

January 2015

Discovery of Molecules that Modulate Protein-Protein Interactions in the Context of Human Proliferating Cell Nuclear Antigen-Associated Processes of DNA Replication and Damage Repair

Matthew David Bartolowits
Purdue University

Follow this and additional works at: https://docs.lib.purdue.edu/open_access_dissertations

Recommended Citation

Bartolowits, Matthew David, "Discovery of Molecules that Modulate Protein-Protein Interactions in the Context of Human Proliferating Cell Nuclear Antigen-Associated Processes of DNA Replication and Damage Repair" (2015). *Open Access Dissertations*. 1089.

https://docs.lib.purdue.edu/open_access_dissertations/1089

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

**PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance**

This is to certify that the thesis/dissertation prepared

By Matthew David Bartolowits

Entitled

DISCOVERY OF MOLECULES THAT MODULATE PROTEIN-PROTEIN INTERACTIONS
IN THE CONTEXT OF HUMAN PROLIFERATING CELL NUCLEAR ANTIGEN-ASSOCIATED
PROCESSES OF DNA REPLICATION AND DAMAGE REPAIR

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Vincent Jo Davisson

Chair

Robert L. Geahlen

Markus A. Lill

Mark A. Lipton

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Vincent Jo Davisson

Approved by: Val J. Watts

Head of the Departmental Graduate Program

10/20/2015

Date

DISCOVERY OF MOLECULES THAT MODULATE PROTEIN-PROTEIN INTERACTIONS
IN THE CONTEXT OF HUMAN PROLIFERATING CELL NUCLEAR ANTIGEN-
ASSOCIATED PROCESSES OF DNA REPLICATION AND DAMAGE REPAIR

A Dissertation
Submitted to the Faculty
of
Purdue University
by
Matthew David Bartolowits

In Partial Fulfillment of the
Requirements for the Degree
of
Doctor of Philosophy

December 2015
Purdue University
West Lafayette, Indiana

To my parents, David and Leslie, who provided the means to pursue my dreams of happiness, and other family members and friends who gave me the strength to persevere.

ACKNOWLEDGEMENTS

The past several years have proved, collectively, to be the biggest challenge I have faced in my life, and without the help and encouragement of others, none of it would have been possible. I have had the pleasure of working alongside numerous excellent graduate students, postdoctoral researchers, technicians, undergraduates and professors, whose individual expertises proved invaluable to my personal and professional development.

I would first like to thank my advisor, Dr. V. Jo Davisson, for his guidance throughout the years, and for providing an environment in which I have been able to grow. He helped me to constantly push my own boundaries and to explore new innovative concepts. The opportunities that he has provided for me to realize and pursue my goals are immensely appreciated, and I consider myself very fortunate to have been able to work with him. I would also like to recognize the support that each of the members of my thesis advisory committee has given me throughout my time as a graduate student: Dr. Markus Lill, Dr. Robert Geahlen, Dr. Laurie Parker and Dr. Mark Lipton.

I would like to acknowledge all of the current and past members of the Davisson laboratory for all of their help in performing various tasks, for their expertise in experimentation and instrumentation, and for helping to make my time in the lab an enjoyable experience: Dr. Qinshou Chen, Raymond Fatig III, Dr. Anthony Pedley, Dr. Andrew Bieberich, Angela Wolf, Dr. Fiona Thomas, Dino Petrov, Kyle Denton, Aaron Lindstrom and Jonathan Gast. I would also like to thank all of the undergraduates that I worked with over the years, and I extend my best wishes to them in finishing their studies and having success in their careers: Kyle Harvey, Anthony Cirricione, Rachel O' Connor, Ashley Hasler, Jinny Zhang and Stephanie Grebinoski.

I would also like to recognize the contributions of several individuals for their help conducting experimentation, and for providing crucial data: Dr. Wells Brown for his extensive contribution to the nuclear targeting project, including Western blotting, and the culturing and imaging of murine cell lines NME and LM1; Jennifer Sturgis for confocal imaging of the cellular uptake of NLys-based compounds; Raymond Fatig III for 2D MTT cell growth assays; Amr Abdallah for his assistance conducting the Bayesian classification and subsequent principal

component analysis on the tripeptoids; Dr. Anthony Pedley for his help learning how to perform molecular dynamics simulations, and for his assistance in incorporating the fitting equations for fluorescence polarization into OriginPro; Rachel O' Conner and Ashley Hasler, whose help in the synthesis of the peptoid library proved to be invaluable; Anthony Cirrincione for his assistance with protein purification. There are many others who may not be individually mentioned, but I want to give them all my great thanks.

Finally, I want to express my immeasurable gratitude to my family and friends. Without their help, I could not have hoped to complete this journey; and without the love and support of my family, I would never have been provided the means to pursue my dreams.

Financial support for this work was provided by the Purdue Research Foundation and Purdue University Graduate School.

“We must be willing to let go of the life we have planned, so as to have the life that is waiting for us.”

- E. M. Forster

TABLE OF CONTENTS

	Page
LIST OF TABLES	ix
LIST OF FIGURES	x
LIST OF SCHEMES.....	xiii
ABBREVIATIONS	xiv
NOMENCLATURE	xvii
ABSTRACT	xix
1. INTRODUCTION	1
1.1. Regulation of Protein Repair Complexes at Sites of DNA Damage	2
1.1.1. Targeting Cancer-Specific Defects in DNA Repair Pathways	3
1.1.1.1. Using Small Molecules to Target Processes of DNA Damage Repair	5
1.2. Considerations of Protein Subpockets in Fragment-Based Drug Design	5
1.2.1. Protein Fragment Recognition in Drug Design	7
1.2.1.1. Concept of Shared Subpockets	7
1.2.1.2. Fragment Diversity and Protein Interactions	8
1.2.2. Identifying Subpockets	11
1.2.2.1. Using Protein Topology to Search for Binding Sites	12
1.2.2.2. Ligand Interaction Approaches to Compare Binding Sites	17
1.2.3. Shared Subpockets in ATP Binding Sites and Relevance to Ligand Selectivity	19
1.2.4. Considerations of Protein Interfaces and Cooperative Interactions	21
1.2.5. Summary of Computational Approaches toward FBDD	25
1.3. Research Scope	26
2. PEPTOIDS AS PROBES FOR DISRUPTING PCNA-PROTEIN COMPLEXES	28
2.1. Protein-Protein Interactions as Therapeutic Targets	29
2.1.1. Targeting Protein Interfaces with Small Molecules	29
2.2. Proliferating Cell Nuclear Antigen is Crucial for the Processes of DNA Damage Repair and Replication	31
2.2.1. PCNA Phosphorylation Controls Its Stability on Chromatin and is Associated with Dysregulation of DNA Mismatch Repair	32

	Page
2.2.2. PCNA as a Molecular Drug Target	34
2.2.2.1. Targeting the PCNA PIP Box Binding Site	35
2.3. Peptoids as a Fragment-Based Screening Tool	36
2.4. Rationale	38
2.5. Experimental	40
2.5.1. Methodology	40
2.5.2. Experimental Materials and Methods	44
2.5.3. Synthesis of Non-Commercially Available Primary Amines and T2AA	53
2.6. Results	61
2.6.1. Preferences for Fragments along the Peptoid Backbone	62
2.6.2. Virtual Incorporation of Second Generation Tripeptoid Ligands	64
2.6.3. In Vitro Screening of Synthesized Tripeptoids	67
2.6.4. Identification of Molecular Recognition Features of PCNA for Peptoid Ligands	71
2.6.5. Analysis of the Diversity and Chemical Classification of Peptoid Inhibitors	76
2.7. Discussion	83
3. NUCLEAR TARGETING OF KNOWN CANCER THERAPEUTICS	86
3.1. Targeting Mechanisms of Developed Resistance to Drug Therapies	87
3.1.1. Subcellular Targeting as a Therapeutic Strategy	89
3.1.1.1. Peptides and Peptoids as Cellular Uptake or Nuclear Targeting Sequences	90
3.2. Rationale	93
3.3. Experimental	94
3.3.1. Methodology	94
3.3.2. Experimental Materials and Methods	96
3.3.3. Synthesis of Primary Amines, Gefitinib Analogue and Peptide-Peptoid Conjugates	100
3.4. Results	104
3.4.1. NLys-Based Conjugates Enhance the Anti-Proliferative Effect of Gefitinib in Resistant Cells	108
3.4.2. Drug-Peptoid Conjugates Alter the Phosphorylation Status of STAT3	111
3.4.3. NArg-Based Compounds Disrupt STAT3 Phosphorylation in a Dose- Dependent Manner	114
3.4.4. Drug-Peptoid Conjugates Downregulate EGFR-Dependent, but not JAK- Dependent STAT3 Phosphorylation	117
3.4.5. Gef-SV40-NArg Has Strong Potential for Oral Absorption	119
3.5. Discussion	121
4. FUTURE DIRECTIONS	124
4.1. Using Chemical Information from Peptoid Hits to Develop More Drug-Like Inhibitors	124
4.2. Further Assessment of the Nuclear Effect of Subcellularly-Targeted Gefitinib, and Exploring the Localization of Alternative Small Molecule Inhibitors	128

	Page
LIST OF REFERENCES	135
APPENDICES	
A. Characterization of PCNA Peptoid Inhibitors	171
B. Supplementary Information for Nuclear-Targeted Therapeutics and Cell Lines	194
VITA	198

LIST OF TABLES

Table		Page
1.1	Examples of Small Molecule Inhibitors of DDR-Associated Pathways in Cancer in Phase I/II/III or Pre-Clinical Trials	4
1.2	Preferential Amino Acid Fragments	10
1.3	Methods or Tools for Analysis of Binding Site Similarity and Pharmaceutical Discovery	13
2.1	Examples of Small Molecule Protein-Protein Interaction Inhibitors of DDR-Associated Pathways in Cancer	30
2.2	Methods for Dysregulating PCNA-Dependent Processes	35
2.3	Examples of Peptoid-Based Small Molecule Inhibitors	37
2.4	Hit Compound IC ₅₀ and K _i Values Measured by Fluorescence Polarization	70
2.5	Peptoid Descriptor Statistics from Bayesian Model	80
2.6	iPPI Descriptor Statistics from Bayesian Model	80
2.7	Non-iPPI Descriptor Statistics from Bayesian Model	80
2.8	Descriptor Statistics for Experimentally Identified Peptoid Inhibitors	82
3.1	Examples of Cell Penetrating or Nuclear Targeting Peptides	91
3.2	Examples of Peptoid Sequences that Enhance Cellular or Intracellular Uptake	92
3.3	Permeability of Gef-SV40-NArg in Caco-2 Cells	120
Appendix Table		
A.1	PCNA Crystal Structures Used for Computational Screening	171
A.2	Characterization of Screened Tripeptoids	173
A.3	Peptoid Number from Initial Tripeptoid FP Screen (Figure 2.8)	177
B.1	Solvent Gradient for HPLC Method	195
B.2	Characterization of Subcellularly-Targeted Peptoid- or Peptide-Based Conjugates	197

LIST OF FIGURES

Figure	Page
1.1	Shared Subpocket Between Two Structurally Distinct Proteins 8
1.2	Conserved Subpockets in Protein Kinases Contribute to their Inherent Promiscuity 20
1.3	Protein Interface Hotspots Contain Inducible Microenvironments that Bind Conserved Fragments Between Molecule Types 23
2.1	PCNA Structural Features and PIP Box Binding Site 33
2.2	Development of Initial Virtual Library 41
2.3	Creation of a Virtual Library Containing T2AA as a Fragment 43
2.4	Tripeptoids Dock at the PCNA PIP Box Binding Site 62
2.5	Percentage that Amine Fragments in the Initial Library Appeared in the Tripeptoid Backbone 63
2.6	Percentage that Fragments Including T2AA Appeared in the Tripeptoid Backbone 65
2.7	Second Generation Peptoid Ligands Dock at the PCNA PIP Box Binding Site 66
2.8	Initial Tripeptoid FP Screen 68
2.9	Dose Response Curves of Peptoid-Ligand Hits 69
2.10	Average Structures of the Final 50 Frames of Molecular Dynamic Simulations 72
2.11	Principle Component Analysis of PCNA Topology Variance 73
2.12	Peptoid Inhibitors Disrupt Key PCNA-PIP Box Interactions 75
2.13	Similarity Clustering Analysis of Peptoid Ligands 76
2.14	Ligand Classification Based On Principle Component Analysis of Descriptors 81
3.1	Enhanced Uptake of 5-Carboxyfluorescein Using a NLS or CPPo 105
3.2	Cell Penetrating Peptoid Sequences and Peptide-Peptoid Drug Conjugates Used in This Study 106
3.3	NArg-Peptoid Conjugates Enhance Nuclear Uptake 107
3.4	Activity of Commercial Gefitinib and Pip Gef in 3D Cell Cultures of NME and LM1 109

Figure	Page
3.5	NLys-Based Drug Conjugates Negatively Impact Growth of Gefitinib-Resistant Cell Lines 110
3.6	Effect of Nuclear-Targeted Conjugates on Phosphorylation Status of EGFR, ERK1/2 and STAT3..... 112
3.7	Analysis of the Phosphorylation Status of ERK1/2 and STAT3 in MDA-MB-468 Cells 113
3.8	Gef-SV40-NArg Disrupts Nuclear Accumulation of Phosphorylated STAT3 115
3.9	Dose Response of Gef-SV40-NArg against STAT3 and ERK1/2 in NME and LM1 Cells 116
3.10	Dose Response of NArg-Based Drug Conjugates in MDA-MB-468 Cells 118
3.11	Drug-Peptoid Conjugates Downregulate EGFR-Dependent Phosphorylation of STAT3 119
3.12	Cytoplasmic and Nuclear Modes of EGFR Signaling 122
4.1	Shared Chemical Features Between Peptoid Hits Can Be Incorporated into Lead Compounds 125
4.2	Set of Fragments for Incorporation into Future Virtual Libraries 126
4.3	Simultaneous Targeting of the Three PIP Box Binding Sites on PCNA 129
4.4	Alternative Reducible Linkers for Drug-Peptoid Conjugates 131
4.5	Variants of Protein Kinase Inhibitors Can Be Incorporated into a Peptide/Peptoid 133

Appendix Figure

A.1	Initial Peptoid Fragment Tally by PCNA Crystal Structure 172
A.2	Z'-Factor Analysis 176
A.3	PCNA Titration for Determining Affinity of FAM-PL 176
A.4	Chemical Structure of NMba 180
A.5	Necessary Protecting Groups on Peptoid Fragments 181
A.6	Dose Response Curves of Individual Peptoid Fragments 182
A.7	Structure of T2AA-Conjugates 183
A.8	Co-Crystal Structure of PCNA-PL Peptide (PDB ID: 1VYJ) 184
A.9	PCNA-T2AA MD Simulation Output 185
A.10	PCNA-NLys-NPip-NBal MD Simulation Output 186
A.11	PCNA-NLys-NTyr-NBal MD Simulation Output 187

Appendix Figure	Page
A.12 PCNA-T2AA-NEal-NPip MD Simulation Ouput	188
A.13 PCNA-T2AA-Gly-NPip MD Simulation Ouput	189
A.14 MD Simulation C α Atom RMSD Fluctuation by Residue Number	190
A.15 MD Simulation Interaction Diagrams	191
A.16 Enrichment Plot for Bayesian Model Prediction of the 2P2I Hunter Compound Set	192
A.17 Distribution of Peptoid Descriptor Statistics	193
B.1 Localization of EGFR in NME and LM1 Cell Lines	196

LIST OF SCHEMES

Scheme	Page
2.1 Synthesis of NArg	53
2.2 Synthesis of NBal	54
2.3 Synthesis of NBza	56
2.4 Synthesis of NTrp	58
3.1 Synthesis of Peptide-Peptoid Gefitinib Conjugates	95
3.2 Synthesis of PiperazinyI Gefitinib	101

ABBREVIATIONS

9-1-1	Rad9-Rad1-Hus1
ATM	Ataxia Telangiectasia Mutated
ATP	Adenosine Triphosphate
ADP	Adenosine Diphosphate
BRCA	Breast Cancer Susceptibility Gene
CPP	Cell Penetrating Peptide
CPPo	Cell Penetrating Peptoid
DCM	Dichloromethane
DDR	DNA Damage Response
DIC	N,N'-Diisopropylcarbodiimide
DIEA	N,N-Diisopropylethylamine
DMAP	4-Dimethylaminopyridine
DMEM	Dulbecco's Modified Eagle's Medium
DMF	N,N-Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic Acid
DNA-PK	DNA-Dependent Protein Kinase
DSB	DNA Double Stranded Break
DTT	Dithiothreitol
EDCI	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
ER	Estrogen Receptor
ERK	Extracellular-Signal-Regulated Kinase
ESI	Electrospray Ionization

FAM	5-Carboxyfluorescein
FBDD	Fragment-Based Drug Design
FBS	Fetal Bovine Serum
FP	Fluorescence Polarization
GPCR	G Protein-Coupled Receptor
HBSS	Hank's Balanced Salt Solution
HCTU	2-(6-Chloro-1H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High Performance Liquid Chromatography
HR	Homologous Recombination
HRMS	High-Resolution Mass Spectrometry
IDCL	(PCNA's) Interdomain Connecting Loop
i-PrOH	Isopropanol
IL-6	Interleukin-6
iPPI	Protein-Protein Interaction Inhibitor
IPTG	Isopropyl β -D-1-thiogalactopyranoside
JAK	Janus Kinase
LM1	Lung Metastatic 1 (derived from NME cells)
MALDI	Matrix Assisted Laser Desorption/Ionization
MD	Molecular Dynamics (simulation)
MeOH	Methanol
MMEJ	Microhomology-Mediated End Joining
nEGFR	Nuclear Epidermal Growth Factor Receptor
NHEJ	Non-Homologous End Joining
NLS	Nuclear Localization Sequence
NME	Normal Mammal Epithelial (cells)
NMR	Nuclear Magnetic Resonance
NMP	N-Methyl-2-pyrrolidone
p-ERK1/2	Phosphorylated ERK1/2
p-STAT3	Phosphorylated (Y705) STAT3
PAGE	Polyacrylamide Gel Electrophoresis
PARP	Poly (ADP-Ribose) Polymerase

PCA	Principal Component Analysis
PCNA	Proliferating Cell Nuclear Antigen
PDB	RCSB Protein Data Bank
PI3K	Phosphoinositide-3 Kinase
PIP	PCNA Interacting Protein
PL	Pogo Ligase
PPI	Protein-Protein Interaction
PVDF	Polyvinylidene Fluoride (membrane)
pY211	Phosphorylated (PCNA residue) Tyrosine 211
RFC	Replication Factor C
RMSD	Root-Mean-Square Deviation
RPA	Replication Protein A
SASA	Solvent Accessible Surface Area
SSB	DNA Single Stranded Break
STAT3	Signal Transducer and Activator of Transcription 3
TEA	Triethylamine
TFA	Trifluoroacetic Acid
TGF- β	Transcription Growth Factor Beta
THF	Tetrahydrofuran
TIS	Triisopropylsilane
TKI	Tyrosine Kinase Inhibitor
TLC	Thin Layer Chromatography
TLS	Translesion Synthesis

NOMENCLATURE

(** In this table, four-letter tripeptoid fragment codes [“NXxx”] describe the respective fragment in its chemical form containing any necessary protecting groups **)

FAM-PL	5-Carboxyfluorescein-labeled POGO ligase peptide
FAM-SV40-NArg	5-Carboxyfluorescein-labeled SV40-NArg peptoid
Gef	Gefitinib
Gef-NLys	Gefitinib-tagged poly-lysine (n=9) peptoid
Gef-SV40-NLys	Gefitinib-tagged poly-lysine (n=9) peptoid with SV40 nuclear localization peptide
Gef-NArg	Gefitinib-tagged poly-arginine (n=9) peptoid
Gef-SV40-NArg	Gefitinib-tagged poly-arginine (n=9) peptoid with SV40 nuclear localization peptide
NAea	Tert-butyl (4-(2-aminoethyl)phenyl)carbamate
NAem	4-(2-Aminoethyl)morpholine
NAIl	Prop-2-en-1-amine
NAmP	Pyridin-4-ylmethanamine
NArg	N-(N-(3-aminopropyl)carbamimidoyl)-2,2,4,6,7-pentamethyl-2,3-dihydrobenzofuran-5-sulfonamide
NAsp	Tert-butyl glycinate
NBal	(4-(((tert-butyldimethylsilyl)oxy)methyl)phenyl)methanamine
NBza	Tert-butyl 4-(aminomethyl)benzoate
NCyp	Cyclopentylamine
NEal	2-(((tert-butyldimethylsilyl)oxy)ethan-1-amine
NEba	Tert-butyl (4-(2-aminoethyl)benzyl)carbamate
NFfa	Furan-2-ylmethanamine
NGIn	3-amino-N-tritylpropanamide

Nile	sec-Butylamine
NLys	Tert-butyl (4-aminobutyl)carbamate
NLys ₇	7-member polylysine peptoid
NLys ₉	9-member polylysine peptoid
NMba	Tert-butyl (4-(aminomethyl)benzyl)carbamate
NMet	2-(methylthio)ethan-1-amine
NPhe	Phenylmethanamine
NPip	Piperonylamine
NTrp	Tert-butyl 3-(2-aminoethyl)-1H-indole-1-carboxylate
NTyr	(4-(tert-butoxy)phenyl)methanamine
NVal	Isopropylamine
Pip Gef	Piperazinyl Gefitinib
Piperazinyl Gefitinib	2-(4-(3-((4-((3-chloro-4-fluorophenyl)amino)-7-methoxyquinazolin-6-yl)oxy)propyl)piperazin-1-yl)acetic acid
SV40	Simian Vacuolating Virus 40 (NLS derived from the viral sequence)
T2AA	(S)-4-(4-(2-amino-3-hydroxypropyl)-2,6-diiodophenoxy)phenol

ABSTRACT

Bartolowits, Matthew David Ph.D., Purdue University, December 2015. Discovery of molecules that modulate protein-protein interactions in the context of human proliferating cell nuclear antigen-associated processes of DNA replication and damage repair. Major Professor: Vincent Jo Davisson.

