA titration screen, effects to the melting temperature of PCNA were not observed until a final FBSM09 concentration of 12.5 μM (+ 1.5 °C). The melting temperature doubled at 25 and 50 μM FBSM09 concentrations to produce significant ΔT_m values of 3.0 °C. However, at the 50 μM FBSM09 concentration a second negative shift peak was also observed indicating that this molecule may also form a destabilizing interaction with PCNA at higher concentrations. This is further supported by the negative shift of -1.0 °C produced at the 100 μM concentration of FBSM09 (**Fig. 3.03D**). The FBSM10 hit did not produce a consistent T_m value, which prevented accurate TSA shift extrapolation. This could be attributed to the hydrophobicity of this compound (cLogP = 3.7) as compared to the other two more hydrophilic molecules, FBSM06 and FBSM09, which have cLogP values of 2.5 and 2.6 respectively. Regardless of solubility, all three molecules are undergoing crystallization studies utilizing both published PCNA crystallization conditions and co-crystallization sparse matrix screens. Of the three hits, the two TSA detected positive shifters have produced co-crystal structures in two individual sparse matrix crystallization conditions each (**Fig. 3.03E and F**). Our TSA results have further validated the predicted *in silico* binding interactions of the hydrophilic small molecule hits that have been acquired thus far. The remaining compounds are being

investigated for potential synthesis to perform the discussed *in vitro* analysis and all hits are being analyzed by our collaborating medicinal chemists to continue to develop PIP-box cavity selective analogs.

3.4.2   Molecular docking analysis of oxyresveratrol binding to caspase-3

This study was initiated by observed oxyresveratrol-induced cancer cell apoptosis that occurred in a caspase-3 independent manner. All-over docking of oxyresveratrol against the entire caspase-3 structure indicated that 44% of the randomly generated poses directed to the catalytic site. Characteristic of this catalytic site is the S1-4 pockets, which compose the interaction groove. We then focused our docking simulations to the catalytic site and determined that 100% of the 100 GA runs bound the S1 pocket to form a hydrogen bonding interaction with the R207 residue lining this pocket (**Fig. 3.04**). The predicted binding energy of this pose was -7.77 kcal/mol, a value within hit range for AutoDock simulations. Therefore, molecular docking studies suggest that oxyresveratrol has direct inhibitory action against the catalytic site of caspase-3 and prevents activation of this enzyme The inhibition of caspase-3 activation by oxyresveratrol is further supported by MTT assays of MDA-MB-231 cells following pre-treatment with pan-caspase inhibitor, QVD-OPH, that did not impact oxyresveratrol-induced cell death. In addition, immunoblot analysis confirmed the absence of caspase-3 fragmentation that is characteristic of enzyme activation, thus concluding that caspase-3 had not been activated in response to oxyresveratrol treatment. The findings of this study indicate that the anti-cancer effects of oxyresveratrol include caspase-3 inhibition to promote AIF pro-apoptotic signaling. Advantageously, the oxyresveratrol compound presented reduced toxicity in non-

95

malignant cells and favorable pharmacokinetics to provide a promising template for therapeutic development. This work has been included in our 2020 collaborative publication to the journal of Biochemical Pharmacology [216].

3.5   Methods

3.5.1   FBLD via SAR-by-catalog with PCNA fragment hits

Molecular docking- The structure of PCNA bound to T3 (PDB:3VKX) was prepared in Pymol (Schrodinger) to remove all heteroatoms and obtain the isolated peptide atomic coordinates. Both the fragment hits and catalog small molecules were built in Marvin Sketch and prepared to produce 3D coordinates. AutoDockTools was used to generate docking input files for both the ligands and PCNA receptor. A grid box of 61 x 53 x 48 on a 0.375-Å spacing was used to analyze the PCNA PIP-box binding site and produce grid parameters for Autdodock4.2. Both the fragments and small molecule ligands were docked using a Lamarckian genetic algorithm, with a population of 100 randomly positioned poses. The maximum energy evaluation was restricted to 2.5 x 10^7, and mutation rotation was limited to 0.02 with a crossover rate of 0.8. Clusters were generated at 2.0 Å RMSD.

3.5.2   Molecular docking analysis of oxyresveratrol binding to caspase-3

Molecular docking- The co-crystal structure of caspase-3 bound to known inhibitor Ac-DEVD-Cho (PDB:2H5I) was prepared in Pymol (Schrodinger) to remove all heteroatoms. The oxyresveratrol structure data file was obtained from PubChem (CID 658108) [217]. Both the ligand and receptor docking input files were generated using AutoDockTools. Initially, an all-over docking was performed with a grid box of 126 x 126 x 126 with 0.46

Å spacing to encapsulate the entire caspase-3 structure. Subsequent docking simulations centered on the catalytic site with a grid box of 46 x 46 x46 on a 0.35 Å grid spacing. Oxyresveratrol docking simulations were performed utilizing a Lamarckian genetic algorithm with a population of 100 randomly positioned poses. The maximum energy limitation was set to 2.5 x 10^7 with a mutation rotation of 0.02 and crossover rate of 0.8. Clustering outputs were generated at 2.0 Å RMSD.

3.6   Supplemental Information
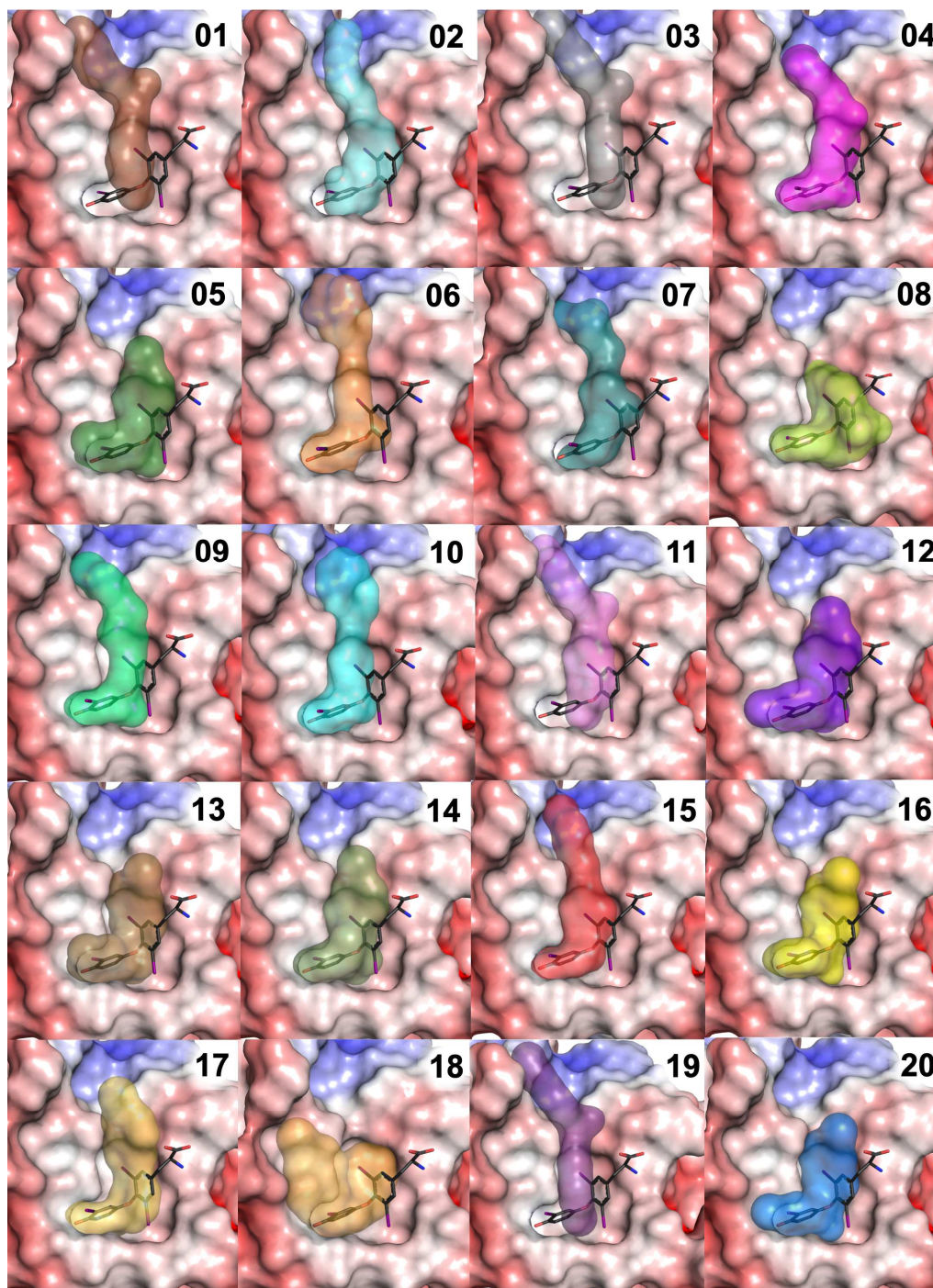
Supplementary Figure 3.00 | Top 20 SAR-by-catalog hits. (A) Highest predicted binding the SAR-by-catalog hits in surface representation of varying colors overlayed with the co-crystal binding orientation of known bioactive, T3 (black). The PCNA Poisson-Boltzmann electrostatic surface potentials are shown in red and blue, corresponding to -5 to +5 kT/e respectively.

Chapter 4
Utilizing macromolecular crystallography in SBDD

Chapter 4: Utilizing macromolecular crystallography in SBDD

4.1   Abstract

Macromolecular crystallography (MX) enables atomic resolution structural characterization to identify the detailed binding interactions between target protein and bound ligand molecule to guide SBDD campaigns. We hypothesized that MX could be utilized to characterize the binding interactions of our *in silico* and *in vitro* screens to identify PIP-box selective binders. Thus far, we have successfully resolved the co-crystal structures of 3 of our AOH1160 analogs and 1 AI-CADD predicted binders to atomic resolutions. The reported AOH1160 analog structures are the first to elucidate the novel binding interactions of the AOH1160 scaffold and reveal a similar binding mode for the metabolically stable AOH1996 analog that has now entered Phase-1 clinical trials. The high-resolution structure of the AI-CADD identified B05 molecule bound to the PIP-box cavity of PCAN has revealed a novel PIP-box selective scaffold that has initiated an alternative approach for the development of caPCNA selective inhibitors. Our extensive crystallization screening studies that have tested over 1,000 diverse crystallization conditions have also led to the identification of two novel *apo*-PCNA crystallization conditions. These new crystallization conditions will allow us to characterize our TSA fragment hits, to identify advantageous chemistries for future analog design in our SBDD campaign to generate caPCNA PIP-box selective anti-cancer therapeutics with optimized ADME properties.

4.2   Structural characterization via macromolecular crystallography

Structure-based drug design (SBDD) provides detailed structural information for the identification of key interactions to help guide the design of selective therapeutics that will promote or inhibit the activation of downstream pathways [218, 219]. A variety of techniques can be utilized to elucidate protein structures and subsequently characterize protein:drug-lead complexes to drive SBDD campaigns. The most common methods of 3D-structure characterization are nuclear magnetic resonance (NMR) and macromolecular crystallography (MX), with cryo-electron microscopy (Cryo-EM) becoming more popular in recent years [218]. The applicability of these methods is dependent on the properties of the target protein and each has their own advantages and disadvantages. For example, both NMR and Cryo-EM are effective methods to identify dynamic structural interactions such as: ligand association/dissociation, alternative binding modes and characterizing flexible regions of a protein target [220, 221]. However, both methods present potential disadvantages when applied to an SBDD campaign. For example, structural characterization via NMR is limited to a protein size of <30 kDa, requires large amounts of protein sample and the multi-dimensional NMR spectra analysis can be challenging to deconvolute [222]. The cryo-EM method is useful in elucidating large complex assemblies of 200 kDa or greater but is generally limited by low signal-to-noise ratios. As a result, cryo-EM is limited by lower resolution data and difficulty in accurate structure determination for smaller (<50 – 70 kDa) or asymmetric protein targets [221]. Alternatively, structural characterization via MX analysis provides atomic resolution data for both small and large protein complexes with favorable signal-to-noise ratios [223].

Unlike NMR and cryo-EM, structural characterization via MX is dependent on the production of a stable protein crystal. The first protein crystals were observed nearly 200 years ago and it wasn't until the 1930's when Bernal and Crowfoot (Hodgkin) reported the first X-ray diffraction image produced from a pepsin crystal [224, 225]. Significant developments in the field of structural biology over the next 30 years provided the foundation for the first structures solved by MX in 1960 with the characterization of hemoglobin and myoglobin by Perutz and Kendrew [226, 227]. Today, modernized MX has become an indispensable tool for the development of selective therapeutics via SBDD by providing a detailed model of protein:ligand interactions to guide lead optimization in the development of selective therapeutics. Here, we will discuss MX theory, application and highlight how MX has been applied to our own efforts in inhibit vital caPCNA PPIs.

Obtaining a Protein Crystal

First, sufficient amounts of purified protein must be obtained, (0.5 – 10.0 mg) for MX experiments [228]. This involves protein engineering, which is focused on designing a DNA construct that will express soluble target protein. The recombinant protein expressed from a heterologous expression system is then subjected to purification to obtain homologous target protein sample. Ideally, the target protein will be isolated to $\geq$ 95% purity to limit intrinsic heterogeneity that will prevent crystallization [229]. The successful production of purified protein sample is a significant hurdle in MX-based structural characterization. However, producing a solid protein crystal from purified sample is considered the rate-limiting step of the entire process. Protein crystals are formed through the establishment of protein nuclei that then associate with surrounding protein molecules

into an ordered array. This is driven by supersaturation, which means that the number of protein molecules in solution exceeds the molecule's solubility limit [230]. The challenge of MX is determining the ideal conditions to reach supersaturation without precipitating the protein sample and instead induce crystal formation. The spontaneous formation of a protein crystal can be described by Gibb's law of free energy, $\Delta G = \Delta H - T\Delta S$. The disordered array of excess protein molecules has significant entropy (ΔS) and few contacts between protein molecules, or low enthalpy (ΔH) at a certain temperature (T) to result in a negative ΔG. Equilibrium is re-established through a simultaneous decrease in ΔS and increase in ΔH that occurs with the formation of crystal contacts or interaction between the protein molecules [231]. A well-ordered array of interacting protein molecules will come out of solution in the form of a solid protein crystal [232, 233].

Sitting-drop **Hanging-drop** **Micro-batch**

$$[\text{precipitant}]_{\text{drop}} = [\text{precipitant}]_{\text{reservoir}} / 2$$
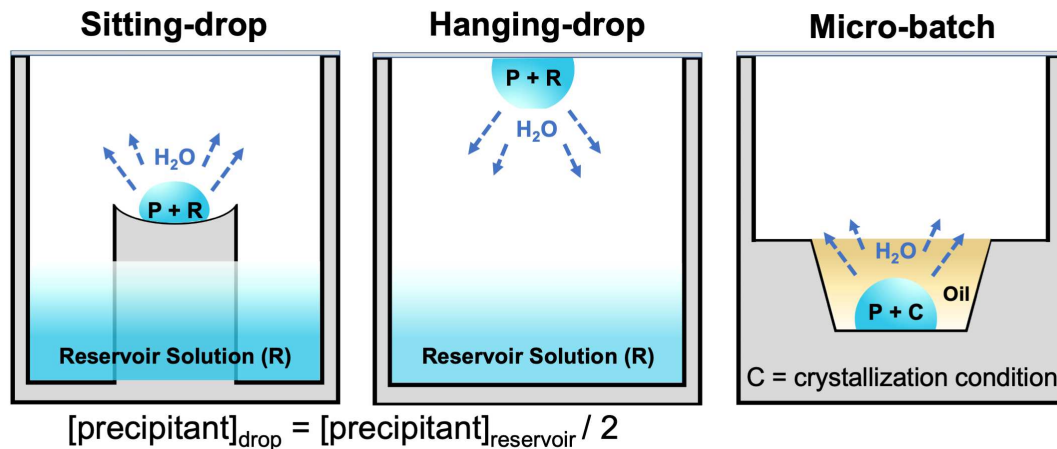
Figure 4.00 | Vapor diffusion methods schematics. (*Left)* The half protein, half reservoir drop sits in a pedestal above a reservoir solution. (*Center*) The drop mixture is suspended on a coverslip above the reservoir solution. The concentration of precipitant in the reservoir is 2x greater than in the droplet. Water molecules will leave the drop via vapor diffusion until equilibrium is achieved. (*Right*) The half protein half, crystallization condition droplet is covered by a layer of oil (i.e. paraffin) to slow evaporation until supersaturation is achieved. All systems are sealed within either sealing film or coverslip.

The most common technique to induce crystallization is vapor diffusion, which can be performed in sitting-drop, hanging-drop, or micro-batch formats (**Fig. 4.00**) [234]. In principle, a droplet of purified protein is mixed with a precipitating solution, or crystallization condition. This mixture is allowed to either equilibrate against a larger reservoir of crystallization condition (sitting- and hanging-drop) or covered with a thin layer of oil to control evaporation (micro-batch) within a closed system. In the two proceeding techniques the concentration of crystallization condition is much lower in the drop than in the reservoir and as the solutions equilibrate water vapor will leave the droplet to increase the concentration of both the precipitant and the protein sample. Ultimately, vapor diffusion allows for controlled supersaturation of the protein sample [232, 235, 236].

Structural biologists will induce spontaneous crystal formation by altering the components of the reaction environment (i.e., protein concentration, crystallization condition and temperature) to influence the rate of vapor diffusion and promote interaction between protein molecules ($\uparrow\Delta H, \downarrow\Delta S$) to facilitate crystal growth in the metastable zone (**Fig. 4.01**) [231, 235]. This is done through crystallization matrix screens that determine the perfect combination of reaction variables that promote crystal growth following

Figure 4.01 | Protein crystallization phase diagram. The diagram is separated into regions based on protein solubility. The relationship between protein and precipitant concentrations will dictate localization on the phase diagram. The bottom left is undersaturated, soluble protein. The three remaining regions to the right of the solubility line indicate states of supersaturation. In the metastable zone protein crystals will grow. In the nucleation zone protein nuclei are established and may grow into crystals. The top right indicates the range that protein will precipitate from the solution as aggregate.

supersaturation. The crystallization condition is typically composed of buffering agent, salts and additional polymers or ions. The buffering agent will determine the pH of the droplet and directly affect the ionization state of the surface amino acid residues of the protein sample to promote interaction between the protein molecules. The salt will generally act as the precipitant and its concentration will increase or decrease the ionic strength and directly influence the rate of vapor diffusion. Bridging ions or polymers added

105

to the condition can promote interaction between the protein molecules or with the solvent components and stimulate precipitation [232, 237-239]. The addition of organic solvents, cross-linking agents and ligands can also affect protein solubility and intrinsic interactions [239, 240]. The development of commercially available sparse matrix crystallization kits, which contain thousands of diverse conditions, have allowed researchers to test a large sampling of the crystallization space to identify ideal conditions to produce diffraction quality crystals or indicate starting points for further optimization [241].

If all attempts at crystallization fail, it is necessary to reanalyze the quality of the protein sample. The slightest contaminant could prevent ordered crystal packing, thus requiring further purification. Alternatively, the purified sample itself may contain disordered or flexible regions that result in varying protein assemblies that also contribute to increase sample heterogeneity [242]. The post-purification technique of reductive methylation of surface lysine residues has been shown to reduce surface residue entropy and rescue protein molecule packing interactions within the crystal lattice [243]. Non-structural proteins or proteins containing overtly flexible regions may require truncation at the construct design phase or via *in situ* proteolysis to digest disordered regions [244].

In the event crystals do not form or form poorly packed structures, a valuable technique to promote nucleation and ordered crystal growth is seeding. Seeding can be performed in two formats, micro- and macro-seeding. Micro-seeding is a technique in which previously formed crystals are crushed in solution to form microscopic crystal fragments. These crystal fragments are then introduced to a fresh protein drop to provide a controlled nucleation point for protein molecules in solution to localize to and begin forming a crystal

lattice [245]. Similarly, macro-seeding also provides an artificial nucleation point to facilitate crystal growth, but instead utilizes small crystals of 5 – 50 μm in diameter [245]. Supplementing the process of *de novo* nucleation increases the likelihood of crystal growth in the metastable zone by attracting more protein molecules to a single site for intermolecular interaction and subsequent growth. Seeding has also been implemented to resolve crystal twinning, or the growth of more than one crystal from a single nucleation site and high mosaicity that results from uneven lattice packing.

In 2004 Ireton and Stoddard introduced micro matrix seeding, the addition of seeds to a matrix of crystallization conditions that differed from the originating condition. Three years later, D'Arcy and collogues paired this approach with commercially available matrix screening kits to develop random micro-seed matrix screening (rMMS), a high throughput structural characterization technique that has revolutionized MX structural characterization by increasing the rate of successful crystal formation [246, 247]. This alternative seeding strategy has been executed utilizing seeds derived from both the target protein and of homologous protein samples, which is referred to as cross-seeding [247]. Cross-seeded rMMS studies by Novartis have proved successful in increasing the number of crystal hits by promoting new crystal packing orientations, the crystallization of previously un-crystallizable targets and overall improved diffraction quality [248].

Following the production of a well-ordered protein crystal, the sample will then be mounted on a cryogenic loop and flash-frozen in liquid nitrogen. Flash-freezing helps protect the crystal during data collection studies in which prolonged exposure to X-ray radiation may result in sample decay and increased disorder within the crystal lattice [249,

250]. Crystals can be flash-frozen in their mother liquor, the droplet solution that remains following crystallization, or transferred into a cryoprotectant mixture prior to freezing. The cryoprotectant is a viscous solution (i.e. glycerol, xylitol, high MW PEGs, etc…) that is added to the crystallization condition to coat the crystal and replace the aqueous solvent within the channels between protein molecules. Upon freezing, an amorphous glass forms and reinforces the crystal lattice against radiation-induced decay [249, 250].

X-ray Diffraction and Structure Determination

The advancements of modern crystallography also include increased accessibility to X-ray sources such as synchrotron facilities that provide state of the art detection technologies and crystal handling methods paired with strong beam intensities to produce quality diffraction data [251]. The synchrotron storage ring houses electrons that are accelerated to extremely high speeds. When surrounding magnets force the electrons to change direction energy is emitted [252]. The emitted photons are concentrated and released to select beamlines placed around the ring structure. At the beamline, the protein crystal is mounted below a constant stream of liquid nitrogen gas on an X-ray goniometer that will position and rotate the sample in the path of the X-ray beam [252, 253].

The X-rays shot at a protein crystal will collide with the electrons of atoms within the crystal structure and scatter or diffract [254]. The diffracted X-rays will strike a detector on the opposite side of the crystal and the detector will record the X-ray beam strikes as a pattern of individual reflections known as a diffraction pattern [225, 255]. The diffraction of X-ray waves follows Bragg's law in which the angle of incident and diffracted beams will form the same angle as compared to the crystal lattice plane [254, 256]. During data

collection, the crystal is rotated to collect the reflections produced in all directions from each atom of the protein molecule [253]. The diffracted beams that are in phase will reinforce and increase the intensity of each reflection; a phenomenon referred to as constructive interference. Alternatively, beams that are out of phase will cancel one another out in destructive interference [254, 257]. Therefore, the intensity of the reflections depends on the three-dimensional, periodic arrangement of atoms within the repeating unit cell of the crystal [257, 258]. The unit cell, of set dimensions (a, b and c) and set angles between them (α, β and γ), is the translationally repeated unit that makes up the three-dimensional crystal lattice [257, 258]. The unit cell contains the asymmetric unit of the protein molecule. The asymmetric unit is the smallest portion of the crystal that undergoes rotation and translation symmetry operations to generate a single unit cell [258-260].

Real Space

Reciprocal Space

Unit Cell

X-ray beam

Protein
Crystal

**Fourier
Transform**

Diffraction Pattern

Inverse Fourier

$$p(xyz) = \frac{1}{V} \sum_{-\infty}^{\infty}_{hkl} |F(hkl)| \cdot e^{-2\pi i[hx+ky+lz-\phi(hkl)]}$$

Amplitudes

Phases

Real Space
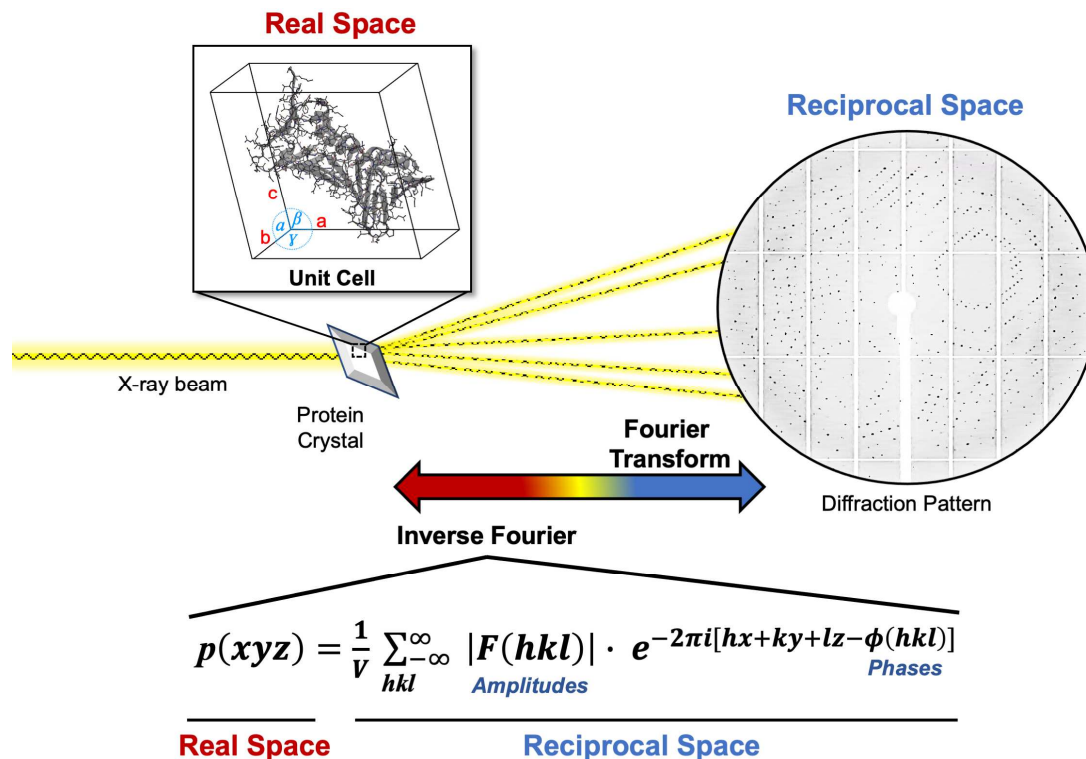
Reciprocal Space

Figure 4.02 | Diagram of Real vs. Reciprocal Space. This diagram demonstrates the Fourier series relationship between atoms of the macromolecule structure within the unit cell (real space) and the diffracted rays in the reciprocal space. The real space electron density distributions can be extrapolated from the diffraction pattern by applying the displayed inverse Fourier transform.

The produced diffraction pattern can be related to a magnified image of a small object. The pattern of reflections is a sum of the waves diffracted from the electron density within the unit cell and is therefore a Fourier transform of the electron density distribution. This is a representation of the electron density in the reciprocal space (*hkl*) and as a lens refocuses a magnified image the diffraction pattern must be refocused to determine the electron density distribution in real space (*xyz*) [261]. Refocusing is done mathematically by applying an inverse Fourier transform to generate the electron density maps necessary to resolve the target protein structure (**Fig. 4.02**) [260]. During data collection studies the intensity (amplitude) and the position of each diffracted wave is measured, but the relative

position in respect to all other waves (phase) cannot be determined from the measured amplitudes [253, 262]. Phasing methods must be implemented to obtain the phasing information and compute electron density maps. These methods include, but are not limited to, anomalous scattering (SAD and MAD) and molecular replacement (MR) [262, 263].

Currently the most common methods to solve novel macromolecular crystal structures is based on the signal produced from anomalous scatterers. Anomalous scatterers are heavy atoms that cause significant X-ray scattering. Typical anomalous scatterers include selenium atoms that can replace natural sulfur or heavy metal atoms [264, 265]. Multiple wavelength anomalous diffraction (MAD) takes advantage of tunable X-ray beams to collect multiple datasets at varying wavelengths around the absorption edge of the heavy atom [266, 267]. The produced anomalous diffraction pattern is then analyzed to determine the positioning of the anomalous scatterer in the protein structure and provides a reference to determine the location of the remaining atoms within the protein structure and calculate phase angles [264, 266, 267]. Single-wavelength anomalous diffraction (SAD) is identical to MAD, but only uses a single set of data from a single wavelength which greatly reduces crystal decay from radiation exposure [264]. However, obtained SAD datasets require further modification to remove phase ambiguities. Solvent flattening is a common modification to adjust the electron density to subtract the signal produced from the low-density solvent region between protein molecules [268, 269].

Alternatively, the MR method is dependent on the existence of a previously solved protein structure that is identical or homologous to the target structure from which the diffraction data has been collected [157]. In addition to phasing information for the

111

computation of electron density maps the selected model also provides structure information such as atom type, atomic coordinates and occupancy [223]. Protein structure databases (PDB) store a vast library of resolved protein structures and their corresponding atomic coordinates that can be utilized in MR structure solutions [157]. Once the model has been selected both rotation and translation functions are applied to identify the maximum likelihood positioning within the unit cell of the target crystal structure [270]. This method requires significantly less time and preparation as compared to anomalous scattering since there is no need to prepare heavy atom protein derivatives or obtain and interpret multiple diffraction datasets. However, working from an existing model to model the target crystal structure introduces model bias [271]. Therefore, selecting an ideal model with greater completeness and low atomic coordinate deviation (RMSD) from the target structure is key to obtaining a more accurate MR solution [270].

A variety of collection statistics are calculated to determine the quality of the diffraction dataset. This information is valuable to not only assess data quality, but also estimate the high-resolution limits in more detail than what is obtained from visual inspection of the diffraction patterns. These parameters include: completeness, mean $I/\sigma(I)$, R-merge and additional reflection analyses (R-symm, R-meas and R-pim) [272]. The mean $I/\sigma(I)$ is the average intensity (I) of the reflection divided by its error ($\sigma(I)$). The intensity of a reflection correlates to the amount of signal detected and its error considers the random noise that can create signal interference. Typically, the high-resolution limit of the dataset is established at least 2x signal-to-noise ratio, or a mean $I/\sigma(I) \geq 2$ [273]. The significant signal is key, but also there should be sufficient completeness, or number of reflections at

this resolution limit [274]. If the quality of the collected dataset is sufficient, then phasing is performed, and the target structure electron density maps are computed. The topographical electron density maps are a time and space averaged representation of the electron cloud surrounding each atom within the unit cell [223]. The map clarity is dependent both on the phasing quality and diffraction resolution limit. Occurrences of increased mosaicity (packing misalignments), structural flexibility and overall structural heterogeneity will result in lower resolution diffraction data [259]. A rough guide to recognizable structural characteristics based on resolution range is provided in **Table 4.00** The two common types of electron density maps generated are the difference map ($F_o - F_c$ map) and the double difference map ($2F_o - F_c$ map). The crystal structure factor (F_{hkl}) represents the phase and amplitude of the diffracted wave as defined in the Fourier transform equation and therefore directly relates to the positioning of atoms in the unit cell [260]. The difference maps are calculated by subtracting the model-derived structure factor amplitudes (F_c)

Table 4.00 | Resolution vs. Map Quality	
Resolution Range	**Electron Density Map Features**
≥ 4.0 Å	Can identify the location secondary structure elements, but their exact orientation is ambiguous.
$3.0 - 4.0$ Å	Clear identification of secondary structure elements and some side chains, but side chains are most likely mismodeled.
$2.5 - 3.0$ Å	Surface loops, long side chain (K, N, E, R, D and Q) and some short side chain amino acids may be mismodeled.
$2.0 - 2.5$ Å	Considerably less mismodeled side chains. Water molecules and small ligands become visible.
$1.5 - 2.0$ Å	Few errors in modeling even in surface loops.
< 1.5 Å	Almost no errors. At resolutions below 1.0 Å individual atoms can be resolved and hydrogens are visible.

from the observed (F_o) to indicate where the model and experimental data differ [272]. Positive $F_o - F_c$ densities indicate missing atoms from the current model and negative densities indicate misplaced atoms. Areas of the model that agree with the experimental

113

data will have no F_o - F_c density. With each additional properly modeled atom the phasing estimates improve and more accurate electron density maps are generated in an iterative process [260]. The $2F_o$-F_c maps weigh the observed data more heavily as compared to the model-derived structure factor amplitudes [272]. Areas of the model that agree with the experimental data will be covered in positive density. Misplaced atoms or more dynamic regions of the protein molecule will lack $2F_o$-F_c density. Therefore, accurate placement of atoms into previously positive F_o-F_c density will convert to $2F_o$-F_c density following a refinement cycle. Electron density maps can be contoured based on sigma (σ) level in reference to noise to enable more precise atom placement [260, 272].

The initial protein model will require multiple rounds of iterative structure refinement to accurately represent the experimental data. Structure refinement refers to the process of editing the atomic coordinates of the atoms within the protein molecule. Each atom is assigned coordinates (*xyz*), displacement factors (B-factors) and occupancy values [260]. The B-factor refers to the how much an atom oscillates around a specific position in the model; larger B-factors indicate greater oscillation and less confidence in the atom positioning. Well-ordered atoms within the crystal lattice will have B-factors of 5 - 20 Å^2 [275]. Typically, atom occupancy is 100% unless the atom is part of a flexible side chain where there is more than one conformational state thereby decreasing occupancy in a single location [260].

The previously discussed model specifications are utilized in validation calculations executed with each iterative cycle of model refinement. These are the refinement statistics and include *R*-factors, Ramachandran scores, RMS Deviations (bonds

and angles), and clash scores [260]. The comparison of calculated R-factors (R_{work} vs. R_{free}) is the most important validation method used in structure refinement. The R_{work} value is a measure of agreement between the crystal model and the experimental data [276]. With each refinement cycle the coordinates, B-factors, and occupancies are minimized based on applied restraints. With each refinement cycle the R_{work} value is minimized and could potentially result in the over-fitting of atomic coordinates [276]. To avoid over-fitting during refinement, and biased model generation, the R_{free} value is referenced as a validation parameter. The R_{free} value is similar to the R_{work} measurement, but instead is calculated using 5 -10% of the data that are excluded from the minimization process. The difference between R_{work} and R_{free} values will be no more than 5% for an agreeable model [276, 277]. The RMS(bonds) and RMS(angles), Ramachandran and clash scores indicate how well the model obeys calculated ideal geometry restraints [260, 278]. After the protein model has been refined, additional densities from solvent molecules or ligands can be interpreted.

At higher resolutions the solvent molecules that become trapped in the channels between the protein molecules become visible in the electron density maps. The solvent water molecules form water networks that contribute to the packing of the protein molecules [279]. These "ordered" waters are modeled as single oxygen atoms and their placement can be validated by referencing their calculated B-factors (ideally < 80) [280]. Other crystal condition components such as the salt ions used to induce supersaturation can also become trapped within the crystal structure and will contribute positive F_o-F_c densities [272]. The addition of ligand atoms to the model will be discussed in the next section.

Crystallography in SBDD

The use of MX in SBDD is dependent on the production of a crystallized complex of protein target bound to ligand. In the previous section, we discussed how to obtain a well-ordered *apo*-crystal structure and resolve its corresponding structure model. The two common methods to generate a macromolecular complex crystal structure are crystal soaking and co-crystallization [281-283]. Typically, the soaking approach incubates pre-formed *apo*-protein crystals with a solubilized ligand. A network of solvent channels surrounds the loosely packed protein molecules of the crystal structure. These solvent channels make up 27-65% of the crystal composition, are approximately 20-100 Å in diameter and contain both crystallization condition and ordered waters [284-286]. Solvent channels provide access to ligands to diffuse into the crystal lattice to form interactions. Therefore, soaking can be utilized in rapid screens of compound libraries to produce complex crystal structures for binding characterization. However, soaking is only effective if the inhibitor binding interaction is compatible with the established crystal lattice [287].

As previously discussed, the integrity of a well-ordered crystal structure is dependent on the intramolecular contacts formed between the protein molecules. Traditionally, ligand binding can induce conformational changes in the protein target. If significant binding-induced conformational changes occur the crystal lattice interactions will be compromised and negatively impact X-ray diffraction studies [282]. In extreme cases, visible cracks in the crystal will form or even complete dissolution will occur. A variation of soaking that may avoid significant conformational changes requires a preformed protein:ligand complex crystal structure. In this approach, the soaked ligand

must outcompete the bound ligand for interaction with the target protein [288]. Since the protein complex exists in a conformational state that accounts for ligand occupancy, it is expected that minimal conformational changes will be required to interact with the newly introduced ligand [282]. Factors effecting successful ligand displacement include relative binding affinities, soaking ligand solubility and soaking period [283]. The soaking approach is very common in the characterization of fragment binding. Small fragment molecules (150 – 250 Da) readily diffuse through solvent channels to binding sites [105]. Studies have shown that the preparation of "fragment cocktails", or mixtures of up to 10 fragments enable the binding characterization of multiple candidates [289]. This is advantageous because MX can be used as a high-throughput method to identify fragment hits for fragment-based lead design (FBLD) studies [105].

However, soaking may produce crystal artifacts since the protein structure is trapped within the crystal lattice. A useful alternative is co-crystallization, which pre-incubates both the protein and ligand in solution prior to crystallization to allow for ligand induced conformational changes prior to crystallization [282]. However, if significant conformational changes take place, the newly generated complex may not pack similarly to the *apo-protein* structure or any previously resolved complex structure. Therefore, *apo-*crystallization condition or co-crystal conditions with varying ligand scaffolds are not guaranteed to generate novel complex co-crystals [282, 283]. The high throughput sparse matrix screening described in the previous section can be utilized to identify co-crystallization conditions when screened in the presence of ligand. As with *apo*-protein crystallization, it is necessary to obtain a homogeneous sample for co-crystallization.

Attempting to co-crystallize a mixed sample of bound and unbound protein molecules may prevent crystal lattice packing or result in reduced ligand occupancy within the crystal structure [282, 283]. Upon addition of an unbound ligand to the target protein an equilibrium is established between the unbound (free protein and ligand) and bound protein:ligand complex. Saturating the reaction environment with free ligand will force this equilibrium to the bound state and ensure maximum ligand-protein occupancy. Generally, the ligand concentration should not be less than 10 times its binding affinity (K_d) [282]. In the event binding affinity information is not available, the excess ligand concentration can be estimated based off molar ratio or number of binding sites. For both soaking and co-crystallization experiments, typical small molecule concentrations range from 0.1 – 1.0 mM and fragment concentrations range from 20 – 50 mM [223].

Advancements in modern MX techniques, such as high-throughput sparse matrix screening to identify high-resolution crystallization conditions and a variety of optimization techniques such as seeding, rMMS and crystal cross-linking provide clear routes for producing well-ordered crystal structures [235]. Although not discussed in detail here, the development of automated crystallization robots to perform both sparse matrix and rMMS screenings can significantly reduce the time taken to obtain protein crystals [290, 291]. With the use of methods such as soaking or co-crystallization, the ligand(s) of interest are introduced to the protein structure to elucidate binding interactions. Advantageously, MX can be used to characterize weak affinity (mM) fragments for fragment screening via X-ray crystallography to initiate FBLD campaigns [105]. Increased accessibility to state-of-the-art synchrotron radiation sources have reduced data collection

times from hours to minutes while providing high quality diffraction datasets. As the rate-limiting factor in MX-driven SBDD, once crystals have been produced, protein:ligand complexes can be resolved and will provide valuable information to establish SAR characterization and guide medicinal chemists in iterative drug design campaigns.

4.2   Structural Characterization of Novel PIP-box Binders

Working towards our long-term goal of developing selective therapeutics against the PIP-box binding site of the caPCNA isoform, we will utilize MX-driven SBDD techniques to first characterize our pool of binders detected in both our *in vitro* and *in silico* studies discussed in the previous chapters. Here, we present the detailed structural information of four resolved co-crystal structures of the PCNA protein bound to novel PIP-box interacting molecules. Three of the four structures are co-crystal complexes of PCNA, bound City of Hope CADD derived small molecules, AOH1160LE, AOH1160DE and the recently synthesized AOH1996LE. Utilizing MX techniques of micro-seeding and co-crystallization we have been able to develop a working crystallization condition to obtain diffraction quality co-crystals to resolutions of 2.85 Å, 3.69 Å and 3.75 Å, respectively. The detailed binding interactions of the City of Hope CADD-derived molecules had not been previously structurally characterized and now the combined structural information has aided in the development of a SAR criteria for the AOH scaffold. In addition, MX studies of our TSA hits have produced a high-resolution (1.90 Å) co-crystal structure of PCNA bound to the AI-CADD identified B05 molecule. The novel B05 molecule has not been previously identified as an inhibitor of PCNA activity. This molecule has provided valuable information for the subsequent design of AOH analogs and presents a potential

alternative approach for therapeutic development. Lastly, we have resolved two *apo*-PCNA crystal structures to 2.44 Å and 2.01 Å, respectively. The first structure was developed from a previously published condition and the second was identified from sparse matrix screens. Both conditions will be used in soaking studies to characterize remaining TSA hits for SBDD and FBLD studies.

4.3   Results

4.3.1   Co-crystallization of City of Hope CADD-based Molecules
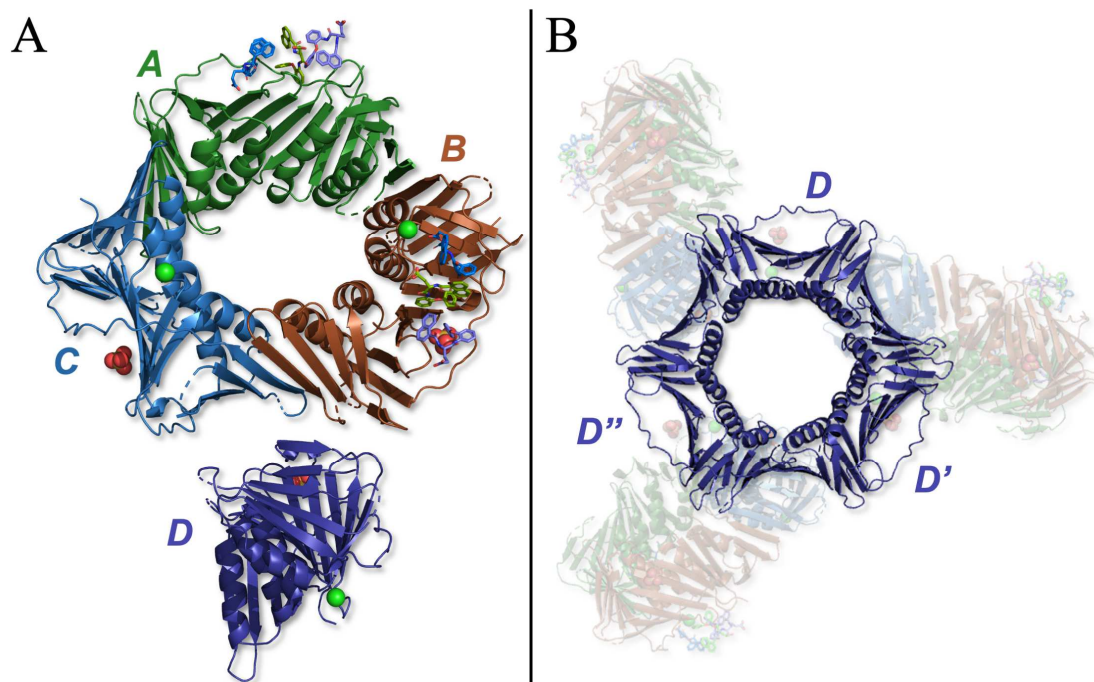
Figure 4.03 | PCNA/AOH1160LE asymmetric unit. (A) Cartoon representation of the asymmetric unit with individual subunit IDs: A (*emerald*), B (*brown*), C (*azure*), and D (*purple*). (B) D-ring biological homotrimeric assembly formed between symmetry mates.

Co-crystallization COH CADD-based molecule, AOH1160LE- The AOH1160LE derivative was soluble at 4 mM in aqueous buffer containing 10% DMSO and incubated with overnight with 9 mg/mL PCNA. Hanging-drop vapor diffusion produced PCNA:AOH1160LE co-crystals, which shared nucleation sites and were utilized to prepare micro-seeds. Seeded co-crystals diffracted to 2.85 Å and were solved via MR using the PBD:3VKX structure as an input model. The final structure was refined with an R_{work} of 19.75% and R_{free} of 25.41% (**Table 4.01**). Four monomers of PCNA were detected in the asymmetric unit. Subunits A, B and C, form the biological homotrimeric assembly (**Fig. 4.03A**). While the D subunit, with adjacent units form their own homotrimer (**Fig. 4.03B**).
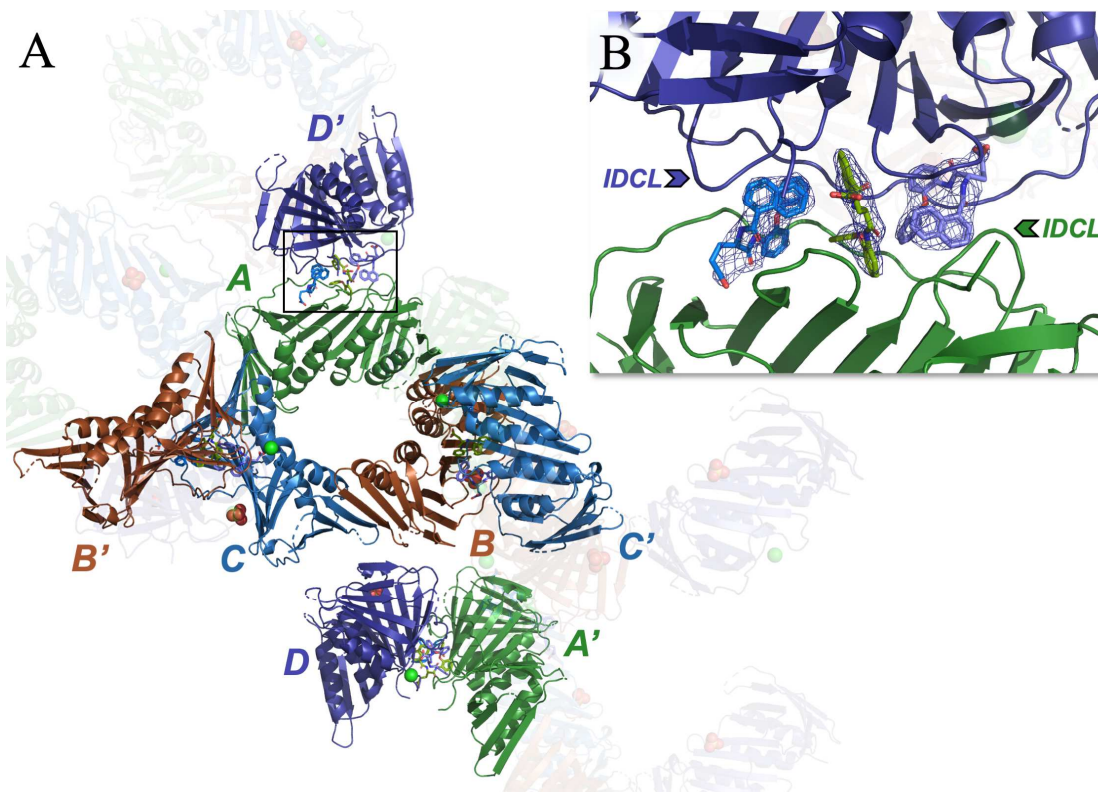
Figure 4.04 | PCNA/AOH1160LE stacking interface. (A) Cartoon representation of the stacking interactions between PCNA/AOH1160LE symmetry mates within the unit cell. Symmetry mate subunits are labeled with an apostrophe ('). (B) The $2F_o$-F_c maps of 3 modeled AOH1160LE molecules at the subunit interface are contoured to 2 σ. The modeled molecules are shown in stick representation and colored based on their positioning in the binding pocket (violet, green and blue).

Packing within the unit cell stacks each subunit against a symmetry related subunit of another ring and follows the pattern: A to D', B to C', C to B' and D to A' (**Fig. 4.04A**). The stacking interaction between each pair of subunits occurs at the PIP-box binding site with IDCL sequences oriented in opposing directions. Observed within the electron density maps are three AOH1160LE molecules that bind in and adjacent to the PIP-box cavity. The bound AOH1160LE molecules bridge the IDCL adjacent pockets of both stacked PCNA subunits (**Fig. 4.04B**). Through symmetry operations, the AOH1160LE molecules are present in the binding sites of each PCNA subunit (**Fig. 4.04B**).
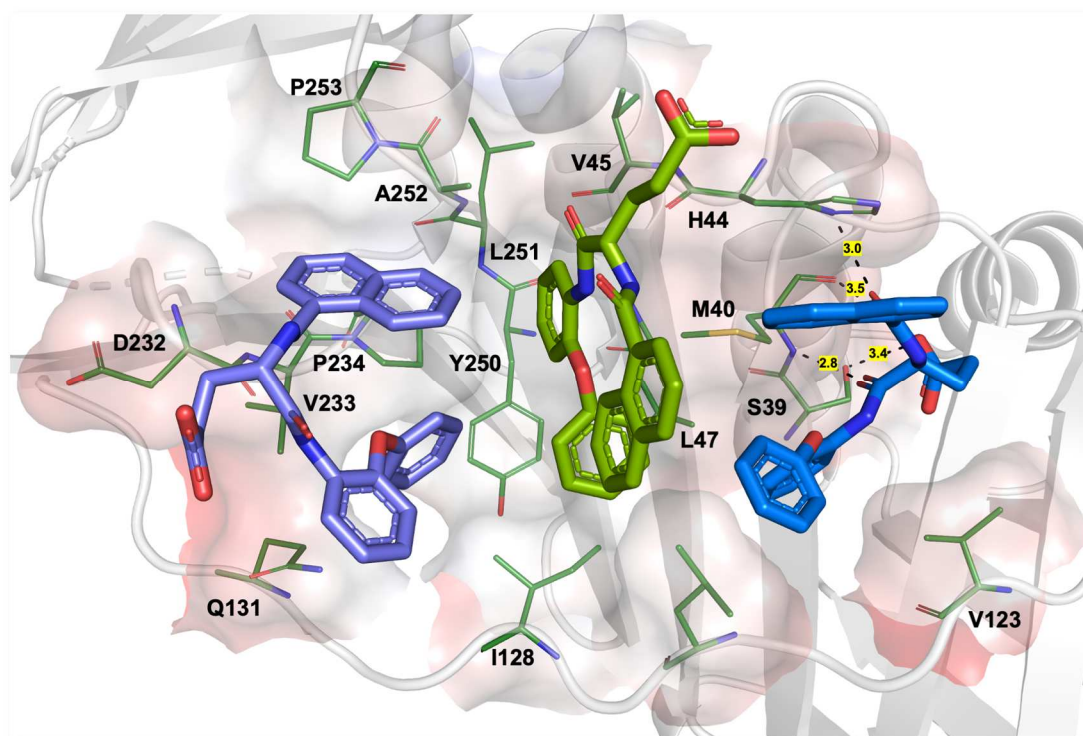
Figure 4.05 | AOH116LE binding site interactions. Cartoon representation of PCNA monomer (*white*) with binding site residues in sticks (*emerald*). The Poisson-Boltzmann electrostatic surface potentials are shown in red and blue, corresponding to -5 to +5 kT/e respectively. Three AOH1160LE molecules are shown as sticks (*violet-left, green-center and blue-right*). Hydrogen bonds are displayed with their corresponding distances (*black dashes*).

The central AOH1160LE molecule (*green*) of the asymmetric unit binds nearly perpendicular to the PIP-box cavity via hydrophobic interactions. The first ring of the di-phenyl ester moiety sits down into a hydrophobic region formed by Y250 and L251. The second ring binds into another hydrophobic pocket formed by H44, V45, L47 and IDCL residues L126 and I128 (**Fig. 4.05**). Symmetry stacking interactions of the subunits, A' and B' to C and D, orient the central molecule with naphthalene moiety bound into the PIP-box cavity of the C and D subunits (**Fig. 4.04A; Supp. Fig. 4.00**). The left-hand molecule (*violet*) binds primarily via its naphthalene moiety to form hydrophobic interactions with residues V233, P234, A252 and P253. In addition, the second ring of its di-phenyl ester

moiety also forms hydrophobic interactions with P234 (**Fig. 4.05**). The third molecule (*blue*) forms interactions with both a PIP-box cavity-adjacent pocket of the asymmetric subunit and extends into the PIP-box cavity of the stacked symmetry mate. Upon further analysis, it was determined that the right-hand molecule binds the stacked symmetry monomer in the same manner that the left-hand molecule binds the asymmetric monomer. In the asymmetric unit, the right-hand molecule via its linker region and glutamate side chain hydrogen bonds with the S39, M40 and H44 residues (**Fig. 4.05**). Binding interactions of the left and right-hand molecules are directly related; the left-hand molecule interactions with the asymmetric monomer are the right-hand molecule interactions with the stacked symmetry mate monomer and vice versa.
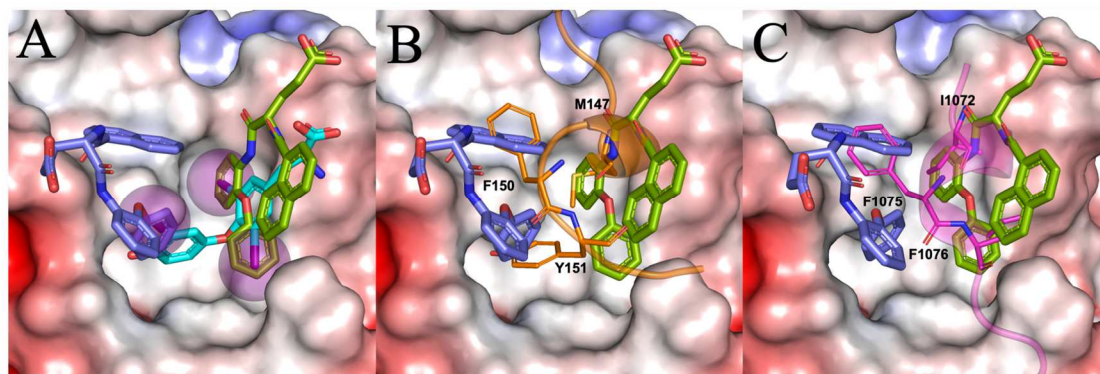
Figure 4.06 | Comparison to known PIP-box binders. Superimpositions of the PCNA:AOH1160LE left (*violet*) and center (*green*) binding modes with (A) T3 (*cyan*), (B) the p21-derived PIP-box peptide (*orange*), and (C) the APIM-derived peptide (*magenta*). The binding site pocket is depicted with Poisson-Boltzmann electrostatic surface potentials shown in red and blue, corresponding to -5 to +5 kT/e respectively. The iodine atoms of T3 are shown as spheres (*purple*). Each of the peptides is shown in cartoon representation with the PIP-box pocket interacting residues shown as sticks.

The PIP-box cavity interactions of the center AOH1160LE molecule is identical to the known small molecule PIP-box inhibitor, T3. The aromatic rings of the central molecule di-phenyl ester moiety binds similar to the iodo-groups of T3 (**Fig. 3.06A**). The second ring of the di-phenyl ester of the left molecule mimics the binding interaction of the terminal iodo-group in T3 (**Fig. 3.06A**). Superimposition of the PCNA:p21-peptide complex (PDB:1AXC) hydrophobic formed by the p21 PIP-box motif residues M147, F150 and Y151 are identical to those formed by the aromatic rings of both the center and left AOH1160LE molecules (**Fig. 3.06B**). Comparison of the APIM-motif derived from PCNA:ZRANB3-peptide complex (PDB:5MLW) with both the left-hand and center molecules reveals that both small molecules once again mimic the hydrophobic interactions of the APIM-motif residues I1072, F1075 and L1076 (**Fig. 3.06C**).
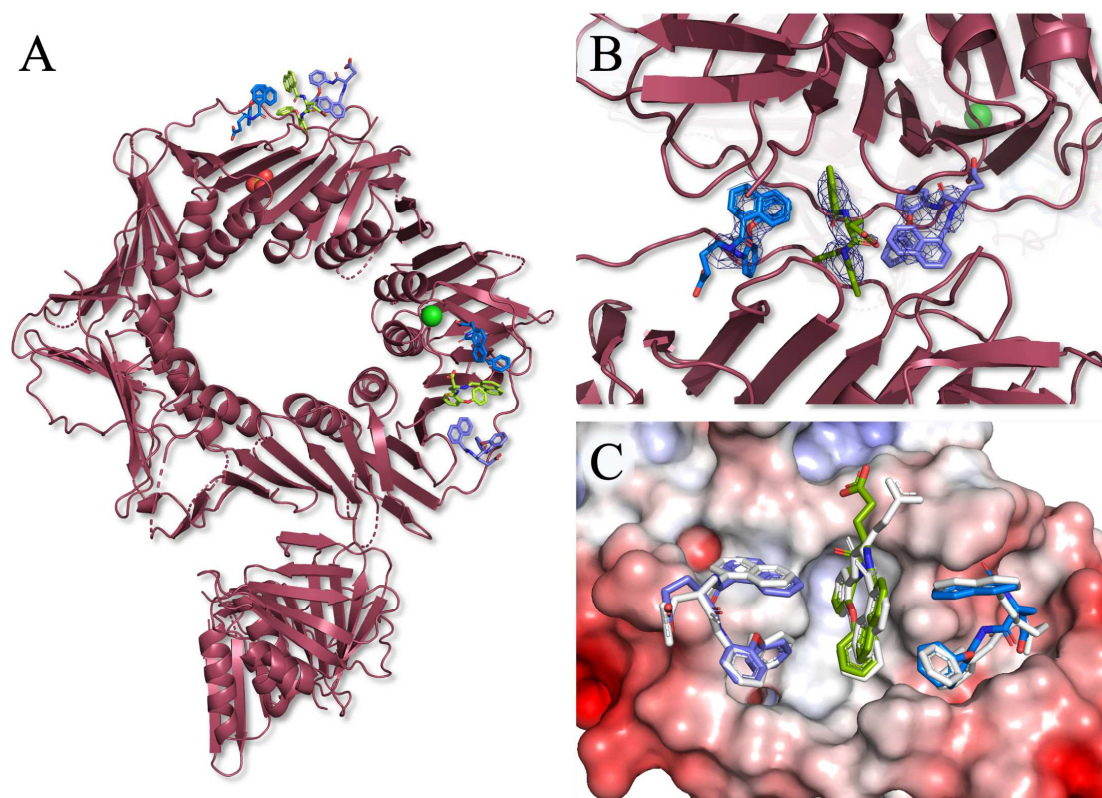
Figure 4.07 | Co-crystal structure of PCNA:AOH1160DE. (A) Cartoon representation of the asymmetric unit (*ruby*). (B) The $2F_o$-F_c *electron* densities of 3 modeled AOH1160DE molecules at the subunit interface, contoured to 2.0 σ. (C) Superimposition of PCNA:AOH1160LE with PCNA:AOH1160LE. The AOH1160LE molecules are shown in stick representation (*white*). The binding pocket is depicted with Poisson-Boltzmann electrostatic surface potentials shown in red and blue, corresponding to -5 to +5 kT/e respectively. The AOH1160LE molecules are depicted as sticks throughout and colored based on their positioning in the binding pocket (*violet-left, green-center, blue-right*).

Co-crystallization COH CADD-based molecule, AOH1160DE- The enantiomer of AOH1160LE, AOH1160DE, was soluble at 4 mM in aqueous buffer containing 10% DMSO. Following an identical crystallization and micro-seeding procedure to the AOH1160LE co-crystal preparation, micro-seeded PCNA:AOH1160DE co-crystals were produced. These co-crystals diffracted to 3.69 Å and the structure was solved via MR. The final model was refined with an R_{work} of 21.78% and R_{free} of 26.34% (**Table 4.01**). The

resolved structure adopts an identical asymmetric unit and lattice packing assembly as the PCNA:AOH1160LE co-crystal structure (**Fig. 4.07A**). Although this is a lower resolution structure the electron density of three individual AOH1160DE molecules can be clearly observed in the PIP-box binding cavities at the interface of two symmetry related subunits (**Fig. 4.07B**). Two sets of three AOH1160DE molecule have been modeled into the A and B subunits of the co-crystal structure. The difference in chirality has not altered the general binding mode of the AOH1160DE molecules as compared to the LE derivative (**Fig. 4.07C**). We have observed two orientations of binding for the central AOH1160DE molecule; either the di-phenyl ester moiety pointed down (A and B subunits), or the naphthalene moiety pointed down into the PIP-box cavity (C and D subunits). The two observed binding orientations are due to the bridging interactions that occur in the stacked symmetry related subunits. In addition, the left molecule binding mode in the A and B subunits is identical to the binding mode of the right molecule in the C' and D' symmetry related subunits. Both the central and right-hand (*blue*) AOH1160DE molecules form a single hydrogen bond interaction with the residues of the PIP-box binding cavity and adjacent pocket. The central AOH160DE in the di-phenyl ester down binding mode forms a hydrogen bond between the linker region and the backbone of H44. The right-hand AOH1160DE molecule forms a hydrogen bond with the side chain of the same residue (**Supp. Fig. 4.01**).
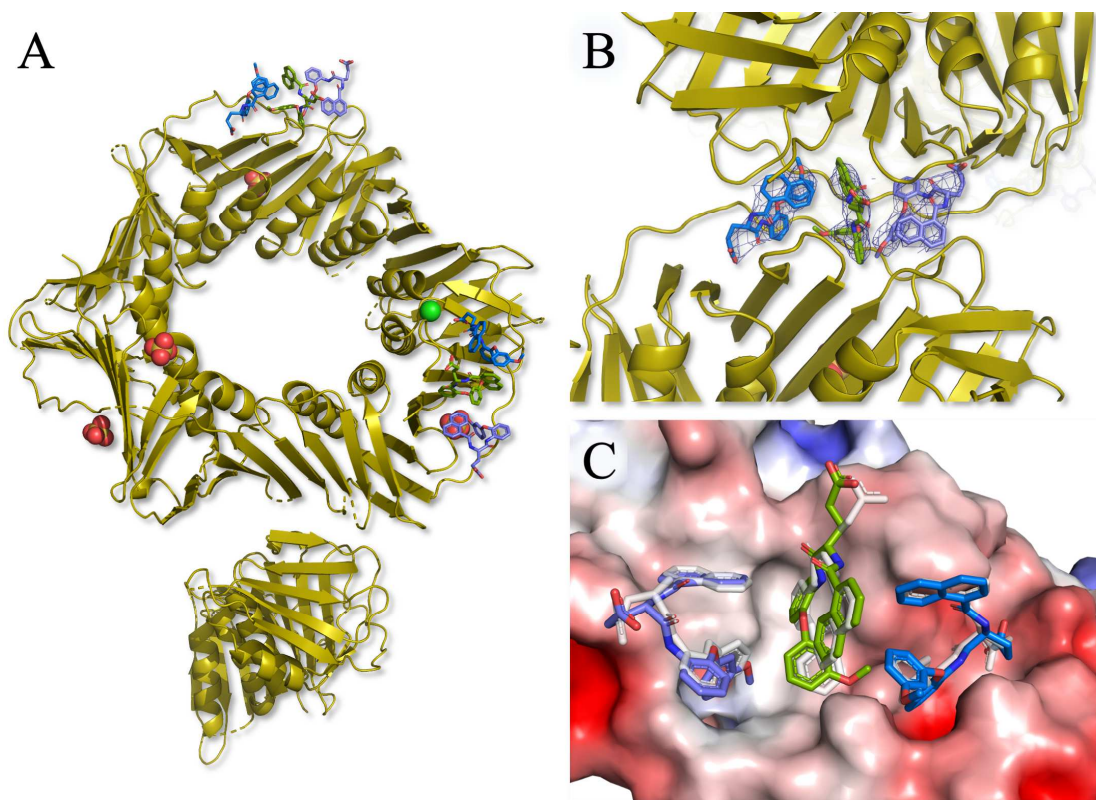
Figure 4.08 | Co-crystal structure of PCNA:AOH1996LE. (A) Cartoon representation of the asymmetric unit (*gold*). (B) The *2F$_o$-F$_c$ electron* densities of 3 modeled AOH1996LE molecules at the subunit interface, contoured to 2 σ. (C) Superimposition of PCNA:AOH1996LE with PCNA:AOH1160LE. The AOH1160LE molecules are shown in stick representation (*white*). The binding pocket is depicted with Poisson-Boltzmann electrostatic surface potentials shown in red and blue, corresponding to -5 to +5 kT/e respectively. The AOH1996LE molecules are depicted throughout in stick representation and colored based on their positioning in the binding pocket (*violet-left, green-center, blue-right*).

Co-crystallization COH CADD-based molecule, AOH1996LE- Following the successful characterization of the AOH1160LE derivative questions remained regarding its sister molecule, AOH1996, which contains an additional methoxy group to increase the compound's stability in serum. Therefore, the AOH1996LE molecule was synthesized to determine if the methoxy group attached at the second di-phenyl ester ring would result in an alternative binding conformation. The AOH1996LE derivative was soluble at 2 mM in

aqueous buffer containing 10% DMSO. The PCNA:AOH1996LE co-crystals were obtained following an identical co-crystallization and seeding procedure as the previously described AOH1160 derivative structures. The produced co-crystals diffracted to 3.75 Å and were resolved utilizing MR. The final model was refined with an R_{work} of 19.70% and R_{free} of 25.61% (**Table 4.01**). The electron density of three individual AOH1996LE molecules could be clearly observed in the PIP-box binding cavities at the interface of two symmetry related subunits following protein refinement (**Fig. 4.08B**). Two sets of three AOH1996LE molecules have been modeled into the A and B subunits of the co-crystal structure. The additional methoxy-group has not significantly alerted the binding interactions of the AOH scaffold as compared to the modeled AOH1160LE molecules (**Fig. 4.08C**). The bridging interactions of the AOH1996LE molecules into the stacked symmetry related subunits result in two orientations of binding for the central molecule, as observed in the co-crystal structures of the two AOH1160 derivatives. Only the right-hand (*blue*) molecule forms a hydrogen bond interaction with the S39 and H44 residues of the PIP-box adjacent pocket. As previously characterized in the stacking interactions of the AOH1160 scaffold complexes the left molecule binding mode matches the binding mode of the right molecule up into the symmetry related subunit (**Supp. Fig. 4.03**).

Table 4.01 | Table of Crystallography Data- PCNA:AOH Derivative Complex

	PCNA:AOH1160LE	PCNA:AOH1160DE	PCNA:AOH1996LE
Data collection			
Space group	H3	H3	H3
a, b, c (Å)	197.14, 197.14, 126.97	196.54, 196.54, 125.98	199.49, 199.49, 129.41
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120
Resolution (Å)	37.96 – 2.85 (2.95-2.85)	44.21 – 3.69 (3.82 – 3.69)	38.51 – 3.75 (3.88 – 3.75)
R_{merge} (%)	0.115 (0.878)	0.111 (0.219)	0.183 (0.986)
I/σ (I)	13.4 (2.3)	8.1 (2.8)	8.61 (1.9)
Completeness (%)	99.48 (97.70)	98.98 (96.09)	99.68 (99.54)
Redundancy	7.5 (7.8)	10.6 (11)	5.8 (5.9)
Refinement			
No. reflections	42,977 (4,247)	19328 (1892)	19,649 (1,962)
R_{work}/R_{free} (%)	31.67 / 39.93 (19.75 / 25.41)	40.11 / 44.58 (21.78 / 26.34)	28.37 / 32.35 (19.70 / 25.61)
No. atoms			
Protein	7,657	7,346	7,378
Ligand	228	216	243
Solvent	8	1	0
B-factors			
Protein	54.47	106.30	89.36
Ligand	53.87	104.06	89.89
Solvent	38.81	113.64	0
R.M.S. Deviations			
Bond lengths (Å)	0.011	0.004	0.004
Bond angles (°)	1.21	0.71	0.74

Highest resolution shell is shown in parenthesis

$R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith observation of reflection *hkl* and $I(hkl)$ is the average over all observations of reflection *hkl*.