Integral to cell viability is the homotrimeric protein complex Proliferating Cell Nuclear Antigen (PCNA) that encircles chromatin-bound DNA and functionally acts as a DNA clamp that provides topological sites for recruitment of proteins necessary for DNA replication and damage repair. PCNA has critical roles in the survival and proliferation of cells, as disease-associated dysregulation of associated functions can have dire effects on genome stability, leading to the formation of various malignancies ranging from non-Hodgkin's lymphoma to skin, laryngeal, ocular, prostate and breast cancers. Here, a strategy was explored with PCNA as a drug target that may have wider implications for targeting protein-protein interactions (PPIs) as well as for fragment-based drug design. A design platform using peptidomimetic small molecules was developed that maps ideal surface binding interaction sites at a PPI interface before considering detailed conformations of an optimal ligand. A novel *in silico* multi-fragment, combinatorial screening approach was used to guide the selection and subsequent synthesis of tripeptoid ligands, which were evaluated in a PCNA-based competitive displacement assay. From the results, some of the peptoid-based compounds that were synthesized displayed the ability to disrupt the interaction between PCNA and a PIP box-containing peptide. The IC_{50} values of these compounds had similar or improved affinity to that of T2AA, an established inhibitor of PCNA-PIP box interactions. The information gained here could be useful for subsequent drug lead candidate identification.

As a second goal of this study, the known anti-tumor agent, EGFR antagonist, gefitinib, was chemically modified and incorporated into subcellularly-targeted peptoid-peptide hybrid complexes. It is hypothesized that targeting this agent into the nucleus of tumor cells can allow for bypassing mechanisms of drug resistance observed in various tumor types, particularly triple

negative breast cancer. Enhanced cellular and nuclear uptake was observed with the drug conjugates. Furthermore, some of the compounds demonstrated the ability to selectively downregulate STAT3 phosphorylation, while not affecting the phosphorylation of ERK1/2, and had enhanced antiproliferative activity in gefitinib-resistant cells. Further studies are needed to assess the mechanism of action of these molecules, but they hold promise as agents that can be used to treat drug-resistant cancers.

1. INTRODUCTION

DNA is the repository in which genetic information is stored in every living cell, and its structural integrity is vital to proper cellular function and survival. At any given time, cells are being exposed to exogenous and endogenous agents that cause as many as tens of thousands to one million DNA lesions per day [1,2]. If the resulting damage is not repaired, it gives rise to genetic instability, and possibly disease [3–9]. Additionally, DNA damage has been observed to be a major factor in cellular senescence and aging [10,11]. In response to these damaging events, a complex process is initiated that comprises the detection and repair of DNA lesions. These make up what is referred to as the DNA damage response (DDR).

The DDR as a whole is composed of a number of distinct mechanisms that are activated in response to specific types of DNA damage. DNA lesions can arise from numerous sources, with examples such as replication errors during DNA synthesis, UV-induced linkages or radiation-/chemically-induced DNA double strand breaks (DSBs) [6]. Damage induced by exogenous agents can result in an unpredictable pattern of DNA lesions, potentially due to decreased efficiencies of the DNA repair process, particularly in the context of DNA damaging chemotherapies [6,12]. In general, cells that have defective mechanisms of DDR display enhanced sensitivity to DNA-damaging agents, and these defects can ultimately lead to disease. The major processes of damage repair and their associated signaling mechanisms have been extensively studied [6,13–15], and what is becoming clear are the implications for dysregulation of the DDR.

A number of human diseases, both inherited and non-inherited, have been associated with an accumulation of DNA lesions. Parkinson's, Huntington's and Alzheimer's diseases, in addition to ataxia, have been associated with an increased amount of DNA damage in neurons [9,16]. Infertility [17,18], metabolic syndrome [19,20] and heritable diseases such as Fragile X, diabetes mellitus type 2, Creutzfeldt-Jakob, amyotrophic lateral sclerosis and Leigh syndrome are also caused by deficiencies in the DDR [21–23]. Perhaps most notoriously, cancer arises from genome instability induced by large numbers of DNA lesions. Correspondingly, the majority of agents classified as carcinogens operate by causing DNA damage, many times resulting in mutations that

can negatively impact numerous cellular signaling and checkpoint pathways [24]. Individuals can also have predispositions to cancer based on inherited defects in gene products involved in the DDR, and a patient's outcome can be further complicated due to these same defects enabling neoplastic lesions to progress into malignancy [25,26]. Ultimately, these conditions highlight the importance of proper cellular response to inflicted DNA damage, as well as the careful maintenance of multiple genome stability pathways.

1.1 Regulation of Protein Repair Complexes at Sites of DNA Damage

Evoking an appropriate cellular response to DNA damage involves a complex series of events including the sensing and recognition of damage, subsequent signaling and repair of those lesions. There are various pathways that act to repair various types of exogenous and endogenous agent-induced damage in humans. The protein MGMT reverts the mutagenic O⁶-methylguanine lesion back to guanine, and is able to prevent mismatch errors during the replication and transcription of DNA [27]. When damage occurs that leads to breaks in a single strand of DNA, three major mechanisms, base excision, nucleotide excision and mismatch repair, are elicited [28]. These mechanisms function by using the opposite intact strand of DNA as a template, removing a damaged nucleotide and replacing it with the proper complementary nucleotide. Single strand breaks are the most common type of DNA damage that occurs [29], so efficient repair of these lesions is essential for cell survival.

Alternatively, DNA DSBs are particularly troublesome since they can result in rearrangements of the genome. DSBs are repaired by three different mechanisms, homologous recombination (HR), non-homologous end joining (NHEJ) and microhomology-mediated end joining (MMEJ) [6]. HR and NHEJ are the primary repair pathways invoked in mammalian DSB repair [30], though MMEJ does occur at relatively reduced levels. NHEJ, in particular, is a versatile system that exhibits a high degree of structural tolerance of DNA end substrate configurations. Owing to its importance, individuals that lack normally functioning NHEJ are sensitive to ionizing radiation, and are also severely immunodeficient [31].

Separate from the other mechanisms of DDR, translesion synthesis (TLS) is a process of DNA damage tolerance that enables DNA replication machinery to replicate past lesions. In general, if replication is initiated before a site of DNA damage is repaired, it can lead to the replication machinery stalling at the damage locus, ultimately leading to cell death. TLS involves the recruitment of specialized DNA polymerases that can bypass the site of the lesion, allowing

DNA synthesis to continue [32]. The switching out of regular polymerases for specialized ones is mediated, among various other factors [33], by RAD6/18-mediated ubiquitination of Proliferating Cell Nuclear Antigen (PCNA) [34]. While this damage tolerance mechanism aids cell survival, due to the nature of these specialized polymerases being error prone, undesired consequences, such as the formation of cancer, can potentially result, particularly if numerous DNA lesions continue to accumulate in a cell's genome [35].

As a protective measure, upon DNA damage cell cycle checkpoints become activated; this allows the cell to repair DNA lesions before continuing to divide. Typically, the type of damage is recognized and the appropriate signaling pathway is initiated, resulting in the formation of repair complexes at sites of lesions [15]. The proteins that initially detect lesions can be classified as “sensors”; this class is made up of DNA scaffolds such as the Rad9-Rad1-Hus1 (9-1-1) complex, replication protein A (RPA), phosphoinositide-3-kinase (PI3K) members ATM, ATR and DNA-PK, and the aforementioned PCNA [6]. Each of these recognizes and interacts with a number of mediatory proteins; this leads to the transduction of amplified signals towards effectors that affect processes such as cell cycle regulation, transcription, DNA repair, chromatin remodeling and apoptosis. In this way, the scaffold proteins act as sites of recruitment, forming complexes that regulate the essential functions associated with the DDR [15].

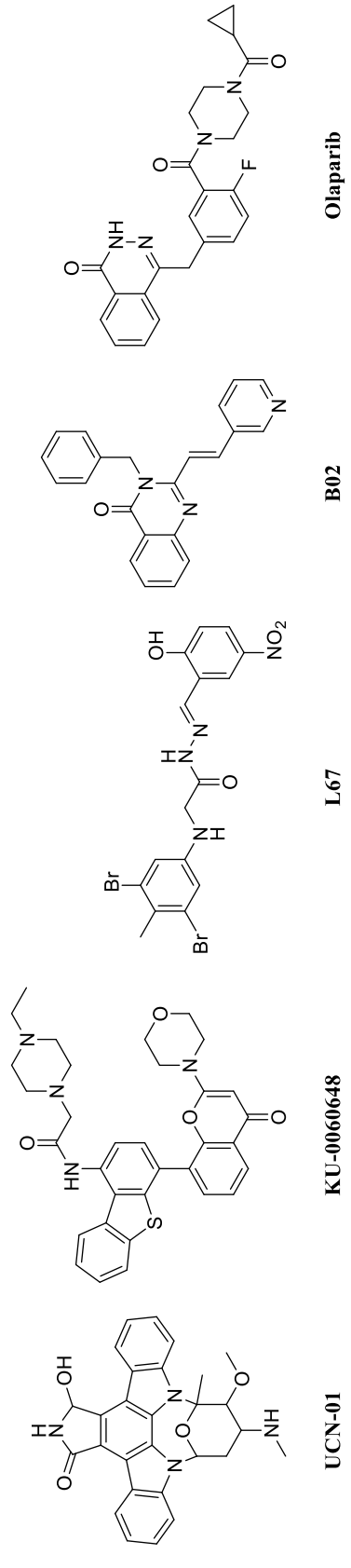
1.1.1 Targeting Cancer-Specific Defects in DNA Repair Pathways

The significance of genome maintenance, and subsequent cancer prevention, in the DDR is highlighted by the observation that genes that encode components of DNA repair pathways are among the most frequently mutated genes in cancer [36]. Studies have shown that as tumorigenesis progresses, cancer cells tend to inactivate the DDR and overcome senescence, potentially indicating that abrogating portions of DDR signaling pathways is necessary for malignant transformation [25,26,37–40]. This likely explains why defects in the DDR are typically seen in human tumors and why genomic instability is increasingly being seen as a hallmark of cancer [36,40].

Though hundreds of mutations may be present in a given cell's genome, only a select few of these may be necessary to induce carcinogenesis [41]. These “driver” mutations cause some sort of key dysregulation in a cell's growth control, and are the cause of specific cancers being ‘addicted’ to these oncogenic drivers. With the increase in knowledge of the mutational landscape of cancer genomes, function-oriented classification of cancer-associated genes has increased. As a result, the ability to classify tumors based on their unique repertoire of DNA repair deficiencies has

Table 1.1
 Examples of Small Molecule Inhibitors of DDR-Associated Pathways in Cancer in Phase I/II/III or Pre-Clinical Trials

Pathway	Target(s)	Drug Name	Application	Type of Cancer or Model Cell Line
Cell Cycle Checkpoint	Chk1	UCN-01	Explored in combination therapy with various DNA-damaging agents and topotecan	Various leukemia, lymphoma, pancreatic, breast, lung, ovarian, prostatic and kidney cancers
NHEJ	DNA-PK and PI3K	KU-0060648	Sensitization to etoposide and doxorubicin	Breast and colon cancer cells
Alternative NHEJ	DNA Ligases I and III α	L67	Sensitization to ionizing radiation, PARP inhibitors and methyl methanesulfonate	Chronic myeloid leukemia cells
Homologous Recombination	RAD51	B02	Sensitization to doxorubicin and cisplatin	Multiple myeloma and breast cancer cells
Base-Excision Repair	PARP	Olaparib (AZD2281)	Explored in combination therapy with various DNA-damaging agents, growth factor receptor inhibitors and enhanced radiotherapy	Various pancreatic, breast, lung, ovarian, gastric, colorectal, glioblastic, cervical and esophageal cancers



improved. Targeting these cellular defects has been a goal of drug development against cancer. Many current and traditional chemotherapies make use of a single drug to treat a specific type of cancer. However, these strategies have not always proven to be effective in the short- or long-term due to cellular toxicity and development of resistance [42]. Because mammals have evolved complex networks of parallel and overlapping pathways to repair diverse types of genotoxic lesions, it is likely that targeting multiple pathways simultaneously is necessary for a robust, sustained response to drug therapy.

1.1.1.1 Using Small Molecules to Target Processes of DNA Damage Repair

Three major classes of inhibitors exist that target processes of DNA repair: protein-protein interaction (see Table 2.1), DNA-protein interaction and enzyme-based inhibitors [5,43,44]. While enzyme inhibition has had success in treating certain diseases, many traditional chemotherapies and molecularly-targeted agents including protein kinases inhibitors have not provided sustained clinical efficacy by themselves [45–49]. Compounds such as DNA damaging agents often display initial benefit in cancer patients, but tumors eventually become resistant to these drugs due to various factors, including upregulation of the DDR [50]. As a result, new strategies are being explored for the treatment of disease including new rational therapeutic drug combinations [42].

A potential approach toward overcoming certain cellular mechanisms of drug resistance is to simultaneously target multiple pathways associated with DNA replication and damage repair. A number of small molecules in pre-clinical or clinical trials have been developed that target DDR-associated pathways such as cell cycle checkpoints, NHEJ, homologous recombination and base-excision repair (Table 1.1) [44]. These compounds are typically used in combination with other types of chemotherapeutics, and show enhanced sensitization to ionizing radiation, DNA damaging agents and growth factor receptor inhibition [44]. Ultimately, this strategy may prove to be far more efficacious, and in many cases, necessary in the treatment of cancer.

1.2 Considerations of Protein Subpockets in Fragment-Based Drug Design

(The following section is reproduced with permission from Bartolowits, M., and Davisson, V.J. (2015) [51]; © 2015 John Wiley & Sons A/S. doi: 10.1111/cbdd.12631.)

Fragment-based drug design (FBDD) is an important strategy in both industry and academia for the discovery of novel ligands and aids the progression toward lead compounds [52]. It is based on

the idea that through the use of low molecular weight chemical fragments, which typically only bind weakly to their intended target, higher affinity lead ligands can be obtained by combining or ‘growing’ these small compounds into larger drug-like molecules. Due to high levels of diversity between biological targets, incorporating FBDD as a high-throughput screening tool can have significant advantages over traditional higher molecular weight chemical libraries [53]. The approach considers factors such as compound availability, ease of synthesis, large chemical space, and limits on steric “bulkiness”, which may otherwise preclude many higher molecular weight ligands from recognizing non-covalent enthalpically-driven affinity factors (e.g. hydrogen bonding, etc.) at a target binding site. Indeed, FBDD can prove to be robust for rational fragment identification in the absence of 3D structural data [54]. However, FBDD still faces several challenges, such as the general lack of accountability for ligand specificity or selectivity [55,56], and the fact that key interactions and geometry of an original fragment hit may need to be changed when incorporated into a lead compound [57]. Furthermore, the role of ligand-dependent receptor conformations has been largely untested.

A number of reviews have discussed recent advances in fragment-based drug design and how these tools can be used to improve the lead design process [55,58–65]. Yet, relatively few evaluations have appeared which address the importance of small localized environments in a protein binding site, and how microenvironments ultimately drive ligand binding and observed effects such as drug non-specificity. The field of drug discovery has long struggled with the accurate prediction of a drug’s cross-pharmacologic profile [66–69] and side effects. However, recent advances offer avenues toward understanding the significance of protein “subpockets”—the physical chemical and geometric properties surrounding an individual residue. Traditional methods of computational analysis to find binding sites have sought to identify the similarity between proteins based on inherent sequence conservation or overall structural similarity. However, the localized chemical environments to be sampled by chemical fragment screens are potentially vast. Further development of generalized methods for discovery of similar protein-ligand binding sites and predicting the interaction profile of molecular fragments remains of high interest [70].

Recently, multiple methods have been developed that compare proteins based on localized environments within binding sites, or the chemical environment around protein-bound ligands derived from PDB database crystal structures. These methods take one of two general approaches: 1) proteins are compared and binding sites organized based on their structural or chemical similarities, or 2) data from known ligand-protein interactions are used to identify potentially similar sites in other proteins that could bind equivalent chemical fragments. This review aims to

provide an overview of the background and emerging computational approaches that can define, and take into account, the significance of protein subpockets in the design of new chemical probes and pharmaceuticals. Many of these same tools offer methods which can also enhance understanding of drug activity. The relevance of new methodological insights to fragment-based drug discovery and the implications for lead development when considering factors such as non-specificity and side effects are also considered.

1.2.1 Protein Fragment Recognition in Drug Design

1.2.1.1 Concept of Shared Subpockets

In many modern cases of structure-based drug discovery, structural information for a target in question drives the ligand development process [71,72]. The RCSB protein data bank (PDB) [73] currently contains more than 89,000 structures solved by X-ray crystallography, with more than 10,000 others solved through other means (NMR, electron microscopy, etc.). This knowledge database, which continues to expand, is a vital tool for understanding the general structures of target proteins and the topology of target ligand binding sites. Among “sibling” members of a protein subfamily, binding sites may have a moderate-to-high level of overall similarity, such as the ATP binding sites in the large family of protein kinases [74]. Although, even in cases where two arbitrarily selected proteins have overall dissimilar ligand-binding sites, they may still bind the same chemical fragments if they share similar or equivalent topological features at the subpocket level (see Figure 1.1 for an example). This explains why two binding sites with substantially different sequences can bind identical chemical fragments, such as the case with the trifluoromethyl- and sulfonyl-binding subpockets between isoenzymes of the carbonic anhydrase family and cyclooxygenase-2 [75]. Convergent evolution in nature is commonly observed to reveal similar enzyme active sites between proteins [76]. Likewise, divergent evolution can result in shared ligand binding sites among proteins that have highly dissimilar overall structures [77]. Between proteins even of distant origin, similar ligand binding motifs containing conserved consecutive residues can be found, indicating that it is quite efficient for nature to reuse localized features to bind similar molecules, even when the functional relationships may not be clear.