$R_{factor} = \sum_{hkl}||F_o| - |F_c||/\sum_{hkl}|F_o|$ for all data excluding the 10% used for cross-validation (R_{free}).

4.3.2 Co-crystallization of Atomwise AI-CADD-based molecule
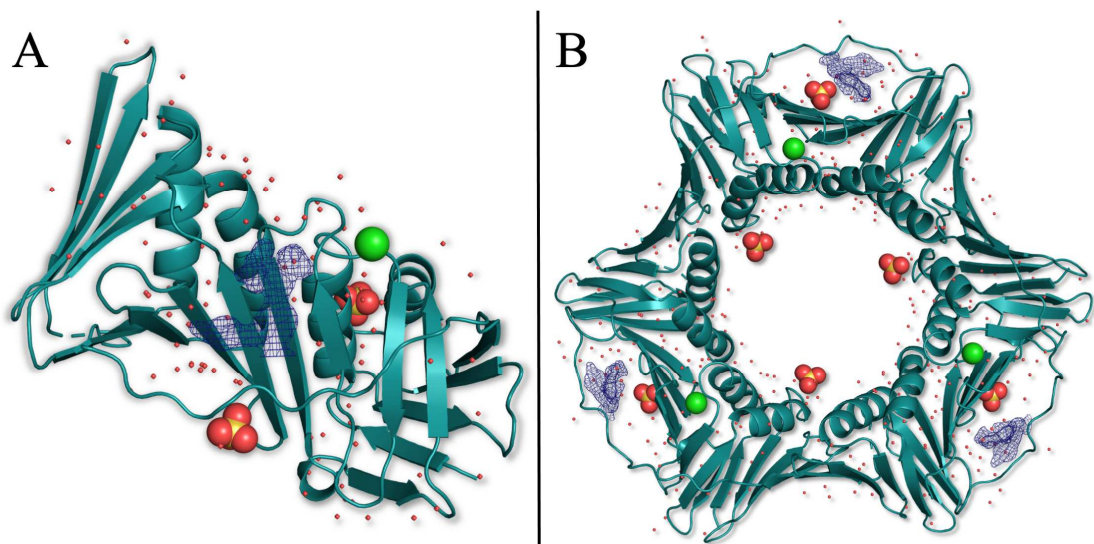
Figure 4.09 | Co-crystal structure of PCNA:B05. (A) Cartoon representation of the asymmetric unit (*teal*) with waters and hetero atoms represented as spheres. (B) Packing of the asymmetric unit within the crystal lattice to form the biological heterotrimeric assembly of PCNA. The $2F_o$-F_c electron density map (*indigo*) of B05 is contoured to 2 σ.

Co-crystallization of Atomwise AI-CADD-based molecule, B05- The B05 molecule was soluble at 2 mM in aqueous buffer containing 10% DMSO and incubated with overnight with 9 mg/mL PCNA. Utilizing hanging drop vapor diffusion and the AOH1160LE complex crystallization condition, PCNA:B05 co-crystals were produced after 1-day. These co-crystals, consequently, suffered from cracking due to rapid growth and improper lattice packing. Therefore, the precipitant concentration was decreased and produced co-crystals no longer possessed cracks. These crystals were collected and produced high-resolution diffraction limits of 1.90 Å. The structure was resolved via MR using the PDB:3VKX structure as an input model and refined with an R_{work} of 20.57% and R_{free} of 23.01% (**Table 4.02**). The asymmetric unit of the PCNA:B05 structure is a single PCNA monomer whose stacking interactions form the homotrimeric ring assembly (**Fig. 4.09**).
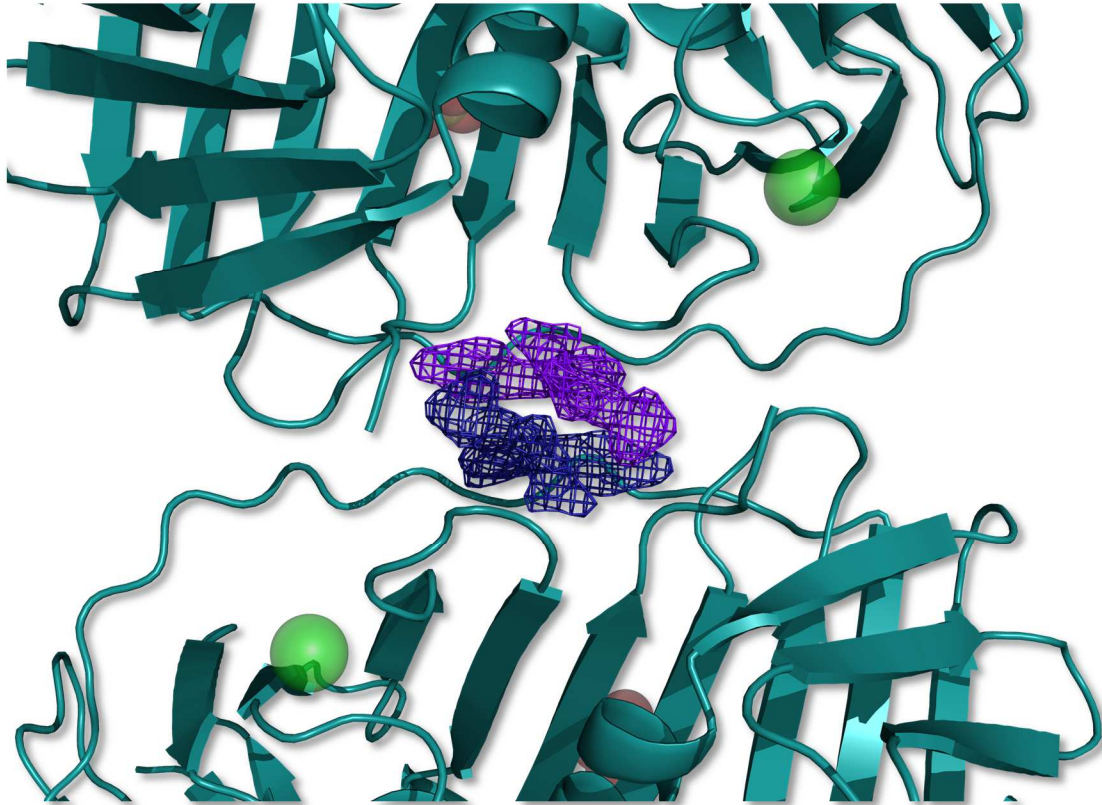
Figure 4.10 | PCNA:B05 stacking interaction. Electron density maps of two B05 molecules (*indigo* and *purple*) modeled in the PIP-box binding sites of the stacked PCNA monomers shown in cartoon representation (*teal*). The $2F_o\text{-}F_c$ maps are contoured to 2 σ.

Following refinement of the protein structure, the electron density of a single B05 molecule was clearly observed in the PIP-box binding cavity (**Fig. 4.09**). The chemical structure of B05 is not shown due to proprietary reasons, and is therefore represented by its electron density. Similar to the PCNA:AOH-derivative co-crystal packing, the PCNA:B05 subunits interface at their PIP-box binding sites and interact via the IDCL regions. However, the B05 molecule does not bridge into the stacked subunit like the AOH-derivatives. Instead, there is a single B05 molecule for every monomer that is also stacked at the subunit interface, similar to the stacking observed in the PCNA:T3 co-crystal structure (**Fig. 4.10**).
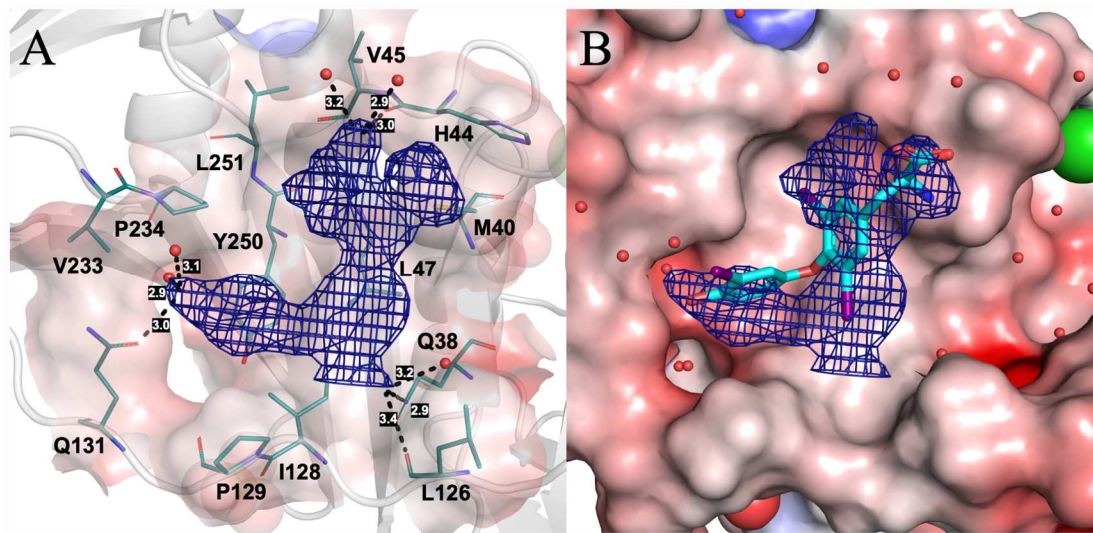
Figure 4.11 | B05 in the PIP-box binding site. (A) Cartoon representation of PCNA monomer (*white*) with binding site residues in stick representation (*teal*). Hydrogen bonds are displayed with their corresponding distances (*black dashes*). (B) Superimposition with the T3 (*cyan*) co-crystal structure. The $2F_o$-F_c electron density maps (*indigo*) of the modeled B05 molecule are contoured to 2 σ. The Poisson-Boltzmann electrostatic surface potentials are shown in red and blue, corresponding to -5 to +5 kT/e respectively.

The B05 molecule occupies the entire PIP-box interaction site through both hydrophobic and polar interactions. The hydrophobic region formed by the amino acid residues M40, H44, V45, L47, I128, P129, V233, P234, Y250 and L251 accounts for most of the binding site interaction with B05. In addition, B05 forms hydrogen bonding interactions with the Q38, H44, L126 and Q131 residues. The high-resolution of data obtained enabled the identification of ordered waters in the binding site. A total of five water molecules have been characterized for form hydrogen bonding interactions with the B05 molecule in the PIP-box binding pocket (**Fig. 4.11A**). A superimposition of PCNA:B05 with the co-crystal structure of the known PIP-box binder, T3, indicates that the novel B05 molecule occupies a larger area of the PIP-box binding site (**Fig. 4.11B**).

Table 4.02 | Table of Crystallography Data- PCNA:AI-CADD Complex

	PCNA:B05
Data collection	
Space group	I 2₁ 3
a, b, c (Å)	141.09, 141.09, 141.09
α, β, γ (°)	90, 90, 90
Resolution (Å)	37.71 – 1.90
	(1.96 – 1.90)
R_{merge} (%)	0.062 (1.417)
I/σ (I)	34.49 (2.29)
Completeness (%)	99.97 (100)
Redundancy	22.8 (22.3)
Refinement	
No. reflections	36,840 (3,655)
R_{work}/R_{free} (%)	25.67 / 27.72
	(20.57 / 23.01)
No. atoms	
Protein	1917
Ligand	44
Solvent	111
B-factors	
Protein	39.94
Ligand	38.60
Solvent	39.96
R.M.S. Deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.06

Highest resolution shell is shown in parenthesis

$R_{merge} = \sum_{hkl}\sum_{i}|I_i(hkl) - \langle I(hkl)\rangle |/\sum_{hkl}\sum_{i}I_i(hkl)$, where $I_i(hkl)$ is the intensity of the *i*th observation of reflection *hkl* and $I(hkl)$ is the average over all observations of reflection *hkl*.

$R_{factor} = \sum_{hkl}||F_o| - |F_c||/\sum_{hkl}|F_o|$ for all data excluding the 10% used for cross-validation (R_{free}).

4.3.3 High Resolution *apo*-PCNA structure for FBLD
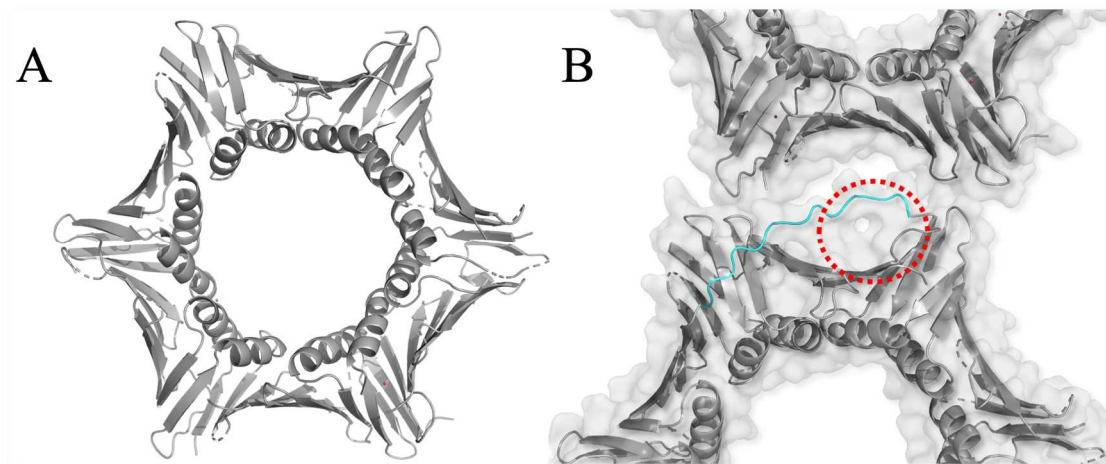
Figure 4.12 | Seeded apo-PCNA structure. (A) Cartoon representation of the asymmetric unit (*silver*). (B) Surface representation (*grey*) at the subunit packing interface with PIP-box binding site indicated (red circle). An overlay of the IDCL extracted from PDB:3VKX (*cyan,* cartoon) is used to represent the unmodeled IDCL in the *apo*-PCNA structure.

Seeded apo-PCNA structure- The co-crystallization conditions did not crystallize in the absence of small molecule binders. To structurally characterize the large pool of fragment hits detected in our TSA screens, we aimed to develop a high-resolution *apo*-PCNA crystallization condition that could be used in soaking or co-crystallization experiments. The previously published *apo*-PCNA (PDB:1VYM) produced small, cracked crystals that diffracted to 3.4 Å. Subsequent optimization paired with micro-seeding improved diffraction resolution to 2.44 Å. The structure was resolved via MR using the PDB:3VKX structure as an input model and refined with an R_{work} of 22.44% and R_{free} of 29.88% (**Table 4.03**). Three monomers of PCNA are modeled in the asymmetric unit. However, the lower resolution dataset prevented accurate modeling of the more dynamic regions of the protein, such as the IDCL (**Fig. 4.12A**). An overlay with the IDCL of the PDB:3VKX structure demonstrates that the PIP-box binding site is open to binding interaction (**Fig. 4.12B**).
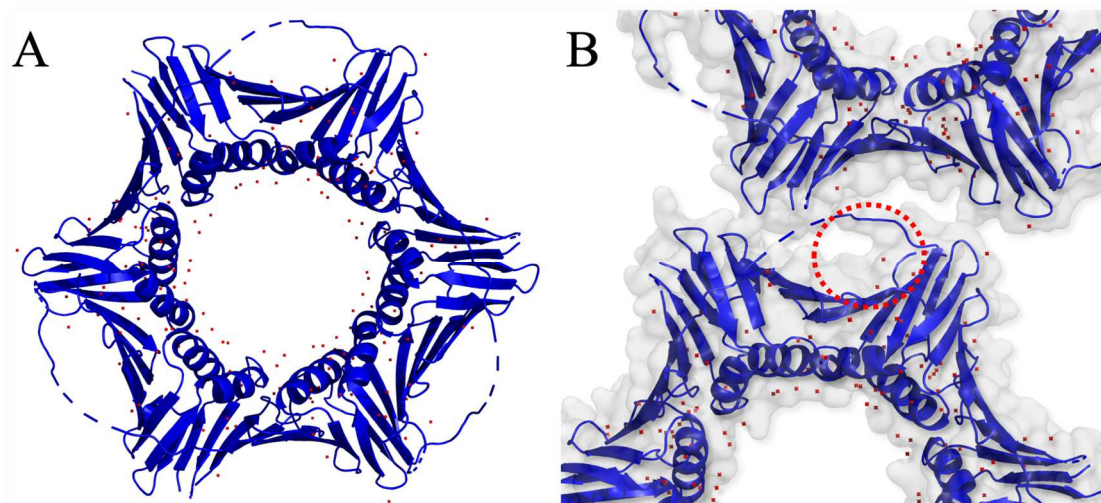
Figure 4.13 | Novel apo-PCNA hit condition structure. (A) Cartoon representation of the asymmetric unit (*sapphire*) with modeled waters (*red spheres*). (B) Surface representation (*grey*) at the subunit packing interface with PIP-box binding site indicated (red circle).

Resolution of a novel apo-PCNA structure- Sparse matrix screens of over 1,000 diverse crystallization conditions were tested in sitting-drop and hanging-drop formats utilizing 10 mg/mL purified PCNA protein. One of the screening hits produced rod-shaped crystals that were further optimized to improve stability and size by adjusting crystallization temperature and precipitant concentration. These crystals were subject to synchrotron-based X-ray data collection, and the structure was determined to a resolution of 2.01 Å. The final model was solved via MR with the PDB:3VKX structure as an input model and the generated electron density maps were used to refine the *apo* structure to a final model with an R_{work} of 19.72% and R_{free} of 23.98% (**Table 4.03**). A single monomer of PCNA is present in the asymmetric unit and stacks to form the biological homotrimeric ring (**Fig. 4.13A**). Packing of these rings within the crystal lattice places the PIP-box binding site of one monomer at the monomer interface of the stacked symmetry mate. This assembly leaves the PIP-box binding site open to potential binders (**Fig. 4.13B**).

Table 4.03 | Table of Crystallography Data- *apo*-PCNA Structures

	Seeded *apo*-PCNA	Novel *apo*-PCNA hit condition
Data collection		
Space group	I 2_1	P 6_3
a, b, c (Å)	70.92, 82.32, 119.82	82.67, 82.67, 71.37
α, β, γ (°)	90, 94.11, 90	90, 90, 120
Resolution (Å)	41.05 – 2.44 (2.527 – 2.44)	35.80 – 2.01 (2.082 - 2.01)
R_{merge} (%)	0.04269 (0.3347)	0.03315 (0.307)
I/σ (*I*)	13.14 (2.28)	14.28 (2.45)
Completeness (%)	99.93 (100)	98.22 (98.98)
Redundancy	2.0 (2.0)	2.0 (2.0)
Refinement		
No. reflections	51,311 (5,109)	19328 (1,892)
R_{work}/R_{free} (%)	28.57 / 38.83 (22.44 / 29.88)	25.25 / 30.81 (19.72 / 23.98)
No. atoms		
Protein	5013	1886
Ligand	0	0
Solvent	6	69
B-factors		
Protein	43.74	32.17
Ligand	0	0
Solvent	35.08	32.31
R.M.S. Deviations		
Bond lengths (Å)	0.008	0.008
Bond angles (°)	1.00	1.02

Highest resolution shell is shown in parenthesis

$R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the *i*th observation of reflection *hkl* and *I(hkl)* is the average over all observations of reflection *hkl*.

$R_{factor} = \sum_{hkl} ||F_o| - |F_c|| / \sum_{hkl} |F_o|$ for all data excluding the 10% used for cross-validation (R_{free}).

4.4 Discussion

4.4.1 Co-crystallization of City of Hope CADD-based Molecules

The co-crystal complexes of three City of Hope CADD-based molecules have been resolved to atomic resolutions. Two of the co-crystal structures elucidate the binding interactions of both AOH1160 glutamate derivative enantiomers. The parent AOH1160 molecule was too hydrophobic and readily precipitated when introduced to aqueous solutions, even up to a 50% DMSO concentration. Thus, glutamate derivatives of this molecule were designed to ionize in aqueous solutions and were soluble at 4 mM in 10% DMSO. Initial co-crystals with both enantiomers produced shared nucleation sites resulting in twinned and low-resolution data sets. The structure of each derivative was solved by implementing a micro-seeding technique, which in turn, corrected the nucleation issue. The PCNA:AOH1160LE structure produced the highest resolution dataset at 2.85 Å. Subsequent MR revealed a unique asymmetric unit assembly of four PCNA monomers as opposed to the monomeric or biological trimeric assembly. Upon further analysis, it was determined that lattice packing interactions create a traditional homotrimeric biological assembly among the symmetry mates of the fourth subunit. Within the PIP-box binding site of each monomer three individual clouds of positive F_o-F_c density and spots of $2F_o$-F_c density were observed. Modeling of three AOH1160LE molecules into the A and B subunits agreed with the detected electron densities (**Fig. 4.04B**). The three AOH1160LE molecules form a unique bridging interaction by binding not only into the A and B subunit PIP-box cavities, but also bind into the PIP-box cavities of the stacked symmetry mate C and D subunits (**Fig. 4.04A**). An identical asymmetric unit stacking assembly and bridging

interactions were observed in both lower resolution AOH1160DE and AOH1996LE co-crystal structures (**Fig. 4.07A and B, 4.08A and B**). Although there are three individual molecules modeled in the resolved co-crystal structures, the 'central' molecule is likely representative of the AOH-scaffold binding mode. The 'central' molecule binds similarly to known PIP-box inhibitor, T3. As observed in an overlay of both the AOH1160LE and T3 co-crystal structures, the aromatic rings of the AOH-scaffold's di-phenyl ester moiety binds similarly to the iodo-groups of T3 and occupy nearly identical three-dimensional volume to the large iodine atoms. In addition, the 'central' binding mode interacts with the residues of the PIP-box cavity that are known to be essential for binding PIP-box and APIM motif peptides (**Fig. 4.06**). The 'left-handed' molecule and the 'right-handed' molecule are mirrors of one another generated through crystal symmetry operations: The partial PIP-box cavity interactions formed by the 'left-handed' molecule to the asymmetric unit are identical to the interactions formed by the 'right-handed' molecule as it extends into the PIP-box cavity of the stacked symmetry related protein subunit.

Studies by our collaborators determined that the parent AOH1160 scaffold was not stable and was prematurely metabolized in human liver microsome. Therefore, one of the designed AOH1160 analog structures included an additional methoxy group extension from the di-phenyl ester moiety to prevent degradation by liver carboxyl esterase enzymes. This newly synthesized molecule, deemed AOH1996, was found to improve microsomal stability (data not shown). However, the AOH1996 scaffold had not been structurally characterized and it was unknown whether the additional methoxy group would result in a different binding mode to the PIP-box binding site. Since this molecule is also very

hydrophobic, our collaborators synthesized an AOH1996LE derivative for further analysis. We then implemented an identical approach to the AOH1160LE co-crystallization studies to the AOH1996LE derivative and were able to obtain PCNA:AOH1996LE co-crystals that diffracted to 3.75 Å. Although this was a lower-resolution data set we observed an identical asymmetric unit assembly and three positive F_o-F_c density clouds in the PIP-box binding cavity of the PCNA monomers. Modeling of three AOH1996LE molecules agreed with the experimentally generated electron density (**Fig. 4.08B**). The significant B-factors (> 80) of the modeled atoms are due to the low-resolution data. To generate a more accurate model of the AOH1996LE co-crystal complex the AOH1160LE co-crystal structure peptide chains were used as a reference in the refinement process. Complete resolution of the model and generated AOH1996LE omit maps (not shown) confirmed similar binding interactions as compared to the AOH1160LE molecules. Therefore, the methoxy modification did not affect binding to the PIP-box binding site (**Fig. 4.08C**). The characterized binding interactions of the AOH derivatives has provided valuable information regarding parent AOH1160 and AOH1996 scaffold binding interactions. These structures are being utilized for the characterization of AOH-molecule SAR criteria and future analog design.

4.4.2   Co-crystallization of Atomwise AI-CADD-based molecule

Utilizing the AOH-molecule co-crystallization condition all the TSA hits detected from our AI-CADD derived small molecule library were screened. The co-crystallization of PCNA with the B05 molecule readily grew crystals after a single day. These crystals were cracked due to rapid growth and therefore their co-crystallization was optimized to slow

growth and produce a more well-ordered structure. These co-crystals diffracted to a high-resolution of 1.90 Å. The subsequent structure refinement revealed significant 2*Fc-Fc* density in the PIP-box binding cavity. The atomic coordinates of the B05 molecule were modeled into the unaccounted-for density and the experimental data agreed with the newly modeled structure. Unlike the AOH-derivative molecules, only a single B05 molecule was modeled into the PIP-box binding site (**Fig. 4.09A**) without bridging interaction with symmetry mate subunits (**Fig. 4.10**). The difference in binding mode could be attributed to the variation in crystallization condition components. The B05 molecule occupies the entire of the PIP-box binding cavity and forms hydrophobic interactions identical to characterized PIP-motif interactions except for the Q-pocket near Y211. Identical to the 'central' AOH-derivative, PIP- and APIM-motif containing peptides, B05 sits in the hydrophobic cavity formed by the β-hairpin residues M40-H44, C-terminal loop amino acids Y250 and L251, and the region of the IDCL that spans residues L126-I128. Unlike the 'central' AOH-derivative molecule, the B05 ligand extends into the remaining portion of the PIP-box binding pocket that is lined by IDCL residues P129-Q131 and the hydrophobic β-hairpin residues P234 and V233.

Analyses performed by our collaborators have revealed that the B05 molecule is nearly 10 times more potent in inducing cell apoptosis in malignant cell models as compared to the AOH1996 molecule (data not shown). However, the therapeutic index is not as significant as the AOH1996 molecule (data not shown). The novel B05 scaffold is being investigated further to determine the source of its relative potency but lesser selectivity compared to AOH1996. The discovery of the B05 scaffold has elucidated a

novel, alternative avenue for the development of caPCNA inhibitors. Advantageously, the high-resolution structure data has enabled the detection of ordered water molecules in the regions surrounding the PIP-box cavity that can be used to identify additional interaction points to the target protein. This information is useful to develop B05 analogs that will target these interactions nad develop selectivity for the PIP-box cavity in subsequent SBDD studies.

4.4.3 High Resolution *apo*-PCNA structure for FBLD

The previously published *apo*-crystallization condition (PDB:1VYM) was prepared to develop a platform for soaking and co-crystallization of TSA hits. However, directly repeating this condition produced crystals that diffracted to low resolutions of 3.40 Å. To improve crystal growth micro-seeding was implemented and found to improve diffraction resolutions to 2.44 Å. The resolved *apo*-PCNA structure did not have sufficient electron density for the IDCL, but this is most likely due to the dynamic nature of this loop region (**Fig. 4.12A**). Based on the crystal lattice packing interactions the PIP-box binding cavity is left open to interaction with potential binders (**Fig. 4.12B**). This resolution could be further improved to greatly aid the characterization of small compounds such as fragments. In addition, micro-seeding is dependent on seed and stock preparations and is a complex procedure that can be challenging to reproduce and time consuming for drug discovery. Therefore, we set out to identify a novel *apo*-PCNA crystallization condition that could produce high-resolution diffraction data. Following sparse matrix screens of over 1,000 diverse crystallization conditions in both sitting- and hanging-drop formats at varying temperatures, we were able to identify a single condition that produced well-ordered PCNA

crystals at 4 °C. These crystals diffracted to 2.01 Å, the highest-resolution limit recorded for an *apo*-PCNA structure as compared to the RCSB PDB. Similar to the seeded *apo*-crystal structure, the stacking interactions of the PCNA monomers within the crystal lattice leaves the PIP-box binding site open to interaction with potential binders (**Fig. 4.13B**). This condition is currently being used to characterize both small molecule and fragment TSA hits for SBDD and FBLD studies.

4.5   Methods

4.5.1   Structural Characterization of COH CADD-based molecules

Co-crystallization of City of Hope CADD-based molecules- Each AOH analog was pre-incubated with purified PCNA protein (section 2.5.1) in 0.09 M HEPES pH 7.4, 90 mM sodium chloride and 9% dimethyl sulfoxide. Precipitant was observed upon addition of inhibitor to protein sample and therefore following overnight incubations each sample were centrifuged at 12,000 xg for 20 mins at 4 °C to remove precipitation. The previously published co-crystallization condition, 0.1 M sodium cacodylate pH 6.5, 0.2 M sodium chloride and 2.0 M ammonium sulfate (PDB:3VKX), was utilized to co-crystallize AOH analogs. The incubated sample of 9 mg/mL PCNA with 4 mM AOH1160LE was added to reservoir solution in a 1.5:1.5 µL drop ratio over a 500 µL reservoir at 20 °C grew to full size (100 µm x 100 µm x 100 µm) after 4 weeks. The incubated sample of 9 mg/mL PCNA with 4 mM AOH1160DE produced co-crystals of identical size and shape after 7-8 weeks. Micro-seeding was implemented to resolve observed shared nucleation in both crystal sets. Micro-seed stocks were prepared from both the PCNA:AOH1160LE and PCNA:AOH1160DE co-crystals following the Hampton Research Seed Bead Kit procedure (HR4-780). Seeded crystals were grown by adding 4 µL protein:compound incubation to 4 µL 1:10,000 seed stock dilution over a reservoir solution of 500 µL 0.1 M sodium cacodylate pH 6.5, 0.2 M sodium chloride and 1.5 M ammonium sulfate. Post-seeding PCNA:AOH1160LE co-crystals grew to full size (200 µm x 200 µm x 200 µm) after 1 week. Post-seeding PCNA:AOH1160_DE co-crystals grew to full size (100 µm x 100 µm x 100 µm) after 4 weeks. The overnight incubation of 9 mg/mL PCNA with 1 mM

AOH1996_LE in 0.09 M HEPES pH 7.4, 90 mM sodium chloride and 9% dimethyl sulfoxide was added to a reservoir solution of 0.1 M sodium cacodylate pH 7.0, 0.2 M sodium chloride and 2.4 M ammonium sulfate in a ratio of 2:2 µL over a 500 µL reservoir at 20 °C. Many, small PCNA:AOH1996_LE co-crystals (25 µm x 25 µm x 25 µm) appeared in the well after 3 weeks. These co-crystals were utilized to prepare a micro-seed stock to improve growth time and size. The incubated sample of 9 mg/mL PCNA with 2 mM AOH1996LE was added to a micro-seed dilution of 1:100,000 in a 2:2 µL drop ratio over a reservoir solution of 500 µL 0.1 M sodium cacodylate pH 7.0, 0.2 M sodium chloride and 2.0 M ammonium sulfate. Seeded PCNA:AOH1996_LE co-crystals grew to full size (50 µm x 50 µm x 50 µm) after 2 weeks.

X-ray diffraction and structure determination- Both the PCNA:AOH1160LE and PCNA:AOH1996LE diffraction data were collected at the Stanford Synchrotron Radiation Light source (SSRL) beamline 9-2 equipped with a Dectris Pilatus 6M detector. The diffraction data for the PCNA:AOH1160DE co-crystal structure was collected at the Advance Light Source (ALS) Synchrotron Radiation Facility beamline 5.0.2 using the Dectris Pilatus3 6M detector. The wavelengths used for each collection was 0.98 Å. All datasets were processed using the XDS package [292] and solved by molecular replacement with Phaser [293] using the atomic coordinates of the *h*PCNA monomer (PDB:3VKX) with heteroatoms and water molecules removed. All structures were refined in Phenix [294] and model building was performed in Coot [295]. Each corresponding ligand restraint was generated using the eLBOW [296] extension in Phenix [294].

4.5.2 Structural Characterization of Atomwise AI-CADD-based molecules

Co-crystallization of Atomwise AI-CADD-based molecule- Each of the Atomwise AI-CADD based molecules that induced a positive shift of 0.5 °C or greater. Hit compounds were incubated overnight at a final concentration of 2 mM with 9 mg/mL PCNA (313 μM) in 9 mM HEPES pH 7.4, 0.09 M sodium chloride and 10% dimethyl sulfoxide and subjected to screening around the AOH1160_LE co-crystallization condition. In addition, overnight incubations of the most significant shifters at 1.0 °C or greater A06, A07, A10, A11, C03, D04 and E01) were screened in the Anatrace Top 96 sparse matrix kit. Of the 19 molecules tested, the co-crystallization of B05 in 0.1 M sodium cacodylate pH 6.5, 0.2 M sodium chloride and 2.0 M ammonium sulfate in sitting-drop format fully formed crystals (200 μm x 200 μm x 200 μm) after 2 days at 20 °C.

X-ray diffraction and structure determination- The diffraction data for PCNA:B05 was collected at the SSRL beamline 9-2 utilizing a Dectris Pilatus 6M detector at a wavelength of 0.98 Å. The collected dataset was processed using the XDS package [292] and solved by molecular replacement with Phaser [293] using the atomic coordinates of the *h*PCNA monomer obtained from the PDB:3VKX structure with heteroatoms and water molecules removed. The ligand coordinate file and corresponding restraints were generated using the eLBOW [296] extension within Phenix [294] and model building was performed in Coot [295]. Following multiple rounds of refinement the representative model produced reasonable R-factors, no Ramachandran outliers and B-factors corresponding to the dataset resolution limits.

4.5.3   Resolution of apo-PCNA structure for FBLD

Seeding apo-PCNA condition- The previously published *apo*-PCNA condition, 0.2 M magnesium acetate, 20% PEG3350 (PDB:1VYM) was buffer optimized with 0.1 M Bis Tris pH 7.2. The addition of 2 µL of PCNA with 2 µL reservoir solution in sitting drop format above a 400 µL reservoir solution at 16°C produced small cubic crystals (50 µm, 50 µm, 50 µm). Micro-seeding was utilized to resolve shared nucleation and cracking issues in addition to improving overall crystal size. Seeded crystals were grown by adding 2 µL 10 mg/mL PCNA to 2 µL 1:1,000 seed dilution in sitting drop format over the previously described reservoir. Seeded crystals grew four times larger (200 µm x 200 µm x 200 µm) following a 4-day growth period. Crystals were then harvested by flash freezing in liquid nitrogen prior to shipping to the synchrotron source.