The principle of protein subpockets requires definition at the level of individual amino acids within a protein. Each residue is surrounded in three dimensions by other amino acid residues, water molecules and/or metal ions. The resultant microenvironments create the recognition features

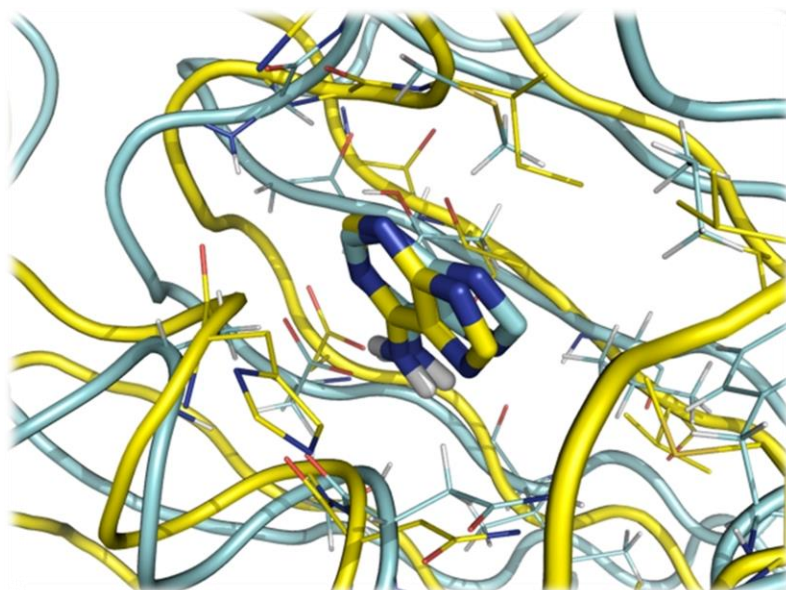


Figure 1.1. Shared Subpocket Between Two Structurally Distinct Proteins. The conformation of the adenine portion of ACP is conserved between two structurally and sequentially diverse proteins, demonstrating a shared subpocket at the binding site (HSP90 N-terminal domain, yellow, PDB ID: 3T10; Chemotaxis protein CheA, blue, PDB ID: 1I5A). Figure is reproduced with permission: © 2015 John Wiley & Sons A/S. doi: 10.1111/cbdd.12631.

for a particular chemical fragment, albeit with likely moderate ranges of affinities. These localized environments created by the protein generally define what are termed “subpockets” within a binding cavity. Even within the limits of 20 different standard amino acids and the amide backbone, a substantial number of distinct subpockets with unique “fingerprints” are possible. These subpockets differ by the relative proximity of specific amino acids to each other and distances between chemical functional groups. The amount of chemical space covered by chemical fragments is vast, but not all of them comprise preferable features for protein binding [78]. Therefore, any given chemical fragment is not guaranteed to bind to a protein subpocket across the entire proteome.

1.2.1.2 Fragment Diversity and Protein Interactions

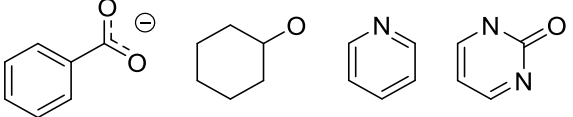
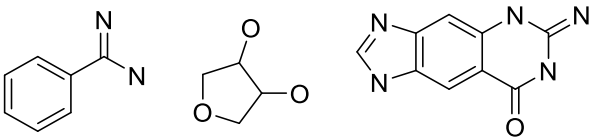
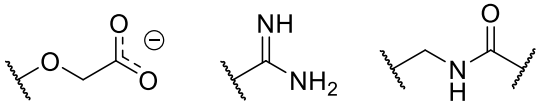
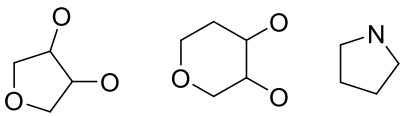
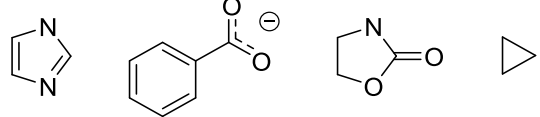
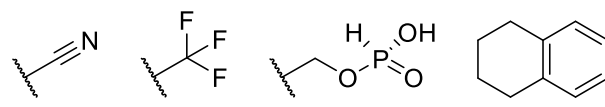
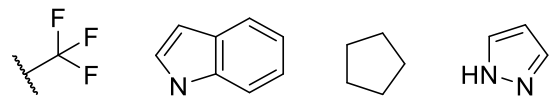
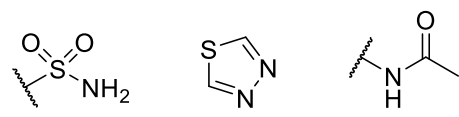
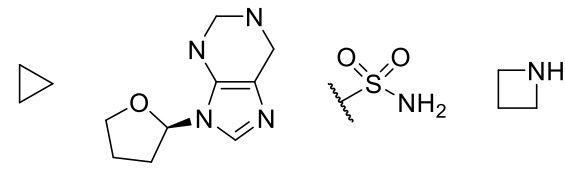
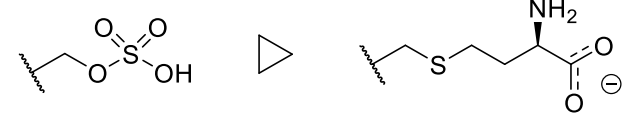
For a chemical fragment to bind with favorable free energy, it must have the appropriate overall shape, proper spacing of its chemical functional groups and generally compatible hydrophobic features to establish meaningful life-times within a subpocket. How these characteristics overlap

with the physiochemical properties associated with the characteristics of drug-like molecules warrants consideration. In principle, chemical fragments can represent high diversity in shape and chemical features, but only a subset of compounds that could be classified as “fragments” are useful for the purposes of screening against biological targets. A study by Zuegg and Cooper [79] analyzed more than eight million unique compounds from more than 100 chemical vendors and found that only 400,000 of these passed the fragment-like filter for the analysis of drug- and lead-likeness based upon “the rule of three” (molecular weight < 300 Da, ClogP < 3, the number of hydrogen bond donors and acceptors < 3, and the number of rotatable bonds < 3) [80]. A large number of fragments are represented in this subgroup, but there is an argument for ignoring a majority of commercially available compounds and enhancing practicality in a fragment-based screening platform. Of course, rules often have exceptions, and there is evidence that simply adhering to “the rule of three” for the design of a fragment library may preclude compounds that would otherwise show up as hits [81].

To further facilitate the fragment library design process, there are a number of studies that have examined the binding preferences of proteins for chemical fragments (Table 1.2). By analyzing fragments from lead-like compounds in the PDB [73], Chan *et al.* found that particular types of fragments are more likely to interact with specific amino acid side chains (Asp, Glu, Arg and His) and engage in hydrogen bonding interactions [82]. Separately, through the use of a developed algorithm, LigFrag-RPM, Wang *et al.* (2011) mapped the interaction profiles of 315 unique fragments—derived from 71,798 different PDB ligands—against 20 naturally-occurring amino acids, also identifying the preferences of fragment-types for particular amino acids [83]. This map could be used to determine whether a given fragment is in a favorable or unfavorable environment, and potentially guide the lead chemistry process.

While this may serve as a quick, efficient way to direct the design of ligand topology, the orientations of individual amino acids have to be taken into account. Even when comparing the same binding site between two crystal structures of the same protein, the microenvironments in that site may differ significantly based on residue orientations, and this can substantially bias docking studies. As an example, Cox, B. *et al.* [84] noted that although co-crystal structures had been obtained of the G-protein-coupled receptor (GPCR) CXC-motif chemokine receptor 4 (CXCR4) bound to the anti-HIV small molecule IT1t and peptide CVX15 [85], docking of IT1t back into the CXCR4:IT1t crystal structure resulted in large RMSD values compared to experimentally solved structures. Similarly, docking other small molecule antagonists into the crystal structure gave poses that lacked critical interactions identified by site-directed mutagenesis [86,87]. It was observed that

Table 1.2
Preferential Amino Acid Fragments

Residue	Examples of Fragments
Arginine	
Aspartic Acid	
Cysteine	
Glutamic Acid	
Histidine	
Methionine	
Phenylalanine	
Threonine	
Tryptophan	
Tyrosine	

Examples of chemical fragments that select amino acids interact with. [82,83]
Table is reproduced with permission: © 2015 John Wiley & Sons A/S. doi:
10.1111/cbdd.12631.

the CXCR4:CVX15 (CVX15: cyclic peptide antagonist) crystal structure produced a result that showed agreement between the computational binding pose of another small molecule CXCR4 inhibitor, AMD11070, and independent site-directed mutagenesis data. When analyzed in more detail, the orientations of residues Gln200 and His203 in the interior of the binding pocket differed substantially between the two co-crystal structures. Though the same amino acids are present at the binding site, changes in side chain orientation can substantially affect predicted binding modes and docking scores in virtual-compound screens, and may result in an unacceptably large false negative rate, particularly when rigid protein docking approaches are used. Therefore, the most accurate representations of protein subpockets must take into account multiple possible rotameric states of amino acid side chains if binding site microenvironments are to be generalized as chemical spaces represented by single “fingerprints”. This presents a potential problem for many approaches to binding site comparison that make use of pre-existing ligand-bound data, since the amino acids that make up the microenvironment surrounding a ligand fragment are already conformationally-biased in a given crystal structure.

While particular amino acids are seen to favor certain chemotypes, there exist some types of compounds that show relatively limited specificity and can negatively affect fragment-based drug screening efforts. These molecules, being related to pan-assay interference compounds, are often widely promiscuous in the number and types of interactions with proteins and serve as artifacts, yielding false signals across many types of assays [56,88–92]. There are varied reasons for the promiscuous nature of these compounds [93,94]; but, they represent a subclass of indirect observations of the potential ‘fuzziness’ in protein microenvironments. In essence, these compounds own a fragment chemical space that can recognize more general protein features that have recurring characteristics which contribute to the high level of promiscuity. When considered in this context, protein subpockets cannot always be rigidly defined by a size and shape boundary since some chemical functionalities are seen to interact with less well-defined features.

1.2.2 Identifying Subpockets

In reality, defining what constitutes a ‘subpocket’ for fragment screening is not always a clear task. A significant number of topological features on a protein could potentially be classified as being part of a subpocket if one were to focus on the environment of a single amino acid. A common approach for defining pockets is to focus on the proximity of protein atoms or residues to a bound

ligand. There have been many studies dealing with the analysis of binding sites, ranging from sequence comparisons, to overall binding site structural similarity and pharmacophore (“fingerprint”) searching [59,95–105]. Until recently, the majority of the work in the field has looked at the relationship between binding sites in the context of an entire defined pocket. Traditional computational methods have attempted to reduce the representation of protein surfaces in an effort to maximize the speed with which a search query can be performed (for examples, see Jordan, *et al.* (2012) [106] and Xie, L. & Bourne, P.E. (2007) [107]). It is likely that higher degrees of accuracy in predicting the similarity of binding sites can be obtained by taking into account more in-depth features such as hydrophobicity, aromaticity and hydrogen-bonding capability at the level of single amino acids. As a result, attention in the field has turned to studying local areas of protein surfaces, with implications for traditional receptor binding sites as well as protein-protein interaction interfaces. Table 1.3 highlights many of the currently available programs, methods or databases used to screen binding sites; and while there is diversity between each of these, most can be generalized into one of two groups: protein topology- and ligand interaction-driven approaches.

1.2.2.1 Using Protein Topology to Search for Binding Sites

A significant number of methods currently exist for comparing protein binding sites [98]. Some of these detect locally conserved residue patterns in order to define similar topological features between proteins and their probability of containing equivalent interaction sites. These approaches can function independently of known ligand-binding data, since the protein structure drives the query. A web server such as ProBiS [108] is useful for detecting either global or local similarities between proteins and identifying structurally conserved binding sites on proteins of interest. This tool takes into account geometric as well as physicochemical properties for aligning the local structure of two proteins; a user can decide to compare the entire surface of a protein, or only a selected surface patch, against a database of more than 35,000 non-redundant structures. The program COFACTOR [109] is able to identify functional sites on a protein as well as predict its biological function by comparing the local and global 3D structure of a protein against the large BioLiP database [110]. Similarly, the IMAAAGINE web server [111] can be used to define an arrangement of 3-8 amino acid residues that are converted into a search pattern to query the PDB and identify structural motifs with equivalent localized environments amongst the arrangement of amino acids within a defined tolerance distance (default of 1.5Å). These examples demonstrate

Table 1.3
Methods or Tools for Analysis of Binding Site Similarity and Pharmaceutical Discovery

Method or Database	Publication or Software URL	Description
CatSId [112,113]	http://catsid.llnl.gov	Searches for matches between catalytic sites and proteins
CavBase (Relibase) [114–118]	http://relibase.ccdc.cam.ac.uk	Uses 3D property descriptors to detect similarities between protein cavities that share little or no sequence homology
COFACTOR [109]	http://zhanglab.ccmb.med.umich.edu/COFACTOR	Structure-based method for biological protein function annotation
CrystalDock [119]	http://nbcrc.ucsd.edu/data/sw/hosted/crystaldock/	Finds suitable fragments that match protein pocket-lining residues
DoGSiteScorer [120–122]	http://dogsite.zbh.uni-hamburg.de	Detects protein subpockets and predicts site druggability
F-SPE-FP-PH3 [123]	http://www.ncbi.nlm.nih.gov/pubmed/20886466	3D fingerprint descriptors for environments surrounding fluorinated ligands in the PDB
FINDSITE ^{LHM} [124]	http://cssb.biology.gatech.edu/findsitelhm	Homology modeling approach to flexible ligand docking
FragFEATURE [125]	http://www.ncbi.nlm.nih.gov/pubmed/24762971	Predicts small fragments preferred by a target protein structure
GIRAF [126,127]	http://pdj.org/giraf/source/distr.tar.gz	Searches and flexibly aligns protein ligand binding interfaces
G-LoSA [128]	http://pubs.acs.org/doi/abs/10.1021/ci300178e	Uses known protein-ligand binding site data to predict/design a ligand for the target protein
HOMOLOBIND [129]	http://pibase.janelia.org/homolobind/	Identifies residues that are similar to structurally characterized binding site
IMAAAGINE [111]	http://mfrlab.org/grafss/imaagine	Searches 3D arrangements of amino acids in PDB structures

Table 1.3 Continued

KRIPO [104]	http://www.ncbi.nlm.nih.gov/pubmed/22830492	Quantifies subpocket similarities to identify bioisosteres
Med-SuMo [130]	http://medit-pharma.com/index.php?page=med-sumo	Locates similar regions on protein surfaces that are linked to certain chemical function
Nucleos [131]	http://nucleos.bio.uniroma2.it/nucleos	Identifies nucleotide-binding sites based on nucleotide modularity
Patch-Surfer [132]	http://kiharalab.org/patchsurfer2.0/	Web tool for predicting binding ligands for a protein
Phosfinder [133]	http://phosfinder.bio.uniroma2.it	Web server for the prediction of phosphate binding sites
PocketFEATURE [134]	https://simtk.org/home/pocketfeature	Uses protein microenvironments to identify similar binding sites; method is able to recognize distant protein relationships
PocketAlign [135]	http://proline.physics.iisc.ernet.in/pocketalign/	Generates structural superpositions of binding sites
PoSSuM [136]	http://possum.cbrc.jp/PoSSuM	Enables rapid exploration of similar binding sites between structurally diverse proteins
PrISE [106]	http://prise.cs.iastate.edu/index.py	Predicts interface residues using local surface structural similarity
ProBiS [108]	http://probis.cmm.ki.si	Enables structural comparison and local alignment of proteins
PROLIX [137]	http://www.ncbi.nlm.nih.gov/pubmed/22582806	Mines protein-ligand interactions in large structure databases
RCSB PDB [73]	http://www.rcsb.org/pdb/home/home.do	Worldwide repository of 3D structural information of large biological molecules
SA-Mot [138,139]	http://sa-mot.mti.univ-paris-diderot.fr	Extracts structural motifs of interest from protein loop structures

Table 1.3 Continued

SCOWLP [140]	http://www.scowlp.org	Enables individual and comparative analysis of protein interactions
SiteBinder [141]	http://ncbr.muni.cz/SiteBinder	Allows the superimposition of large sets of protein structural motifs
SiteComp [142]	http://sitecomp.sanchezlab.org	Compares binding sites, evaluates residue contribution to binding and identifies sub-sites with distinct molecular interaction properties
SubCav [143]	http://www.ncbi.nlm.nih.gov/pubmed/23327721	Similarity searching and alignment of protein subpockets
TargetHunter [144]	http://www.cbligand.org/TargetHunter	Web-based target prediction tool
TRAPP [145]	http://www.mcm-hits.org/trapp/	Analyzes binding pocket variation along protein motion trajectory and assesses the druggability of a target
TrixP [146]	http://www.ncbi.nlm.nih.gov/pubmed/23390978	Fast protein binding site comparison and function prediction
Wallach, I. & Lilien, R.H. algorithm [147]	http://bioinformatics.oxfordjournals.org/content/25/12/i296.full	Uses pharmacophoric features to predict subcavity binding preferences

Table is reproduced with permission: © 2015 John Wiley & Sons A/S. doi: 10.1111/cbdd.12631.

larger-scale (whole-protein) methods of comparison that identify conserved features of binding site topologies between proteins.

While these strategies can prove to be quite effective in certain circumstances, some of these ‘globalized’ methods may not be able to recognize distant similarities, particularly in cases where the dynamics of a protein allow a ligand to bind at sites that may be very structurally diverse. Recent tools have been developed to approach protein comparisons and binding site identification by taking a “subpocket-focused” approach. Based on the 3D coordinates of a protein, DoGSiteScorer [120–122] is able to detect potential pockets on the surface and split each of them into subpockets. The program uses properties such as volume, depth, surface, ellipsoid main axes, site lining atoms and residues, as well as functional groups present to calculate the predicted pockets. This tool scores each of these pockets for their potential druggability, yielding accuracies of 88% when trained and tested against a dataset of 1069 different structures [121].

Among some of the other methods [135,148–150] used to compare binding sites are CavBase [114–118] and PrISE [106]. CavBase derives 3D descriptors that characterize the surface properties of a binding cavity. The individual amino acids lining a cavity are analyzed to form descriptors of the localized chemical environment. Dummy atoms (“pseudocenters”) are placed on the surface, representing the overall chemical property expressed by the surrounding exposed atoms in that area. Therefore, cavities described by a series of pseudocenters are compared with a database (Relibase [151]) of cavities from PDB protein structures. Similarity is determined through the matching of 3D property descriptors (pseudocenters) between the different sites to provide a “pharmacophore-esque” search that takes advantage of localized environments within a protein binding cavity. Additionally, a new evaluation formalism for entries in Cavbase has recently been reported that improves efficiency of large-scale mining for similar protein binding pockets [152]. Separately, PrISE is able to predict interface residues of protein-protein interactions by defining structural elements consisting of a central amino acid residue and its surrounding residues on the protein surface. PrISE deconstructs the surface of a query protein into its structural elements and compares those to similar elements in a database that contains elucidated structural elements in PDB format. Elements are labeled as interface or non-interface based on the characteristics of the central residue. The labels are then used to predict whether the central residue of each structural element of the query protein is an interface residue. Ultimately, although different, CavBase and PrISE are able to use the chemical environment around a specific amino acid residue to predict the similarity between sites of interaction on different proteins.

For analyzing similarities between binding sites in terms of sub-cavities, the PocketFEATURE algorithm [134] takes a modified approach by analyzing “microenvironments” to assess overall similarity between proteins for prediction of shared ligands. The method does not rely on the sequence or relative pattern of amino acids in a binding site. Instead, Liu & Altman define a “microenvironment” as a local, spherical region within a protein structure that may encompass amino acid residues that are discontinuous in sequence and structure. A set of 80 physiochemical properties [153] is calculated over six concentric spherical shells, centered on a predefined functional center, with the total radius of the “microenvironment” being 7.5Å. Two sites are compared using an adjusted Tanimoto coefficient based on the presence of similar properties. Advantageously, this algorithm relies weakly on geometric requirements, instead using biophysical and biochemical measurements to characterize a subregion of a binding pocket. This approach allows for dynamic conformational changes in both the ligand at a binding site and the pocket itself. In this way, the algorithm may most closely align itself with a true subpocket-type search when compared to the aforementioned computational methods.

1.2.2.2 Ligand Interaction Approaches to Compare Binding Sites

As opposed to protein structure-driven approaches, ligand-driven approaches utilize known interaction data of ligands (small molecules and chemical fragments alike) with proteins to develop binding models that can be used to compare binding sites. Traditionally, many of these methods analyze the environment surrounding a bound ligand in a protein’s binding site, encoding physiochemical and/or geometric features to form a general pharmacophore. These descriptors can then be screened against databases of known structures to detect similar sites. A program such as G-LoSA [128] uses this general strategy to predict or design a ligand for a target protein given known interaction data between ligands and predicted similar binding sites. The region surrounding a bound ligand is compared to localized regions in other proteins, and those containing a large set of aligned residues are identified. Similarly, GIRAF [126,127] utilizes a database of known ligand-protein structures to create an index of the geometric features of the surrounding atomic environment. Similar ligand-binding sites can then be identified and aligned with the query structure, independent of sequence homology, or protein fold. PROLIX [137] uses fingerprints of ligand-protein interaction patterns to rapidly mine large crystal structure databases for similar patterns. The PoSSuM database [136] enables the rapid exploration of similar binding sites between proteins based on physiochemical and geometric similarities, sorted using the neighbor-search

algorithm SketchSort [154]. Each of these methods is useful for finding similar binding sites to a query protein in a small amount of time based on already-known interaction data, but these generally take whole-site approaches as opposed to screening small microenvironments of the binding site itself.

As with protein topology-driven approaches for binding site analysis, recent methods utilizing ligand-bound data are advancing the analysis of subpockets. The web tool Patch-Surfer [132] represents a binding pocket as a set of small, localized surface patches. Each of these patches is further characterized using 3D Zernike descriptors [155], enabling the identification of corresponding regions in pockets on other proteins, even if the overall pocket shapes are different. This method is able to compare a queried pocket to known ligand binding pockets, and predicts binding ligands for the query. In its most recent version, Patch-Surfer was tested against a large dataset of more than 6000 non-redundant pockets, with 2707 different ligands, where it displayed better predictive performance than many other currently available methods to predict protein pockets. Another program, CrystalDock [119] is able to take a ligand-receptor complex, break the ligand into its constituent molecular parts, use the microenvironments surrounding the ligand fragments, and perform a geometric comparison to identify similar microenvironments in ligand-bound PDB structures. This information can be used to predict chemical fragments that would bind a site of interest. FragFEATURE [125] has a similar function in that it also compares the structural environments within a target protein to those in databases to find statistically preferred fragments at a binding site. KRIPPO [104] is able to use microenvironment pharmacophore fingerprints to identify similar binding sites between proteins, and potential bioisosteric replacements for queried molecular fragments. Combining a sub-cavity comparison search with pharmacophoric analysis, Wallach and Lilien developed a method [147] to cluster similar binding site sub-cavities to predict patterns of binding between proteins that do not share any structural similarity with known systems.

Taking a modified approach to ligand-driven subpocket analysis, Kalliokoski, T. *et al.* developed SubCav [143], a tool for comparing and aligning protein subpockets. In this method, a modified version of the 3D pharmacophore fingerprint descriptor F-SPE-FP-PH3 [123] was used to define pharmacophoric features of all protein atoms within 4.5 Å of a bound ligand. These atoms were described according to eight different chemical features. Fingerprints composed of 7680 elements incorporating pharmacophoric features and the three dimensional triangular distances between them, were generated for a grouping of subpockets within a binding site. Normalized Tanimoto scores were then used to assess the similarity between two fingerprints. Subpockets were aligned using the methodology described by Kabsch [156] implemented in BioPython [157]. An

advantage of this method is that it can identify similar fragment binding sites independent of protein structural or sequence similarity. This focus on local equivalent microenvironments enables more comprehensive predictions of the pharmacological selectivity for on- and off-target binding events for fragments in ligands.

1.2.3 Shared Subpockets in ATP Binding Sites and Relevance to Ligand Selectivity

A current challenge for small molecule lead development is represented by target site similarities within the superfamily of human protein kinases. Major efforts for drug discovery targeting these kinases have focused on their ATP binding sites. Although these proteins have evolved distinct cellular functions and hence substrates, there exists a significant level of similarity at the ATP binding sites. One of the challenges in developing inhibitors to these proteins is the optimization of selectivity for one or more related kinases [158]. A study by Anastassiadis, *et al.* (2011) showed that a large proportion of commercially available kinase inhibitors display significant levels of cross-reactivity, with some having the ability to inhibit the catalytic activity by more than 50% in nearly 30% of all tested kinases [159]. The clinical implications for these observations are not always clear, but offers a platform for understanding the role of subpockets in rendering predictable inhibitor kinase cross reactivity.

Within kinase ATP binding sites, there exists a conserved phenylalanine residue of the Asp-Phe-Gly (DFG) motif that is buried in a hydrophobic pocket, which is positioned in a groove between the two ‘lobes’ of the kinase. This motif is targeted by type I kinase inhibitors [160,161], being generally more promiscuous than type II and III inhibitors, which typically take advantage of an allosteric binding site that becomes available upon a structural change of the DFG motif [160–162]. The idea of a protein subpocket finds significant application here given the large similarities in the binding pockets for type I inhibitors. An example can be seen with the JNK3 human kinase, where the residues Glu147, Met149 and Val196 form a microenvironment (Figure 1.2c) that recognizes shared chemical features between a natural ligand, adenosine monophosphate (AMP) [163], and a dihydroanthrapyrazole-based inhibitor [164], though their structures, overall, are quite different.

These binding site subpockets are used advantageously when designing mimics of natural ligands as inhibitors, but the similarities in these microenvironments between kinases can also give rise to promiscuous chemical landscapes. Staurosporine is a classical ATP-competitive inhibitor that is notoriously promiscuous [165], with there already being 48 crystal structures available in

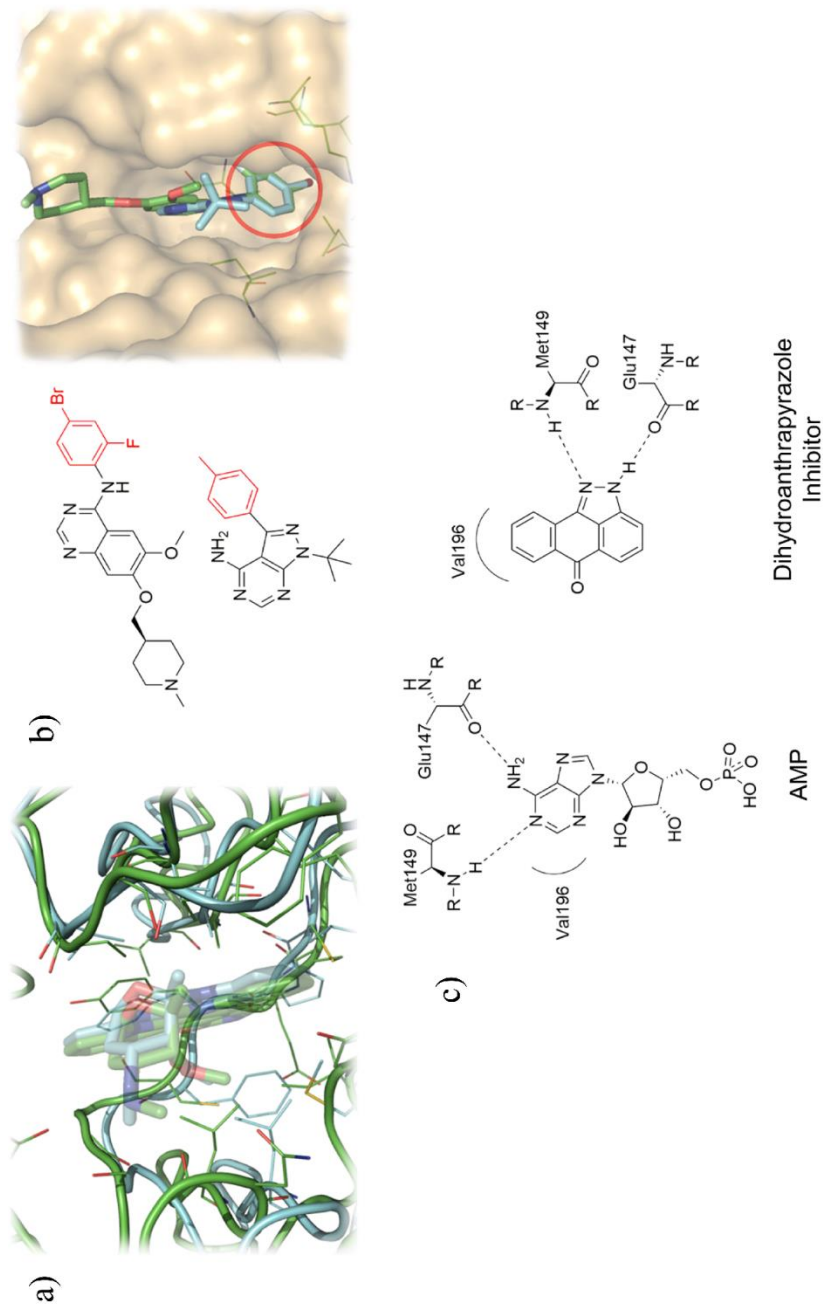


Figure 1.2. Conserved Subpockets in Protein Kinases Contribute to their Inherent Promiscuity. (a) Staurosporine (shown as partially transparent) is known to have inhibitory activity against a plethora of protein kinases from different families [164]. Conserved subpockets between the TAO2 MAP3-level kinase (green loop) and serine-threonine kinase 16 (STK16) (blue loop) allow staurosporine to bind with comparable affinities to both proteins (TAO2, PDB ID: 2GCD; STK16, PDB ID: 2BUJ). (b) The promiscuous RET tyrosine kinase (orange surface) contains a binding site subpocket that allows for the binding of overall structurally distinct inhibitors. The red highlighted portion of each inhibitor overlaps in the subpocket, shown with the red circle on the right (green, PDB ID: 2IVU ; blue, PDB ID: 2IVV). (c) Within the binding pocket of JNK3 MAP kinase, residues Glu147, Met149 and Val196 form a microenvironment that binds comparable chemical features between a natural ligand, adenosine monophosphate (AMP) (PDB ID: 4KKE), and inhibitors such as the dihydroanthrapyrazole-based antagonist shown on the right (PDB ID: IPMV). Figure is reproduced with permission: © 2015 John Wiley & Sons A/S. doi: 10.1111/cbdd.12631.

the PDB that show it bound to a variety of human kinases. With roughly 32% sequence similarity, the TAO2 and serine-threonine kinase 16 (STK16) kinases both bind staurosporine with good affinity. The sequence of residues at the ligand binding site differs between the two proteins, but TAO2 [166] and STK16 [167] both fold in a way that causes chemically similar amino acids to overlap, forming shared subpockets that result in a favorable binding site for the competitive inhibitor (Figure 1.2a).

There are additional proteins in this family that display high levels of promiscuity as well. The RET tyrosine kinase has been observed to be significantly inhibited by a variety of competitive inhibitors [159]. While many of these ligands have structurally distinct features and vary in size, one conserved binding orientation [168] makes use of an overall hydrophobic subpocket surrounding Val804, as can be seen in Figure 1.2b. Even though the inhibitors shown in this example are not vastly dissimilar, they do demonstrate that when designing ligands, consideration of the fact that some proteins recognize more generalized chemical features. This insight is especially important when considering kinases as biological targets; but, importantly the example is instructive to apply to other protein families as well. Furthermore, documented ligand-kinase associations offer many cases to bench mark tests of methodologies to identify and utilize fragment subpockets for the desired selectivity or cross-reactivity.

1.2.4 Consideration of Protein Interfaces and Cooperative Interactions

In contrast to the types of ligand interaction sites considered so far, protein-protein interaction (PPI) interfaces give fewer clues on a residue by residue basis for what is required for a specific binding event. Protein surfaces most often are relatively flat, and the interface at which two proteins engage typically makes up a large surface area (1000-2000 Å²) with multiple contact points [169]. Drug discovery efforts, particularly in the pharmaceutical industry, traditionally avoided exploring PPIs as drug targets, considering them “undruggable” until recent challenges of this classification [170]. Within the last decade, research into targeting PPIs as a therapeutic strategy for multiple diseases has continued to accelerate. Previous work has shown that not all of the residues at a PPI interface are critical for the interaction. In the best studied cases to date, there exist regions of “hot spots” that confer most of the binding energy [170]. These “hot spots” are adept at binding protein or peptide-esque molecules and have been suitable targets for peptides and a number of mimetic compounds [171]. However, the practical utility of the hot spot theory requires additional consideration with respect to the design of specific protein ligand interactions.

Previous studies have concluded that amino acids such as Ile, Leu, Met, Phe, Trp, Tyr and Val appear more frequently at PPI interfaces relative to other amino acids, and their average distribution throughout the genome [172,173]. Some residues at interface hot spots form “anchor sites”, where they serve as critical recognition features that drive the binding process [174]. In this way, these sites are analogous to the previously discussed concept of a subpocket in that the localized environment and relative orientation of particular amino acids in anchoring sites are important for the recognition of specific chemotypes. The feasibility of targeting anchor sites was demonstrated when developing small-molecule PPI inhibitors [173]. The surface of the E3 ubiquitin-protein ligase MDM2 forms a natural binding groove that the Phe19, Trp23 and Leu25 residues of p53 favorably bind (Figure 1.3a). Even when visualizing the unbound form of MDM2, the pocket where Trp23 of p53 ultimately binds is visible, suggesting that this site is important for the recognition and binding of p53. Phe19 and Leu25, upon binding, appear to induce the formation of hydrophobic pockets on MDM2 that result in favorable binding environments for the two residues. In this case, the recognition features of a subpocket are not necessarily prearranged for a meaningful binding event to occur. Rather, a molecular fragment may induce formation of a new subpocket with optimal chemical environment for that fragment to bind, when it otherwise would not if considered a static system.

In a similar fashion, proliferating cell nuclear antigen (PCNA) contains a region on its surface that recognizes specific chemical features. PCNA acts as a homotrimeric scaffold protein that binds many different proteins associated with DNA replication and damage repair [175]. Many of these proteins contain a conserved eight-member sequence known as the PCNA-Interacting Protein (PIP box) motif [176]. While four out of the eight amino acids in this sequence are highly variable, the remaining four residues represent conserved chemical functionalities that are critical for proteins to bind at the PIP box-binding site on PCNA (Figure 1.3b). The conserved aromatic residues (tyrosine or phenylalanine) in the seventh and eighth positions of the PIP box are part of a short 3_{10} helix that binds in a highly flexible [177] hydrophobic pocket on the surface of PCNA. This site generally recognizes hydrophobic molecular fragments exemplified by the discovery of a small molecule, T2AA, that binds in that pocket and inhibits the interaction of PCNA with other PIP box-containing proteins [178]. Upon comparison between the structures of PCNA bound to p21 (PDB ID: 1AXC) [179] and T3 (an analogue of T2AA; PDB ID: 3VKX) [178], a subpocket formed by Ile128, Tyr133, Tyr250, Pro234 and Val236 binds an aromatic fragment of both T3 and p21's PIP box tyrosine (Figure 1.3c). Furthermore, the inhibitor's core iodine atoms induce a structural widening of the binding pocket, exposing a larger hydrophobic surface, resulting in a

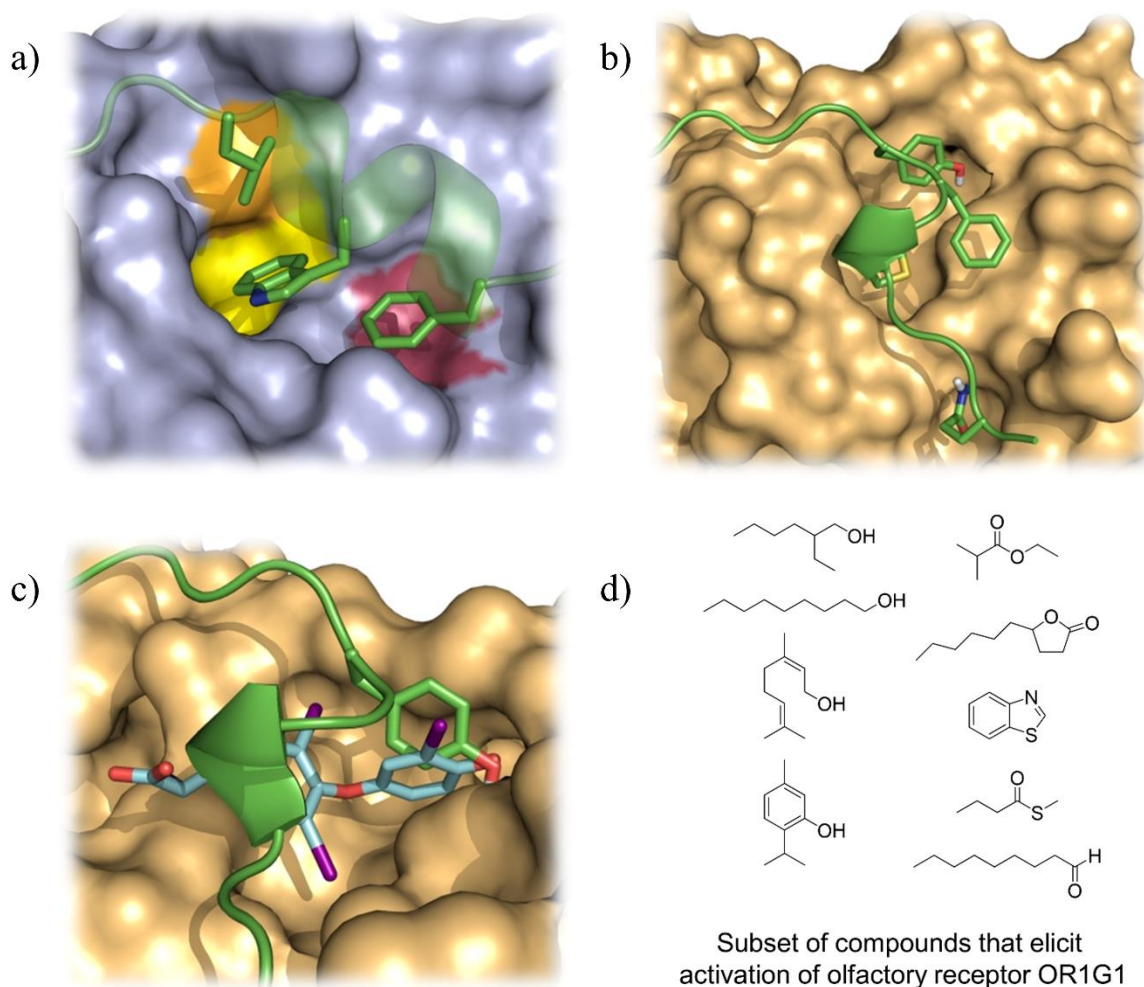


Figure 1.3. Protein Interface Hotspots Contain Inducible Microenvironments that Bind Conserved Fragments Between Molecule Types. (a) Three residues of p53 (F19, W23 and L25; green sticks) become buried in the surface of MDM2 (blue surface), inducing the formation of hydrophobic subcavities. Residue W23, in particular, likely acts as an anchoring residue, substantially contributing to the binding affinity (PDB ID: 1YCR). (b) Four key residue positions (green sticks) in the highly conserved PIP box sequence are essential for the binding of PIP box-containing proteins (p21 shown as green loop) to PCNA (orange surface). The fifth, seventh and eighth residues in the sequence bind at a surface pocket on PCNA made up of several hydrophobic microenvironments (PDB ID: 1AXC). (c) Within the hydrophobic PIP box binding site on PCNA (orange surface), the subpocket defined by the relative orientation of I128, Y133, Y250, P234 and V236 to one another has affinity for an aromatic ring moiety, with a tyrosine residue (green sticks) of p21's PIP Box sequence (green loop) anchored in the same location as the tyrosine-analogous fragment of the small molecule inhibitor T3 [178] (PDB ID: 3VKX, p21 peptide from 1AXC). (d) Some receptors exist that themselves accept a variety of structurally diverse substrates, the best example being the large family of GPCR olfactory receptor proteins. OR1G1, a member of this family, becomes activated upon exposure to numerous diversified odorants [180]. Figure is reproduced with permission: © 2015 John Wiley & Sons A/S. doi: 10.1111/cbdd.12631.

stably bound planar orientation of the molecule. The ligand-dependence of the subpocket within the PIP box binding region is another example of how the plasticity of protein interfaces must be taken into consideration.

Because drug-like compounds are often composed of several segmental fragments, any one substructure of a molecule could have affinity for a sub-pocket fingerprint shared between two or more proteins. Of course, the observable binding affinity of the drug molecule depends on all of its interacting components with a protein; a single interaction within a microenvironment may not be nearly sufficient enough to stabilize a specific bound conformation to elicit a pharmacological effect. However, protein systems do exist that have evolved to accept a variety of chemical structures, whether through flexible binding sites, or weak requirements for a binding event to produce a desired effect. The large family of GPCRs that make up the mammalian olfactory system is one example. Many individual members of this group become activated upon interaction with a plethora of different odorants [180–183] (Figure 1.3d), implicating the existence of ‘fuzzy subcavities’ making up interaction sites. Receptor OR1G1, and others like it, thus exhibit flexible recognition of general chemotypes, rather than a specific orientation of atoms.