Sparse matrix screening apo-PCNA- To identify a high-resolution crystallization condition that could be used in subsequent hit characterization studies, thousands of conditions were screened and subsequently optimized. Initially, screens were set up with 1:1 µL ratios of 10 mg/mL of protein to reservoir against over 1,000 conditions obtained from the sparse matrix screening kits: Hampton Research Index, PEG/Ion 1-2, PEG Rx, Salt Rx, Grid Screen, Crystal Screen 1-2, PEG/pH 1-2, Low Ionic Strength Screens, Anatrace Top96, MCSG 1-4, PurePEGs, Molecular Dimensions JCSG, BCS and PGA screens. The hit condition grew hexagonal rods (150 x 50 x 50 µm) after 3 days. Crystals were submerged in a cryoprotectant solution (1:1 (v/v) reservoir and 50% glycerol) prior to flash freezing in liquid nitrogen to shipping to the synchrotron source.

X-ray diffraction and structure determination- Both the micro-seeded *apo*-PCNA and novel sparse matrix hit condition crystals' diffraction data were collected from the ALS synchrotron source at beamlines 5.0.1 and 5.0.3, respectively. The *apo* structure data collection at ALS 5.0.1 beamline was done utilizing a CCD ADSC Q315R detector while the ALS 5.0.3 beamline collection utilized a Pilatus 3 2M detector. Both energy wavelengths were approximately 0.98 Å. The micro-seeded *apo*-crystal diffraction data was processed in HKL2000 [297]. The data collected for crystals produced from the novel sparse matrix hit condition was processed using the XDS package [292]. Both structures were solved via molecular replacement with Phaser [293] and subjected to multiple rounds of refinement in Phenix [294] paired with model building in Coot [295]. The novel condition crystals were resolved to reasonable R-factors.

4.6   Supplemental Information
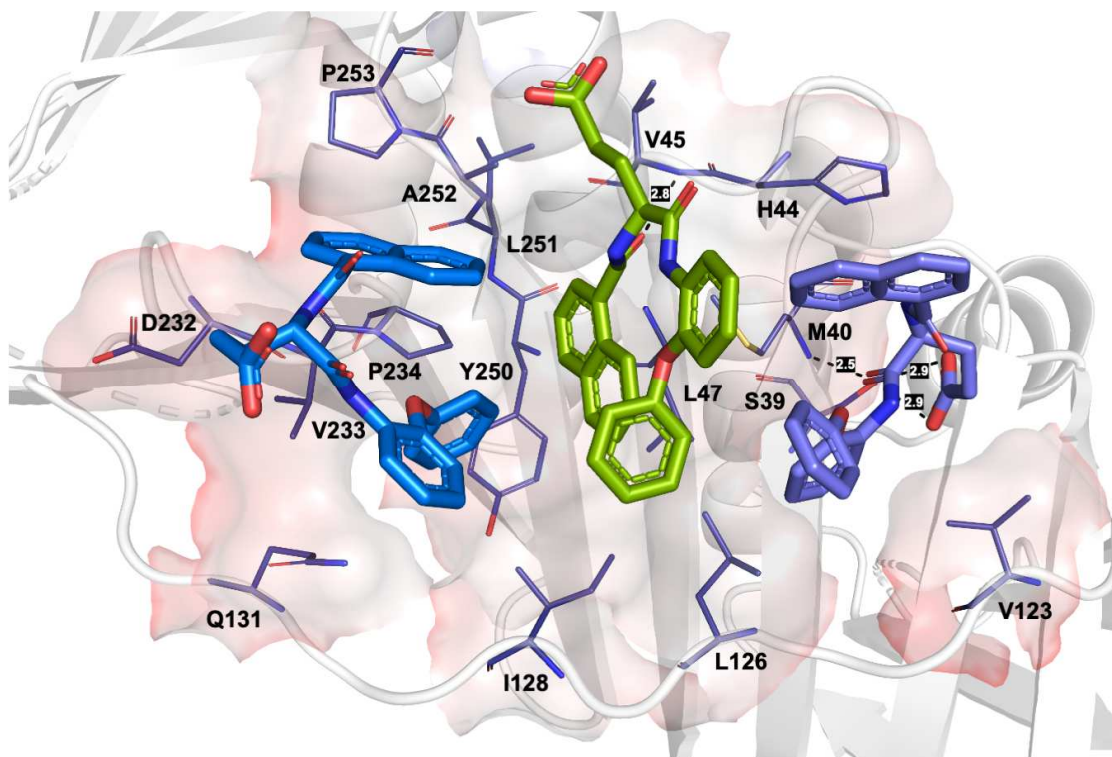
Supplementary Figure 4.00 | AOH160LE 'naphthalene down' binding interactions.
Cartoon representation of PCNA symmetry mate monomer (*white*) with binding site
residues in sticks (*emerald*). The Poisson-Boltzmann electrostatic surface potentials are
shown in red and blue, corresponding to -5 to +5 kT/e respectively. Three AOH1160LE
molecules are shown as sticks (*violet-left, green-center and blue-right*). The central
AOH1160LE molecule is shown in the with the naphthalene moiety bound down in the
PIP-box binding cavity. Hydrogen bonds are displayed with their corresponding
distances (*black dashes*).
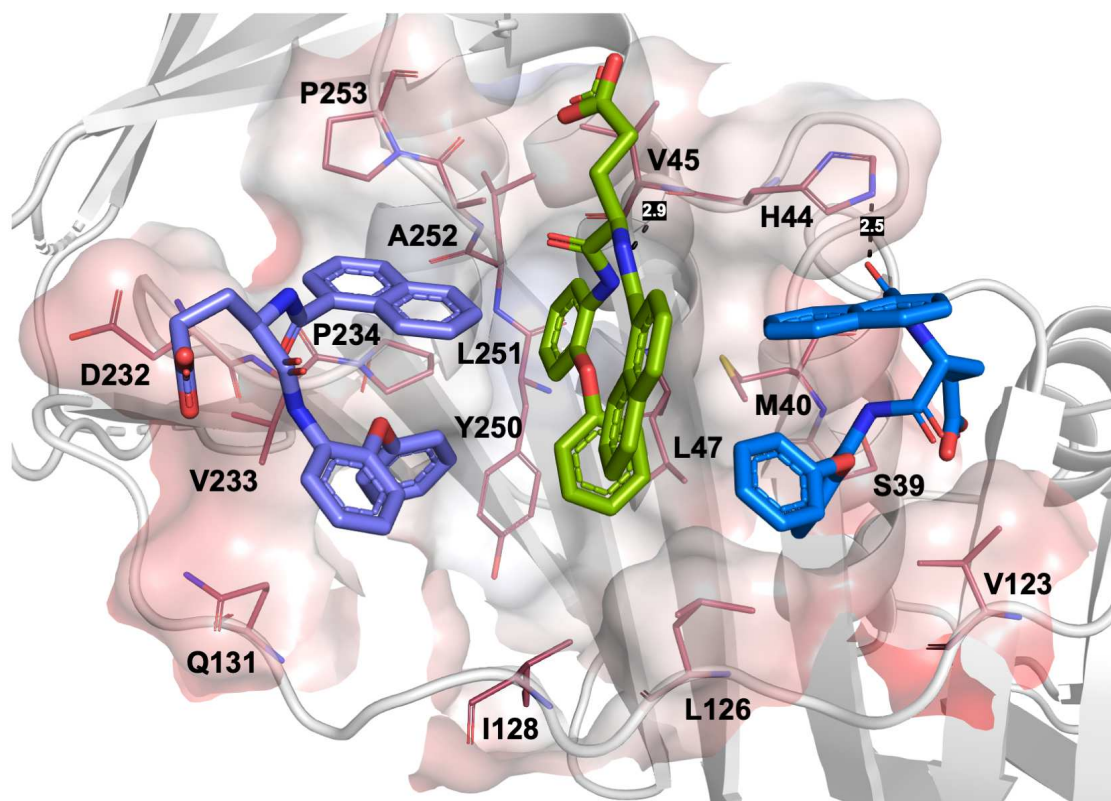
Supplementary Figure 4.01 | AOH1160DE binding interactions. Cartoon representation of PCNA monomer (*white*) with binding site residues in sticks (*emerald*). The Poisson-Boltzmann electrostatic surface potentials are shown in red and blue, corresponding to -5 to +5 kT/e respectively. Three AOH1160DE molecules are shown as sticks (*violet-left, green-center and blue-right*). Hydrogen bonds are displayed with their corresponding distances (*black dashes*).
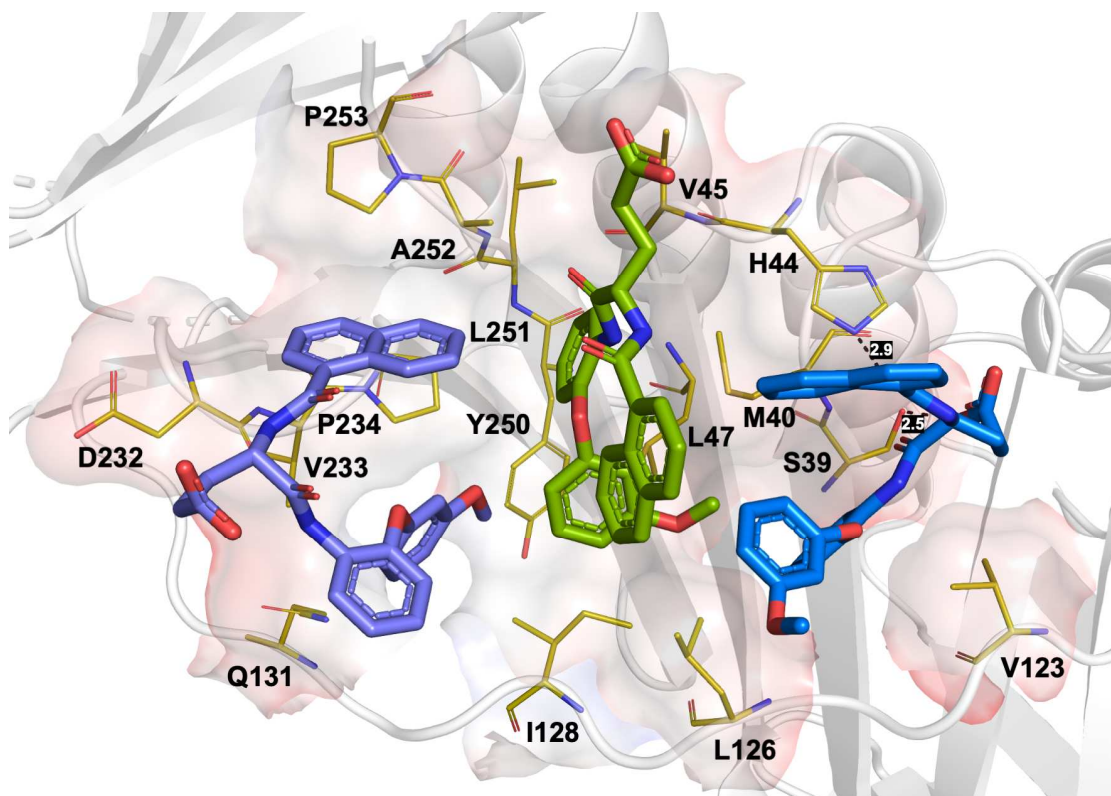
Supplementary Figure 4.02 | AOH1996LE binding interactions. Cartoon representation of PCNA monomer (*white*) with binding site residues in sticks (*emerald*). The Poisson-Boltzmann electrostatic surface potentials are shown in red and blue, corresponding to -5 to +5 kT/e respectively. Three AOH1996LE molecules are shown as sticks (*violet-left, green-center and blue-right*). Hydrogen bonds are displayed with their corresponding distances (*black dashes*).

Chapter 5
Discussion

Chapter 5: Discussion

5.1   Conclusions

The DNA-sliding clamp, PCNA, acts as a vital scaffold to organize numerous components for DNA replication, repair, chromatin formation, and cell cycle progression. Recent studies have expanded upon the nuclear function of PCNA with the identification of a cytosolic form of PCNA. Cytosolic PCNA acts as a scaffold for glycolytic enzymes, to promote energy metabolism and survival of cancer cells. Notably, our collaborators have identified the caPCNA isoform that is highly expressed in cancer cells, but not at significant levels in non-malignant cells. This isoform lacks the PTMs typical of non-malignant, *wt*PCNA, potentially making the PIP-box binding site more accessible to protein-protein interactions. Therefore, the caPCNA isoform and essential binding pocket offer an attractive target to inhibit the proliferative mechanisms of cancer cells for the development of anti-cancer therapeutics. Following this discovery, a caPCNA selective drug discovery campaign was initiated by our collaborators, Dr. Hickey and Dr. Malkas, and this led to the development of the 'first-in-class' small molecule, AOH1160, which inhibits caPCNA hub activity *in vitro*, *in cellulo* and *in vivo* at nanomolar concentrations, while not affecting non-malignant PCNA activity. Further characterization in human liver microsome studies indicated that AOH1160 was not metabolically stable, and thus needed further optimization to develop it into a drug-like lead.

We hypothesized that structural-based drug design (SBDD) techniques could be used to characterize the AOH1160 scaffold binding interactions at atomic resolutions and help guide future caPCNA inhibitor design. Testing our hypothesis, we implemented an

iterative research cycle that began with *in vitro* and *in silico* screening assays to readily detect potential binders for later structural characterization studies. The set of 12 CADD-derived AOH1160 analogs and 63 AI-CADD identified molecules assayed were originally identified from virtual screening studies performed by our collaborators against the PIP-box binding site. The combined virtual screening efforts tested a total of 13 million compounds database that had been pre-filtered for optimized ADME properties. Our *in vitro* TSA screens detected a total of 24 CADD and AI-CADD-based small molecule hits and an additional 1,183 fragment hits that increased protein thermal stability by 0.5 °C or greater. This significant fragment hit rate at 33% could arise from the 3 potential binding sites of the PCNA trimer. To focus on the most promising hits for initial structural characterization studies we have applied a more significant shift cut-off of 2 °C or greater to result in a pool of 245 fragments. These hits have now advanced to crystallization studies alongside our small molecule hits. The efficacy of our TSA screening approach is exemplified in our collaborative XIAP project, which aimed to characterize a set of covalent warheads designed to target surface amino acid residues of the XIAP-BIR3 PPI site. Our analysis readily detected covalent adduct formation, which was indicated by significant shift values of > 20 °C. These studies helped guide chemical development of reactive and stable covalent warheads, which has contributed to the design of novel PPI inhibitors.

To leverage our fragment hit chemistries into rapid small molecule lead generation, we also performed SAR-by-catalog studies to detect 150 small molecule superstructures that were subsequently evaluated via molecular docking. The designed docking analysis

against the PIP-box binding site was first validated by docking the known PIP-box selective inhibitor, T3, which produced a predicted binding energy of -7.55 kcal/mol and 100% clustering. We then selected a test set of 58 brominated fragment hits ($\Delta T_m \geq 1$ °C) as potential SAR-by-catalog candidates. Molecular docking was performed to detect predicted fragment binding to focus on fragments that produced the lowest relative predicted binding energies, which resulted in a total of 9 fragments. Parallel substructure and similarity searches against the eMolecules and ZINC compound databases (775 million) produced 150 small molecule superstructures that were then docked and subsequently ranked. We have presented a list of top 20 hits that produced a binding energy of $\leq$ -7.95 kcal/mol and > 30% clustering. Three of the top 20 docking hits were readily available for purchase and tested via TSA and 2 produced significant positive shifts of > 2.0 °C. Notably, one of these positive shifters produced the most significant shift detected thus far in our small molecule TSA studies, which was + 3.0 °C. Based on the stabilizing interactions detected from our initial SAR-by-catalog studies, we have demonstrated this docking-validated approach can provide a quick hit-to-lead method to develop our fragments into high affinity small molecule binders. Characterization of inhibitor binding interactions through molecular docking approaches was also demonstrated in our collaborative studies with Dr. Nair at Amrita University. We utilized docking analysis to explore the mechanism of oxyresveratrol induced, caspase-independent cancer cell apoptosis. It was hypothesized that oxyresveratrol inhibited caspase-3 activity to promote apoptosis via a caspase-independent signaling pathway in malignant cells. Our docking studies predicted binding energies of -7.77 kcal/mol with 100% clustering of

oxyresveratrol to the S1 pocket of the caspase-3 catalytic site. This predicted energy, which falls within hit range for Autodock molecular docking characterization, indicates binding interaction and potential inhibition of caspase-3 activation. This *in silico* data was supported by *in cellulo* assays performed by our collaborators, which did not detect the formation of caspase-3 fragments following treatment with oxyresveratrol, thus indicating caspase-3 activation was inhibited in the presence of oxyresveratrol.

Utilizing macromolecular crystallography, we have structurally characterized the co-crystal complexes of three AOH1160 analogs and one AI-CADD identified small molecule binder. Due to the hydrophobic characteristics of the City of Hope CADD-based molecules, exhaustive sparse matrix screening studies were performed to obtain stable crystal structures, however a majority of the analogs were too hydrophobic for crystallization. Advantageously, the AOH1160 derivatives, LE and DE were soluble in 10% DMSO, which was effective for crystallization and enabled successful production of co-crystal complexes for both molecules that were resolved to 2.85 and 3.69 Å respectively. These structures are the first to elucidate the novel AOH1160 scaffold interaction with the PIP-box binding site. Of the three individual molecules modeled in the resolved co-crystal structures, the 'central' molecule is likely representative of the AOH-scaffold binding mode. The 'central' molecule's di-phenyl ester moiety binds similarly to and occupies identical three-dimensional volume to the iodo-groups of known PIP-box inhibitor, T3. In addition, the 'central' binding mode interacts with the hydrophobic residues of the PIP-box cavity that are known to be essential for binding PIP-box and APIM

motif peptides. While the 'left-handed' and 'right-handed' binding modes are mirrors of one another that are generated through crystal symmetry operations.

The CADD-derived AOH1160 analog, AOH1996, includes a methoxy group extension from the di-phenyl ester moiety of the parent scaffold to improve metabolic stability. Since the parent scaffold structure had been altered, our collaborators synthesized a AOH1996-LE derivative for structural characterization to determine the effect on compound binding interaction. Utilizing an identical co-crystallization approach, we have successfully resolved the PCNA:AOH1996LE complex to 3.7 Å. Notably, the additional methoxy did not significantly affect the binding mode. Identical to the AOH1160 derivative co-crystal assemblies, the PCNA:AOH1996-LE structure was resolved with 3 individual compounds bound in the PIP-box and adjacent pockets to facilitate crystal lattice stacking interactions. The fourth co-crystal structure characterized the binding interaction of the AI-CADD generated B05 molecule and was resolved to 1.9 Å. Unlike the AOH derivative molecule binding mode, only one B05 molecule associates the PIP-box binding cavity and does not extend into the adjacent pocket, very likely due to variation in co-crystallization conditions between the two scaffolds. As indicated by the electron density maps, B05 occupies a majority of the hydrophobic PIP-box cavity to form the characterized PIP-motif interactions typically formed by the hydrophobic PIP-box and APIM- motifs, and a majority of the interactions formed by the central AOH-molecule. In addition to our characterized co-crystal structures, we have resolved two *apo*-crystal structures that diffracted to resolutions of 2.44 and 2.01 Å respectively. Notably, both structures formed a crystal lattice assembly that leaves the PIP-box binding pocket unoccupied for potential

binding interaction, and thus will be utilized in the near-term for our structural characterization studies of our fragment hits.

5.2   Future Directions

The structural characterization of the AOH1160 derivatives and AOH1996LE analog molecules has provided valuable information for the generation of a reliable SAR model for subsequent lead design. Subsequent studies performed by our collaborators have revealed that the AOH1996 scaffold is metabolically stable as compared to the AOH1160 parent molecule while also maintaining potency, indicating superior therapeutic activity (data not shown). Notably the AOH1996 molecule has recently entered Phase I clinical trials. However, there is a high failure rate (~90%) of compounds heading into the clinic, so we are continuing investigation to improve upon both the AOH1160 and AOH1996 scaffolds to generate additional analogs that can be fed into the now established path to the clinic. Notably, the AI-CADD generated B05 molecule presents a novel chemical scaffold for selective PIP-box cavity interaction. This molecule was tested alongside the AOH1996 compound in cell viability assays and produced IC50s that indicated approximately 20-times greater potency (data not shown). However, comparison of the therapeutic index, or ratio of drug concentration that has therapeutic effect as compared to the concentration that induces toxicity, is worse for the B05 molecule (data not shown). Based on the potency of this molecule our collaborators at Atomwise are now performing follow-up *in silico* investigation to generate B05 analogs that will be funneled into an alternative lead design campaign from the AOH-scaffold investigation. However, both avenues will supply information in regards to SAR and identify chemical characteristics that contribute to the

specificity of the AOH molecules as compared to the B05 scaffold. From our extensive *in vitro* and *in silico* studies have provided a large pool of small molecule and fragment structures that will be used to either further optimize our AOH-scaffold and B05-scaffold analogs, or even investigate the development of PIP-box selective inhibitors in different areas of chemical space.

We are still working to develop a detailed binding assay that will be compatible with the hydrophobicity of our PIP-box selective inhibitors. Common binding affinity assays such as ITC require compound solubility in an aqueous environment to buffer match to the protein target of interest. Since the AOH and B05 molecules are very hydrophobic (cLogP= 4.5 - 6) we are not able to implement ITC at this time. Although not shown here, we have tested dissociation-enhanced lanthanide fluorescence immunoassays (DELFIA) assays to assess compound:target interaction through competitive interaction with a PIP-box motif peptide. However, initial studies indicated that the compounds were not stable in the 1% DMSO reaction environment and precipitated. In an attempt to promote compound solubility, we have increased the final DMSO concentration 4-fold, but still observed compound precipitation. Working at DMSO concentrations higher than this disrupted the assay, potentially due to PCNA protein being less stable/unfolding, thus indicating this assay could not be used to test our hydrophobic inhibitors. We aim to leverage the more soluble AOH-derivatives and investigate B05 derivative synthesis to determine if improved compound solubility characteristics will remedy this issue and provide us with binding affinity data to influence our future analog design. However, *in cellulo* IC50 values have been conducted on all of these compounds by our collaborators

at the City of Hope. Currently, we are working with our collaborating medicinal chemists at City of Hope and Atomwise to produce these new molecules. With the help of our collaborators, we will continue to combined SBDD data with the on-going *in cellulo* characterization studies to develop potent drug-like leads to contribute to the development of further novel anti-cancer therapeutics targeting caPCNA.

APPENDICES
Appendix A: Melt Curve and Derivative plots for Table 2.04.

wt XIAP-BIR3/Compound 2

2h, 25 °C Incubation

apo T_m = 39.5 °C
$T_{m\,2}$ = 70.5 °C
ΔT_m = 31 °C

6h, 37 °C Incubation
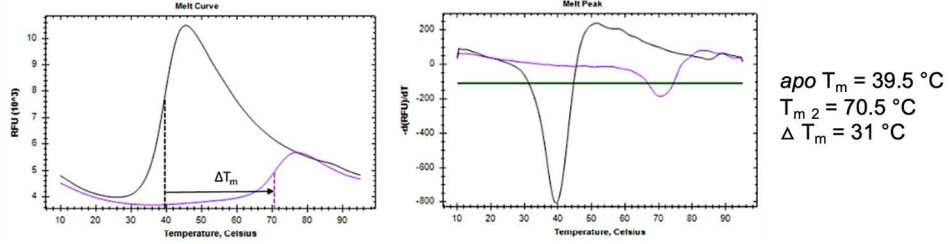
apo T_m = 39.5 °C
$T_{m\,2}$ = 76.5 °C
ΔT_m = 37 °C
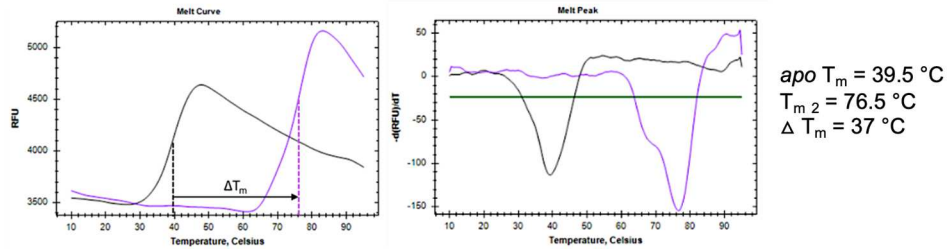
wt XIAP-BIR3/Compound 3

2h, 25 °C Incubation

apo T_m = 39.5 °C
$T_{m\,2}$ = 45/75 °C
ΔT_m = 5.5/35.5 °C

6h, 37 °C Incubation

apo T_m = 39.5 °C
$T_{m\,2}$ = 46.5/74.5 °C
ΔT_m = 7/35 °C

161
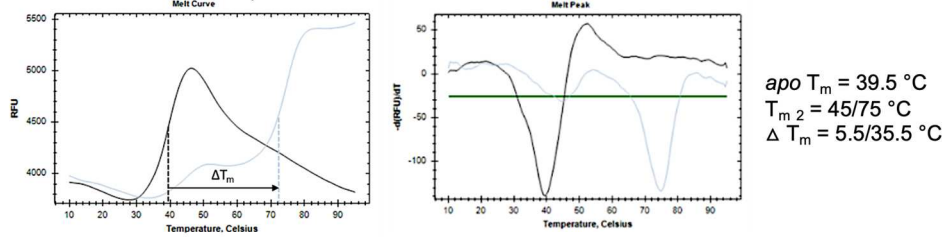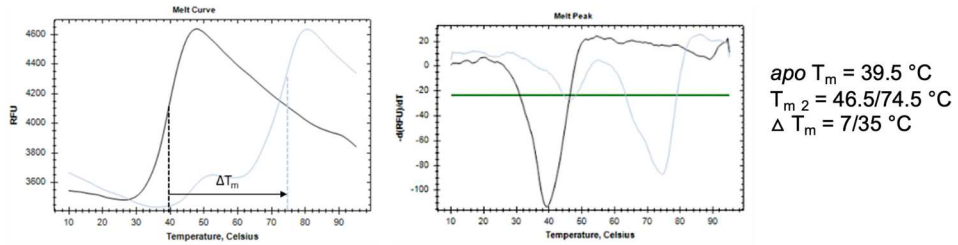
wt XIAP-BIR3/Compound 4

2h, 25 °C Incubation
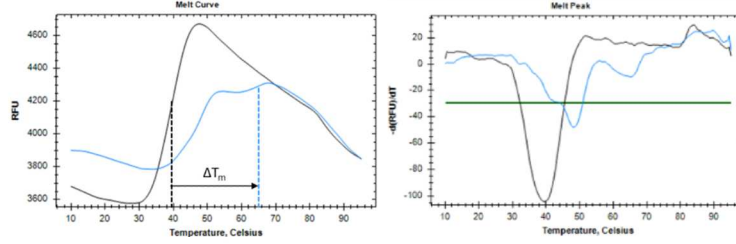
apo T_m = 39.5 °C
$T_{m\,2}$ = 48/65 °C
ΔT_m = 8.5/25.5 °C

6h, 37 °C Incubation
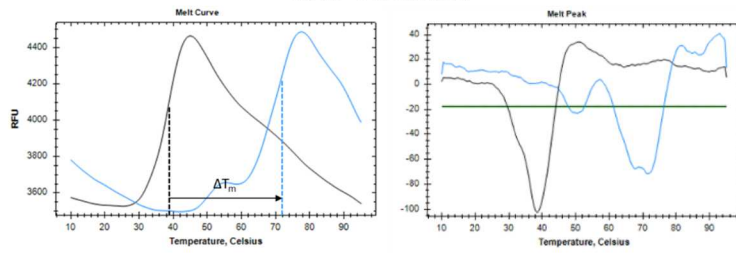
apo T_m = 38.5 °C
$T_{m\,2}$ = 51/71.5 °C
ΔT_m = 12.5/33 °C

wt XIAP-BIR3/Compound 5

2h, 25 °C Incubation
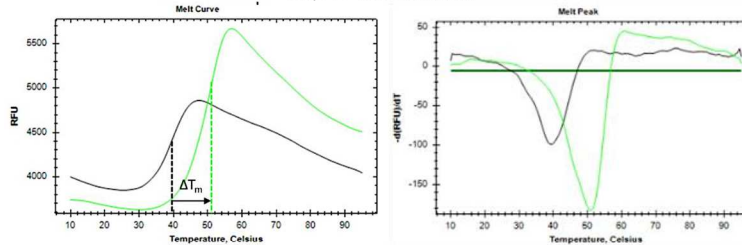
apo T_m = 39.5 °C
$T_{m\,2}$ = 51.0 °C
ΔT_m = 11.5 °C
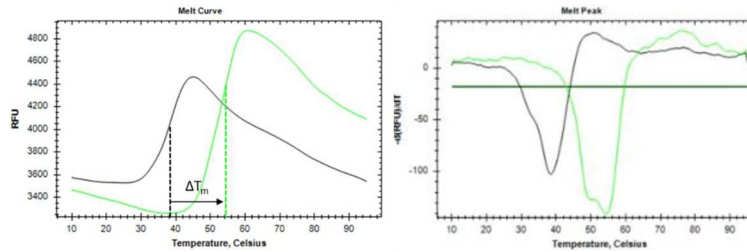
6h, 37 °C Incubation

apo T_m = 38.5 °C
$T_{m\,2}$ = 50/54.5 °C
ΔT_m = 11.5/16 °C

162

wt XIAP-BIR3/Compound 6

2h, 25 °C Incubation

apo T_m = 39.5 °C
$T_{m\ 2}$ = 51.0 °C
ΔT_m = 11.5 °C

6h, 37 °C Incubation
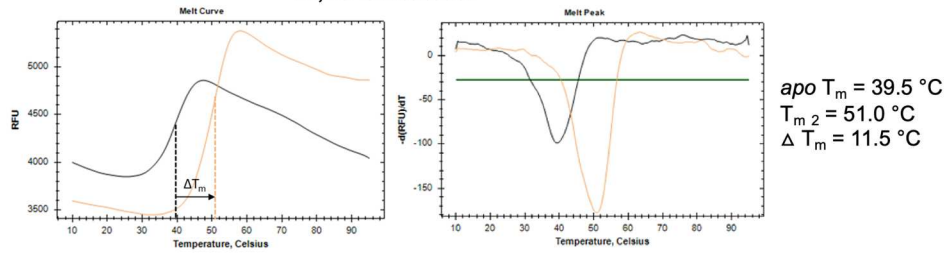
apo T_m = 38.5 °C
$T_{m\ 2}$ = 53.5 °C
ΔT_m = 15 °C
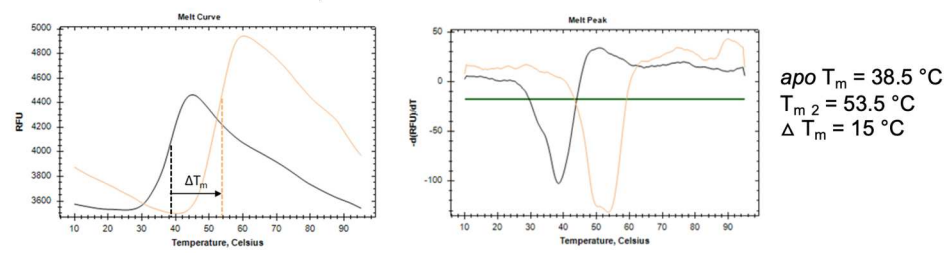
wt XIAP-BIR3/Compound 7

2h, 25 °C Incubation

apo T_m = 39.5 °C
$T_{m\ 2}$ = 44/70 °C
ΔT_m = 4.5/30.5 °C

6h, 37 °C Incubation

apo T_m = 38.5 °C
$T_{m\ 2}$ = 48/69.5 °C
ΔT_m = 9.5/31 °C

163
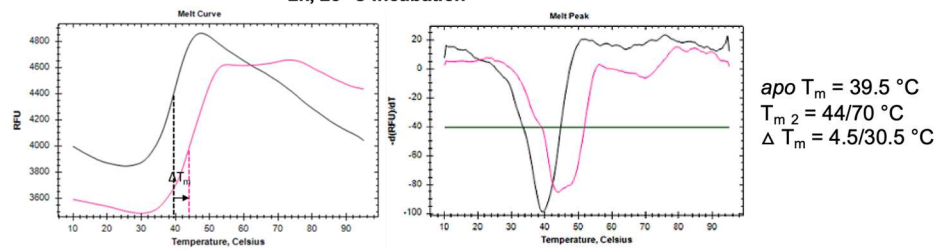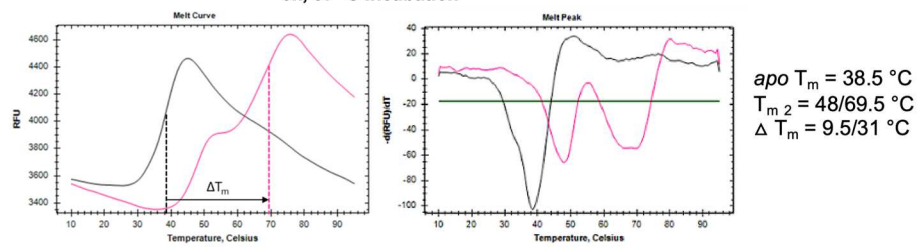
wt XIAP-BIR3/Compound 8