The aforementioned examples demonstrate that PPI interfaces containing subpockets are analogous to those found in traditional small ligand binding sites. However, these characteristics indicate the transient aspects of these subpockets and articulate both opportunities and challenges for discovery of selective ligands with useful pharmacological effects. Many cases highlight limitations of structure-based virtual screening, and improved methods to achieve overall accuracy continue to emerge [184–189]. Computational approaches to address the issues of ligand and receptor dynamics has been an area of continued advancement [61,184,185,187,190,191]. The distinctions are likely that induced-fit and conformational selection play more significant roles in formation of subpockets and the resulting affinity of a molecule composed of multiple substructures [61,192–194]. These mechanistic principles have recently been exemplified in the formation of transient binding sites for protein interface inhibitors where evidence for both local conformational selection and induced-fit pathways have been evaluated [192].

An understanding of how subpockets are influenced by other microenvironments in proximity is a general characteristic for consideration in FBDD. When a fragment of a small molecule binds in a given subpocket, the binding event can induce structural changes that modulate the features of surrounding regions. This means that in a fragment-based screen, suitable binding in a particular chemical microenvironment may not occur in the absence of nearby transient binding events of other fragments, resulting in false negatives. A recent case study addressed how current

FBDD methods are limited in detection of fragments that exhibit cooperative binding [195]. However, the deployment of these methods to provide improved accuracy in fragment-based screens is less than complete.

Several computational approaches to meet these limitations are emerging and show significant promise for expanding successes. Mahasenan, K.V. & Li, C. were able to develop an approach that incorporates multiple protein conformers via ligand-directed modeling [196,197]. Substantial improvements for sampling and free energy scores in docking studies can be achieved when taking into account multiple fragment-ligands simultaneously [198,199]. Novel inhibitors of STAT3 and the IL-6/GP130 protein interface were discovered by considering multiple subpockets, and simultaneous ligand screening [200–203]. Type II kinase inhibitors were also developed for the MELK kinase by taking into account an induced conformational shift of the protein [204]. These examples implicate the importance for ligand-dependent subpockets when using FBDD approaches; they may represent a general process that can allow for cooperative binding into subpockets. The full extent to which this impacts current drug discovery efforts and computational screening methods remains to be fully established, but opens avenues for improvement in the utilization of computational and experimental approaches to discover multiple fragment sites at protein interfaces.

1.2.5 Summary of Computational Approaches toward FBDD

There are an increasing number of computational tools available to complement experimental fragment-based screens and ligand design. Advantages to these approaches are that they enable a potential general solution to the challenges of optimal fragment selectivity and linkages but these goals remain to be fulfilled. The prediction of protein structure microenvironments and dynamics requires consideration and methodological advances are emerging. As previously highlighted, the similarities of protein microenvironments between evolutionarily related or convergent proteins could argue for increased chances of off-target effects. To the extent that this may be a concern at protein interfaces still remains largely unknown [170,205]. However, the potentially unique combinations of transient subpockets that arise upon the binding of a ligand may ultimately result in higher degrees of selectivity.

The occurrence of subpockets in unrelated proteins is generally recognized as a potential contribution to drug effects; however, an account of these features when selecting specific fragments in drug development has been traditionally based upon empirical deduction.

Experimental and computational approaches that can account for potential cooperative interactions are subjects of considerable importance for structure-based drug design. It is increasingly feasible using computational approaches to consider a classification process to stratify both risk and benefit of new molecular candidates. These tools can also be applied to molecular fragments and their associated libraries. As opposed to promiscuous scaffolds, fragments with high propensity for subpocket interactions can be argued to represent important tools for lead identification when considered in an appropriate knowledge framework. Whether fragment libraries can be qualified on a proteome-wide scale remains a future challenge, but the combinations of experimental and computational methods offers increasing promise.

1.3 Research Scope

The number of new drug applications over the past several years has paled in comparison to the more than 40 per year that were observed during the late 1990's. Though the human genome project promised to usher in a new era of drug discovery, it has not completely lived up to its lofty goals. Indeed, much of protein research still focuses on the 10% of proteins that were already known before the mapping of the human genome [206]. Much of the difficulty in finding new drugs may be associated with the content of libraries currently used for high-throughput and virtual screening, since these libraries tend to be historically biased toward traditional drug targets [207–209]. Although much progress has been made in the elucidation of regulatory mechanisms of key intracellular processes, such as the DDR, there remains a need to better understand the necessary molecular features for conferring selectivity between major regulators of these pathways.

Amongst the redundancies between alternative pathways in the DDR, DNA scaffolds play key roles in that they recruit, stabilize and orient other repair-facilitating proteins that make up complexes at sites of DNA damage. Due to specific conserved structural features between binding partners of DNA scaffolds, it stands to reason that targeting the site of recruitment of these proteins on a scaffold could produce a substantial pharmacological response in a given context. Protein-protein interfaces, however, have proved to be challenging drug targets, and often require unique classes of molecules as inhibitors [210]. Additionally, the high cost and often limited diversity of high-throughput screening libraries both exist as prohibitory factors in many drug discovery efforts. As a result, a method for pairing down the necessary compounds to screen would be beneficial, particularly when little is known about a specific target's ideal drug chemotype.

PCNA serves as an interesting drug target given its importance in DNA replication and damage repair. Studies have shown that the scaffold's PIP Box binding site can be targeted using small molecules to modulate its interactions with its natural binding partners. However, SAR studies have been unable to significantly improve upon the relatively limited chemotype identified that binds at that site. The first goal of this research is the development of a peptoid-based screening platform that enables the reduction of cost and time in the initial synthesis of an early-stage compound library that can be used to identify generic features of a protein-protein interaction antagonist for a given site. This approach is used in the discovery of new inhibitors of PCNA-PIP Box protein interactions. Understanding the key interactions that may be driving the propensity for ligands to bind at that site could be useful in creating more drug-like compounds that target disease-specific complexes in order to control the progression of cancer.

The second goal of this research is to explore the differential modulation of signaling pathways based on intracellular compartmentalization of known cancer therapeutics, and whether this could serve as a viable strategy for targeting traditionally drug-resistant cancer types. It was seen that with nuclear-targeted kinase inhibitors, phosphorylation of intracellular STAT3 was preferentially modulated, as opposed to either ERK1/2 or EGFR. Alone, this may not have a substantial improved effect over commercial gefitinib, but when used in combination with DNA damaging agents or inhibitors of DDR-related processes, it could produce a substantial phenotypic response in drug-resistant tumor types reliant on abnormal expression or localization of EGFR.

2. PEPTOIDS AS PROBES FOR DISRUPTING PCNA-PROTEIN COMPLEXES

Integral to the majority of biological functions, protein-protein interactions (PPIs) mediate a large number of important regulatory pathways. There are estimated to be around 400,000 different PPIs within the human proteome [211] that are involved in various processes such as signal transduction, cell metabolism, apoptosis, growth, membrane transport and cellular motility. PPIs are regulated by various factors ranging from the simple affinity of binding between proteins or other ligands, to things such as protein concentration, ligand concentration, presence of ions and covalent modification, and these interactions can be classified as being either stable or transient. Over the last couple decades, significant research has resulted in large-scale protein interaction networks that have allowed a better understanding of pathway functionality, as well as the association between genetic variation and disease states [212]. Research is still ongoing to understand how the dysregulation of PPI networks leads to the formation of specific diseases, and this ultimately may lead to the discovery of a generation of new drugs.

Study of the molecular structure of proteins can provide details on the site of interaction between different proteins or at the interface between two substructures of the same molecule. Assembly of protein complexes can result in the formation of homologous and heterologous structures made up of potentially many different macromolecules. In the case of a molecule such as PCNA, the protein itself forms a homo-oligomeric structure around chromatin-bound DNA, and recruits many other regulatory factors to it that can make up large hetero-oligomeric complexes implicated in DNA replication and damage repair [175]. Proteins generally contain structural domains that allow for the recognition and interaction with specific amino acid sequences, or other structural features, on other proteins. Common examples of these include PDZ, SH2, SH3, SAM, LIM and PTB domains [213], and in the context of proteins that interact with PCNA, the PIP Box [214] and AlkB homologue 2 PCNA interacting motifs (APIM) [215]. Because these structural features are often heavily conserved among different proteins, their importance has long been recognized, and some motif recognition factors are viewed as potential targets for therapeutic intervention.

2.1 Protein-Protein Interactions as Therapeutic Targets

Traditionally, PPIs were seen as being “undruggable” by commercial practitioners, and drug discovery efforts avoided these types of targets [216]. Typically, most endeavors comprised the design of compounds that mimicked endogenous ligands of enzymes and other proteins [217]. A typical protein-protein interface can encompass an area of 1,500-3,000 Å² with 750-1,500 Å² of buried surface area [170,218], and a binding event between two proteins may involve multiple points of contact, also referred to as “hot spots” [219]. Simply blocking one of these points of contact may be insufficient to elicit a pharmacological response, though this would be heavily context dependent. These features present significant challenges for developing inhibitors of PPIs (iPPis), but the technological hurdles associated with PPI targets are becoming less daunting [220].

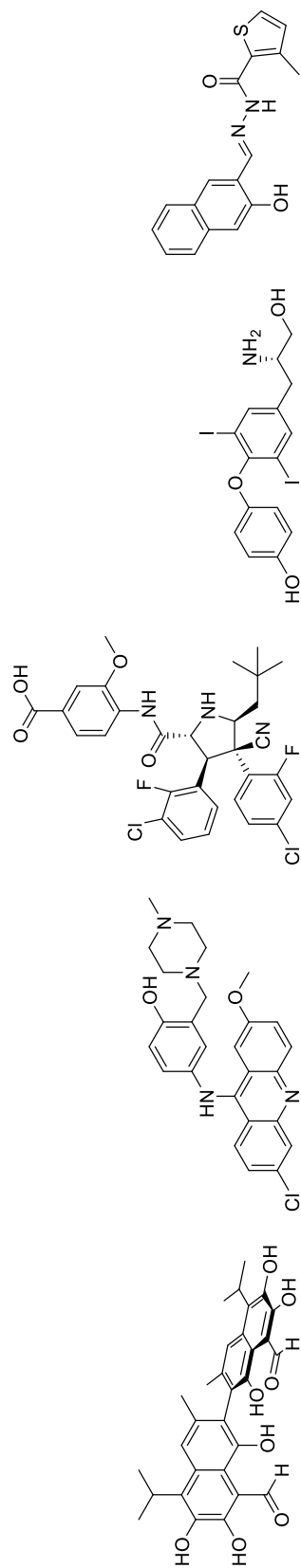
Various strategies are being explored to develop positive therapeutic outcomes for disease by targeting PPIs, and these have mostly consisted of peptidomimetic and small molecule-based approaches. iPPis can be generally be classified as either orthosteric inhibitors, directly interfering with the protein-protein interface, or allosteric inhibitors, which bind at some distal site and induce or prevent some conformational change in the protein that hinders complex formation. Potentially ideal inhibitors of macromolecular interactions would be stable peptides or peptidomimetics that adopt a defined secondary structure. However, smaller peptides do not tend to form defined conformations in solution. As a result, research on using peptides as iPPis has focused on developing analogue mimics of β -turns [221] as well as stabilized α -helical [222] and β -turn peptides [223–225], and foldamers [226]. Other major strategies of developing iPPis have involved designing secondary structure and surface mimetics. Each of these approaches shows promise, and the number of iPPis discovered using these methodologies continues to grow, especially where inhibition of large surface contact regions is necessary.

2.1.1 Targeting Protein Interfaces with Small Molecules

Similar to the case of more traditional drug targets, such as enzymes, small molecules have been used as antagonists of protein-protein interactions. However, predicting a targetable site for inhibition at a protein interface is not technically straight forward since all of the amino acids at that site do not contribute equally to binding. Typically, only a small subset of the residues contributes the majority of free binding energy to binding; these hot spots are defined as a significant increase in the free energy of binding ($\Delta\Delta G_{\text{binding}} > 1.5$ kcal/mol) when a residue is

Table 2.1
Examples of Small Molecule Protein-Protein Interaction Inhibitors of DDR-Associated Pathways in Cancer

Targeted Protein-Protein Interaction	Compound Name	Affinity or Inhibition Constant	Methodology	References
(±)-Gossypol	NSC 130813			
PARP1-Hsp70 & PARP1-OGG1	(±)-Gossypol	$K_d = 4.11 \pm 0.21 \mu\text{M}$ (PARP1 BRCT Domain)	Isothermal titration calorimetry	[227]
ERCC1-XPF	NSC 130813	$IC_{50} = 2.9 \pm 0.9 \mu\text{M}$	MDA-MB-231 cells	[228]
p53-MDM2	RG7388	$IC_{50} = 0.03 \mu\text{M}$	MTT with SJSA1, RKO and HCT116 (p53 ^{wT}) cell lines	[229,230]
PCNA-PIP Box Proteins	T2AA	$IC_{50} = 1.3 \mu\text{M}$	Fluorescence polarization (PCNA and PL Peptide)	[178,231,232]
PCNA-PCNA (monomer interface)	PCNA-II	$K_d = 200\text{-}400 \text{ nM}$	Surface plasmon resonance	[233-235]



mutated to alanine [219]. Identified concave hot spot regions are often the subject of investigation for their potential druggability, though other factors must be considered when targeting those sites [236].

The major challenge of developing small molecular weight iPPIs, though, has been that inhibition of PPIs typically requires coverage of a large number of surface contacts; but for a compound to fall under the criteria for Lipinski's guidelines [237], it may not possess the necessary features that would make it a good iPPI. However, it is not clear that these guidelines, which normally apply to small molecule-enzyme interactions, should be applicable to protein-protein interfaces [238]. Studies have shown that iPPIs have unusual physiochemical properties when compared to inhibitors of traditional targets (e.g. enzymes, surface receptors, etc.), such as higher molecular weights, higher hydrophobicity and more aromatic rings [239,240]. Furthermore, iPPIs typically demonstrate higher degrees of globularity, lower distribution of hydrophilic regions, smaller proportions of exposed hydrophilic regions and stronger capacities to bind hydrophobic patches at the core of protein-protein interfaces [210]. Due to these factors, conventional methods used in drug discovery have proved unsuccessful for the most part in finding small molecule iPPIs, but researchers have turned to alternative methods [220,241] that have thus far made progress in discovering new inhibitors [238,242], with compounds currently in clinical trials for six different targets [205].

While discovery of new iPPIs seems to be increasing, finding new inhibitors through traditional screening methods has been limited by the chemistry available. One of the main problems has been that high-throughput screening has often relied on compound libraries that are biased toward previous drug discovery efforts (i.e. targeting enzymes and surface receptors), and this has hindered the success of these screening efforts with PPIs [208,243–245], whose chemotypes are generally not well-represented in current libraries [239,246]. Additionally, drug discovery has been focused on a relatively limited number of targets [247–249], which as of 2006 stood at only 324 [250]. When contrasted with the approximately 400,000 predicted PPIs in humans [211], the need for expanding the explored druggable space is clear.

2.2 Proliferating Cell Nuclear Antigen is Crucial for the Processes of DNA Damage Repair and Replication

PCNA is a nuclear homotrimeric protein complex that encircles chromatin-bound DNA and acts as a processivity factor in DNA damage repair and replication [251]. Functionally, it serves as a DNA clamp in archaea and eukaryotes, and is composed of three PCNA molecules that associate to form

a tight ring around duplex DNA. It is expressed in the nuclei of cells during the S phase of the cell cycle [252], and serves as a topological site of recruitment on DNA for various proteins that are necessary for DNA replication or repair, such as DNA polymerase δ [253]. Upon destabilization of reservoir PCNA complexes, PCNA is loaded onto DNA in a replication factor C (RFC)-dependent manner with subsequent cell cycle inhibition [254,255]. Once bound to chromatin, PCNA mediates protein complex formation that is involved in the excision of DNA lesions, and promotes DNA synthesis [175,176,256]. A large number of proteins are currently known to interact with PCNA, implicating the involvement of PCNA in all facets of the DDR. In addition to being crucial for the base excision, mismatch, nucleotide excision and translesion synthesis repair mechanisms, PCNA acts as a regulator of cell cycle progression, chromatin remodeling and transcription [256]. Its expression is controlled by E2F transcription factor-containing complexes [257], with p53 playing a major role in helping to decide the ultimate fate of the cell [258]. Altogether, these regulatory mechanisms protect the cell from increased levels of genetic instability, and guard against uncontrolled proliferation.

2.2.1 PCNA Phosphorylation Controls Its Stability on Chromatin and is Associated with Dysregulation of DNA Mismatch Repair

In order to modulate its functions, PCNA is post-translationally modified by multiple biochemical events, each directing a different functional consequence. For example, in the context of DNA damage repair, mono-ubiquitination at lysine-164 leads to activation of Rad6-dependent translesion synthesis [259]. Extensions of this mono-ubiquitin chain are thought to activate the template switch pathway; alternatively, modification by the Small Ubiquitin-like Modifier (SUMO) protein inhibits the template switch pathway [259]. Separately, phosphorylation of PCNA at tyrosine 211 (Y211) is associated with stabilization of chromatin-bound PCNA, and by extension, subsequent PCNA-associated functions. Recent observations suggest nuclear-located tyrosine kinases, epidermal growth factor receptor (EGFR) and c-Abl, facilitate association of PCNA with chromatin and are required for the formation of nuclear foci of PCNA in DNA-damaged cells [260–263].

PCNA has critical roles in the survival and proliferation of cells, and disease-associated dysregulation of associated functions can have dire effects on genome stability [264]. Abnormally high levels of PCNA expression are associated with various malignancies, ranging from non-Hodgkin's lymphoma to skin, laryngeal, ocular, prostate and breast cancers [265–272]. PCNA has also proven to be a useful marker in the detection of increased levels of cellular proliferation and overall prognosis in breast cancer patients [273–275]. PCNA Y211 phosphorylation (pY211) levels

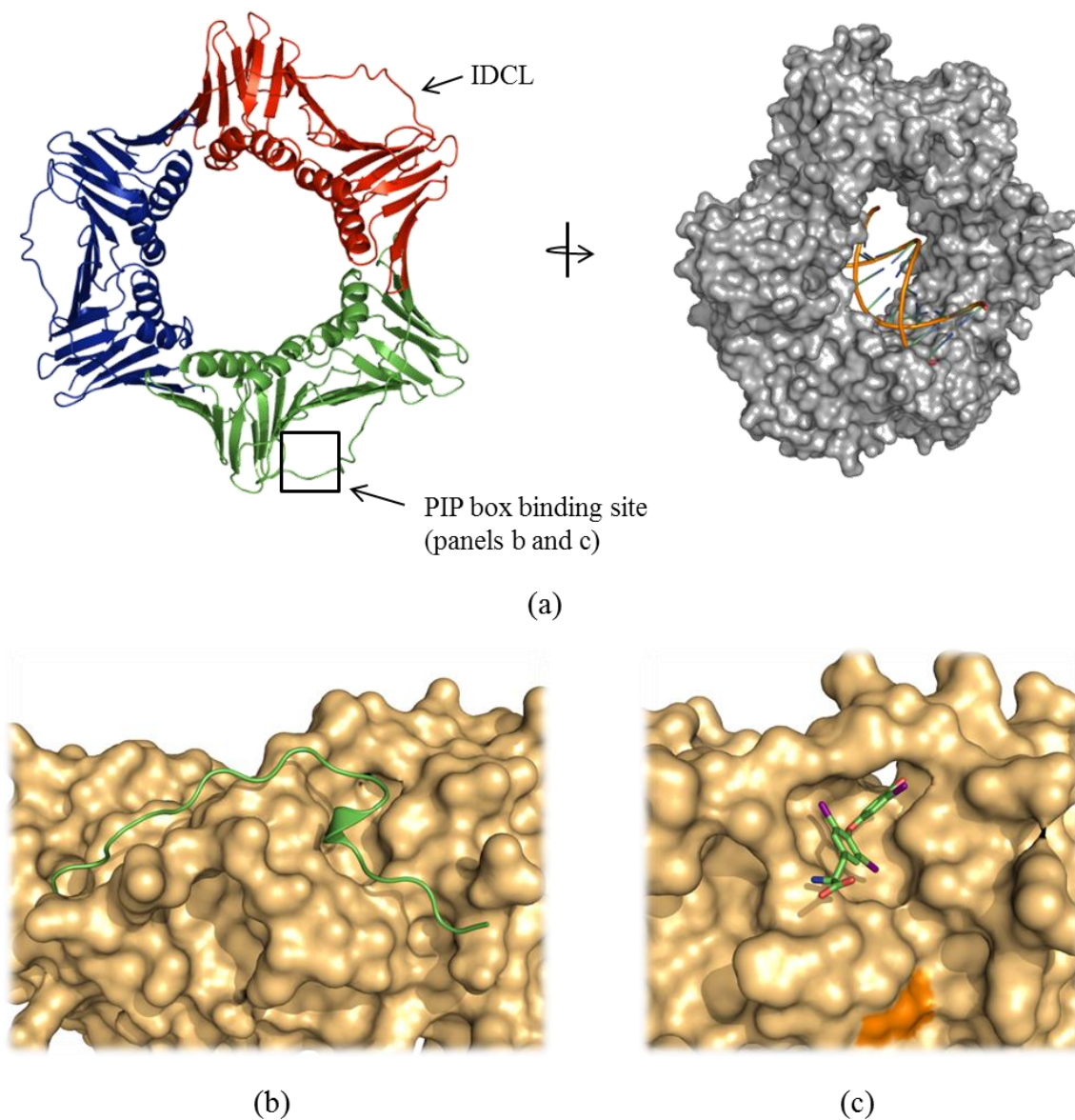


Figure 2.1. PCNA Structural Features and PIP Box Binding Site. (a) *left*: Cartoon representation of human PCNA with the interdomain connecting loop (IDCL) and PIP box binding site highlighted (PDB ID: 1AXC); *right*: surface rendering of *S. Cerevisiae* PCNA bound to DNA (PDB ID: 3K4X). (b) The 3_{10} -helical structure of the p21 peptide's (green) PIP box binds in a hydrophobic surface patch on PCNA (tan surface; PDB ID: 1AXC). (c) The small molecule antagonist T3 binds in the hydrophobic surface patch of the PIP box binding region on PCNA and disrupts the association of PCNA with PIP box-containing proteins (PDB ID: 3VKX); the proximity of this "hydrophobic binding site" to PCNA residue Y211 is shown (dark orange residue).

correlate with disease, such as the observation that elevated levels of pY211 are found in half of all human prostate cancers [261], and high levels of Y211 phosphorylation better correlate with poor survival of patients with breast cancer than does the total level of PCNA [260].