2h, 25 °C Incubation
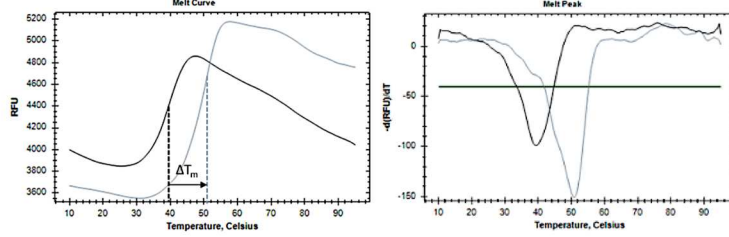
apo T_m = 39.5 °C
$T_{m\,2}$ = 51 °C
ΔT_m = 11.5 °C
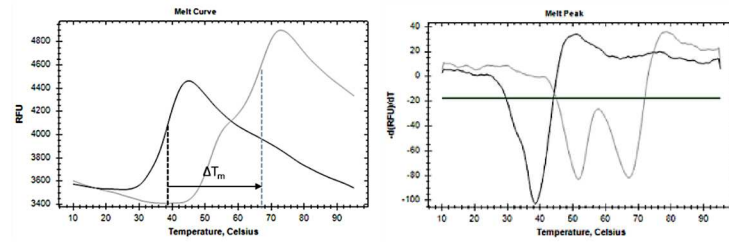
6h, 37 °C Incubation

apo T_m = 38.5 °C
$T_{m\,2}$ = 51.5/67 °C
ΔT_m = 13/28.5 °C

XIAP-BIR3 K311A/Compound 2

2h, 25 °C Incubation
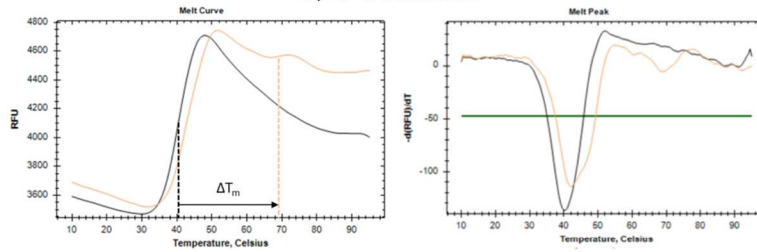
apo T_m = 40.5 °C
$T_{m\,2}$ = 42/69 °C
ΔT_m = 1.5/28.5 °C

6h, 37 °C Incubation

apo T_m = 39 °C
$T_{m\,2}$ = 47.5 °C
ΔT_m = 8.5 °C

XIAP-BIR3 K311A/Compound 3

2h, 25 °C Incubation
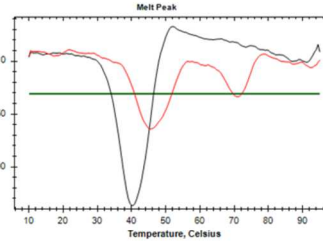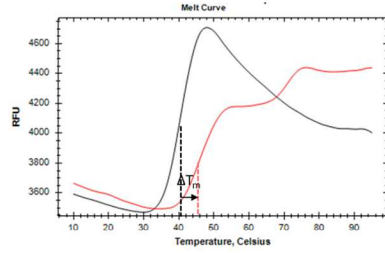
apo T_m = 40.5 °C
$T_{m\,2}$ = 45.5/71 °C
ΔT_m = 5/30.5 °C

6h, 37 °C Incubation
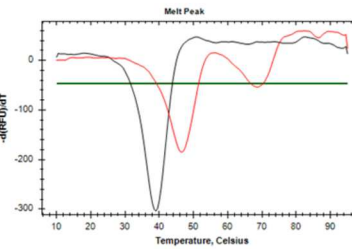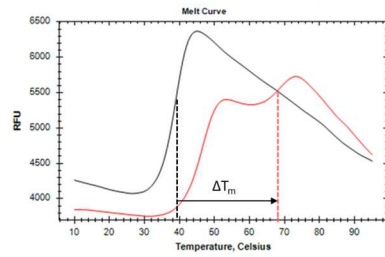
apo T_m = 39 °C
$T_{m\,2}$ = 46.5/68 °C
ΔT_m = 7.5/38 °C

XIAP-BIR3 K311A/Compound 4

2h, 25 °C Incubation

apo T_m = 40.5 °C
$T_{m\,2}$ = 51.5 °C
ΔT_m = 11 °C

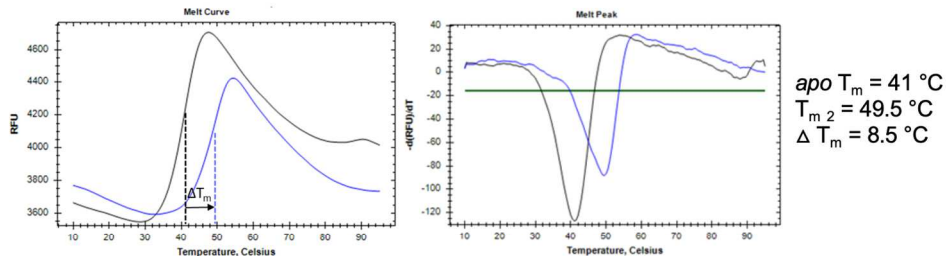## 6h, 37 °C Incubation

apo T_m = 39 °C
$T_{m\,2}$ = 52.5 °C
ΔT_m = 13.5 °C
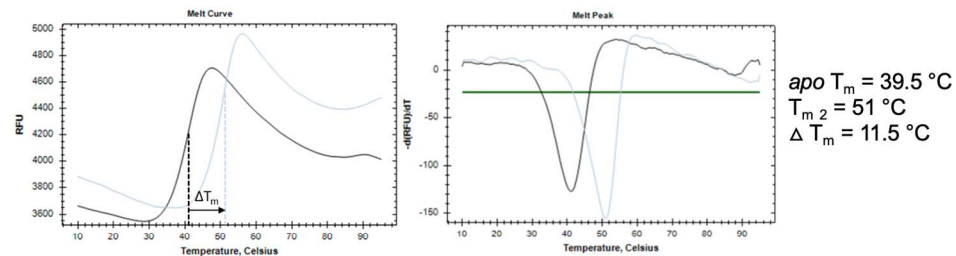
XIAP-BIR3 K311A/Compound 5

2h, 25 °C Incubation
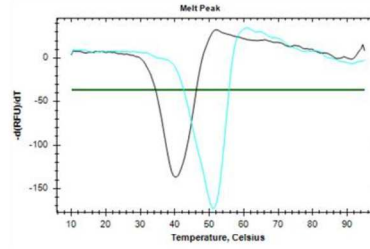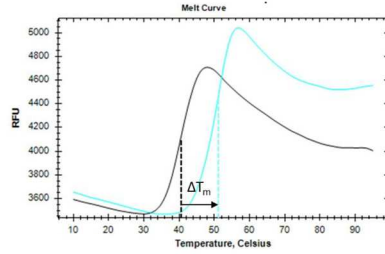
apo T_m = 40.5 °C
$T_{m\,2}$ = 51 °C
$\Delta\,T_m$ = 10.5 °C
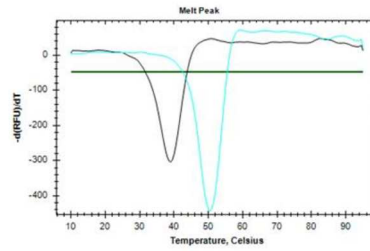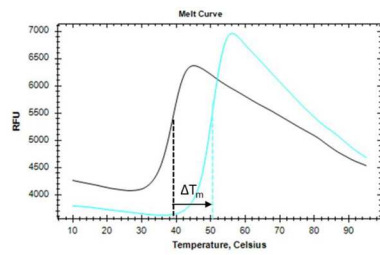
6h, 37 °C Incubation

apo T_m = 39 °C
$T_{m\,2}$ = 50.5 °C
$\Delta\,T_m$ = 11.5 °C

XIAP-BIR3 K311A/Compound 6

2h, 25 °C Incubation

apo T_m = 40.5 °C
$T_{m\,2}$ = 51 °C
$\Delta\,T_m$ = 10.5 °C

6h, 37 °C Incubation

apo T_m = 39 °C
$T_{m\,2}$ = 50.5 °C
$\Delta\,T_m$ = 11.5 °C

166

XIAP-BIR3 K311A/Compound 7

2h, 25 °C Incubation
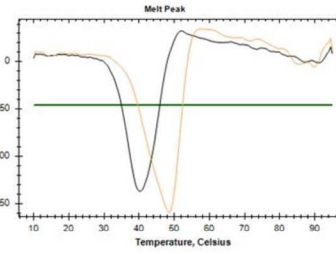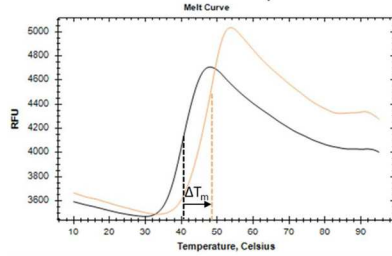
apo T_m = 40.5 °C
$T_{m\,2}$ = 58.5 °C
$\triangle\,T_m$ = 8 °C

6h, 37 °C Incubation
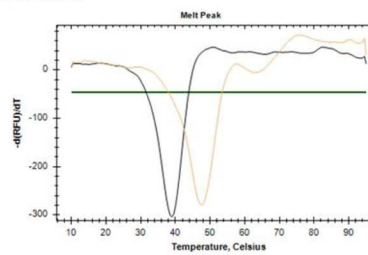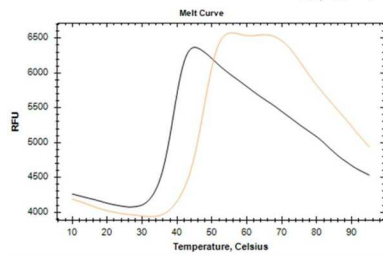
apo T_m = 39 °C
$T_{m\,2}$ = 47.5 °C
$\triangle\,T_m$ = 8.5 °C

XIAP-BIR3 K311A/Compound 8

2h, 25 °C Incubation
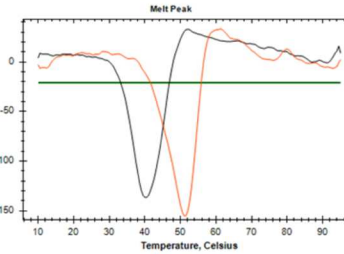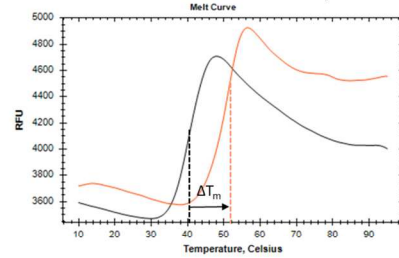
apo T_m = 40.5 °C
$T_{m\,2}$ = 51.5 °C
$\triangle\,T_m$ = 11 °C

6h, 37 °C Incubation
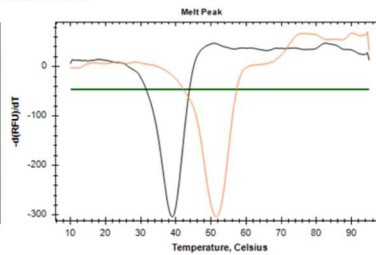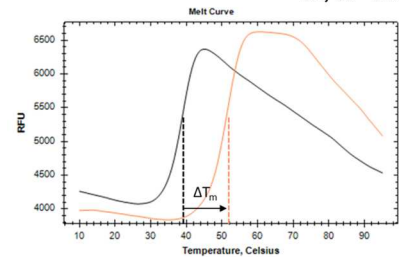
apo T_m = 39 °C
$T_{m\,2}$ = 51.5 °C
$\triangle\,T_m$ = 12.5 °C

XIAP-BIR3 K311Y/Compound 2

2h, 25 °C Incubation
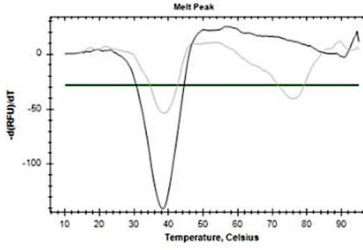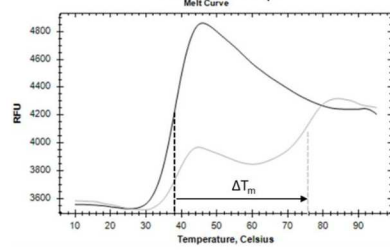
apo T_m = 38.5 °C
$T_{m\,2}$ = 76 °C
Δ T_m = 37.5 °C

6h, 37 °C Incubation
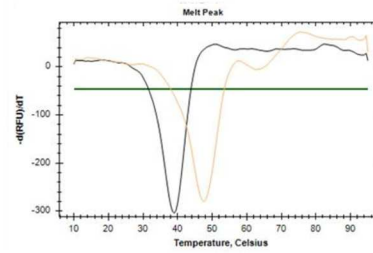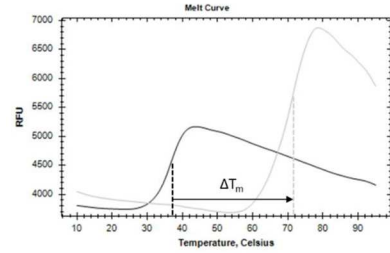
apo T_m = 37 °C
$T_{m\,2}$ = 71.5 °C
Δ T_m = 34.5 °C

XIAP-BIR3 K311Y/Compound 3

2h, 25 °C Incubation
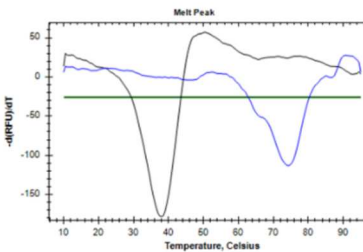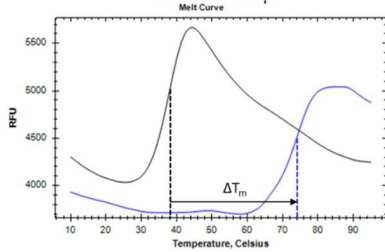
apo T_m = 38 °C
$T_{m\,2}$ = 74 °C
Δ T_m = 36 °C
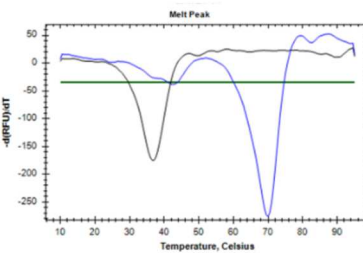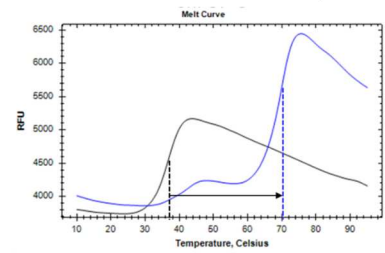
6h, 37 °C Incubation

apo T_m = 37 °C
$T_{m\,2}$ = 42.5/70 °C
Δ T_m = 5.5/33 °C

XIAP-BIR3 K311Y/Compound 4

2h, 25 °C Incubation
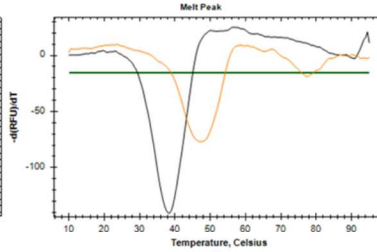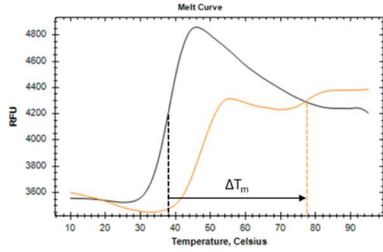
apo T_m = 38 °C
$T_{m\,2}$ = 74 °C
ΔT_m = 36 °C
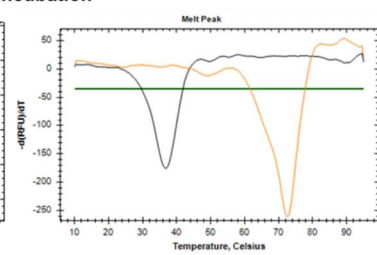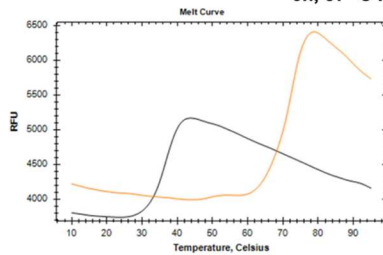
6h, 37 °C Incubation

apo T_m = 37 °C
$T_{m\,2}$ = 72.5 °C
ΔT_m = 35.5 °C

XIAP-BIR3 K311Y/Compound 5

2h, 25 °C Incubation

apo T_m = 38.5 °C
$T_{m\,2}$ = 49 °C
ΔT_m = 10.5 °C

6h, 37 °C Incubation

apo T_m = 37 °C
$T_{m\,2}$ = 48.5 °C
ΔT_m = 11.5 °C

XIAP-BIR3 K311Y/Compound 6

2h, 25 °C Incubation

apo T_m = 38.5 °C
$T_{m\,2}$ = 49 °C
ΔT_m = 10.5 °C

6h, 37 °C Incubation

apo T_m = 37 °C
$T_{m\,2}$ = 48.5 °C
ΔT_m = 11.5 °C

XIAP-BIR3 K311Y/Compound 7

2h, 25 °C Incubation
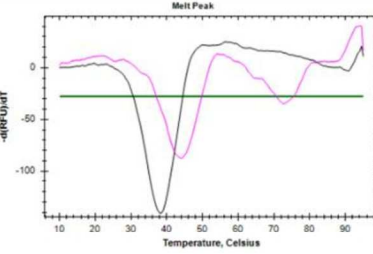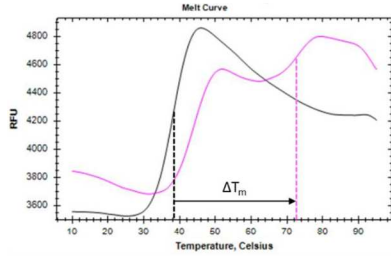
apo T_m = 38.5 °C
$T_{m\,2}$ = 44/72.5 °C
ΔT_m = 5.5/34 °C

6h, 37 °C Incubation

apo T_m = 37 °C
$T_{m\,2}$ = 69 °C
ΔT_m = 32 °C

170
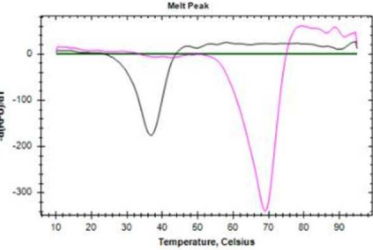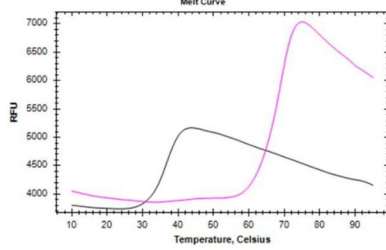
XIAP-BIR3 K311Y/Compound 8

2h, 25 °C Incubation
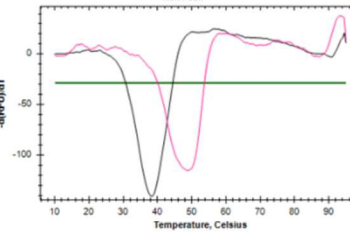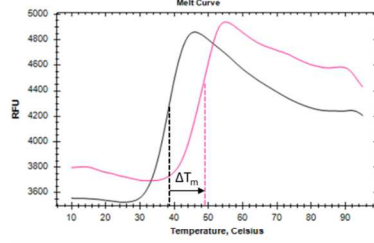
apo T_m = 38.5 °C
$T_{m\,2}$ = 49 °C
$\triangle T_m$ = 10.5 °C

6h, 37 °C Incubation
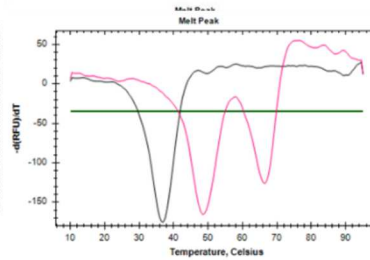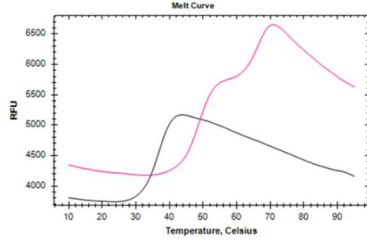
apo T_m = 37 °C
$T_{m\,2}$ = 48.5/66.5 °C
$\triangle T_m$ = 11.5/29.5 °C

XIAP-BIR3 K311H/Compound 2

2h, 25 °C Incubation
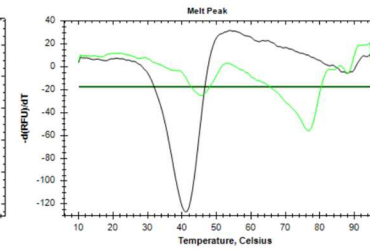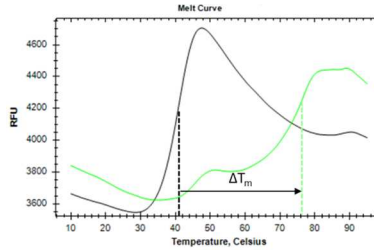
apo T_m = 41 °C
$T_{m\,2}$ = 45.5/76.5 °C
$\triangle T_m$ = 4.5/35.5 °C

6h, 37 °C Incubation
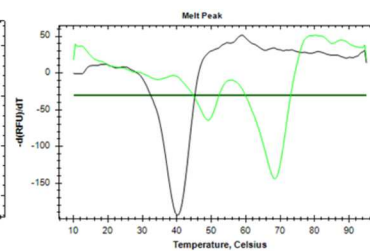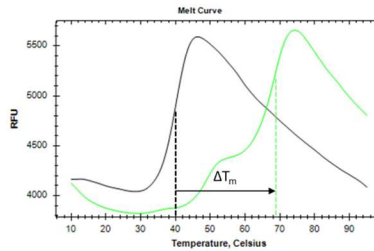
apo T_m = 37 °C
$T_{m\,2}$ = 49/68.5 °C
$\triangle T_m$ = 9/28.5 °C

171

XIAP-BIR3 K311H/Compound 3

2h, 25 °C Incubation
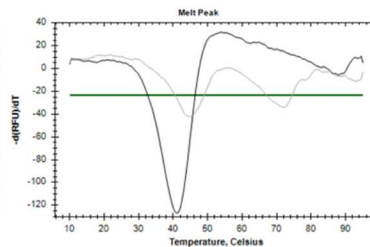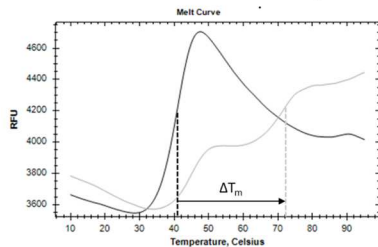
apo T_m = 41 °C
$T_{m\,2}$ = 44.5/72 °C
ΔT_m = 3.5/31 °C

6h, 37 °C Incubation
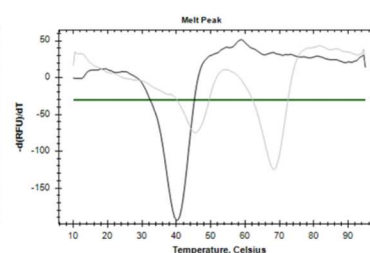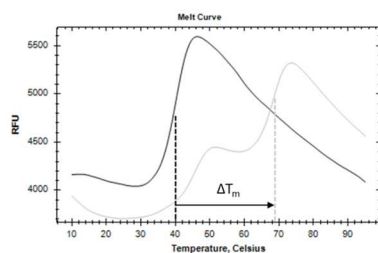
apo T_m = 40 °C
$T_{m\,2}$ = 45/68.5 °C
ΔT_m = 5/28.5 °C

XIAP-BIR3 K311H/Compound 4

2h, 25 °C Incubation

apo T_m = 41 °C
$T_{m\,2}$ = 52 °C
ΔT_m = 11 °C

6h, 37 °C Incubation

apo T_m = 40 °C
$T_{m\,2}$ = 52.5 °C
ΔT_m = 12.5 °C

172
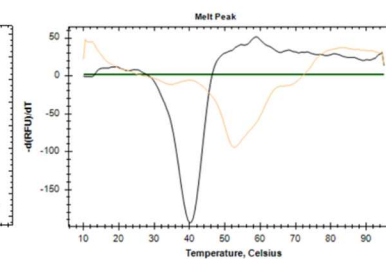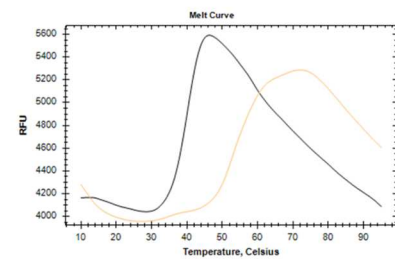
XIAP-BIR3 K311H/Compound 5

2h, 25 °C Incubation

Melt Curve

Melt Peak

apo T_m = 41 °C
$T_{m\,2}$ = 49.5 °C
$\triangle T_m$ = 8.5 °C

6h, 37 °C Incubation

Melt Curve

Melt Peak

apo T_m = 40 °C
$T_{m\,2}$ = 48.5 °C
$\triangle T_m$ = 8.5 °C
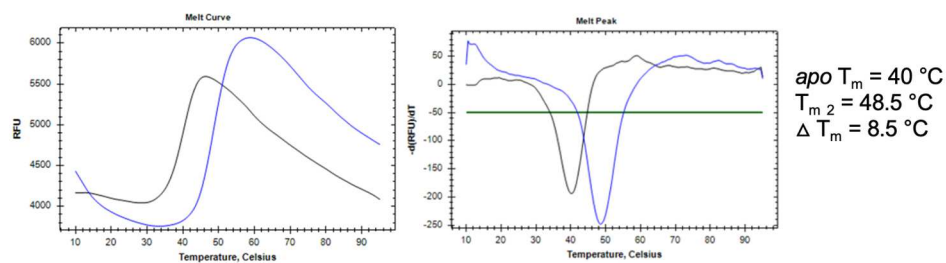
XIAP-BIR3 K311H/Compound 6

2h, 25 °C Incubation

Melt Curve

Melt Peak

apo T_m = 39.5 °C
$T_{m\,2}$ = 51 °C
$\triangle T_m$ = 11.5 °C

6h, 37 °C Incubation

Melt Curve

Melt Peak

apo T_m = 40 °C
$T_{m\,2}$ = 52 °C
$\triangle T_m$ = 12 °C
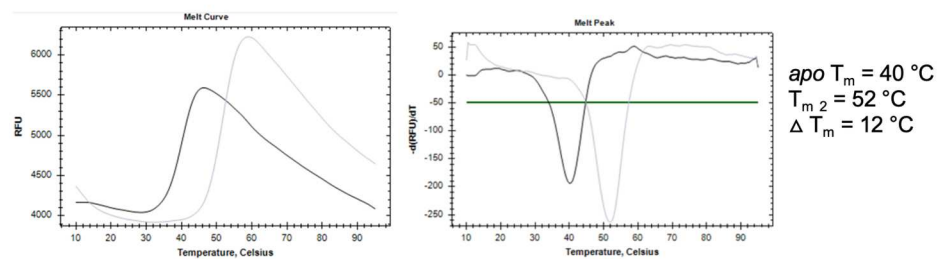
XIAP-BIR3 K311H/Compound 7

2h, 25 °C Incubation
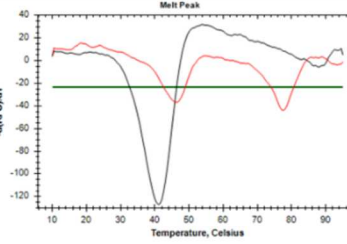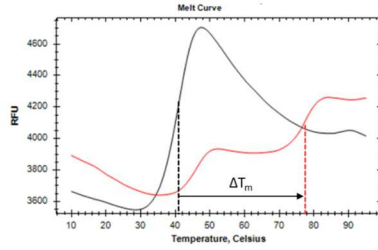
apo T_m = 41 °C
$T_{m\,2}$ = 46.5/77.5 °C
ΔT_m = 5.5/36.5 °C

6h, 37 °C Incubation
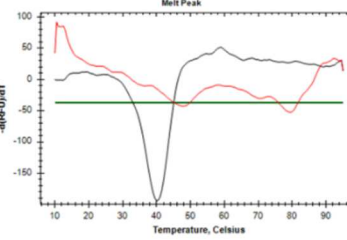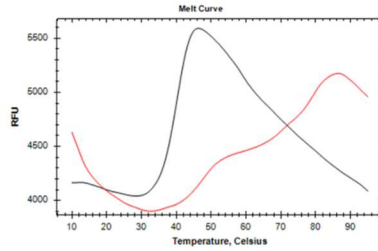
apo T_m = 40 °C
$T_{m\,2}$ = 47.5/79.5 °C
ΔT_m = 7.5/39.5 °C

XIAP-BIR3 K311H/Compound 8

2h, 25 °C Incubation
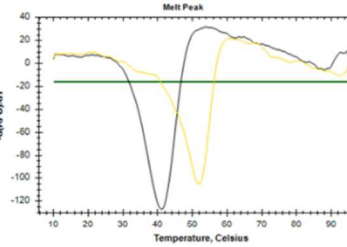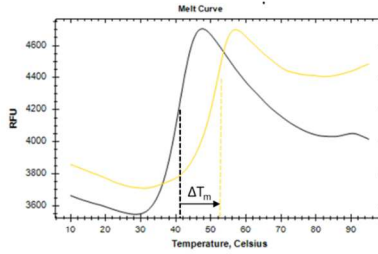
apo T_m = 41 °C
$T_{m\,2}$ = 52 °C
ΔT_m = 11 °C
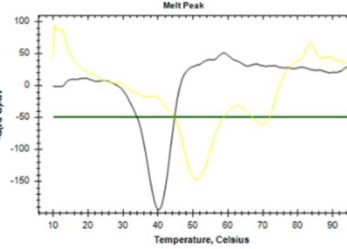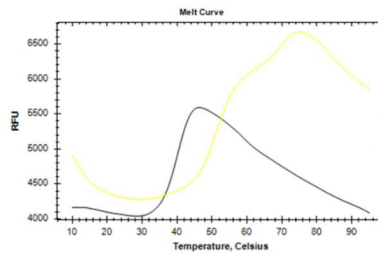
6h, 37 °C Incubation

apo T_m = 40 °C
$T_{m\,2}$ = 51/69.5 °C
ΔT_m = 11/29.5 °C

XIAP-BIR3 K311T/Compound 2

2h, 25 °C Incubation
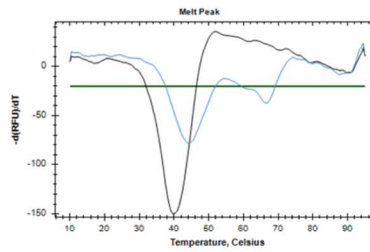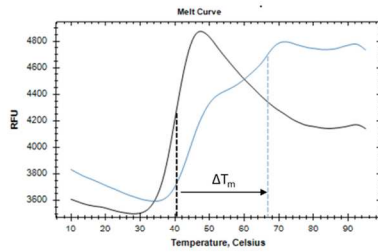
apo T_m = 40 °C
$T_{m\,2}$ = 44/66.5 °C
$\Delta\, T_m$ = 4/26.5 °C

6h, 37 °C Incubation
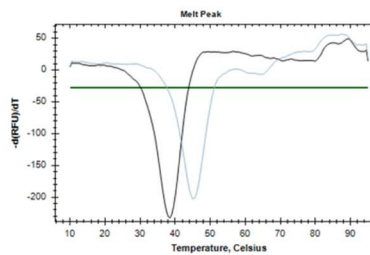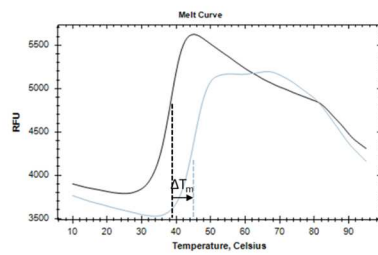
apo T_m = 38.5 °C
$T_{m\,2}$ = 45 °C
$\Delta\, T_m$ = 6.5 °C

XIAP-BIR3 K311T/Compound 3

2h, 25 °C Incubation
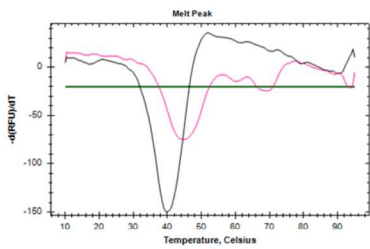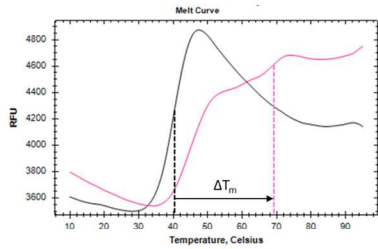
apo T_m = 40 °C
$T_{m\,2}$ = 45/69 °C
$\Delta\, T_m$ = 5/29 °C

6h, 37 °C Incubation
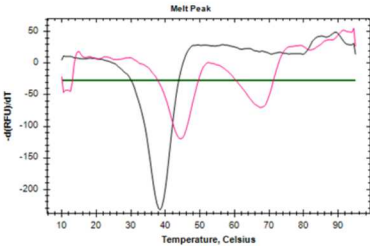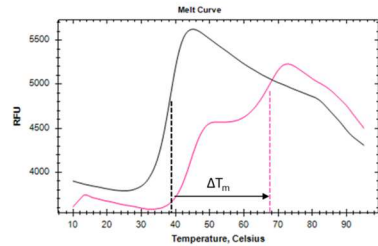
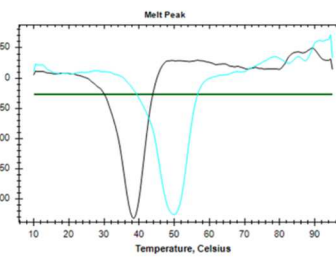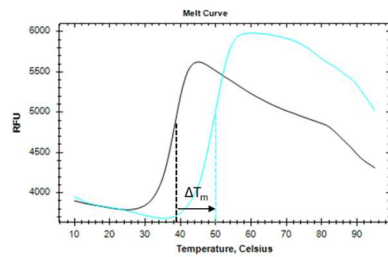*apo* T_m = 38.5 °C
$T_{m\,2}$ = 44.5/67.5 °C
$\Delta\, T_m$ = 6/29 °C

XIAP-BIR3 K311T/Compound 4

2h, 25 °C Incubation
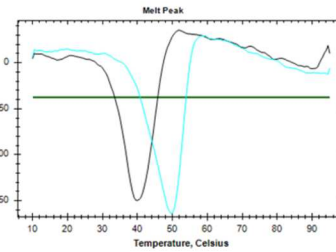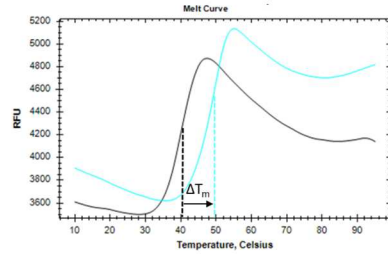
apo T_m = 40 °C
$T_{m\,2}$ = 49.5 °C
ΔT_m = 9.5 °C

6h, 37 °C Incubation

apo T_m = 38.5 °C
$T_{m\,2}$ = 50 °C
ΔT_m = 11.5 °C

XIAP-BIR3 K311T/Compound 5

2h, 25 °C Incubation

apo T_m = 40 °C
$T_{m\,2}$ = 48.5 °C
ΔT_m = 8.5 °C

6h, 37 °C Incubation

apo T_m = 38.5 °C
$T_{m\,2}$ = 48 °C
ΔT_m = 9.5 °C

XIAP-BIR3 K311T/Compound 6
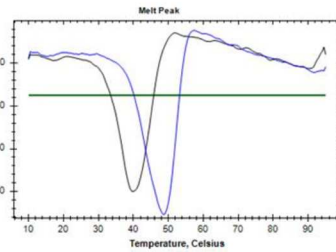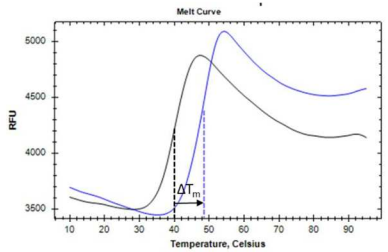
2h, 25 °C Incubation
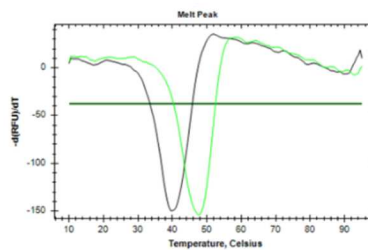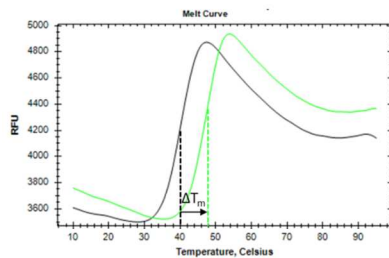
apo T_m = 40 °C
$T_{m\,2}$ = 47.5 °C
$\Delta\,T_m$ = 7.5 °C

6h, 37 °C Incubation
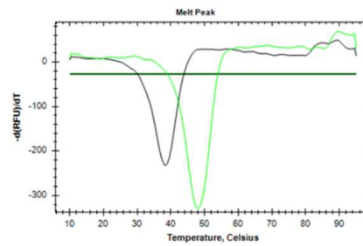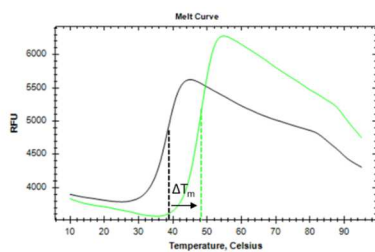
apo T_m = 38.5 °C
$T_{m\,2}$ = 48 °C
$\Delta\,T_m$ = 9.5 °C

XIAP-BIR3 K311T/Compound 7

2h, 25 °C Incubation
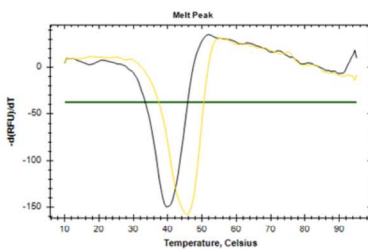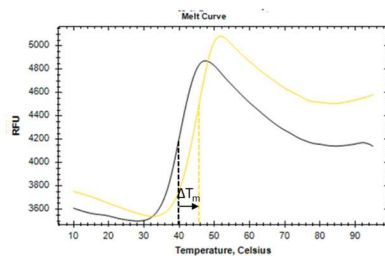
apo T_m = 40 °C
$T_{m\,2}$ = 45.5 °C
$\Delta\,T_m$ = 5.5 °C

6h, 37 °C Incubation
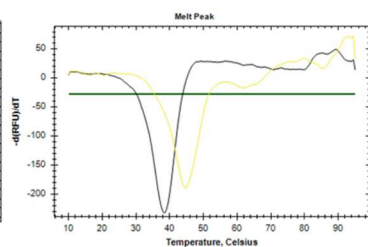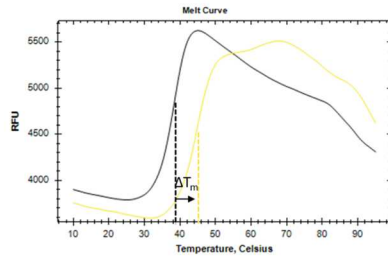
apo T_m = 38.5 °C
$T_{m\,2}$ = 45 °C
$\Delta\,T_m$ = 6.5 °C

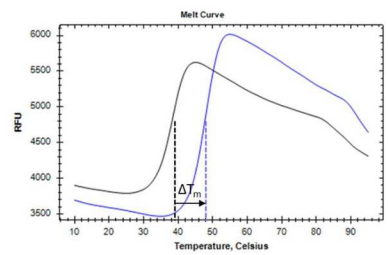# XIAP-BIR3 K311T/Compound 8

2h, 25 °C Incubation

apo T_m = 40 °C
$T_{m\,2}$ = 48.5 °C
ΔT_m = 8.5 °C

6h, 37 °C Incubation
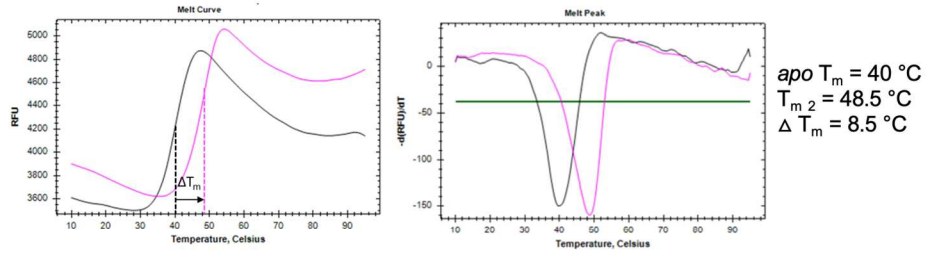
apo T_m = 38.5 °C
$T_{m\,2}$ = 48.5 °C
ΔT_m = 10 °C
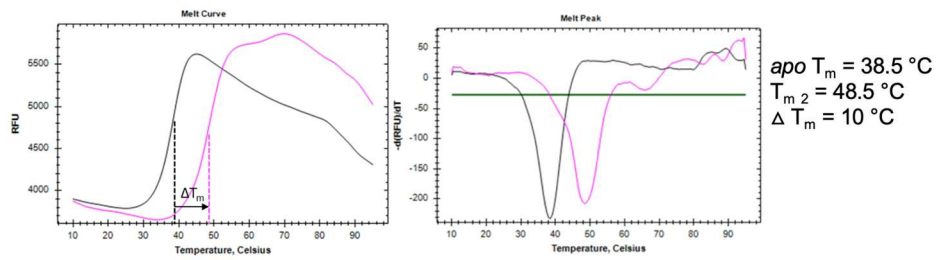
XIAP-BIR3 K311S/Compound 2

2h, 25 °C Incubation

apo T_m = 39 °C
$T_{m\,2}$ = 44/66.5 °C
ΔT_m = 5/27.5 °C

6h, 37 °C Incubation
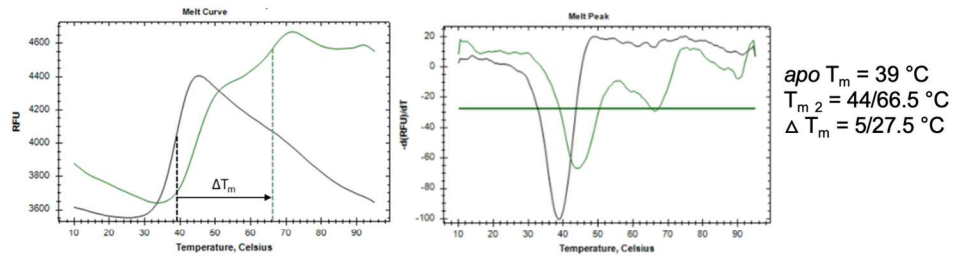
apo T_m = 38.5 °C
$T_{m\,2}$ = 49.5 °C
ΔT_m = 11 °C
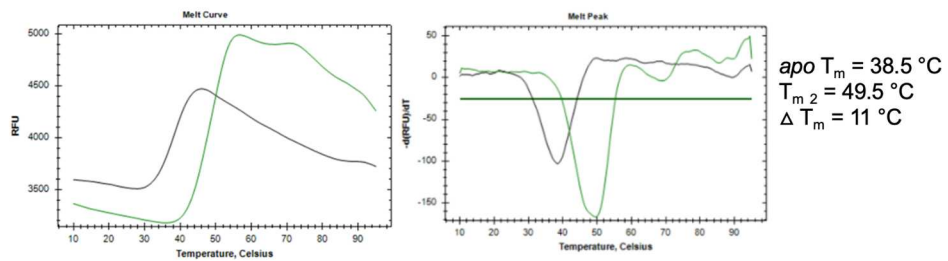
XIAP-BIR3 K311S/Compound 3

2h, 25 °C Incubation
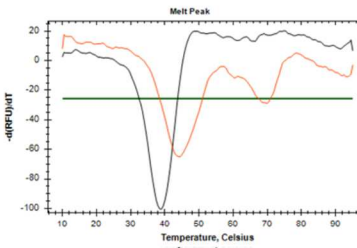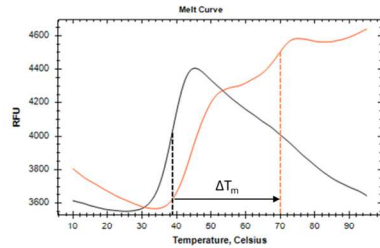
apo T_m = 39 °C
$T_{m\,2}$ = 44/70 °C
$\triangle T_m$ = 5/31 °C

6h, 37 °C Incubation
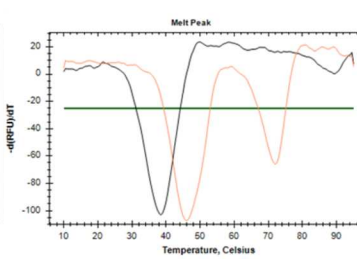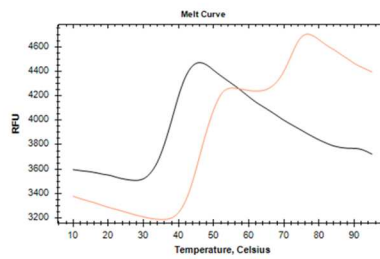
apo T_m = 38.5 °C
$T_{m\,2}$ = 46/72 °C
$\triangle T_m$ = 7.5/33.5 °C

XIAP-BIR3 K311S/Compound 4

2h, 25 °C Incubation
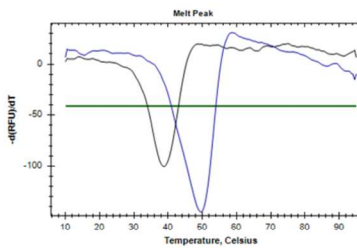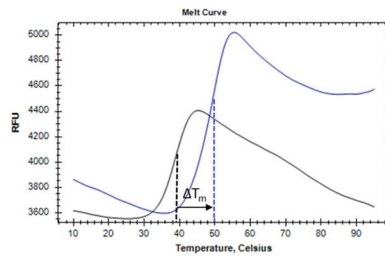
apo T_m = 39 °C
$T_{m\,2}$ = 49.5 °C
$\triangle T_m$ = 10.5 °C
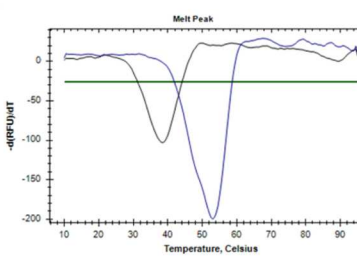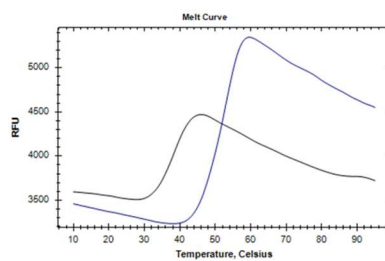
6h, 37 °C Incubation

apo T_m = 38.5 °C
$T_{m\,2}$ = 53 °C
$\triangle T_m$ = 14.5 °C

XIAP-BIR3 K311S/Compound 5

2h, 25 °C Incubation
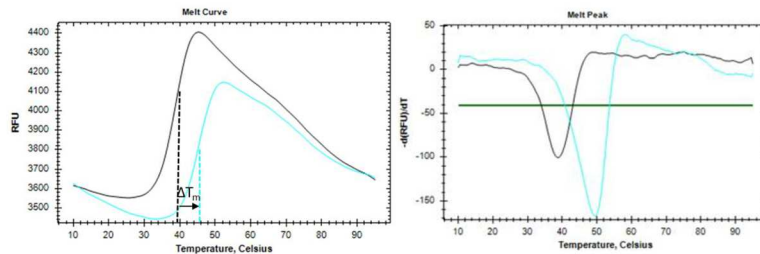
apo T_m = 39 °C
$T_{m\,2}$ = 45.5 °C
ΔT_m = 6.5 °C

6h, 37 °C Incubation
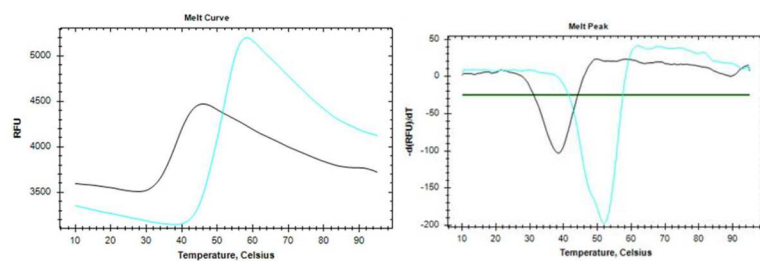
apo T_m = 38.5 °C
$T_{m\,2}$ = 52 °C
ΔT_m = 13.5 °C

XIAP-BIR3 K311S/Compound 6

2h, 25 °C Incubation
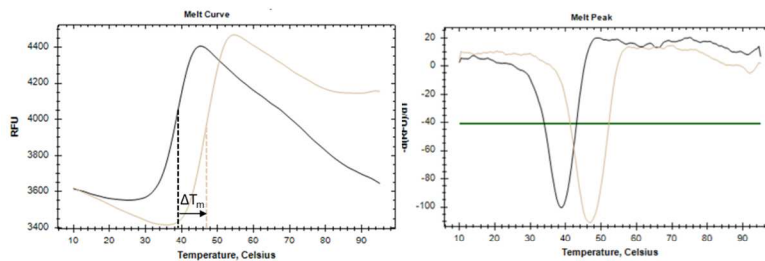
apo T_m = 39 °C
$T_{m\,2}$ = 47 °C
ΔT_m = 8 °C
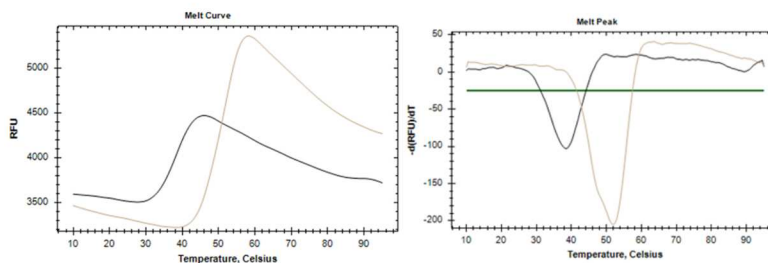
6h, 37 °C Incubation

apo T_m = 38.5 °C
$T_{m\,2}$ = 52 °C
ΔT_m = 13.5 °C

XIAP-BIR3 K311S/Compound 7

2h, 25 °C Incubation
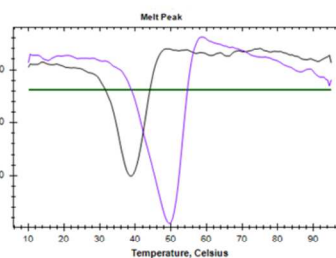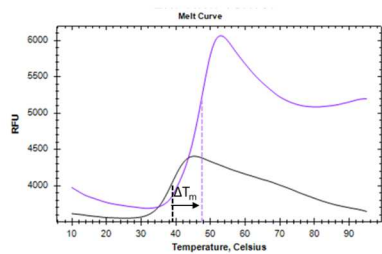
apo T_m = 39 °C
$T_{m\,2}$ = 47.5 °C
$\Delta\,T_m$ = 8.5 °C

6h, 37 °C Incubation
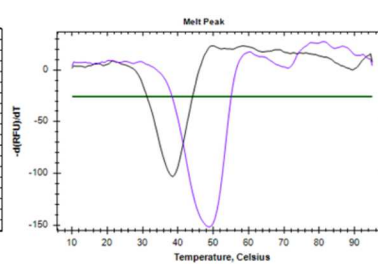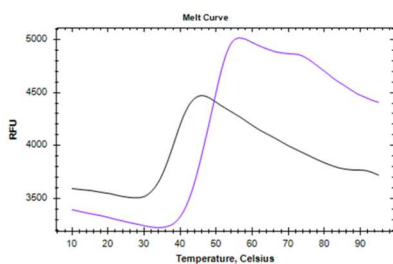
apo T_m = 38.5 °C
$T_{m\,2}$ = 49 °C
$\Delta\,T_m$ = 10.5 °C

XIAP-BIR3 K311S/Compound 8

2h, 25 °C Incubation
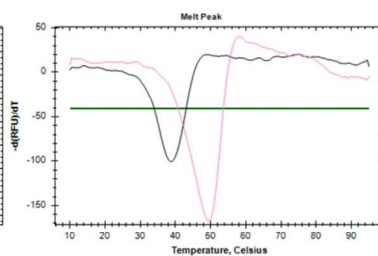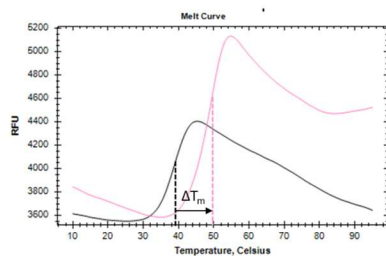
apo T_m = 39 °C
$T_{m\,2}$ = 49.5 °C
$\Delta\,T_m$ = 10.5 °C

6h, 37 °C Incubation
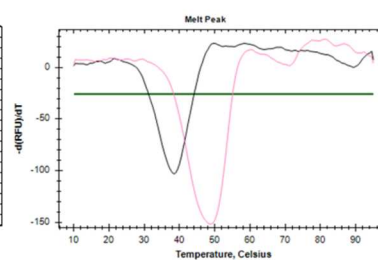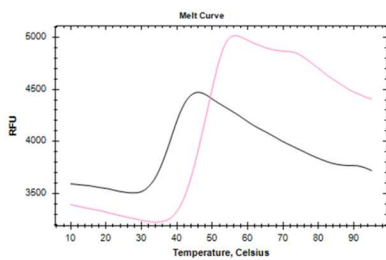
apo T_m = 38.5 °C
$T_{m\,2}$ = 49 °C
$\Delta\,T_m$ = 10.5 °C

REFERENCES

[1] Moldovan, G. L., Pfander, B., & Jentsch, S. (2007). PCNA, the maestro of the replication fork. *Cell*, *129*(4), 665-679. https://doi.org/10.1016/j.cell.2007.05.003

[2] González-Magaña, A., & Blanco, F. J. (2020). Human PCNA Structure, Function and Interactions. *Biomolecules*, *10*(4). https://doi.org/10.3390/biom10040570

[3] Kelman, Z., & O'Donnell, M. (1995). Structural and functional similarities of prokaryotic and eukaryotic DNA polymerase sliding clamps. *Nucleic Acids Res*, *23*(18), 3613-3620. https://doi.org/10.1093/nar/23.18.3613

[4] Chatr-Aryamontri, A., Breitkreutz, B. J., Oughtred, R., Boucher, L., Heinicke, S., Chen, D., . . . Tyers, M. (2015). The BioGRID interaction database: 2015 update. *Nucleic Acids Res*, *43*(Database issue), D470-478. https://doi.org/10.1093/nar/gku1204

[5] Kurki, P., Lotz, M., Ogata, K., & Tan, E. M. (1987). Proliferating cell nuclear antigen (PCNA)/cyclin in activated human T lymphocytes. *J Immunol*, *138*(12), 4114-4120.

[6] Park, S. Y., Jeong, M. S., Han, C. W., Yu, H. S., & Jang, S. B. (2016). Structural and Functional Insight into Proliferating Cell Nuclear Antigen. *J Microbiol Biotechnol*, *26*(4), 637-647. https://doi.org/10.4014/jmb.1509.09051

[7] Naryzhny, S. N., & Lee, H. (2010). Proliferating cell nuclear antigen in the cytoplasm interacts with components of glycolysis and cancer. *FEBS Lett*, *584*(20), 4292-4298. https://doi.org/10.1016/j.febslet.2010.09.021

[8] Ohayon, D., De Chiara, A., Chapuis, N., Candalh, C., Mocek, J., Ribeil, J. A., . . . Witko-Sarsat, V. (2016). Cytoplasmic proliferating cell nuclear antigen connects glycolysis and cell survival in acute myeloid leukemia. *Sci Rep*, *6*, 35561. https://doi.org/10.1038/srep35561

[9] Ohayon, D., De Chiara, A., Dang, P. M., Thieblemont, N., Chatfield, S., Marzaioli, V., . . . Witko-Sarsat, V. (2019). Cytosolic PCNA interacts with p47phox and controls NADPH oxidase NOX2 activation in neutrophils. *J Exp Med*, *216*(11), 2669-2687. https://doi.org/10.1084/jem.20180371

[10] Tang, D., Liu, X., Chen, K., Li, Z., Dai, Y., Xu, J., . . . Liu, L. (2020). Cytoplasmic PCNA is located in the actin belt and involved in osteoclast differentiation. *Aging (Albany NY)*, *12*(13), 13297-13317. https://doi.org/10.18632/aging.103434

[11] Nishitani, H., & Lygerou, Z. (2004). DNA replication licensing. *Front Biosci*, *9*, 2115-2132. https://doi.org/10.2741/1315

[12] Montanari, M., Macaluso, M., Cittadini, A., & Giordano, A. (2006). Role of geminin: from normal control of DNA replication to cancer formation and progression? *Cell Death Differ*, *13*(7), 1052-1056. https://doi.org/10.1038/sj.cdd.4401932

[13] Pellegrini, L. (2012). The Pol α-primase complex. *Subcell Biochem*, *62*, 157-169. https://doi.org/10.1007/978-94-007-4572-8_9

[14] Boehm, E. M., Gildenberg, M. S., & Washington, M. T. (2016). The Many Roles of PCNA in Eukaryotic DNA Replication. *Enzymes*, *39*, 231-254. https://doi.org/10.1016/bs.enz.2016.03.003

[15] Pluciennik, A., Dzantiev, L., Iyer, R. R., Constantin, N., Kadyrov, F. A., & Modrich, P. (2010). PCNA function in the activation and strand direction of MutLα endonuclease in mismatch repair. *Proc Natl Acad Sci U S A*, *107*(37), 16066-16071. https://doi.org/10.1073/pnas.1010662107

[16] Stodola, J. L., & Burgers, P. M. (2017). Mechanism of Lagging-Strand DNA Replication in Eukaryotes. *Adv Exp Med Biol*, *1042*, 117-133. https://doi.org/10.1007/978-981-10-6955-0_6

[17] Ghosal, G., & Chen, J. (2013). DNA damage tolerance: a double-edged sword guarding the genome. *Transl Cancer Res*, *2*(3), 107-129. https://doi.org/10.3978/j.issn.2218-676X.2013.04.01

[18] Waters, L. S., Minesinger, B. K., Wiltrout, M. E., D'Souza, S., Woodruff, R. V., & Walker, G. C. (2009). Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance. *Microbiol Mol Biol Rev*, *73*(1), 134-154. https://doi.org/10.1128/MMBR.00034-08

[19] Hoege, C., Pfander, B., Moldovan, G. L., Pyrowolakis, G., & Jentsch, S. (2002). RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature*, *419*(6903), 135-141. https://doi.org/10.1038/nature00991

[20] Bienko, M., Green, C. M., Crosetto, N., Rudolf, F., Zapart, G., Coull, B., . . . Dikic, I. (2005). Ubiquitin-binding domains in Y-family polymerases regulate translesion synthesis. *Science*, *310*(5755), 1821-1824. https://doi.org/10.1126/science.1120615

[21] Bi, X. (2015). Mechanism of DNA damage tolerance. *World J Biol Chem*, *6*(3), 48-56. https://doi.org/10.4331/wjbc.v6.i3.48

[22] Kannouche, P. L., & Lehmann, A. R. (2004). Ubiquitination of PCNA and the polymerase switch in human cells. *Cell Cycle*, *3*(8), 1011-1013.

[23] Friedberg, E. C., Wagner, R., & Radman, M. (2002). Specialized DNA polymerases, cellular survival, and the genesis of mutations. *Science*, *296*(5573), 1627-1630. https://doi.org/10.1126/science.1070236

[24] Zhang, H., & Lawrence, C. W. (2005). The error-free component of the RAD6/RAD18 DNA damage tolerance pathway of budding yeast employs sister-strand recombination. *Proc Natl Acad Sci U S A*, *102*(44), 15954-15959. https://doi.org/10.1073/pnas.0504586102

[25] Saugar, I., Ortiz-Bazán, M., & Tercero, J. A. (2014). Tolerating DNA damage during eukaryotic chromosome replication. *Exp Cell Res*, *329*(1), 170-177. https://doi.org/10.1016/j.yexcr.2014.07.009

[26] Ripley, B. M., Gildenberg, M. S., & Washington, M. T. (2020). Control of DNA Damage Bypass by Ubiquitylation of PCNA. *Genes (Basel)*, *11*(2). https://doi.org/10.3390/genes11020138

[27] Lambert, S., Watson, A., Sheedy, D. M., Martin, B., & Carr, A. M. (2005). Gross chromosomal rearrangements and elevated recombination at an inducible site-specific replication fork barrier. *Cell*, *121*(5), 689-702. https://doi.org/10.1016/j.cell.2005.03.022

[28] Prado, F. (2018). Homologous Recombination: To Fork and Beyond. *Genes (Basel)*, *9*(12). https://doi.org/10.3390/genes9120603

[29] Pfander, B., Moldovan, G. L., Sacher, M., Hoege, C., & Jentsch, S. (2005). SUMO-modified PCNA recruits Srs2 to prevent recombination during S phase. *Nature*, *436*(7049), 428-433. https://doi.org/10.1038/nature03665

[30] Parker, J. L., & Ulrich, H. D. (2012). A SUMO-interacting motif activates budding yeast ubiquitin ligase Rad18 towards SUMO-modified PCNA. *Nucleic Acids Res*, *40*(22), 11380-11388. https://doi.org/10.1093/nar/gks892

[31] Papouli, E., Chen, S., Davies, A. A., Huttner, D., Krejci, L., Sung, P., & Ulrich, H. D. (2005). Crosstalk between SUMO and ubiquitin on PCNA is mediated by recruitment of the helicase Srs2p. *Mol Cell*, *19*(1), 123-133. https://doi.org/10.1016/j.molcel.2005.06.001

[32] Prudden, J., Pebernard, S., Raffa, G., Slavin, D. A., Perry, J. J., Tainer, J. A., . . . Boddy, M. N. (2007). SUMO-targeted ubiquitin ligases in genome stability. *Embo j*, *26*(18), 4089-4101. https://doi.org/10.1038/sj.emboj.7601838

[33] Perry, J. J., Tainer, J. A., & Boddy, M. N. (2008). A SIM-ultaneous role for SUMO and ubiquitin. *Trends Biochem Sci*, *33*(5), 201-208. https://doi.org/10.1016/j.tibs.2008.02.001

[34] Gary, R., Ludwig, D. L., Cornelius, H. L., MacInnes, M. A., & Park, M. S. (1997). The DNA repair endonuclease XPG binds to proliferating cell nuclear antigen (PCNA) and shares sequence elements with the PCNA-binding regions of FEN-1 and cyclin-dependent kinase inhibitor p21. *J Biol Chem*, *272*(39), 24522-24529. https://doi.org/10.1074/jbc.272.39.24522

[35] Havens, C. G., & Walter, J. C. (2011). Mechanism of CRL4(Cdt2), a PCNA-dependent E3 ubiquitin ligase. *Genes Dev*, *25*(15), 1568-1582. https://doi.org/10.1101/gad.2068611

[36] Cazzalini, O., Sommatis, S., Tillhon, M., Dutto, I., Bachi, A., Rapp, A., . . . Prosperi, E. (2014). CBP and p300 acetylate PCNA to link its degradation with nucleotide excision repair synthesis. *Nucleic Acids Res*, *42*(13), 8433-8448. https://doi.org/10.1093/nar/gku533

[37] Krokan, H. E., & Bjørås, M. (2013). Base excision repair. *Cold Spring Harb Perspect Biol*, *5*(4), a012583. https://doi.org/10.1101/cshperspect.a012583

[38] Friedman, J. I., & Stivers, J. T. (2010). Detection of damaged DNA bases by DNA glycosylase enzymes. *Biochemistry*, *49*(24), 4957-4967. https://doi.org/10.1021/bi100593a

[39] Robertson, A. B., Klungland, A., Rognes, T., & Leiros, I. (2009). DNA repair in mammalian cells: Base excision repair: the long and short of it. *Cell Mol Life Sci*, *66*(6), 981-993. https://doi.org/10.1007/s00018-009-8736-z

[40] Jacobs, A. L., & Schär, P. (2012). DNA glycosylases: in DNA repair and beyond. *Chromosoma*, *121*(1), 1-20. https://doi.org/10.1007/s00412-011-0347-4

[41] Moore, L. D., Le, T., & Fan, G. (2013). DNA methylation and its basic function. *Neuropsychopharmacology*, *38*(1), 23-38. https://doi.org/10.1038/npp.2012.112

[42] Fan, J., Otterlei, M., Wong, H. K., Tomkinson, A. E., & Wilson, D. M. (2004). XRCC1 co-localizes and physically interacts with PCNA. *Nucleic Acids Res*, *32*(7), 2193-2201. https://doi.org/10.1093/nar/gkh556

[43] Li, G. M. (2008). Mechanisms and functions of DNA mismatch repair. *Cell Res*, *18*(1), 85-98. https://doi.org/10.1038/cr.2007.115

[44] Flores-Rozas, H., Clark, D., & Kolodner, R. D. (2000). Proliferating cell nuclear antigen and Msh2p-Msh6p interact to form an active mispair recognition complex. *Nat Genet*, *26*(3), 375-378. https://doi.org/10.1038/81708

[45] Peters, J. M., & Nishiyama, T. (2012). Sister chromatid cohesion. *Cold Spring Harb Perspect Biol*, *4*(11). https://doi.org/10.1101/cshperspect.a011130

[46] Woodcock, C. L., Safer, J. P., & Stanchfield, J. E. (1976). Structural repeating units in chromatin. I. Evidence for their general occurrence. *Exp Cell Res*, *97*, 101-110. https://doi.org/10.1016/0014-4827(76)90659-5

[47] Kornberg, R. D. (1977). Structure of chromatin. *Annu Rev Biochem*, *46*, 931-954. https://doi.org/10.1146/annurev.bi.46.070177.004435

[48] Olins, D. E., & Olins, A. L. (2003). Chromatin history: our view from the bridge. *Nat Rev Mol Cell Biol*, *4*(10), 809-814. https://doi.org/10.1038/nrm1225

[49] Moggs, J. G., Grandi, P., Quivy, J. P., Jónsson, Z. O., Hübscher, U., Becker, P. B., & Almouzni, G. (2000). A CAF-1-PCNA-mediated chromatin assembly pathway triggered by sensing DNA damage. *Mol Cell Biol*, *20*(4), 1206-1218. https://doi.org/10.1128/MCB.20.4.1206-1218.2000

[50] Gaillard, P. H. L., Martini, E. M., Kaufman, P. D., Stillman, B., Moustacchi, E., & Almouzni, G. (1996). Chromatin assembly coupled to DNA repair: a new role for chromatin assembly factor I. *Cell*, *86*(6), 887-896. https://doi.org/10.1016/s0092-8674(00)80164-6

[51] Becker, P. B., & Hörz, W. (2002). ATP-dependent nucleosome remodeling. *Annu Rev Biochem*, *71*, 247-273. https://doi.org/10.1146/annurev.biochem.71.110601.135400

[52] Poot, R. A., Bozhenok, L., van den Berg, D. L., Hawkes, N., & Varga-Weisz, P. D. (2005). Chromatin remodeling by WSTF-ISWI at the replication site: opening a window of opportunity for epigenetic inheritance? *Cell Cycle*, *4*(4), 543-546. https://doi.org/10.4161/cc.4.4.1624

[53] Poot, R. A., Bozhenok, L., van den Berg, D. L., Steffensen, S., Ferreira, F., Grimaldi, M., . . . Varga-Weisz, P. D. (2004). The Williams syndrome transcription factor interacts with PCNA to target chromatin remodelling by ISWI to replication foci. *Nat Cell Biol*, *6*(12), 1236-1244. https://doi.org/10.1038/ncb1196

[54] Handy, D. E., Castro, R., & Loscalzo, J. (2011). Epigenetic modifications: basic mechanisms and role in cardiovascular disease. *Circulation*, *123*(19), 2145-2156. https://doi.org/10.1161/CIRCULATIONAHA.110.956839

[55] Milutinovic, S., Zhuang, Q., & Szyf, M. (2002). Proliferating cell nuclear antigen associates with histone deacetylase activity, integrating DNA replication and chromatin modification. *J Biol Chem*, *277*(23), 20974-20978. https://doi.org/10.1074/jbc.M202504200