As previously alluded to, controlling PCNA-associated processes is extremely important for ensuring cell viability and proper growth control. The protein is post-translationally modified in different ways at specific times during the cell cycle in response to a variety of cellular events, and in this way its activity is tightly controlled. Disruption of these regulated, time-sensitive, processes can have profound effects on the cell. A recent study by Ortega, *et al.* [276] showed that high levels of EGFR, such as is seen in multiple cancers, can lead to abnormally high levels of PCNA-pY211. The phosphorylation at Y211 alters PCNA's interaction with mismatch-recognition proteins MutS α and MutS β , and interferes with the PCNA-dependent activation of MutL α endonuclease. As a result, DNA mismatch repair becomes inhibited at the initiation step, suggesting that PCNA-pY211 facilitates error-prone DNA replication by suppressing the mismatch repair function, thus promoting cancer development and progression. Given that phosphorylation at Y211 is important for the stability of PCNA on chromatin, it can be seen that subtle disruptions of the ideal levels of PCNA present in a specific form can have a substantial effect on a cell's outcome.

2.2.2 PCNA as a Molecular Drug Target

PCNA has been identified as a potential target for the treatment of certain cancers. Recently, small molecule antagonists of PCNA have shown useful pharmacologic effects when used either alone or in conjugation with a DNA damaging agent [178,231,233,277]. Additionally, targeting the phosphorylation event at Y211 has exhibited pharmacological relevance in breast and prostate cancer cell lines, as well as *in vivo* [261,263]. Correspondingly, rationale for PCNA as a disease target has emerged, but the nature of PCNA-protein interactions does not provide a traditional view of a drug target and therefore generates motivation for a new approach to modulate PPIs. Perturbing the association of PCNA with PIP (PCNA Interacting Protein) box-containing proteins, or disrupting Y211 phosphorylation would ultimately impair the cell's ability to repair or replicate DNA. Four major strategies for affecting this system (Table 2.2) include inhibiting the association of PCNA with its interaction partners [178,231,278–283], stabilizing the interaction between these species [284,285], dysregulate the stability of the PCNA homotrimer [233,234], or disrupting the ability of kinases to phosphorylate Y211 on PCNA [261,262,286]. It is hypothesized that a peptide or peptidomimetic molecule that is analogous to the PIP Box of known PCNA binding partners

Table 2.2
Methods for Dysregulating PCNA-Dependent Processes

Method	Description	References
PIP Box antagonist	small molecule or peptide that binds at the PCNA PIP box binding site and disrupts the association of PCNA with partner proteins	[178,231,281–283] ^{ab}
PPI stabilization	modifying the interaction of PCNA binding partners with PCNA's IDCL can stabilize the binding event, dysregulating time-dependent PCNA-critical processes	[284,285] ^c
dysregulate stability of PCNA homotrimer	small molecule that binds at the PCNA monomer-monomer interface and stabilizes PCNA trimers, reducing the amount of chromatin-associated PCNA	[233,234] ^a
disrupt Y211 phosphorylation	hinder tyrosine kinases from phosphorylating PCNA residue Y211, reducing PCNA's chromatin stability	[261,262,286] ^b
compete with PCNA for binding to partner proteins	peptide constituting a small region on PCNA mimics the protein and acts as a competitive inhibitor, reducing the pool of available PCNA binding partners	[278–280] ^b

^a small molecule-based

^b peptide based

^c not yet demonstrated as a therapeutic strategy

could bind to PCNA with good affinity at the PIP Box binding site. This would be sufficient to either block the association of PCNA with other proteins at that site, or the peptide/peptidomimetic would act as an adhesive, securing another protein on PCNA. Either method would dysregulate PCNA-dependent processes, and could be useful as therapeutic strategies when cell survival depends on upregulated levels of DNA damage repair.

2.2.2.1 Targeting the PCNA PIP Box Binding Site

Evidence shows that the majority of proteins that bind to PCNA have a conserved sequence motif known as the PCNA Interacting Protein (PIP) Box [214,287–290]. The sequence is composed of

eight residues with the sequence QXX ϕ XX $\Omega\Omega$, with Q being glutamine, X being any amino acid, ϕ being a hydrophobic residue and Ω being aromatic. These proteins bind at a shared site on PCNA, the PIP Box binding domain. From this, it can be inferred that this is a critical site for the regulation of various processes tied to PCNA. This is supported by observations that deletion of the PIP Box within c-Abl disrupts the interaction of PCNA with c-Abl, and ultimately increases the nuclear c-Abl apoptotic function in DNA-damaged cells [291]. Furthermore, peptides lacking the PIP Box motif or specific critical residues within that sequence showed substantial decreases in overall binding to PCNA [292–294].

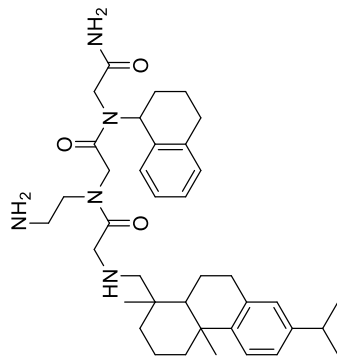
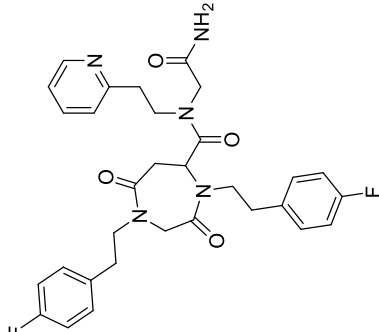
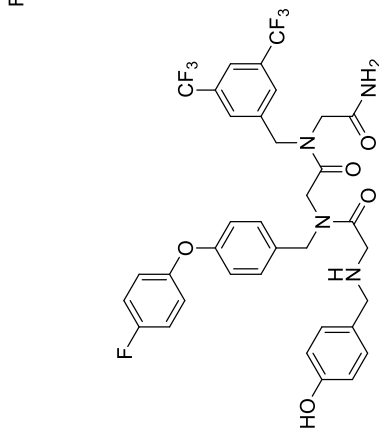
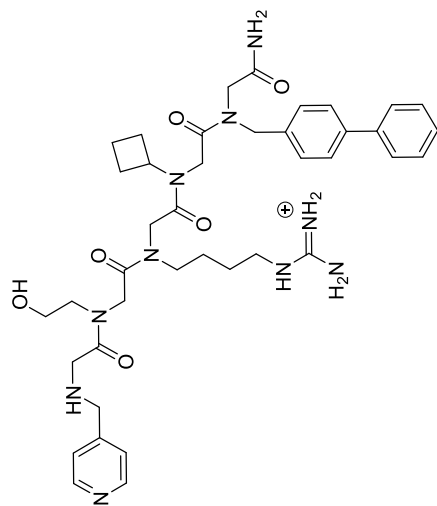
Recently, it was demonstrated *in vitro* that a small molecule could be used to disrupt the association of PCNA with a PIP Box-containing peptide [178,231]. When introduced to HeLa and U2OS cells, the compound T2AA and some of its synthesized variants were able to decrease cell growth and diminish DNA replication and translesion DNA synthesis. An obtained co-crystal structure of T3 (additional 3'-iodo substituent, relative to T2AA) and PCNA showed that the small molecule bound in the hydrophobic surface pocket of the PIP Box binding site (see Figure 2.1). Using a fluorescence polarization assay, the interaction between the POGO Ligase (PL) peptide and PCNA could be disrupted with low micromolar IC₅₀ values for a number of variants of T2AA. While this discovery provided an important proof of concept, there is still much potential improvement to be made with the small molecules' binding affinities and overall molecular surface area covered.

2.3 Peptoids as a Fragment-Based Screening Tool

Peptoids, being analogous to peptides, make up a class of peptidomimetics where side chains are appended to the backbone nitrogen atom, rather than the α -carbon. Because of this substitution pattern, these poly-N-substituted glycines lack the amide nitrogen which gives rise to many secondary structure elements in both peptides and proteins. The use of peptidomimetics is a popular approach in certain drug discovery efforts due to the disadvantages associated with peptides such as their proneness to proteolytic degradation and poor oral bioavailability [295]. Peptoids, as opposed to peptides, have a generally better inherent ability to penetrate cell membranes, have higher resistance to proteolysis, are more resistant to solvent, temperature and chemical denaturants, and are generally cheaper to produce [296–299]. When synthesized using the sub-monomer method [300], thousands of commercially-available primary amines could be used for incorporation into the backbone. To date, more than 200 different amines have been used as peptoid

Table 2.3
Examples of Peptoid-Based Small Molecule Inhibitors

Drug Target	Compound Name	Affinity or Inhibition Constant	Methodology	References
RNA Hairpin (Apical loop of pre-miR-21)	Compound 12	K _d = 12 ± 11 μM	Fluorescence	[301]
α/β-Tubulin	P26	IC ₅₀ = 5.7 ± 1.2 μM	HeLa cell growth inhibition	[302]
Ubc13-Uev1 (Protein-Protein Interaction)	Compound 1a	K _d = 4.4 pM (Ubc13)	Surface plasmon resonance	[303]
Cell Membrane (Gram- +/- Bacteria)	CHIR29498	MIC 6 μg/mL	Staphylococcus aureus cell death	[304,305]



Compound 12

P26

Compound 1a

CHIR29498

side chains [306], resulting in substantial chemical diversity at each substitution position. Peptoids have been used in a number of different biomedical applications from drugs [307,308] (Table 2.3) to synthetic lung surfactants [309], hydrophobic nanosheets [310,311] and antibody biomarkers for Alzheimer's Disease [312]. Peptoid-based small molecule-esque ligands have been discovered that target proteins such as vascular endothelial growth factor receptor 2 [313], Src Homology 3 and WW domains [314], Karopherins and Tubulin [302]. Additionally, peptoids have been useful as antibiotics against both gram-positive and -negative bacteria [315–317]. Due to their versatility in numerous applications, they have become a popular screening tool since large combinatorial libraries of peptoids, of approximate molecular weights between 450-750 Da, can be created with diverse types of chemical substitutions in relatively short fashion.

Peptoid screening has proven successful in identifying numerous types of ligands targeting many types of protein targets. Many of these efforts make use of trimeric peptoid libraries, which are generally small enough to be considered small molecule-like. Using iterative re-synthesis, Zuckerman, *et al.* (1994) created a library of 3500 peptoid trimers that was utilized in the discovery of nanomolar ligands for the opiate and α_1 -adrenergic receptors [318]. Screening of trimeric peptoid libraries has also resulted in the discovery of antimicrobial compounds [305] and mimics of the Agouti-related protein [319]. Positional scanning techniques have been used to discover noncompetitive antagonists of the vanilloid receptor subunit 1 (VR1) [320] as well as inhibitors of Karopherins and Tubulin [302]. Moreover, in certain cases, large tripeptoid libraries of more than 300,000 compounds have been used, and this has resulted in the finding of antagonists of the nicotinic acetylcholine [321], melanocortin type 1 (MC1) and gastrin-releasing peptide/bombesin receptors [322]. Because of their relative ease of synthesis, short peptoids can be valuable tools in the discovery of molecules that may be useful inhibitors in and of themselves, or as screening hits that can be further refined into lead drug-like compounds.

2.4 Rationale

Traditional chemotherapies and molecularly-targeted agents including protein kinases inhibitors have not provided sustained clinical efficacy [46–49,323]. Selective inhibition of kinases has been a major issue, with current inhibitors often showing high levels of cross-kinase inhibition [159]—though, the pharmacological significance of this promiscuity remains unclear. Targeting upstream processes in general has inherent limitations, since the activation of many proteins results in propagation (signaling cascades) and redundancy of intracellular signals. As an alternative,

targeting certain downstream proteins that function in the nucleus, such as PCNA, could have significant advantages as a strategy, particularly when the protein is targeted at the same time as other molecular targets, including proteins involved in the DNA damage response or signaling.

PCNA is known to interact with more than 200 different proteins that require binding to PCNA in order to regulate specific pathways of DNA replication or the DDR. While PCNA has been a subject of detailed investigation for more than a decade, there is still much that is not well understood. Despite the gaps in knowledge, two types of inhibitors have been discovered that target PCNA, as previously discussed in Chapter 1. Though high-throughput screening efforts led to the discovery of the thyroxine class of antagonists that target PCNA's PIP Box binding site [178], structure-activity relationship studies were not able to yield substantial improvements to the reported IC₅₀ values [231]. Although these inhibitors demonstrate the potential for modulating PCNA-protein interactions, it is clear that additional information on the molecular surface features of PCNA is necessary in order to design higher affinity ligands.

A potential approach toward the rational design of inhibitors of PCNA-protein interactions would be to use computational and biochemical screening methods to facilitate a fragment based drug design effort. For the development of iPPIs, it would be advantageous to use a system where molecular mimetics could rapidly map ideal surface binding interactions sites in tandem before considering detailed conformations of an optimal drug-like ligand. While this challenge can be met with large peptides, the translation to drug-like molecules is not clear. Structures of PIP Box-containing peptides/proteins in complex with PCNA solved to date indicate the core PIP Box recognition sequence is involved in a 3₁₀-helix that binds in a hydrophobic surface pocket on PCNA. However, small molecule mimics of this topology are not easily defined. It is hypothesized that determining the chemical recognition features of PCNA-protein interactions would allow for the design of small molecule inhibitors that could be used to target PCNA-associated processes of the DDR. While various strategies currently exist for targeting DNA repair pathways, PCNA would serve as a novel target in the inhibition of the DNA damage tolerance pathway, post-replication repair [324], specifically disrupting RAD6-dependent translesion synthesis as well as the "template switch" pathway [259]. The discovery of such an inhibitor would have potential as a sensitizing or synergistic agent in the development of new combination therapies for the treatment of disease.

In this chapter, a versatile chemical platform based on moderate molecular weight peptoids (450-1000 Da) will be evaluated and discussed for its application in the future development of small molecule protein-protein interaction modulators [325]. This system of computational-driven peptoid library design could be applied to other model protein systems where traditional screening

methods fail, or little is known about the protein(s) of interest. The advantages of such an approach would be a reduction in cost and time by greatly reducing the number of compounds to be synthesized, in lieu of creating large (>50,000) ligand libraries. Additionally in this analysis, key PCNA-ligand interactions were identified; together with the apparent flexibility of the surface of PCNA, this may indicate the adaptability of the protein toward multiple forms of antagonists. This chapter will conclude with the discussion of how this information could be used in the design of drug-like small molecules that target PCNA.

2.5 Experimental

2.5.1 Methodology

The process toward discovering peptoid-based inhibitors of PCNA-protein interactions was a multi-step system where PCNA was first interrogated computationally, followed by synthesis and subsequent biochemical screening of each peptoid. The details of the methodology follow.

Step 1: Generate focused library of peptoid-based antagonists using computational tools. The PIP Box binding site on PCNA was examined in an effort to gain basic information on favorable chemical features of potential small molecule inhibitors. First, a set of 20 primary amines was selected that represented variation in hydrophobicity, aromaticity, ability to form hydrogen bonds, and contained substructures present in clinically available drugs. Many of the amines from this set are commercially available, though some require different degrees of synthetic preparation. In preparation for virtual screening, the CombiGLIDE application within Maestro (Schrödinger) [326] was used to generate a virtual library of trimeric peptoids that contained each of the 20 primary amines, plus hydrogen as a potential substituent, in a combinatorial fashion to give 9,261 total compounds. The ligands were ionized (along with desalting and tautomeric generation) using Epik to generate all possible states within a pH range of 7.0 ± 2.0 , and these were minimized using the OPLS-2005 force field.

In general, using a single rigid protein structure for virtual docking tends to produce skewed results due to geometric constraints and functional group directionality at the binding site. This is exacerbated when taking into account a protein such as PCNA, which has a significant degree of plasticity at the PIP Box site [177], since there are likely to be many false negatives when docking rigidly. While a flexible docking method such as Schrödinger Maestro's induced-fit [327] is more

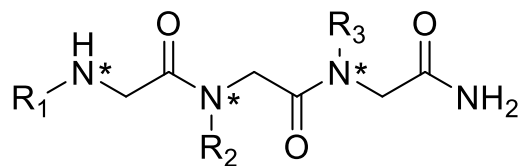
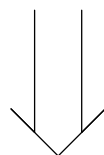
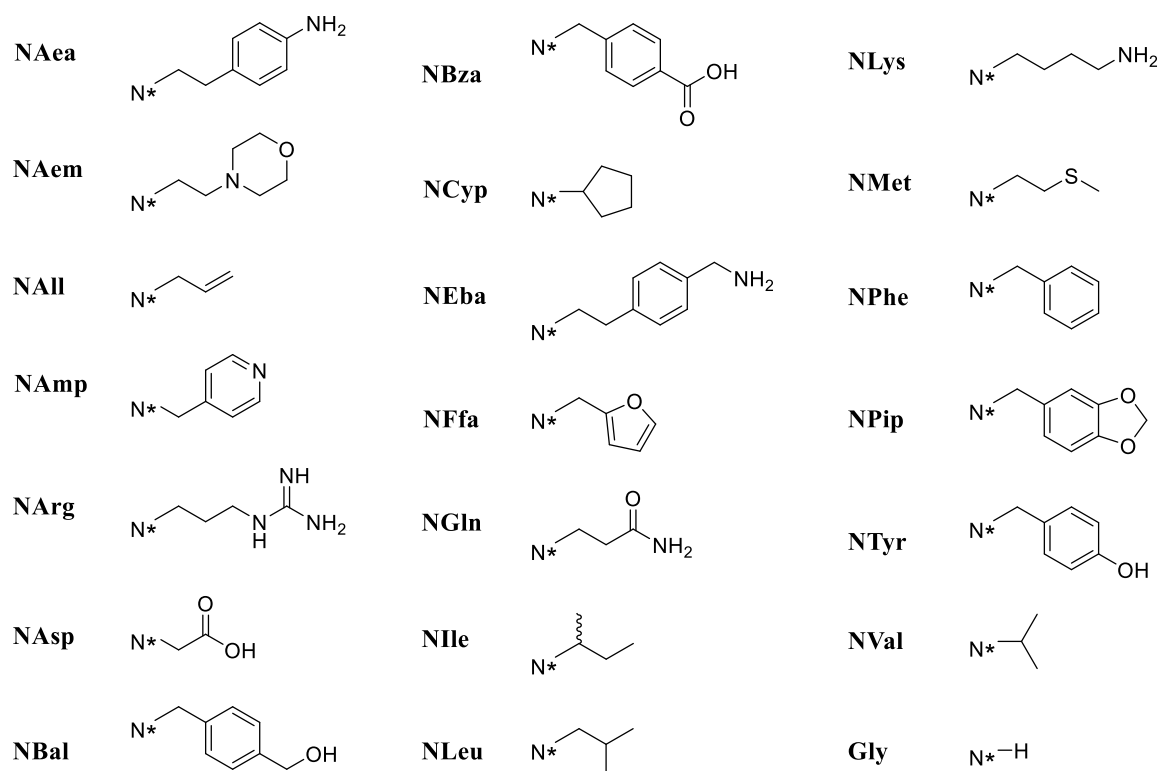


Figure 2.2 Development of Initial Virtual Library. (*top*) 20 commercially and non-commercially available primary amines were selected for their variation in hydrophobicity, aromaticity, ability to form hydrogen bonds, and/or whether they contained substructures present in clinically available drugs. N* indicates the location of the $-NH_2$ group, which is the position of substitution into a peptoid backbone (*bottom*). R₁, R₂ and R₃ indicate the three substitution positions in a resultant tripeptoid, and are referred to here as the first, second and third positions, respectively. Hydrogen was also considered as a potential substituent (Gly), and would therefore effectively constitute the incorporation of a glycine amino acid at wherever hydrogen was to be ‘substituted’.

ideal for screening ligands virtually, as there should theoretically be fewer false positives and negatives, the considerable amount of computational time required for such a method makes it impractical for screening more than a few ligands at a time, especially those with high degrees of freedom such as linear peptides and peptoids. To partially compensate for this, instead of a single rigid structure of PCNA used for screening, four crystal structures of PCNA bound to four different ligands were selected, each having a significant degree of structural divergence surrounding the PIP Box binding site. These structures included PCNA complexed to the C-terminal tail of p21 (PDB ID: 1AXC), PCNA complexed to residues 331-350 of FEN1 (PDB ID: 1U7B), PCNA complexed with the T3 ligand (PDB ID: 3VKX) and PCNA complexed with a fragment of DNA polymerase η (PDB ID: 2ZVK).