[56] Jimenji, T., Matsumura, R., Kori, S., & Arita, K. (2019). Structure of PCNA in complex with DNMT1 PIP box reveals the basis for the molecular mechanism of the interaction. *Biochem Biophys Res Commun*, *516*(2), 578-583. https://doi.org/10.1016/j.bbrc.2019.06.060

[57] Koundrioukoff, S., Jónsson, Z. O., Hasan, S., de Jong, R. N., van der Vliet, P. C., Hottiger, M. O., & Hübscher, U. (2000). A direct interaction between proliferating cell nuclear antigen (PCNA) and Cdk2 targets PCNA-interacting proteins for phosphorylation. *J Biol Chem*, *275*(30), 22882-22887. https://doi.org/10.1074/jbc.M001850200

[58] Karimian, A., Ahmadi, Y., & Yousefi, B. (2016). Multiple functions of p21 in cell cycle, apoptosis and transcriptional regulation after DNA damage. *DNA Repair (Amst)*, *42*, 63-71. https://doi.org/10.1016/j.dnarep.2016.04.008

[59] Vairapandi, M., Azam, N., Balliet, A. G., Hoffman, B., & Liebermann, D. A. (2000). Characterization of MyD118, Gadd45, and proliferating cell nuclear antigen (PCNA) interacting domains. PCNA impedes MyD118 AND Gadd45-mediated negative growth control. *J Biol Chem*, *275*(22), 16810-16819. https://doi.org/10.1074/jbc.275.22.16810

[60] Azam, N., Vairapandi, M., Zhang, W., Hoffman, B., & Liebermann, D. A. (2001). Interaction of CR6 (GADD45gamma) with proliferating cell nuclear antigen impedes negative growth control. *J Biol Chem*, *276*(4), 2766-2774. https://doi.org/10.1074/jbc.M005626200

[61] Smith, M. L., Chen, I. T., Zhan, Q., Bae, I., Chen, C. Y., Gilmer, T. M., . . . Fornace, A. J. (1994). Interaction of the p53-regulated protein Gadd45 with proliferating cell nuclear antigen. *Science*, *266*(5189), 1376-1380. https://doi.org/10.1126/science.7973727

[62] Smith, M. L., Ford, J. M., Hollander, M. C., Bortnick, R. A., Amundson, S. A., Seo, Y. R., . . . Fornace, A. J. (2000). p53-mediated DNA repair responses to UV radiation: studies of mouse cells lacking p53, p21, and/or gadd45 genes. *Mol Cell Biol*, *20*(10), 3705-3714. https://doi.org/10.1128/MCB.20.10.3705-3714.2000

[63] Bouayad, D., Pederzoli-Ribeil, M., Mocek, J., Candalh, C., Arlet, J. B., Hermine, O., . . . Witko-Sarsat, V. (2012). Nuclear-to-cytoplasmic relocalization of the proliferating cell nuclear antigen (PCNA) during differentiation involves a chromosome region maintenance 1 (CRM1)-dependent export and is a prerequisite for PCNA antiapoptotic activity in mature neutrophils. *J Biol Chem*, *287*(40), 33812-33825. https://doi.org/10.1074/jbc.M112.367839

[64] Yin, L., Xie, Y., Yin, S., Lv, X., Zhang, J., Gu, Z., . . . Liu, S. (2015). The S-nitrosylation status of PCNA localized in cytosol impacts the apoptotic pathway in a Parkinson's disease paradigm. *PLoS One*, *10*(2), e0117546. https://doi.org/10.1371/journal.pone.0117546

[65] Nathan, C. (2006). Neutrophils and immunity: challenges and opportunities. *Nat Rev Immunol*, *6*(3), 173-182. https://doi.org/10.1038/nri1785

[66] Witko-Sarsat, V., Mocek, J., Bouayad, D., Tamassia, N., Ribeil, J. A., Candalh, C., . . . Cassatella, M. A. (2010). Proliferating cell nuclear antigen acts as a cytoplasmic platform controlling human neutrophil survival. *J Exp Med*, *207*(12), 2631-2645. https://doi.org/10.1084/jem.20092241

[67] Chiara, A. D., Pederzoli-Ribeil, M., Burgel, P. R., Danel, C., & Witko-Sarsat, V. (2012). Targeting cytosolic proliferating cell nuclear antigen in neutrophil-dominated inflammation. Frontiers in immunology, 3, 311. https://doi.org/10.3389/fimmu.2012.00311

[68] Krishna, T. S., Kong, X. P., Gary, S., Burgers, P. M., & Kuriyan, J. (1994). Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell*, *79*(7), 1233-1243. https://doi.org/10.1016/0092-8674(94)90014-0

[69] Schurtenberger, P., Egelhaaf, S. U., Hindges, R., Maga, G., Jónsson, Z. O., May, R. P., . . . Hübscher, U. (1998). The solution structure of functionally active human proliferating cell nuclear antigen determined by small-angle neutron scattering. *J Mol Biol*, *275*(1), 123-132. https://doi.org/10.1006/jmbi.1997.1435

[70] Dieckman, L. M., Freudenthal, B. D., & Washington, M. T. (2012). PCNA structure and function: insights from structures of PCNA complexes and post-translationally modified PCNA. *Subcell Biochem*, *62*, 281-299. https://doi.org/10.1007/978-94-007-4572-8_15

[71] Pedley, A. M., Lill, M. A., & Davisson, V. J. (2014). Flexibility of PCNA-protein interface accommodates differential binding partners. *PLoS One*, *9*(7), e102481. https://doi.org/10.1371/journal.pone.0102481

[72] Maga, G., & Hubscher, U. (2003). Proliferating cell nuclear antigen (PCNA): a dancer with many partners. *J Cell Sci*, *116*(Pt 15), 3051-3060. https://doi.org/10.1242/jcs.00653

[73] Warbrick, E. (2000). The puzzle of PCNA's many partners. *Bioessays*, *22*(11), 997-1006. https://doi.org/10.1002/1521-1878(200011)22:11<997::AID-BIES6>3.0.CO;2-#

[74] Prestel, A., Wichmann, N., Martins, J. M., Marabini, R., Kassem, N., Broendum, S. S., Otterlei, M., Nielsen, O., Willemoës, M., Ploug, M., Boomsma, W., & Kragelund, B. B. (2019). The PCNA interaction motifs revisited: thinking outside the PIP-box. Cellular and molecular life sciences : CMLS, 76(24), 4923–4943. https://doi.org/10.1007/s00018-019-03150-0

[75] Gilljam, K. M., Feyzi, E., Aas, P. A., Sousa, M. M., Müller, R., Vågbø, C. B., Catterall, T. C., Liabakk, N. B., Slupphaug, G., Drabløs, F., Krokan, H. E., & Otterlei, M. (2009). Identification of a novel, widespread, and functionally important PCNA-binding motif. The Journal of cell biology, 186(5), 645–654. https://doi.org/10.1083/jcb.200903138

[76] Hara, K., Uchida, M., Tagata, R., Yokoyama, H., Ishikawa, Y., Hishiki, A., & Hashimoto, H. (2018). Structure of proliferating cell nuclear antigen (PCNA) bound to an APIM peptide reveals the universality of PCNA interaction. *Acta Crystallogr F Struct Biol Commun*, *74*(Pt 4), 214-221. https://doi.org/10.1107/S2053230X18003242

[77] Xu, H., Zhang, P., Liu, L., & Lee, M. Y. (2001). A novel PCNA-binding motif identified by the panning of a random peptide display library. *Biochemistry*, *40*(14), 4512-4520. https://doi.org/10.1021/bi010103+

[78] Ortega, J., Li, J. Y., Lee, S., Tong, D., Gu, L., & Li, G. M. (2015). Phosphorylation of PCNA by EGFR inhibits mismatch repair and promotes misincorporation during DNA synthesis. *Proc Natl Acad Sci U S A*, *112*(18), 5667-5672. https://doi.org/10.1073/pnas.1417711112

[79] Billon, P., Li, J., Lambert, J. P., Chen, Y., Tremblay, V., Brunzelle, J. S., . . . Côté, J. (2017). Acetylation of PCNA Sliding Surface by Eco1 Promotes Genome Stability through Homologous Recombination. *Mol Cell*, *65*(1), 78-90. https://doi.org/10.1016/j.molcel.2016.10.033

[80] Dillehay, K. L., Seibel, W. L., Zhao, D., Lu, S., & Dong, Z. (2015). Target validation and structure-activity analysis of a series of novel PCNA inhibitors. *Pharmacol Res Perspect*, *3*(2), e00115. https://doi.org/10.1002/prp2.115

[81] Esteva, F. J., Sahin, A. A., Smith, T. L., Yang, Y., Pusztai, L., Nahta, R., Buchholz, T. A., Buzdar, A. U., Hortobagyi, G. N., & Bacus, S. S. (2004). Prognostic significance of phosphorylated P38 mitogen-activated protein kinase and HER-2 expression in lymph node-positive breast carcinoma. Cancer, 100(3), 499–506. https://doi.org/10.1002/cncr.11940

[82] Bechtel, P. E., Hickey, R. J., Schnaper, L., Sekowski, J. W., Long, B. J., Freund, R., Liu, N., Rodriguez-Valenzuela, C., & Malkas, L. H. (1998). A unique form of proliferating cell nuclear antigen is present in malignant breast cells. Cancer research, 58(15), 3264–3269.

[83] Malkas, L. H., Herbert, B. S., Abdel-Aziz, W., Dobrolecki, L. E., Liu, Y., Agarwal, B., . . . Hickey, R. J. (2006). A cancer-associated PCNA expressed in breast cancer has implications as a potential biomarker. *Proc Natl Acad Sci U S A*, *103*(51), 19472-19477. https://doi.org/10.1073/pnas.0604614103

[84] Wang, X., Hickey, R. J., Malkas, L. H., Koch, M. O., Li, L., Zhang, S., . . . Cheng, L. (2011). Elevated expression of cancer-associated proliferating cell nuclear antigen in high-grade prostatic intraepithelial neoplasia and prostate cancer. *The Prostate*, *71*(7), 748-754. https://doi.org/https://doi.org/10.1002/pros.21291

[85] Gu, L., Smith, S., Li, C., Hickey, R. J., Stark, J. M., Fields, G. B., . . . Malkas, L. H. (2014). A PCNA-derived cell permeable peptide selectively inhibits neuroblastoma cell growth. *PLoS One*, *9*(4), e94773. https://doi.org/10.1371/journal.pone.0094773

[86] Gu, L., Lingeman, R., Yakushijin, F., Sun, E., Cui, Q., Chao, J., . . . Malkas, L. H. (2018). The Anticancer Activity of a First-in-class Small-molecule Targeting PCNA. *Clin Cancer Res*, *24*(23), 6053-6065. https://doi.org/10.1158/1078-0432.CCR-18-0592

[87] Mehrotra, N., Gupta, M., Kovar, A., & Meibohm, B. (2007). The role of pharmacokinetics and pharmacodynamics in phosphodiesterase-5 inhibitor therapy. *Int J Impot Res*, *19*(3), 253-264. https://doi.org/10.1038/sj.ijir.3901522

[88] Xiong, Y., Hannon, G. J., Zhang, H., Casso, D., Kobayashi, R., & Beach, D. (1993). p21 is a universal inhibitor of cyclin kinases. *Nature*, *366*(6456), 701-704. https://doi.org/10.1038/366701a0

[89] Waga, S., Hannon, G. J., Beach, D., & Stillman, B. (1994). The p21 inhibitor of cyclin-dependent kinases controls DNA replication by interaction with PCNA. *Nature*, *369*(6481), 574-578. https://doi.org/10.1038/369574a0

190

[90] Warbrick, E., Lane, D. P., Glover, D. M., & Cox, L. S. (1995). A small peptide inhibitor of DNA replication defines the site of interaction between the cyclin-dependent kinase inhibitor p21WAF1 and proliferating cell nuclear antigen. *Curr Biol*, *5*(3), 275-282. https://doi.org/10.1016/s0960-9822(95)00058-3

[91] Zheleva, D. I., Zhelev, N. Z., Fischer, P. M., Duff, S. V., Warbrick, E., Blake, D. G., & Lane, D. P. (2000). A quantitative study of the in vitro binding of the C-terminal domain of p21 to PCNA: affinity, stoichiometry, and thermodynamics. *Biochemistry*, *39*(25), 7388-7397. https://doi.org/10.1021/bi992498r

[92] Kontopidis, G., Wu, S.-Y., Zheleva, D. I., Taylor, P., McInnes, C., Lane, D. P., . . . Walkinshaw, M. D. (2005). Structural and biochemical studies of human proliferating cell nuclear antigen complexes provide a rationale for cyclin association and inhibitor design. *Proceedings of the National Academy of Sciences*, *102*(6), 1871-1876. https://doi.org/doi:10.1073/pnas.0406540102

[93] Müller, R., Misund, K., Holien, T., Bachke, S., Gilljam, K. M., Våtsveen, T. K., . . . Otterlei, M. (2013). Targeting proliferating cell nuclear antigen and its protein interactions induces apoptosis in multiple myeloma cells. *PLoS One*, *8*(7), e70430. https://doi.org/10.1371/journal.pone.0070430

[94] Søgaard, C. K., Blindheim, A., Røst, L. M., Petrović, V., Nepal, A., Bachke, S., . . . Otterlei, M. (2018). "Two hits - one stone"; increased efficacy of cisplatin-based therapies by targeting PCNA's role in both DNA repair and cellular signaling. *Oncotarget*, *9*(65), 32448-32465. https://doi.org/10.18632/oncotarget.25963

[95] Gravina, G. L., Colapietro, A., Mancini, A., Rossetti, A., Martellucci, S., Ventura, L., . . . Festuccia, C. (2022). ATX-101, a Peptide Targeting PCNA, Has Antitumor Efficacy Alone or in Combination with Radiotherapy in Murine Models of Human Glioblastoma. *Cancers (Basel)*, *14*(2). https://doi.org/10.3390/cancers14020289

[96] Smith, S. J., Gu, L., Phipps, E. A., Dobrolecki, L. E., Mabrey, K. S., Gulley, P., . . . Malkas, L. H. (2015). A Peptide Mimicking a Region in Proliferating Cell Nuclear Antigen Specific to Key Protein Interactions Is Cytotoxic to Breast Cancer. *Molecular Pharmacology*, *87*(2), 263-276. https://doi.org/10.1124/mol.114.093211

[97] Punchihewa, C., Inoue, A., Hishiki, A., Fujikawa, Y., Connelly, M., Evison, B., . . . Fujii, N. (2012). Identification of small molecule proliferating cell nuclear antigen (PCNA) inhibitor that disrupts interactions with PIP-box proteins and inhibits DNA replication. *J Biol Chem*, *287*(17), 14289-14300. https://doi.org/10.1074/jbc.M112.353201

[98] Inoue, A., Kikuchi, S., Hishiki, A., Shao, Y., Heath, R., Evison, B. J., . . . Fujii, N. (2014). A small molecule inhibitor of monoubiquitinated Proliferating Cell Nuclear Antigen (PCNA) inhibits repair of interstrand DNA cross-link, enhances DNA double strand break, and sensitizes cancer cells to cisplatin. *J Biol Chem*, *289*(10), 7109-7120. https://doi.org/10.1074/jbc.M113.520429

[99] Wang, R., Lenoir, W. F., Wang, C., Su, D., McLaughlin, M., Hu, Q., . . . Li, L. (2020). DNA polymerase ι compensates for Fanconi anemia pathway deficiency by countering DNA replication stress. *Proc Natl Acad Sci U S A*, *117*(52), 33436-33445. https://doi.org/10.1073/pnas.2008821117

[100] Rocha, C., Silva, M. M., Quinet, A., Cabral-Neto, J. B., & Menck, C. (2018). DNA repair pathways and cisplatin resistance: an intimate relationship. Clinics (Sao Paulo, Brazil), 73(suppl 1), e478s. https://doi.org/10.6061/clinics/2018/e478s

[101] Tan, Z., Wortman, M., Dillehay, K. L., Seibel, W. L., Evelyn, C. R., Smith, S. J., . . . Dong, Z. (2012). Small-molecule targeting of proliferating cell nuclear antigen chromatin association inhibits tumor cell growth. *Mol Pharmacol*, *81*(6), 811-819. https://doi.org/10.1124/mol.112.077735

[102] Li, W., Zhou, Y., Tang, G., Wong, N. K., Yang, M., Tan, D., & Xiao, Y. (2018). Chemoproteomics Reveals the Antiproliferative Potential of Parkinson's Disease Kinase Inhibitor LRRK2-IN-1 by Targeting PCNA Protein. *Mol Pharm*, *15*(8), 3252-3259. https://doi.org/10.1021/acs.molpharmaceut.8b00325

[103] Rees, D. C., Congreve, M., Murray, C. W., & Carr, R. (2004). Fragment-based lead discovery. *Nature Reviews Drug Discovery*, *3*(8), 660-672. https://doi.org/10.1038/nrd1467

[104] Lipinski C. A. (2000). Drug-like properties and the causes of poor solubility and poor permeability. Journal of pharmacological and toxicological methods, 44(1), 235–249. https://doi.org/10.1016/s1056-8719(00)00107-6

[105] Kirsch, P., Hartman, A. M., Hirsch, A. K. H., & Empting, M. (2019). Concepts and Core Principles of Fragment-Based Drug Design. *Molecules*, *24*(23), 4309.

[106] Leach, A. R., Hann, M. M., Burrows, J. N., & Griffen, E. J. (2006). Fragment screening: an introduction [10.1039/B610069B]. *Molecular BioSystems*, *2*(9), 429-446. https://doi.org/10.1039/B610069B

[107] Murray, C. W., & Rees, D. C. (2009). The rise of fragment-based drug discovery. *Nature Chemistry*, *1*(3), 187-192. https://doi.org/10.1038/nchem.217

[108] Bembenek, S. D., Tounge, B. A., & Reynolds, C. H. (2009). Ligand efficiency and fragment-based drug discovery. *Drug Discovery Today*, *14*(5), 278-283. https://doi.org/https://doi.org/10.1016/j.drudis.2008.11.007

[109] Grygorenko, O. O., Volochnyuk, D. M., Ryabukhin, S. V., & Judd, D. B. (2020). The Symbiotic Relationship Between Drug Discovery and Organic Chemistry. *Chemistry*, *26*(6), 1196-1237. https://doi.org/10.1002/chem.201903232

[110] Wouters, O. J., McKee, M., & Luyten, J. (2020). Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA*, *323*(9), 844-853. https://doi.org/10.1001/jama.2020.1166

[111] Pantoliano, M. W., Petrella, E. C., Kwasnoski, J. D., Lobanov, V. S., Myslik, J., Graf, E., . . . Salemme, F. R. (2001). High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J Biomol Screen*, *6*(6), 429-440. https://doi.org/10.1177/108705710100600609

[112] Lo, M. C., Aulabaugh, A., Jin, G., Cowling, R., Bard, J., Malamas, M., & Ellestad, G. (2004). Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal Biochem*, *332*(1), 153-159. https://doi.org/10.1016/j.ab.2004.04.031

[113] Gao, K., Oerlemans, R., & Groves, M. R. (2020). Theory and applications of differential scanning fluorimetry in early-stage drug discovery. *Biophys Rev*, *12*(1), 85-104. https://doi.org/10.1007/s12551-020-00619-2

[114] Vedadi, M., Niesen, F. H., Allali-Hassani, A., Fedorov, O. Y., Finerty, P. J., Wasney, G. A., . . . Edwards, A. M. (2006). Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proceedings of the National Academy of Sciences*, *103*(43), 15835-15840. https://doi.org/doi:10.1073/pnas.0605224103

[115] Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N., & Nordlund, P. (2006). Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal Biochem*, *357*(2), 289-298. https://doi.org/10.1016/j.ab.2006.07.027

[116] Biter, A. B., de la Peña, A. H., Thapar, R., Lin, J. Z., & Phillips, K. J. (2016). DSF Guided Refolding As A Novel Method Of Protein Production. *Scientific Reports*, *6*(1), 18906. https://doi.org/10.1038/srep18906

[117] Silvestre, H. L., Blundell, T. L., Abell, C., & Ciulli, A. (2013). Integrated biophysical approach to fragment screening and validation for fragment-based lead discovery. *Proceedings of the National Academy of Sciences*, *110*(32), 12984-12989. https://doi.org/doi:10.1073/pnas.1304045110

[118] Matulis, D., Kranz, J. K., Salemme, F. R., & Todd, M. J. (2005). Thermodynamic Stability of Carbonic Anhydrase: Measurements of Binding Affinity and Stoichiometry Using ThermoFluor. *Biochemistry*, *44*(13), 5258-5266. https://doi.org/10.1021/bi048135v

[119] Holdgate, G., Embrey, K., Milbradt, A., & Davies, G. (2019). Biophysical methods in early drug discovery. *ADMET DMPK*, *7*(4), 222-241. https://doi.org/10.5599/admet.733

[120] Coyle, J., & Walser, R. (2020). Applied Biophysical Methods in Fragment-Based Drug Discovery. *SLAS Discovery*, *25*(5), 471-490. https://doi.org/10.1177/2472555220916168

[121] Ciulli, A. (2013). Biophysical screening for the discovery of small-molecule ligands. *Methods Mol Biol*, *1008*, 357-388. https://doi.org/10.1007/978-1-62703-398-5_13

[122] Huynh, K., & Partch, C. L. (2015). Analysis of Protein Stability and Ligand Interactions by Thermal Shift Assay. *Current Protocols in Protein Science*, *79*(1), 28.29.21-28.29.14. https://doi.org/https://doi.org/10.1002/0471140864.ps2809s79

[123] Moreau, M. J. J., Morin, I., & Schaeffer, P. M. (2010). Quantitative determination of protein stability and ligand binding using a green fluorescent protein reporter system [10.1039/C002001J]. *Molecular BioSystems*, *6*(7), 1285-1292. https://doi.org/10.1039/C002001J

[124] Alexander, C. G., Wanner, R., Johnson, C. M., Breitsprecher, D., Winter, G., Duhr, S., . . . Ferguson, N. (2014). Novel microscale approaches for easy, rapid determination of protein stability in academic and commercial settings. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1844*(12), 2241-2250. https://doi.org/https://doi.org/10.1016/j.bbapap.2014.09.016

[125] Ghisaidoobe, A. B. T., & Chung, S. J. (2014). Intrinsic Tryptophan Fluorescence in the Detection and Analysis of Proteins: A Focus on Förster Resonance Energy Transfer Techniques. *International Journal of Molecular Sciences*, *15*(12), 22518-22538.

[126] Hawe, A., Sutter, M., & Jiskoot, W. (2008). Extrinsic Fluorescent Dyes as Tools for Protein Characterization. *Pharmaceutical Research*, *25*(7), 1487-1499. https://doi.org/10.1007/s11095-007-9516-9

[127] Niesen, F. H., Berglund, H., & Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols*, *2*(9), 2212-2221. https://doi.org/10.1038/nprot.2007.321

[128] Chilton, M., Clennell, B., Edfeldt, F., & Geschwindner, S. (2017). Hot-Spotting with Thermal Scanning: A Ligand- and Structure-Independent Assessment of Target Ligandability. *J Med Chem*, *60*(12), 4923-4931. https://doi.org/10.1021/acs.jmedchem.7b00208

[129] Kranz, J. K., & Schalk-Hihi, C. (2011). Protein thermal shifts to identify low molecular weight fragments. *Methods Enzymol*, *493*, 277-298. https://doi.org/10.1016/b978-0-12-381274-2.00011-x

[130] McMahon, R. M., Scanlon, M. J., & Martin, J. L. (2013). Interrogating Fragments Using a Protein Thermal Shift Assay. *Australian Journal of Chemistry*, *66*(12), 1502-1506. https://doi.org/https://doi.org/10.1071/CH13279

[131] Larsson, E. A., Jansson, A., Ng, F. M., Then, S. W., Panicker, R., Liu, B., . . . Nordlund, P. (2013). Fragment-Based Ligand Design of Novel Potent Inhibitors of Tankyrases. *Journal of Medicinal Chemistry*, *56*(11), 4497-4508. https://doi.org/10.1021/jm400211f

[132] Redhead, M., Satchell, R., Morkūnaitė, V., Swift, D., Petrauskas, V., Golding, E., . . . Unitt, J. (2015). A combinatorial biophysical approach; FTSA and SPR for identifying small molecule ligands and PAINs. *Anal Biochem*, *479*, 63-73. https://doi.org/10.1016/j.ab.2015.03.013

[133] Baell, J. B., & Nissink, J. W. M. (2018). Seven Year Itch: Pan-Assay Interference Compounds (PAINS) in 2017—Utility and Limitations. *ACS Chemical Biology*, *13*(1), 36-44. https://doi.org/10.1021/acschembio.7b00903

[134] Duckett, C. S. (2005). IAP proteins: sticking it to Smac. *Biochem J*, *385*(Pt 1), e1-2. https://doi.org/10.1042/bj20041800

[135] Srinivasula, S. M., Hegde, R., Saleh, A., Datta, P., Shiozaki, E., Chai, J., . . . Alnemri, E. S. (2001). A conserved XIAP-interaction motif in caspase-9 and Smac/DIABLO regulates caspase activity and apoptosis. *Nature*, *410*(6824), 112-116. https://doi.org/10.1038/35065125

[136] Huang, Y., Rich, R. L., Myszka, D. G., & Wu, H. (2003). Requirement of both the second and third BIR domains for the relief of X-linked inhibitor of apoptosis protein (XIAP)-mediated caspase inhibition by Smac. *J Biol Chem*, *278*(49), 49517-49522. https://doi.org/10.1074/jbc.M310061200

[137] Flygare, J. A., Beresini, M., Budha, N., Chan, H., Chan, I. T., Cheeti, S., . . . Fairbrother, W. J. (2012). Discovery of a Potent Small-Molecule Antagonist of Inhibitor of Apoptosis (IAP) Proteins and Clinical Candidate for the Treatment of Cancer (GDC-0152). *Journal of Medicinal Chemistry*, *55*(9), 4101-4113. https://doi.org/10.1021/jm300060k

[138] West, A. C., Martin, B. P., Andrews, D. A., Hogg, S. J., Banerjee, A., Grigoriadis, G., . . . Shortt, J. (2016). The SMAC mimetic, LCL-161, reduces survival in aggressive MYC-driven lymphoma while promoting susceptibility to endotoxic shock. *Oncogenesis*, *5*(4), e216-e216. https://doi.org/10.1038/oncsis.2016.26

[139] Cai, Q., Sun, H., Peng, Y., Lu, J., Nikolovska-Coleska, Z., McEachern, D., . . . Wang, S. (2011). A Potent and Orally Active Antagonist (SM-406/AT-406) of Multiple Inhibitor of Apoptosis Proteins (IAPs) in Clinical Development for Cancer Treatment. *Journal of Medicinal Chemistry*, *54*(8), 2714-2726. https://doi.org/10.1021/jm101505d

[140] Gambini, L., Baggio, C., Udompholkul, P., Jossart, J., Salem, A. F., Perry, J. J. P., & Pellecchia, M. (2019). Covalent Inhibitors of Protein–Protein Interactions Targeting Lysine, Tyrosine, or Histidine Residues. *Journal of Medicinal Chemistry*, *62*(11), 5616-5627. https://doi.org/10.1021/acs.jmedchem.9b00561

[141] Baggio, C., Udompholkul, P., Gambini, L., Salem, A. F., Jossart, J., Perry, J. J. P., & Pellecchia, M. (2019). Aryl-fluorosulfate-based Lysine Covalent Pan-Inhibitors of Apoptosis Protein (IAP) Antagonists with Cellular Efficacy. *J Med Chem*, *62*(20), 9188-9200. https://doi.org/10.1021/acs.jmedchem.9b01108

[142] Danishuddin, & Khan, A. U. (2016). Descriptors and their selection methods in QSAR analysis: paradigm for drug design. *Drug Discovery Today*, *21*(8), 1291-1302. https://doi.org/https://doi.org/10.1016/j.drudis.2016.06.013

[143] Maia, E. H. B., Assis, L. C., de Oliveira, T. A., da Silva, A. M., & Taranto, A. G. (2020). Structure-Based Virtual Screening: From Classical to Artificial Intelligence [Review]. *Frontiers in Chemistry*, *8*. https://doi.org/10.3389/fchem.2020.00343

[144] Hall, R. J., Murray, C. W., & Verdonk, M. L. (2017). The Fragment Network: A Chemistry Recommendation Engine Built Using a Graph Database. *Journal of Medicinal Chemistry*, *60*(14), 6440-6450. https://doi.org/10.1021/acs.jmedchem.7b00809

[145] Gurung, A. B., Ali, M. A., Lee, J., Farah, M. A., & Al-Anazi, K. M. (2021). An Updated Review of Computer-Aided Drug Design and Its Application to COVID-19. *BioMed Research International*, *2021*, 8853056. https://doi.org/10.1155/2021/8853056

[146] Martin, Y. C. (1992). 3D database searching in drug design. *Journal of Medicinal Chemistry*, *35*(12), 2145-2154. https://doi.org/10.1021/jm00090a001

[147] Sastry, G. M., Dixon, S. L., & Sherman, W. (2011). Rapid Shape-Based Ligand Alignment and Virtual Screening Method Based on Atom/Feature-Pair Similarities and Volume Overlap Scoring. *Journal of Chemical Information and Modeling*, *51*(10), 2455-2466. https://doi.org/10.1021/ci2002704

[148] Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., . . . Shenkin, P. S. (2004). Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry*, *47*(7), 1739-1749. https://doi.org/10.1021/jm0306430

[149] Trott, O., & Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, *31*(2), 455-461. https://doi.org/https://doi.org/10.1002/jcc.21334

[150] Trevizani, R., Custódio, F. L., dos Santos, K. B., & Dardenne, L. E. (2017). Critical Features of Fragment Libraries for Protein Structure Prediction. *PLOS ONE*, *12*(1), e0170131. https://doi.org/10.1371/journal.pone.0170131

[151] Tabeshpour, J., Sahebkar, A., Zirak, R. M., Zeinali, M., Hashemzaei, M., Rakhshani, S., & Rakhshani, S. (2018). Computer-aided Drug Design and Drug Pharmacokinetic Prediction: A Mini-review. *Current Pharmaceutical Design*, *24*(26), 3014-3019. https://doi.org/http://dx.doi.org/10.2174/1381612824666180903123423

[152] Hou, T., & Xu, X. (2004). Recent Development and Application of Virtual Screening in Drug Discovery: An Overview. *Current Pharmaceutical Design*, *10*(9), 1011-1033. https://doi.org/http://dx.doi.org/10.2174/1381612043452721

[153] Swami, R., & Shahiwala, A. (2013). Impact of physiochemical properties on pharmacokinetics of protein therapeutics. *European Journal of Drug Metabolism and Pharmacokinetics*, *38*(4), 231-239. https://doi.org/10.1007/s13318-013-0126-0

[154] Kaserer, T., Beck, K. R., Akram, M., Odermatt, A., & Schuster, D. (2015). Pharmacophore Models and Pharmacophore-Based Virtual Screening: Concepts and Applications Exemplified on Hydroxysteroid Dehydrogenases. *Molecules*, *20*(12), 22799-22832.

[155] Bulusu, K. C., Guha, R., Mason, D. J., Lewis, R. P. I., Muratov, E., Kalantar Motamedi, Y., . . . Bender, A. (2016). Modelling of compound combination effects and applications to efficacy and toxicity: state-of-the-art, challenges and perspectives. *Drug Discovery Today*, *21*(2), 225-238. https://doi.org/https://doi.org/10.1016/j.drudis.2015.09.003

[156] Cavasotto, C. N. (2011). Homology models in docking and high-throughput docking. *Curr Top Med Chem*, *11*(12), 1528-1534. https://doi.org/10.2174/156802611795860951

[157] Berman, H., Henrick, K., & Nakamura, H. (2003). Announcing the worldwide Protein Data Bank. *Nature Structural & Molecular Biology*, *10*(12), 980-980. https://doi.org/10.1038/nsb1203-980

[158] Irwin, J. J., & Shoichet, B. K. (2005). ZINC − A Free Database of Commercially Available Compounds for Virtual Screening. *Journal of Chemical Information and Modeling*, *45*(1), 177-182. https://doi.org/10.1021/ci049714+

[159] Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., & Ferrin, T. E. (1982). A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology*, *161*(2), 269-288. https://doi.org/https://doi.org/10.1016/0022-2836(82)90153-X

[160] Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, *30*(16), 2785-2791. https://doi.org/https://doi.org/10.1002/jcc.21256

[161] Ruiz-Tagle, B., Villalobos-Cid, M., Dorn, M., & Inostroza-Ponta, M. (2017, 16-20 Oct. 2017). Evaluating the use of local search strategies for a memetic algorithm for the protein-ligand docking problem. 2017 36th International Conference of the Chilean Computer Science Society (SCCC)

[162] Huang, S.-Y., & Zou, X. (2010). Advances and Challenges in Protein-Ligand Docking. *International Journal of Molecular Sciences*, *11*(8), 3016-3034.