Each crystal structure of PCNA was prepared using the Protein Preparation Wizard in Maestro, with each protein minimized in complex with its respective ligand/peptide using the OPLS-2005 force field to an RMSD of 0.30Å. Cubic grid boxes were created using either T3 (3VKX), or a PIP Box peptide amino acid at the hydrophobic site on PCNA, as the centroid, with a length of 30Å. The prepared ligands were flexibly docked into each form of PCNA using the standard precision (SP) model in Glide [328], and the top 10% of the ligand hits from each docking run were flexibly re-docked into the respective form of PCNA using the extra precision (XP) Glide algorithm, which places harsher penalties on energy of solvation, solvent-exposed surface area and buried charge. Upon completion of each docking exercise, the results were compiled for each of the top 50 hits in each of the docking runs. The frequency with which a specific side chain was present at a particular location along the tripeptoid backbone was tallied for each crystal structure of PCNA, and the top 50 hits from each run were compiled into a total list of 200 top hits.

To further explore the use of peptoids as antagonists of PCNA-PIP Box interactions, the known PCNA inhibitor, T2AA, was investigated for its use as a peptoid fragment. Functional groups on T2AA potentially allow the compound to be incorporated into a peptoid backbone with the same basic methodology used for tripeptoid synthesis (i.e. submonomer peptoid synthesis [300]). In this way, T2AA could serve as an anchoring residue, potentially helping to direct small peptoid molecules to the PIP Box binding site on PCNA. To explore this possibility, a virtual combinatorial library was created using the same methodology as previously discussed, with a set of 37 peptoid side chains (17 new primary amines in addition to the original set) and two forms of T2AA, with defined attachment points at either the primary amine or amide groups on the molecule. This library was then systematically screened using the crystal structure of PCNA bound to the T3 ligand (PDB ID: 3VKX). All ligands were prepared in Maestro using the same previous

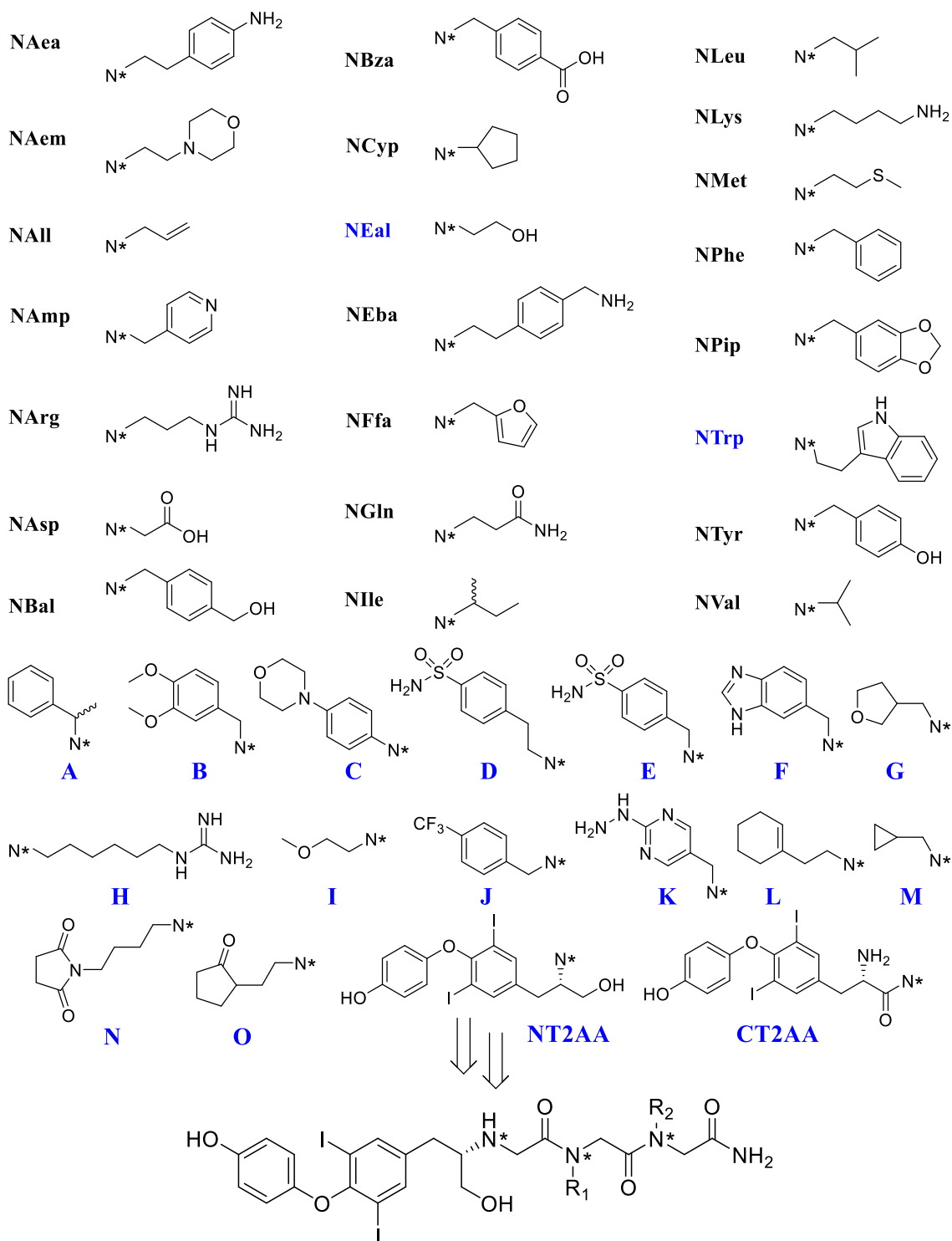


Figure 2.3 Creation of a Virtual Library Containing T2AA as a Fragment. A list of 37 primary amines, along with T2AA, was used to create a combinatorial virtual library of peptoid-based compounds. N* indicates the location of the -NH_2 group, which is the position of substitution into a peptoid backbone. R₁ and R₂ indicate the two substitution positions in a resultant tripeptide. Blue labels denote new fragments.

methodology. The high throughput virtual screening (HTVS) algorithm in Glide was used to screen the initial set of compounds. The top 50% best hit compounds were screened again using Glide's SP model, and the top 7,500 hits from the SP screen were docked again using Glides XP model. As before, the frequency that a particular side chain was present at a given substitution location on the peptoid backbone was tallied for the top 50 hits from the XP screen.

Step 2: Select and Synthesize Set of Tripeptoid Ligands. Based on the results from the virtual docking study, particular side chains appeared to be favored at specific positions along the tripeptoid backbone (see Figure 2.4). Amines that appeared in approximately 10% or more of the top 50 hits at a given position were selected for incorporation via synthesis. Peptoids were synthesized using traditional submonomer solid phase practices [300], purified by HPLC and characterized by ESI mass spectrometry. Side chains that are not commercially available were synthesized in-house using acid-labile protecting groups.

Step 3: Screen Synthesized Tripeptoids Using Fluorescence Polarization (FP). Synthesized peptoids were formulated in DMSO at concentrations of 10 mM. Synthesized peptoids were initially screened using a fluorescence polarization assay (see Experimental Materials and Methods section for details) at concentrations of 1 mM and 250 μ M. Ligands that demonstrated sufficient ability to disrupt binding between PCNA and the PL peptide were identified, and then further characterized by generating dose response curves for each compound to obtain IC_{50} and K_i values.

Step 4: Computational Analysis of Select Top-Hit Compounds. Peptoids that emerged as hits from the fluorescence polarization assay were examined using molecular dynamic (MD) simulations to identify key interactions that may be driving their affinity for PCNA. A principal component analysis (PCA) of each trajectory was performed and compared to MD trajectories from several peptides, as well as T2AA, all of which are known to bind to PCNA.

2.5.2 Experimental Materials and Methods

Reagents and solvents were purchased from Sigma Aldrich unless otherwise noted. Materials were repurified via recrystallization or distillation as necessary before use. NMR experiments were performed on Bruker (Bruker Corp., Billerica, MA) ARX300 (300 MHz), ARX400 (400 MHz) or DRX500 (500 MHz) instruments. Low resolution electrospray ionization (ESI) and atmospheric

pressure chemical ionization (APCI) studies were carried out on an Agilent 6320 Ion Trap (Agilent Labs, Santa Clara, CA) mass spectrometer. High resolution mass measurements were obtained on a LTQ Orbitrap XL mass spectrometer (ThermoScientific Corp.) utilizing electrospray ionization (ESI). Molecular masses and sequences of peptides or peptoids were validated on an Applied Biosystems (Framingham, MA) MALDI-TOF/TOF 4800 mass analyzer, or Applied Biosystems Voyager DE PRO mass spectrometer using either 2,5-dihydroxy benzoic acid or α -cyano-4-hydroxy cinnamic acid matrices. TLC analyses were performed on Merck aluminum-backed F254 silica gel plates. Protein and peptide concentrations were determined by UV absorbance at 280 nm. Fluorescent peptide concentrations were determined by absorbance at 494 nm. Stock solutions of each polypeptoid were made by measuring the dry mass of each in pre-dried, pre-weighed screw-cap vials, and adding the volume of DMSO necessary to give 10 mM solutions. Stock solutions of compounds containing N-terminal 5-carboxyfluorescein were made by measuring the absorbance at 494 nm, using an extinction coefficient of $79,000 \text{ L mol}^{-1} \text{ cm}^{-1}$ and Beer's Law ($A=\epsilon bc$) to calculate concentration. Data analyses and graphical representations were performed in Microsoft Excel, GraphPad Prism 6 or OriginPro 2015.

Ligation Independent Cloning of N-terminal His-tag PCNA Construct. Ligation independent cloning compatible expression vector pEV-L8 containing an N-terminal His-tag and TEV protease recognition site was linearized by digestion with Ssp1 (New England Biolabs), purified by gel filtration, and treated with T4 DNA polymerase (Novagen) in the presence of dGTP (New England Biolabs) for 30 minutes at 22°C, followed by heat inactivation at 75°C for 20 minutes. The PCNA fragment was amplified by PCR from a template plasmid (Genecopeia) using a high-fidelity polymerase Platinum Pfx DNA polymerase (Invitrogen). The resulting PCR products were treated with T4 DNA polymerase in the presence of dCTP to generate 5' overhangs necessary for annealing. A total of 0.2 pmol of each insert was incubated with 0.01 pmol of pEV-L8 vector in 3 μL reaction mix at 22°C for 10 minutes followed by addition of 1 μL of 25 mM EDTA at 22°C for 5 minutes. Annealing reaction products were transformed into X10Gold competent cells (Stratagene) and plated on LB agar containing 50 $\mu\text{g}/\text{mL}$ kanamycin. Individual colonies were grown and the constructs were assessed by PCR for insert size and verified by sequencing before propagating the plasmid.

Expression and Induction of N-terminal His-tag PCNA. 10 μL aliquots of chemically competent BL21 (DE3) *E. coli* cells (Agilent) were transformed via heat shock with 1 μL of purified plasmid

encoding the fusion protein, N-terminal (His)₆-PCNA for 30 seconds at 42°C. The cells were then immediately placed on ice for 2 min, and 140 µL of SOC medium was added. Transformed cells were allowed to grow for 1 hour at 37°C before streaking on a LB agar plate containing 50 µg/mL kanamycin. Single isolated colonies were picked and grown at 37°C to an OD of 0.7-1.0 in the presence of 50 µg/mL kanamycin. Transformed cells were induced with 0.4 mM IPTG for 4 hours at 37°C. Transformed cells were pelleted at 4,000 x g for 20 minutes at 4°C, and stored at -80°C until lysis.

Purification of N-terminal His-tagged PCNA. Two pellets of transformed BL21 (DE3) *E. coli* cells were each resuspended in 20 mL of ice-cold lysis buffer (50 mM Tris HCl at pH 8.0, 0.15 mM NaCl), lysed by sonication at a 30% amp output for 3 minutes (20 second pulses), and centrifuged at 4,000 x g for 20 minutes at 4°C. Each supernatant was decanted and combined. Recombinant (His)₆-PCNA fusion protein was then purified from the soluble fraction by affinity column chromatography using Ni-NTA resin at 4°C. After charging the column resin with the entire soluble protein fraction, the column was washed with 20 mM imidazole in Tris buffer at pH 8.0 to remove nonspecific binding protein. (His)₆-PCNA was then eluted with 10 mL of 1M imidazole in Tris buffer at pH 8.0. The eluted protein was diluted two-fold with dialysis buffer (25 mM HEPES at pH 7.4, 10% glycerol, 0.01% Triton X-100), DTT and EDTA were added to a final concentration of 2 mM, and the entire solution was diluted two-fold with 2 M ammonium sulfate in 25 mM Tris buffer at pH 8.0 to give a final (NH₄)₂SO₄ concentration of 1 M. The solution was agitated for 1 hour at 4°C, and during that time (His)₆-PCNA precipitated from solution. The precipitated protein was pelleted by centrifugation (5,000 x g for 10 minutes), the supernatant decanted, and the protein pellet dissolved into 10 mL of dialysis buffer. The protein concentration was immediately assessed via measuring its absorbance at 280 nm (using an extinction coefficient of 16,000 M⁻¹cm⁻¹), and the protein solution was diluted as necessary to give a stock concentration of 4 µM. This was then dialyzed for 24 hours, swapping the dialysis buffer twice with fresh 25 mM HEPES at pH 7.4, 10% glycerol, 0.01% Triton X-100. After dialysis, the protein concentration was re-confirmed by measurement of its absorbance at 280 nm.

General Method for Synthesis of Synthetic Peptides. Peptides were synthesized via solid phase synthesis using 0.10 mmol of Rink Amide AM or MBHA resin (EMD Millipore). Each synthetic step was performed in a glass fritted peptide reaction vessel placed on an orbital shaker. Before the addition of Fmoc-protected amino acids (Anaspec), the resin was first swelled with

dimethylformamide (DMF) for 30 minutes, and deprotected with a solution 20% piperidine in DMF for 30 minutes. Amino acids were coupled using standard peptide synthesis conditions (1 mmol of Fmoc-protected amino acid, 2.1 mL of 0.45 M 2-(6-Chloro-1H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate (HCTU; Anaspec) in DMF and 500 μ L of 4M N,N-diisopropylthethylamine (DIEA) in N-methyl-2-pyrrolidone (NMP)) for one hour at room temperature, and were deprotected using 20% piperidine in DMF for 30 minutes at room temperature. Between coupling and deprotection steps, the resin was washed with DMF (6x) and dichloromethane (DCM) (3x). Reaction progress was monitored using a ninhydrin (Kaiser) test for the presence of primary amines. For fluorescent N-terminal labeling using 5-carboxyfluorescein (5-FAM; Anaspec), Fmoc-6-aminohexanoic acid was coupled to the completed resin-bound peptide and subsequently deprotected, followed by the addition of a solution of 2 equivalents of 5-FAM, 1.95 equivalents of HCTU and 4 equivalents of DIEA dissolved in 2 mL of DMF, allowing the resin to incubate at room temperature overnight in the dark. Peptides were cleaved from resin using a solution of trifluoroacetic acid (TFA)/triisopropylsilane (TIS)/water (95:2.5:2.5), incubating the resin at room temperature for either 1 hour, or 3 hours in the case of peptides containing glutamine, arginine, tryptophan or histidine. They were then precipitated into ice cold diethyl ether and collected by centrifugation at 4,000 x g for 10 minutes at 4°C. The peptides were purified via HPLC (Beckman Coulter System Gold 166 or 168) using an increasing gradient of acetonitrile (ACN)/water with 0.1% TFA (5:95) to (100:0) over 30 minutes on an Agilent ZORBAX SB-C18 reverse phase semi-preparative column. Molecular masses and sequences were validated via MALDI-TOF/TOF. Purities were determined by HPLC using absorbencies at 219 or 280 nm.

Synthesis of Fluorescein-labeled POGO Ligase Peptide (FAM-PL). POGO Ligase peptide (sequence: SAVLQKKITDYFHPKK) was synthesized by GenScript USA Inc. and provided uncleaved on Rink Amide MBHA resin. 0.10 mmol of this resin was transferred to a glass fritted peptide reaction vessel and swelled in DMF for 30 minutes, followed by washing with DCM (3x). Sufficient deprotection of the final amino acid residue was confirmed by a ninhydrin (Kaiser's) test for primary amines. An aminohexanoic acid linker was added (1 mmol Fmoc-6-Ahx-OH, 2.1 ml of 0.45M HCTU, 500 μ L of 4M DIEA in NMP; 2 hours at room temperature) to separate the fluorescent dye from the peptide sequence. Following subsequent washing of the resin (DMF (6x) and DCM (3x)) and Fmoc deprotection (20% piperidine in DMF; 30 minutes at room temperature), the resin was again washed (DMF (6x) and DCM (3x)) and then transferred to a glass scintillation

vial wrapped in foil. A solution of 75.3 mg of 5-FAM, 80.7 mg of HCTU and 46 mg of DIEA in 2 mL of DMF was added, and the resin was then placed on an orbital shaker overnight at room temperature. After 20 hours of incubation time, the resin was washed with DMF (6x) and DCM (3x), dried over vacuum, transferred to a glass scintillation vial and cleaved from resin using a solution of trifluoroacetic acid (TFA)/triisopropylsilane (TIS)/water (95:2.5:2.5) for 3 hours at room temperature in the dark. The 5-FAM-labeled POGO ligase peptide (FAM-PL) was then precipitated into ice cold diethyl ether and collected by centrifugation at 4,000 x g for 10 minutes at 4°C. It was purified via HPLC (Beckman Coulter System Gold 168) using an increasing gradient of acetonitrile/water with 0.1% TFA (5:95) to (100:0) over 30 minutes. The molecular mass and sequence were validated via MALDI-TOF/TOF. Purity was determined by HPLC using absorbencies at 219 and 280 nm. HRMS (LCMS): calculated mass (C₁₁₆H₁₆₄N₂₅O₂₉) [M-H]¹⁻: 2372.6980, mass found *m/z*: 2372.7221 [M-H]¹⁻.