[163] Karplus, M., & McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, *9*(9), 646-652. https://doi.org/10.1038/nsb0902-646

[164] ten Brink, T., & Exner, T. E. (2009). Influence of Protonation, Tautomeric, and Stereoisomeric States on Protein−Ligand Docking Results. *Journal of Chemical Information and Modeling*, *49*(6), 1535-1546. https://doi.org/10.1021/ci800420z

[165] Li, J., Fu, A., & Zhang, L. (2019). An Overview of Scoring Functions Used for Protein–Ligand Interactions in Molecular Docking. *Interdisciplinary Sciences: Computational Life Sciences*, *11*(2), 320-328. https://doi.org/10.1007/s12539-019-00327-w

[166] Wang, J. C., & Lin, J. H. (2013). Scoring functions for prediction of protein-ligand interactions. Current pharmaceutical design, 19(12), 2174–2182. https://doi.org/10.2174/1381612811319120005

[167] Liu, J., & Wang, R. (2015). Classification of Current Scoring Functions. *Journal of Chemical Information and Modeling*, *55*(3), 475-482. https://doi.org/10.1021/ci500731a

[168] Yang, Y., Lightstone, F. C., & Wong, S. E. (2013). Approaches to efficiently estimate solvation and explicit water energetics in ligand binding: the use of WaterMap. *Expert Opinion on Drug Discovery*, *8*(3), 277-287. https://doi.org/10.1517/17460441.2013.749853

[169] Raha, K., Peters, M. B., Wang, B., Yu, N., Wollacott, A. M., Westerhoff, L. M., & Merz, K. M. (2007). The role of quantum mechanics in structure-based drug design. *Drug Discovery Today*, *12*(17), 725-731. https://doi.org/https://doi.org/10.1016/j.drudis.2007.07.006

[170] Chaskar, P., Zoete, V., & Röhrig, U. F. (2014). Toward On-The-Fly Quantum Mechanical/Molecular Mechanical (QM/MM) Docking: Development and Benchmark of a Scoring Function. *Journal of Chemical Information and Modeling*, *54*(11), 3137-3152. https://doi.org/10.1021/ci5004152

[171] Li, Y., Liu, Z., Li, J., Han, L., Liu, J., Zhao, Z., & Wang, R. (2014). Comparative Assessment of Scoring Functions on an Updated Benchmark: 1. Compilation of the Test Set. *Journal of Chemical Information and Modeling*, *54*(6), 1700-1716. https://doi.org/10.1021/ci500080q

[172] Gohlke, H., Hendlich, M., & Klebe, G. (2000). Knowledge-based scoring function to predict protein-ligand interactions11Edited by R. Huber. *Journal of Molecular Biology*, *295*(2), 337-356. https://doi.org/https://doi.org/10.1006/jmbi.1999.3371

[173] Khamis, M. A., Gomaa, W., & Ahmed, W. F. (2015). Machine learning in computational docking. *Artif Intell Med*, *63*(3), 135-152. https://doi.org/10.1016/j.artmed.2015.02.002

[174] Chen, R., Liu, X., Jin, S., Lin, J., & Liu, J. (2018). Machine Learning for Drug-Target Interaction Prediction. *Molecules*, *23*(9), 2208.

[175] Park, H., Eom, J.-W., & Kim, Y.-H. (2014). Consensus Scoring Approach To Identify the Inhibitors of AMP-Activated Protein Kinase α2 with Virtual Screening. *Journal of Chemical Information and Modeling*, *54*(7), 2139-2146. https://doi.org/10.1021/ci500214e

[176] Aliebrahimi, S., Montasser Kouhsari, S., Ostad, S. N., Arab, S. S., & Karami, L. (2018). Identification of Phytochemicals Targeting c-Met Kinase Domain using Consensus Docking and Molecular Dynamics Simulation Studies. *Cell Biochemistry and Biophysics*, *76*(1), 135-145. https://doi.org/10.1007/s12013-017-0821-6

[177] Hawkins, P. C. D., Warren, G. L., Skillman, A. G., & Nicholls, A. (2008). How to do an evaluation: pitfalls and traps. *Journal of Computer-Aided Molecular Design*, *22*(3), 179-190. https://doi.org/10.1007/s10822-007-9166-3

[178] Truchon, J.-F., & Bayly, C. I. (2007). Evaluating Virtual Screening Methods: Good and Bad Metrics for the "Early Recognition" Problem. *Journal of Chemical Information and Modeling*, *47*(2), 488-508. https://doi.org/10.1021/ci600426e

[179] Jain, A. N. (2008). Bias, reporting, and sharing: computational evaluations of docking methods. *Journal of Computer-Aided Molecular Design*, *22*(3), 201-212. https://doi.org/10.1007/s10822-007-9151-x

[180] de Magalhães, C. S., Barbosa, H. J. C., & Dardenne, L. E. (2004, 2004//). Selection-Insertion Schemes in Genetic Algorithms for the Flexible Ligand Docking Problem. Genetic and Evolutionary Computation – GECCO 2004, Berlin, Heidelberg.

[181] Lätti, S., Niinivehmas, S., & Pentikäinen, O. T. (2016). Rocker: Open source, easy-to-use tool for AUC and enrichment calculations and ROC visualization. *Journal of Cheminformatics*, *8*(1), 45. https://doi.org/10.1186/s13321-016-0158-y

[182] Hamza, A., Wei, N.-N., & Zhan, C.-G. (2012). Ligand-Based Virtual Screening Approach Using a New Scoring Function. *Journal of Chemical Information and Modeling*, *52*(4), 963-974. https://doi.org/10.1021/ci200617d

[183] Chen, X., & Reynolds, C. H. (2002). Performance of Similarity Measures in 2D Fragment-Based Similarity Searching: Comparison of Structural Descriptors and Similarity Coefficients. *Journal of Chemical Information and Computer Sciences*, *42*(6), 1407-1414. https://doi.org/10.1021/ci025531g

[184] Riniker, S., & Landrum, G. A. (2013). Open-source platform to benchmark fingerprints for ligand-based virtual screening. *Journal of Cheminformatics*, *5*(1), 26. https://doi.org/10.1186/1758-2946-5-26

[185] Bajorath, J. (2001). Selected Concepts and Investigations in Compound Classification, Molecular Descriptor Analysis, and Virtual Screening. *Journal of Chemical Information and Computer Sciences*, *41*(2), 233-245. https://doi.org/10.1021/ci0001482

[186] Ballester, P. J., Finn, P. W., & Richards, W. G. (2009). Ultrafast shape recognition: Evaluating a new ligand-based virtual screening technology. *Journal of Molecular Graphics and Modelling*, *27*(7), 836-845. https://doi.org/https://doi.org/10.1016/j.jmgm.2009.01.001

[187] Banegas-Luna, A.-J., Cerón-Carrasco, J. P., & Pérez-Sánchez, H. (2018). A review of ligand-based virtual screening web tools and screening algorithms in large molecular databases in the age of big data. *Future Medicinal Chemistry*, *10*(22), 2641-2658. https://doi.org/10.4155/fmc-2018-0076

[188] Fernández-de Gortari, E., García-Jacas, C. R., Martinez-Mayorga, K., & Medina-Franco, J. L. (2017). Database fingerprint (DFP): an approach to represent molecular databases. *Journal of Cheminformatics*, *9*(1), 9. https://doi.org/10.1186/s13321-017-0195-1

[189] Willett, P. (2011). Similarity Searching Using 2D Structural Fingerprints. In J. Bajorath (Ed.), *Chemoinformatics and Computational Chemical Biology* (pp. 133-158). Humana Press. https://doi.org/10.1007/978-1-60761-839-3_5

[190] Bender, A., Jenkins, J. L., Scheiber, J., Sukuru, S. C. K., Glick, M., & Davies, J. W. (2009). How Similar Are Similarity Searching Methods? A Principal Component Analysis of Molecular Descriptor Space. *Journal of Chemical Information and Modeling*, *49*(1), 108-119. https://doi.org/10.1021/ci800249s

[191] Todeschini, R., & Consonni, V. (2000). *Handbook of molecular descriptors. WileyVCH, Weinheim* (Vol. 11). https://doi.org/10.1002/9783527613106

[192] Butina, D. (1999). Unsupervised Data Base Clustering Based on Daylight's Fingerprint and Tanimoto Similarity: A Fast and Automated Way To Cluster Small and Large Data Sets. *Journal of Chemical Information and Computer Sciences*, *39*(4), 747-750. https://doi.org/10.1021/ci9803381

[193] Thomas, S. E., Mendes, V., Kim, S. Y., Malhotra, S., Ochoa-Montaño, B., Blaszczyk, M., & Blundell, T. L. (2017). Structural Biology and the Design of New Therapeutics: From HIV and Cancer to Mycobacterial Infections: A Paper Dedicated to John Kendrew. *Journal of Molecular Biology*, *429*(17), 2677-2693. https://doi.org/https://doi.org/10.1016/j.jmb.2017.06.014

[194] Chen, Y., & Shoichet, B. K. (2009). Molecular docking and ligand specificity in fragment-based inhibitor discovery. *Nature Chemical Biology*, *5*(5), 358-364. https://doi.org/10.1038/nchembio.155

[195] Zhao, H., Gartenmann, L., Dong, J., Spiliotopoulos, D., & Caflisch, A. (2014). Discovery of BRD4 bromodomain inhibitors by fragment-based high-throughput docking. *Bioorganic & Medicinal Chemistry Letters*, *24*(11), 2493-2496. https://doi.org/https://doi.org/10.1016/j.bmcl.2014.04.017

[196] de Souza Neto, L. R., Moreira-Filho, J. T., Neves, B. J., Maidana, R. L. B. R., Guimarães, A. C. R., Furnham, N., . . . Silva, F. P. (2020). In silico Strategies to Support Fragment-to-Lead Optimization in Drug Discovery [Review]. *Frontiers in Chemistry*, *8*. https://doi.org/10.3389/fchem.2020.00093

[197] Patel, D., Bauman, J. D., & Arnold, E. (2014). Advantages of crystallographic fragment screening: functional and mechanistic insights from a powerful platform for efficient drug discovery. *Prog Biophys Mol Biol*, *116*(2-3), 92-100. https://doi.org/10.1016/j.pbiomolbio.2014.08.004

[198] Torres, P. H. M., Sodero, A. C. R., Jofily, P., & Silva-Jr, F. P. (2019). Key Topics in Molecular Docking for Drug Design. *International Journal of Molecular Sciences*, *20*(18), 4574.

[199] Cukuroglu, E., Engin, H. B., Gursoy, A., & Keskin, O. (2014). Hot spots in protein–protein interfaces: Towards drug discovery. *Progress in Biophysics and Molecular Biology*, *116*(2), 165-173. https://doi.org/https://doi.org/10.1016/j.pbiomolbio.2014.06.003

[200] Choudhury, C. (2021). Fragment tailoring strategy to design novel chemical entities as potential binders of novel corona virus main protease. *Journal of Biomolecular Structure and Dynamics*, *39*(10), 3733-3746. https://doi.org/10.1080/07391102.2020.1771424

[201] Lamoree, B., & Hubbard, R. E. (2017). Current perspectives in fragment-based lead discovery (FBLD). *Essays Biochem*, *61*(5), 453-464. https://doi.org/10.1042/ebc20170028

[202] Bancet, A., Raingeval, C., Lomberget, T., Le Borgne, M., Guichou, J.-F., & Krimm, I. (2020). Fragment Linking Strategies for Structure-Based Drug Design. *Journal of Medicinal Chemistry*, *63*(20), 11420-11435. https://doi.org/10.1021/acs.jmedchem.0c00242

[203] Davis, B. J., & Roughley, S. D. (2017). Chapter Eleven - Fragment-Based Lead Discovery. In R. A. Goodnow (Ed.), *Annual Reports in Medicinal Chemistry* (Vol. 50, pp. 371-439). Academic Press. https://doi.org/https://doi.org/10.1016/bs.armc.2017.07.002

[204] Wang, T., Wu, M. B., Chen, Z. J., Chen, H., Lin, J. P., & Yang, L. R. (2015). Fragment-based drug discovery and molecular docking in drug design. *Curr Pharm Biotechnol*, *16*(1), 11-25. https://doi.org/10.2174/1389201015666141122204532

[205] Hall, R. J., Murray, C. W., & Verdonk, M. L. (2017). The Fragment Network: A Chemistry Recommendation Engine Built Using a Graph Database. Journal of medicinal chemistry, 60(14), 6440–6450. https://doi.org/10.1021/acs.jmedchem.7b00809

[206] Hubbard, R. (2008). Fragment approaches in structure-based drug discovery. *Journal of Synchrotron Radiation*, *15*(3), 227-230. https://doi.org/doi:10.1107/S090904950705666X

[207] Baurin, N., Baker, R., Richardson, C., Chen, I., Foloppe, N., Potter, A., . . . Hubbard, R. E. (2004). Drug-like Annotation and Duplicate Analysis of a 23-Supplier Chemical Database Totalling 2.7 Million Compounds. *Journal of Chemical Information and Computer Sciences*, *44*(2), 643-651. https://doi.org/10.1021/ci034260m

[208] Cheung, K.-M. J., Matthews, T. P., James, K., Rowlands, M. G., Boxall, K. J., Sharp, S. Y., . . . Workman, P. (2005). The identification, synthesis, protein crystal structure and in vitro biochemical evaluation of a new 3,4-diarylpyrazole class of Hsp90 inhibitors. *Bioorganic & Medicinal Chemistry Letters*, *15*(14), 3338-3343. https://doi.org/https://doi.org/10.1016/j.bmcl.2005.05.046

[209] Rudling, A., Gustafsson, R., Almlöf, I., Homan, E., Scobie, M., Warpman Berglund, U., . . . Carlsson, J. (2017). Fragment-Based Discovery and Optimization of Enzyme Inhibitors by Docking of Commercial Chemical Space. *Journal of Medicinal Chemistry*, *60*(19), 8160-8169. https://doi.org/10.1021/acs.jmedchem.7b01006

[210] Bruns, R. F., & Watson, I. A. (2012). Rules for Identifying Potentially Reactive or Promiscuous Compounds. *Journal of Medicinal Chemistry*, *55*(22), 9763-9772. https://doi.org/10.1021/jm301008n

[211] Wallach, I., Dzamba, M., & Heifets, A. (2015). AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery.

[212] Sunilkumar, D., Bose, C., Sanu, S. K., Pandurangan, N., Kumar, G., Banerji, A., & Nair, B. (2017). Coconut shell derived bioactive compound Oxyresveratrol mediates regulation of Matrix metalloproteinase 9. *International Journal of Pharma and Bio Science*, *8*. https://doi.org/10.22376/ijpbs.2017.8.1.p202-210

[213] Rahman, M. A., Bishayee, K., Sadra, A., & Huh, S.-O. (2017). Oxyresveratrol activates parallel apoptotic and autophagic cell death pathways in neuroblastoma

cells. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1861*(2), 23-36. https://doi.org/https://doi.org/10.1016/j.bbagen.2016.10.025

[214] van den Brand, A. D., Villevoye, J., Nijmeijer, S. M., van den Berg, M., & van Duursen, M. B. M. (2019). Anti-tumor properties of methoxylated analogues of resveratrol in malignant MCF-7 but not in non-tumorigenic MCF-10A mammary epithelial cell lines. *Toxicology*, *422*, 35-43. https://doi.org/https://doi.org/10.1016/j.tox.2019.04.009

[215] Hasriadi, & Limpeanchob, N. (2018). *In vitro* cytotoxicity of *Artocarpus lakoocha* aqueous extract and oxyresveratrol in SH-SY5Y cells. *Journal of Physics: Conference Series*, *1028*, 012009. https://doi.org/10.1088/1742-6596/1028/1/012009

[216] Sunilkumar, D., Drishya, G., Chandrasekharan, A., Shaji, S. K., Bose, C., Jossart, J., . . . Nair, B. G. (2020). Oxyresveratrol drives caspase-independent apoptosis-like cell death in MDA-MB-231 breast cancer cells through the induction of ROS. *Biochemical Pharmacology*, *173*, 113724. https://doi.org/https://doi.org/10.1016/j.bcp.2019.113724

[217] Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., . . . Bolton, E. E. (2020). PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Research*, *49*(D1), D1388-D1395. https://doi.org/10.1093/nar/gkaa971

[218] Batool, M., Ahmad, B., & Choi, S. (2019). A Structure-Based Drug Discovery Paradigm. *Int J Mol Sci*, *20*(11). https://doi.org/10.3390/ijms20112783

[219] Anderson, A. C. (2003). The process of structure-based drug design. *Chem Biol*, *10*(9), 787-797. https://doi.org/10.1016/j.chembiol.2003.09.002

[220] Dubey, A., Takeuchi, K., Reibarkh, M., & Arthanari, H. (2020). The role of NMR in leveraging dynamics and entropy in drug design. *J Biomol NMR*, *74*(10-11), 479-498. https://doi.org/10.1007/s10858-020-00335-9

[221] Robertson, M. J., Meyerowitz, J. G., & Skiniotis, G. (2022). Drug discovery in the era of cryo-electron microscopy. *Trends Biochem Sci*, *47*(2), 124-135. https://doi.org/10.1016/j.tibs.2021.06.008

[222] Sugiki, T., Furuita, K., Fujiwara, T., & Kojima, C. (2018). Current NMR Techniques for Structure-Based Drug Discovery. *Molecules*, *23*(1). https://doi.org/10.3390/molecules23010148

[223] Maveyraud, L., & Mourey, L. (2020). Protein X-ray Crystallography and Drug Discovery. *Molecules*, *25*(5). https://doi.org/10.3390/molecules25051030

[224] F. L. Hünefeld, Chemistry in animal organization: Physiological-chemical investigations of the material changes, or of the educational life in the animal organism, in particular of the process of blood formation, of the nature of the corpuscles and their nuclei. Leipzig: Brockhaus, 1840.

[225] Bernal, J. D., & Crowfoot, D. (1934). X-Ray Photographs of Crystalline Pepsin. *Nature*, *133*(3369), 794-795. https://doi.org/10.1038/133794b0

[226] Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G., & North, A. C. (1960). Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. *Nature*, *185*(4711), 416-422. https://doi.org/10.1038/185416a0

[227] Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C., & Shore, V. C. (1960). Structure of myoglobin: A three-dimensional Fourier synthesis at 2 A. resolution. *Nature*, *185*(4711), 422-427. https://doi.org/10.1038/185422a0

[228] Hogg, T., & Hilgenfeld, R. (2007). Protein Crystallography in Drug Discovery. In J. B. T. Taylor, David J. B. T. (Ed.), *Comprehensive Medicinal Chemistry II* (Vol. 3, pp. 875-900). https://doi.org/10.1016/B0-08-045044-X/00111-5

[229] Giegé , R., Dock , A. C., Kern , D., Lorber , B., Thierry , J. C., & Moras , D. (1986). The role of purification in the crystallization of proteins and nucleic acids. *Journal of Crystal Growth*, *76*(3), 554-561. https://doi.org/10.1016/0022-0248(86)90172-7

[230] Chernov, A. A. (2003). Protein crystals and their growth. *Journal of Structural Biology*, *142*(1), 3-21. https://doi.org/10.1016/S1047-8477(03)00034-0

[231] Vekilov, P. G., & Chernov, A. A. (2003). The physics of protein crystallization. *Solid State Physics*, *57*, 1-147. https://doi.org/10.1016/S0081-1947(08)60179-4

[232] Forsythe, E. L., Maxwell, D. L., & Pusey, M. (2002). Vapor diffusion, nucleation rates and the reservoir to crystallization volume ratio. *Acta Crystallogr D Biol Crystallogr*, *58*(Pt 10 Pt 1), 1601-1605. https://doi.org/10.1107/s0907444902014208

[233] Rosenberger, F. E. (1979). *Fundamentals of Crystal Growth I* (X-532, Ed. 1 ed.). https://doi.org/10.1007/978-3-642-81275-0

[234] McPherson, A. (2004). Introduction to protein crystallization. *Methods*, *34*(3), 254-265. https://doi.org/10.1016/j.ymeth.2004.03.019

[235] McPherson, A., & Gavira, J. A. (2014). Introduction to protein crystallization. *Acta Crystallogr F Struct Biol Commun*, *70*(Pt 1), 2-20. https://doi.org/10.1107/S2053230X13033141

[236] Govada, L., & Chayen, N. E. (2019). Choosing the Method of Crystallization to Obtain Optimal Results. *Crystals*, *9*(2), 106.

[237] Roosen-Runge, F., Zhang, F., Schreiber, F., & Roth, R. (2014). Ion-activated attractive patches as a mechanism for controlled protein interactions. *Sci Rep*, *4*, 7016. https://doi.org/10.1038/srep07016

[238] McPherson, A., & Cudney, B. (2014). Optimization of crystallization conditions for biological macromolecules. *Acta Crystallogr F Struct Biol Commun*, *70*(Pt 11), 1445-1467. https://doi.org/10.1107/S2053230X14019670

[239] Luft, J. R., Newman, J., & Snell, E. H. (2014). Crystallization screening: the influence of history on current practice. *Acta Crystallogr F Struct Biol Commun*, *70*(Pt 7), 835-853. https://doi.org/10.1107/S2053230X1401262X

[240] Hassell, A. M., An, G., Bledsoe, R. K., Bynum, J. M., Carter, H. L., Deng, S. J., . . . Shewchuk, L. M. (2007). Crystallization of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr*, *63*(Pt 1), 72-79. https://doi.org/10.1107/S0907444906047020

[241] Wooh, J. W., Kidd, R. D., Martin, J. L., & Kobe, B. (2003). Comparison of three commercial sparse-matrix crystallization screens. *Acta Crystallogr D Biol Crystallogr*, *59*(Pt 4), 769-772. https://doi.org/10.1107/s0907444903002919

[242] Dale, G. E., Oefner, C., & D'Arcy, A. (2003). The protein as a variable in protein crystallization. *Journal of Structural Biology*, *142*(1), 88-97. https://doi.org/https://doi.org/10.1016/S1047-8477(03)00041-8

[243] Kim, Y., Quartey, P., Li, H., Volkart, L., Hatzos, C., Chang, C., . . . Joachimiak, A. (2008). Large-scale evaluation of protein reductive methylation for improving protein crystallization. *Nat Methods*, *5*(10), 853-854. https://doi.org/10.1038/nmeth1008-853

[244] Wernimont, A., & Edwards, A. (2009). In situ proteolysis to generate crystals for structure determination: an update. *PLoS One*, *4*(4), e5094. https://doi.org/10.1371/journal.pone.0005094

[245] Bergfors, T. (2003). Seeds to crystals. *Journal of Structural Biology*, *142*(1), 66-76. https://doi.org/https://doi.org/10.1016/S1047-8477(03)00039-X

[246] Ireton, G. C., & Stoddard, B. L. (2004). Microseed matrix screening to improve crystals of yeast cytosine deaminase. *Acta Crystallogr D Biol Crystallogr*, *60*(Pt 3), 601-605. https://doi.org/10.1107/S0907444903029664

[247] D'Arcy, A., Villard, F., & Marsh, M. (2007). An automated microseed matrix-screening method for protein crystallization. *Acta Crystallogr D Biol Crystallogr*, *63*(Pt 4), 550-554. https://doi.org/10.1107/S0907444907007652

[248] D'Arcy, A., Bergfors, T., Cowan-Jacob, S. W., & Marsh, M. (2014). Microseed matrix screening for optimization in protein crystallization: what have we learned? *Acta Crystallogr F Struct Biol Commun*, *70*(Pt 9), 1117-1126. https://doi.org/10.1107/S2053230X14015507

[249] Jang, K., Kim, H. G., Hlaing, S. H. S., Kang, M., Choe, H.-W., & Kim, Y. J. (2022). A Short Review on Cryoprotectants for 3D Protein Structure Analysis. *Crystals*, *12*(2), 138.

[250] Garman, E. F., & Owen, R. L. (2006). Cryocooling and radiation damage in macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, *62*(Pt 1), 32-47. https://doi.org/10.1107/S0907444905034207

[251] Huxley, H. E., & Holmes, K. C. (1997). Development of Synchrotron Radiation as a High-Intensity Source for X-ray Diffraction. *J Synchrotron Radiat*, *4*(Pt 6), 366-379. https://doi.org/10.1107/S0909049597011618

[252] Su, X. D., Zhang, H., Terwilliger, T. C., Liljas, A., Xiao, J., & Dong, Y. (2015). Protein Crystallography from the Perspective of Technology Developments. *Crystallogr Rev*, *21*(1-2), 122-153. https://doi.org/10.1080/0889311X.2014.973868

[253] Smyth, M. S., & Martin, J. H. J. (2000). X-Ray crystallography. *Molecular Pathology*, *53*(1), 8-14. https://doi.org/10.1136/mp.53.1.8 Smyth, M. S., & Martin, J. H. J. (2000). X-Ray crystallography. *Molecular Pathology*, *53*(1), 8-14. https://doi.org/10.1136/mp.53.1.8

[254] Bragg, W. H., & Bragg, W. L. (1913). The reflection of X-rays by crystals. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, *88*(605), 428-438. https://doi.org/doi:10.1098/rspa.1913.0040

[255] Gruner, S. M. (1994). X-ray detectors for macromolecular crystallography. *Current Opinion in Structural Biology*, *4*(5), 765-769. https://doi.org/https://doi.org/10.1016/S0959-440X(94)90177-5

[256] Ferrari, C., & Bocchi, C. (2008). 4 - Strain and composition determination in semiconducting heterostructures by high-resolution X-ray diffraction. In C. Lamberti (Ed.), *Characterization of Semiconductor Heterostructures and Nanostructures* (pp. 93-132). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-444-53099-8.00004-X

[257] Cassetta, A. (2015). X-Ray Diffraction (XRD). In E. Drioli & L. Giorno (Eds.), *Encyclopedia of Membranes* (pp. 1-3). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-40872-4_1102-2

[258] Richards, F. M. (1977). AREAS, VOLUMES, PACKING, AND PROTEIN STRUCTURE. *Annual Review of Biophysics and Bioengineering*, *6*(1), 151-176. https://doi.org/10.1146/annurev.bb.06.060177.001055

[259] Dauter, Z. (2013). Placement of molecules in (not out of) the cell. *Acta Crystallogr D Biol Crystallogr*, *69*(Pt 1), 2-4. https://doi.org/10.1107/S0907444912044794

[260] Carvalho, A. L., Trincão, J., & Romão, M. J. (2010). X-Ray Crystallography in Drug Discovery. In A. C. A. Roque (Ed.), *Ligand-Macromolecular Interactions in Drug Discovery: Methods and Protocols* (pp. 31-56). Humana Press. https://doi.org/10.1007/978-1-60761-244-5_3

[261] Bragg, W. H. (1915). The Distribution of the Electrons in Atoms. *Nature*, *95*(2378), 344-344. https://doi.org/10.1038/095344a0

[262] Rupp, B., & Kantardjieff, K. (2010). *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*. Garland Science.

[263] Taylor, G. L. (2010). Introduction to phasing. *Acta Crystallogr D Biol Crystallogr*, *66*(Pt 4), 325-338. https://doi.org/10.1107/S0907444910006694

[264] Dauter, Z. (2013). SAD/MAD Phasing. In R. Read, A. G. Urzhumtsev, & V. Y. Lunin, *Advancing Methods for Biomolecular Crystallography* Dordrecht.

[265] Rose, J. P., & Wang, B.-C. (2016). SAD phasing: History, current impact and future opportunities. *Archives of Biochemistry and Biophysics*, *602*, 80-94. https://doi.org/https://doi.org/10.1016/j.abb.2016.03.018

[266] Karle, J. (1980). Some developments in anomalous dispersion for the structural investigation of macromolecular systems in biology. *International Journal of Quantum Chemistry*, *18*(S7), 357-367. https://doi.org/https://doi.org/10.1002/qua.560180734

[267] Hendrickson, W. A. (1991). Determination of Macromolecular Structures from Anomalous Diffraction of Synchrotron Radiation. *Science*, *254*(5028), 51-58. https://doi.org/doi:10.1126/science.1925561

[268] Wang, B.-C. (1985). Resolution of phase ambiguity in macromolecular crystallography. In *Methods in Enzymology* (Vol. 115, pp. 90-112). Academic Press. https://doi.org/https://doi.org/10.1016/0076-6879(85)15009-3

[269] Terwilliger, T. (1999). Reciprocal-space solvent flattening. *Acta Crystallographica Section D*, *55*(11), 1863-1871. https://doi.org/doi:10.1107/S0907444999010033

[270] Evans, P., & McCoy, A. (2008). An introduction to molecular replacement. *Acta Crystallographica Section D*, *64*(1), 1-10. https://doi.org/doi:10.1107/S0907444907051554

[271] Adams, P. D., Pannu, N. S., Read, R. J., & Brunger, A. T. (1999). Extending the limits of molecular replacement through combined simulated annealing and maximum-likelihood refinement. *Acta Crystallographica Section D*, *55*(1), 181-190. https://doi.org/https://doi.org/10.1107/S0907444998006635

[272] Lamb, A. L., Kappock, T. J., & Silvaggi, N. R. (2015). You are lost without a map: Navigating the sea of protein structures. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1854*(4), 258-268. https://doi.org/https://doi.org/10.1016/j.bbapap.2014.12.021

[273] Chen, Z. P., Morris, J., Martin, E., Hammond, R. B., Lai, X., Ma, C., . . . Bytheway, R. (2005). Enhancing the Signal-to-Noise Ratio of X-ray Diffraction Profiles by Smoothed Principal Component Analysis. *Analytical Chemistry*, *77*(20), 6563-6570. https://doi.org/10.1021/ac050616c

[274] Dauter, Z. (2017). Collection of X-Ray Diffraction Data from Macromolecular Crystals. In A. Wlodawer, Z. Dauter, & M. Jaskolski (Eds.), *Protein Crystallography: Methods and Protocols* (pp. 165-184). Springer New York. https://doi.org/10.1007/978-1-4939-7000-1_7

[275] Carugo, O. (2018). How large B-factors can be in protein crystal structures. *BMC Bioinformatics*, *19*(1), 61. https://doi.org/10.1186/s12859-018-2083-8

[276] Laskowski, R. A., MacArthur, M. W., & Thornton, J. M. (1998). Validation of protein models derived from experiment. *Current Opinion in Structural Biology*, *8*(5), 631-639. https://doi.org/https://doi.org/10.1016/S0959-440X(98)80156-5

[277] Kleywegt, G. (2000). Validation of protein crystal structures. *Acta Crystallographica Section D*, *56*(3), 249-265. https://doi.org/doi:10.1107/S0907444999016364

[278] Engh, R. A., & Huber, R. Structure quality and target parameters. In *International Tables for Crystallography* (pp. 474-484). https://doi.org/https://doi.org/10.1107/97809553602060000857

[279] Papoian, G. A., Ulander, J., Eastwood, M. P., Luthey-Schulten, Z., & Wolynes, P. G. (2004). Water in protein structure prediction. *Proceedings of the National Academy of Sciences*, *101*(10), 3352-3357. https://doi.org/doi:10.1073/pnas.0307851100

[280] Carugo, O. (1999). Correlation between occupancy and B factor of water molecules in protein crystal structures. *Protein Engineering, Design and Selection*, *12*(12), 1021-1024. https://doi.org/10.1093/protein/12.12.1021

[281] Turnbull, A. P., & Emsley, P. (2013). Studying Protein–Ligand Interactions Using X-Ray Crystallography. In M. A. Williams & T. Daviter (Eds.), *Protein-Ligand Interactions: Methods and Applications* (pp. 457-477). Humana Press. https://doi.org/10.1007/978-1-62703-398-5_17

[282] Danley, D. (2006). Crystallization to obtain protein-ligand complexes for structure-aided drug design. *Acta Crystallographica Section D*, *62*(6), 569-575. https://doi.org/doi:10.1107/S0907444906012601

[283] Muller, I. (2017). Guidelines for the successful generation of protein-ligand complex crystals. *Acta Crystallographica Section D*, *73*(2), 79-92. https://doi.org/doi:10.1107/S2059798316020271

[284] Chruszcz, M., Potrzebowski, W., Zimmerman, M. D., Grabowski, M., Zheng, H., Lasota, P., & Minor, W. (2008). Analysis of solvent content and oligomeric states in protein crystals—does symmetry matter? *Protein Science*, *17*(4), 623-632. https://doi.org/https://doi.org/10.1110/ps.073360508

[285] Matthews, B. W. (1968). Solvent content of protein crystals. *Journal of Molecular Biology*, *33*(2), 491-497. https://doi.org/https://doi.org/10.1016/0022-2836(68)90205-2

[286] Vilenchik, L. Z., Griffith, J. P., St. Clair, N., Navia, M. A., & Margolin, A. L. (1998). Protein Crystals as Novel Microporous Materials. *Journal of the American Chemical Society*, *120*(18), 4290-4294. https://doi.org/10.1021/ja973449+

[287] Weichenberger, C. X., Afonine, P. V., Kantardjieff, K., & Rupp, B. (2015). The solvent component of macromolecular crystals. *Acta Crystallographica Section D*, *71*(5), 1023-1038. https://doi.org/doi:10.1107/S1399004715006045

[288] Skarzynski, T., & Thorpe, J. (2006). Industrial perspective on X-ray data collection and analysis. *Acta Crystallographica Section D*, *62*(1), 102-107. https://doi.org/doi:10.1107/S0907444905034281

[289] Hartshorn, M. J., Murray, C. W., Cleasby, A., Frederickson, M., Tickle, I. J., & Jhoti, H. (2005). Fragment-Based Lead Discovery Using X-ray Crystallography. *Journal of Medicinal Chemistry*, *48*(2), 403-413. https://doi.org/10.1021/jm0495778

[290] Snook, C. F., Purdy, M. D., & Wiener, M. C. (2000). Use of a crystallization robot to set up sitting-drop vapor-diffusion crystallization and in situ crystallization screens. *Journal of Applied Crystallography*, *33*(2), 344-349. https://doi.org/doi:10.1107/S0021889800001151

[291] Till, M., Robson, A., Byrne, M. J., Nair, A. V., Kolek, S. A., Shaw Stewart, P. D., & Race, P. R. (2013). Improving the success rate of protein crystallization by random microseed matrix screening. Journal of visualized experiments : JoVE, (78), 50548. https://doi.org/10.3791/50548

[292] Kabsch, W. (2010). XDS. *Acta Crystallogr D Biol Crystallogr*, *66*(Pt 2), 125-132. https://doi.org/10.1107/S0907444909047337

[293] McCoy, A. J., Nicholls, R. A., & Schneider, T. R. (2013). SCEDS: protein fragments for molecular replacement in Phaser. *Acta Crystallographica Section D*, *69*(11), 2216-2225. https://doi.org/doi:10.1107/S0907444913021811

[294] Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I., . . . Adams, P. D. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica Section D*, *75*(10), 861-877. https://doi.org/doi:10.1107/S2059798319011471

[295] Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, *66*(Pt 4), 486-501. https://doi.org/10.1107/S0907444910007493

[296] Moriarty, N. W., Grosse-Kunstleve, R. W., & Adams, P. D. (2009). electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallographica Section D*, *65*(10), 1074-1080. https://doi.org/doi:10.1107/S0907444909029436

[297] Otwinowski, Z., & Minor, W. (1997). [20] Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology* (Vol. 276, pp. 307-326). Academic Press. https://doi.org/https://doi.org/10.1016/S0076-6879(97)76066-X