General Method for Synthesis of Peptoid Trimers. Trimeric peptoids were synthesized using an adapted procedure for submonomer [300] peptoid synthesis. Briefly, 0.05 mmol of Rink Amide AM or MBHA resin was transferred to a 25 mL glass fritted peptide reaction vessel and was swelled with DMF for 30 minutes. The resin was then deprotected using two 2.5 mL portions of 20% piperidine in DMF with incubation times of 15 minutes for each addition at room temperature. Following washing of the resin with DMF (6x) and DCM (3x), deprotection was confirmed by a ninhydrin (Kaiser's) test for primary amines. A solution of 1.5 mL of 1 M bromoacetic acid (30 equiv.) in DMF and 230 µL (29.4 equiv.) of N,N'-diisopropylcarbodiimide (DIC) was added, and the resin was placed on an orbital shaker for 1 hour at 37°C. At that time, the resin was washed with DMF (6x) and DCM (3x), and a solution of 1 M respective primary amine (2 M for commercially available primary amines) in DMF was added, with incubation on an orbital shaker for 2 hours at 37°C. These steps were repeated with washing steps in-between to produce the desired peptoid sequence. For the coupling of T2AA, peptoids were first synthesized up to the final bromoacetic acid addition. A solution of 500 mg (19.5 equiv.) of T2AA and 34 µL (39 equiv.) of DIEA in 2.5 mL of DMF was added, and the resin was incubated overnight at room temperature on an orbital shaker. Peptoids were cleaved from resin using a solution of TFA/TIS/water (95:2.5:2.5), incubating the resin at room temperature for either 1 hour, or 3 hours in the case of peptoids containing NArg, NEal or NBal side chains. TFA was removed with a steady stream of blowing air, and the remaining residue was dissolved in ACN/H₂O (50:50) with 0.1% TFA, frozen and lyophilized. The peptoids were purified via HPLC (Beckman Coulter System Gold 166 or 168)

using an increasing gradient of ACN/H₂O with 0.1% TFA (5:95) to (100:0) over 30 minutes on an Agilent ZORBAX SB-C18 reverse phase semi-preparative column. Molecular masses were validated via low resolution ESI or APCI experiments, and exact masses were obtained by high resolution ESI or MALDI.

Fluorescence Polarization Z'-Factor Analysis. 10 μ L of 20 nM FAM-PL in FP binding buffer (25 mM HEPES at pH 7.4, 10% glycerol, 0.01% Triton X-100) was combined with either 10 μ L of 200 nM recombinant (His)₆-PCNA protein in binding buffer or 10 μ L of binding buffer in each of 48 wells on a ProxiPlate-384 F Plus (PerkinElmer) low volume, black, opaque plate (24 replicates for each set). The plate was allowed to incubate at room temperature in the dark for 30 minutes prior to fluorescent measurement. Fluorescence polarization and resultant anisotropy were measured on a BioTek Synergy 4 Multi-Detection Microplate Reader using an excitation filter of 485 nm and an emission filter of 530 nm, each with a 20 nm band-pass. The average of each control set was calculated along with the standard deviation. The Z'-factor was calculated using equation 2.1,

$$Z' = 1 - \frac{3\sigma_+ + 3\sigma_-}{|\mu_- - \mu_+|} \quad (2.1)$$

where σ_+ is the standard deviation of the positive control (FAM-PL peptide in the presence of PCNA protein), σ_- is the standard deviation of the negative control (FAM-PL peptide in the absence of PCNA protein), and μ_+ and μ_- are the mean anisotropy values of the positive and negative controls, respectively.

Fluorescence Polarization Binding Assay. Increasing amounts of recombinant (His)₆-PCNA protein were prepared in FP binding buffer (25 mM HEPES at pH 7.4, 10% glycerol, 0.01% Triton X-100), with an 11-step 2-fold dilution series, and a top concentration of 30 μ M. 10 μ L of each solution was combined with 10 μ L of 20 nM FAM-PL peptide formulated in FP binding buffer in a single well of a ProxiPlate-384 F Plus low volume, black, opaque plate. Each concentration of protein was plated in a replicate of four, and the plate was allowed to incubate at room temperature in the dark for 30 minutes prior to fluorescent measurement. Fluorescence polarization and resultant anisotropy were measured on a BioTek Synergy 4 Multi-Detection Microplate Reader using an excitation filter of 485 nm and an emission filter of 530 nm, each with a 20 nm band-pass. The

parallel and perpendicular intensity values for each sample (n=4) were used to calculate fractional occupancy (FO) of the FAM-PL peptide bound to monomeric PCNA using equation 2.2.

$$f_b = \frac{r - r_f}{(r_b - r)Q + r - r_f} \quad (2.2)$$

where

$$Q = \frac{q_b}{q_f}$$

$$q_f = \parallel_f + 2 \cdot \perp_f$$

$$q_b = \parallel_b + 2 \cdot \perp_b$$

and f_b is the fraction of FAM-PL bound to PCNA, r is the observed anisotropy value, r_f is the anisotropy of free un-bound FAM-PL peptide, r_b is the anisotropy of FAM-PL peptide saturated with PCNA protein, Q is the ratio of quantum yield of bound (q_b) to free (q_f) FAM-PL peptide, \parallel_f and \parallel_b are the parallel intensities of free un-bound and saturated FAM-PL peptide, respectively, and \perp_f and \perp_b are the perpendicular intensities of free un-bound and saturated FAM-PL peptide, respectively.

FO values were analyzed using non-linear regression statistics in OriginPro 2015, representing them as the mean \pm standard error of the mean (Y), and plotting them as a function of the monomeric PCNA protein concentration (X). From this, equation 2.3 was used to obtain a dissociation constant (K_d) for FAM-PL.

$$Y = Y_{max} \cdot \frac{X^n}{K_d^n + X^n} \quad (2.3)$$

where n is the Hill slope.

Fluorescence Polarization Competition Assay. Solutions of competitive ligand were formulated from DMSO stocks (10 mM for tripeptoids, 20 mM for T2AA) into FP binding buffer (25 mM HEPES at pH 7.4, 10% glycerol, 0.01% Triton X-100) at appropriate 2X concentrations, relative to the desired effective screening concentration. 10 μ L of each competitive ligand was combined

with 5 μL of 4 μM recombinant (His)₆-PCNA protein in binding buffer and 5 μL of 40 nM FAM-PL in binding buffer into each well of a ProxiPlate-384 F Plus low volume, black, opaque plate, in replicates of four. DMSO at an equivalent concentration in binding buffer to the concentration of DMSO in the competitive ligand sample was used as a negative control; T2AA at 250 μM was used as a positive control. The plate was allowed to incubate at room temperature in the dark for 30 minutes prior to fluorescent measurement. Fluorescence polarization and resultant anisotropy were measured on a BioTek Synergy 4 Multi-Detection Microplate Reader using an excitation filter of 485 nm and an emission filter of 530 nm, each with a 20 nm band-pass. Anisotropy values were converted to fractional occupancy using equation 2.2, and IC₅₀ values were calculated by fitting the data to equation 2.4.

$$Y = Y_{min} + \frac{Y_{max} - Y_{min}}{1 + 10^{(\log IC_{50} - X)n}} \quad (2.4)$$

where n is the Hill slope.

Inhibition constants (K_i) for the competitive ligands were determined using equation 2.5, which is a modified form of the Cheng-Prusoff equation, previously reported for fluorescence polarization assays [329].

$$K_i = \frac{[I]_{50}}{\left(\frac{[L]_{50}}{K_d} + \frac{[P]_0}{K_d} + 1\right)} \quad (2.5)$$

where $[I]_{50}$ is the concentration of each competitive peptoid at 50% inhibition, $[L]_{50}$ is the concentration of the FITC-PL peptide at 50% inhibition, $[P]_0$ is the concentration of monomeric PCNA protein at 0% inhibition, and K_d is the dissociation constant calculated from equation 2.3.

Molecular Dynamic Simulations. In preparation for molecular dynamic simulations, selected hit peptoids were flexibly docked into the PIP Box binding site of the co-crystal structure of PCNA-Pogo Ligase (PL) peptide (PDB ID: 1VYJ), with the peptide itself removed, using Schrödinger Glide's induced-fit docking model. The PCNA crystal structure was prepared using the Protein Preparation Wizard in Maestro, with PCNA minimized in complex with the PL peptide using the OPLS-2005 force field to an RMSD of 0.30Å. PCNA-peptoid complexes were solvated in

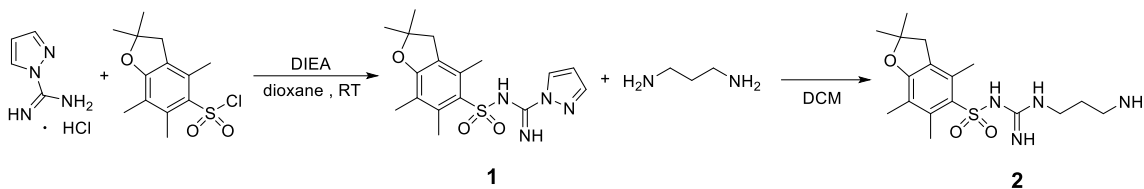
Schrödinger's Desmond [330] using the TIP3P water model in the presence of 0.15 M sodium chloride buffer to generate orthorhombic water boxes that contained a 10 Å buffer region. Each system was then minimized with the OPLS-2005 force field to an RMSD of 0.30Å.

The molecular dynamic simulations were performed in the same way as described in Pedley, *et al.* (2014) [177], using Desmond and the OPLS-2005 force field. In summary, long-range electrostatic interactions were determined using a smooth particle mesh Ewald method with a grid spacing of 0.8 Å. For non-bonded van der Waals interactions, a cut off of 9.0 Å was set. All simulations were performed for 5.0 ns, except in cases where simulations did not fully converge after 5.0 ns (simulations were extended by 2.5 ns in those situations), using the Desmond NPT method with a six step slow relaxation protocol prior to the molecular dynamics run: (i) 2000 step limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) minimization with a loose convergence restraint of 50 kcal/mol/Å; (ii) 2000 step L-BFGS minimization with a convergence constraint of 5 kcal/mol/Å; (iii) a 12 ps Berendsen NVT simulation at a temperature of 10 K with restraints on solute heavy atoms; (iv) a 12 ps Berendsen NPT ensemble at a temperature of 10 K and pressure at 1.01325 bar with restraints on solute heavy atoms; (v) a 24 ps Berendsen NPT ensemble at a temperature of 300 K and a pressure at 1.01325 bar with restraints on solute heavy atoms; (vi) a 24 ps Berendsen NPT ensemble at a temperature of 300 K and a pressure at 1.01325 bar with restraints on residues beyond 15 Å of the restrained ligand. The 5.0 ns molecular dynamic simulation run was performed using NPT ensemble. Temperature of the simulation was kept at 300 K using a Nosé-Hoover thermostat. Pressure was maintained at 1.01325 bar using the Martyna-Tobias-Klein method. Energy and trajectory data was recorded at every 1.2 ps and 5.0 ps, respectively.

Upon completion of each simulation, PCNA trajectory data were processed in VMD [331] after removal of each peptoid ligand. For each trajectory, the protein backbone C α atoms were aligned to the first frame of the simulation to generate RMSD and C α fluctuations (RMSF) values. Simulations were determined to be converged once RMSD values had stabilized (average slope of the RMSD curve over the period of the final 0.5-1.0 ns \approx 0). In preparation for principal component analysis, trajectories for each PCNA-peptoid system were overlaid using the alignment tools in VMD. Additionally, trajectories previously generated from Pedley, *et al.* (2014), including the systems where PCNA is in complex with the Polymerase δ , PL, p85 α , p21, Apo, Akt or Abl peptides, were overlaid with the PCNA-peptoid trajectories for purposes of comparative analysis. Principal component analyses were performed using the Bio3D package [332] in R to analyze the conformational differences between the aligned trajectories over the period of the final 0.5 ns for

each simulation (100 snapshots per PCNA-ligand system). The first two orthogonal eigenvectors (principal components—PC1 and PC2) were plotted on the same set of axes. Average trajectory coordinates for the final 50 frames of each simulation were performed in VMD to generate average overall conformations of PCNA in complex with each peptoid.

2.5.3 Synthesis of Non-Commercially Available Primary Amines and T2AA

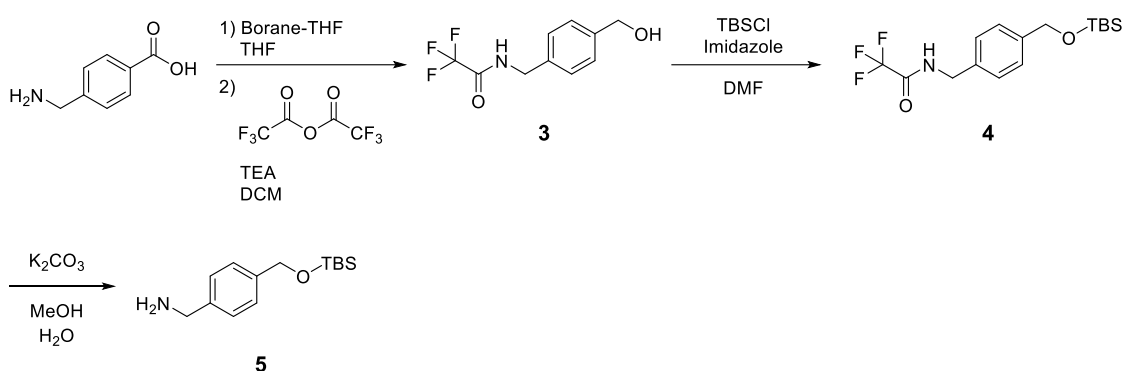


Scheme 2.1
Synthesis of **NArg**

Synthesis of N-((2,2,4,6,7-pentamethyl-2,3-dihydrobenzofuran-5-yl)sulfonyl)-1H-pyrazole-1-carboximidamide (1). 16.16 g (55.96 mmol) of 2,2,5,7,8-pentamethylchroman-6-sulfonyl chloride (Combi-Blocks, San Diego, CA) was dissolved in dioxane (200 mL), and 9.07 g (61.88 mmol) of 1H-pyrazole-1-carboximidamide HCl dissolved in dioxane (200 mL) was added to the solution followed by 22 mL (2 eq.) of DIEA. The reaction mixture was stirred at room temperature for 48 hours, at which time all of the 1H-pyrazole-1-carboximidamide HCl had been consumed as confirmed via TLC. The dioxane was evaporated in vacuo and the remaining brown oil was redissolved into 200 mL DCM. The organic layer was washed with water (3 x 200 mL) and brine (200 mL). The organic layer was mostly evaporated in vacuo, and the remaining mixture was poured over wet silica gel to remove the polar impurity that appeared via TLC under conditions of ethyl acetate/hexane (25:75), but could not be removed with liquid-phase extraction. After flushing the silica gel through with ethyl acetate/hexanes (25:75), the filtrate was evaporated, and the resulting solid was recrystallized from ethanol to give 13.17 g (64.9% yield) of **1** as a white crystalline solid. ¹H NMR (500 MHz, CDCl₃) δ 8.21 (dd, *J* = 2.8, 0.7 Hz, 1H), 7.67 (dd, *J* = 1.6, 0.7 Hz, 1H), 6.40 (dd, *J* = 2.9, 1.6 Hz, 1H), 2.98 (s, 2H), 2.62 (s, 3H), 2.56 (s, 3H), 2.11 (s, 3H), 1.47 (s, 6H). ¹³C NMR (126 MHz, CDCl₃) δ 12.34, 17.86, 19.16, 28.48, 42.99, 86.67, 109.47, 117.74, 124.83, 128.92, 130.88, 132.98, 139.13, 143.44, 148.66, 159.41.

Synthesis of N-((N-(3-aminopropyl)carbamimidoyl)-2,2,4,6,7-pentamethyl-2,3-dihydrobenzofuran-5-sulfonamide (2 ; “NArg”). 5.88 g (17.5 mmol) of **1** was dissolved into DCM (100 mL) and was

added dropwise to 6.75 g (91.1 mmol; 5.2 equiv.) of 1,3-diaminopropane dispersed in DCM (100 mL) at room temperature while stirring. The reaction was allowed to stir for 24 hours, at which time TLC showed that all of **1** had been consumed. The reaction mixture was washed with water (3 x 100 mL) and brine (100 mL), the organic layer was dried over sodium sulfate, and the organics were evaporated in vacuo to give an off-white solid. The product was recrystallized using ethyl acetate/hexanes to give 3.57 g (59.7% yield) of **2** as a white solid. ^1H NMR (500 MHz, CDCl_3) δ 3.30 (t, $J = 6.09$ Hz, 2H), 3.25 (q, $J = 6.20$ Hz, 1H, NH), 3.20 (q, $J = 6.65$ Hz, 1H, NH), 2.93 (s, 2H), 2.88 (t, $J = 7.63$ Hz, 2H), 2.54 (s, 3H), 2.47 (s, 3H), 2.07 (s, 3H), 1.82 – 1.70 (m, 2H), 1.66 (t, $J = 6.66$ Hz, 1H, NH), 1.45 (s, 6H). ^{13}C NMR (126 MHz, CDCl_3) δ 158.71, 138.15, 132.32, 132.12, 124.62, 117.50, 109.22, 86.35, 43.10, 39.78, 37.98, 28.50, 19.23, 17.85, 12.39. HRMS (ESI): calculated mass ($\text{C}_{17}\text{H}_{29}\text{N}_4\text{O}_3\text{S}$) $[\text{M}+\text{H}]^{1+}$: 369.1961, mass found m/z : 369.1980 $[\text{M}+\text{H}]^{1+}$.



Scheme 2.2
Synthesis of NBal

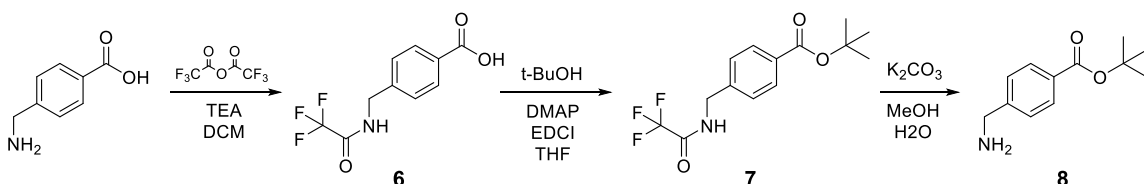
Synthesis of 2,2,2-trifluoro-N-(4-(hydroxymethyl)benzyl)acetamide (3). 20.0 g (132 mmol) of 4-(aminomethyl)benzoic acid was dispersed in 100 mL of anhydrous THF, and 350 mL (2.65 equiv.) of 1 M borane-THF in THF was added dropwise. The reaction mixture was heated to reflux for eight hours, and then allowed to cool to room temperature. 100 mL of MeOH was added to quench the remaining borane-THF, and the reaction was stirred for an additional 15 minutes. The reaction solution was filtered over celite and evaporated in vacuo to give a light yellow solid. 250 mL of DCM was then added to the reaction flask containing the crude intermediate product, followed by 37 mL (2 equiv.) of triethylamine. The reaction was cooled to 0°C , 21 mL (1.1 equiv.) of trifluoroacetic anhydride was added dropwise and the reaction was allowed to stir overnight, allowing it to gradually reach room temperature. After 22 hours of reaction time, all of the starting

material had been consumed as evidenced by TLC. 250 mL of water was added to the reaction and the organic layer was separated. The aqueous layer was extracted once more with 250 mL of DCM, the combined organics were washed with water (100 mL), and were dried over sodium sulfate, filtered and evaporated in vacuo to give a thick yellow oil. This oil was purified via automated flash chromatography (EPCLC W-Prep 2XY, Yamazen Corp., Yodogawa-Ku Osaka, Japan) using DCM/MeOH as the eluents. The fractions containing the desired product were combined and evaporated to give 15.87 g (51.44% yield for both steps) of **3** as a white solid. ¹H NMR (300 MHz, CDCl₃) δ 7.34 (d, J = 8.17 Hz, 2H), 7.26 (d, J = 8.16 Hz, 2H), 6.86 (s, br, 1H), 4.66 (s, 2H), 4.49 (d, J = 5.86 Hz, 2H), 2.02 (s, 1H).

Synthesis of N-(4-(((tert-butyl)dimethylsilyloxy)methyl)benzyl)-2,2,2-trifluoroacetamide (4). 15.68 g (67.24 mmol) of **3** was dissolved in DMF (200 mL), followed by the addition of 9.17 g (2 equiv.) of imidazole and 11.29 g (1.11 equiv.) of tert-butyl dimethylchlorosilane. The reaction was allowed to stir overnight at room temperature. After 20 hours of reaction time, the reaction mixture was evaporated in vacuo to half of its original volume. 100 mL of water was added, and the solution was extracted with ethyl acetate (2 x 200 mL), the combined organics washed with water (200 mL) and brine (100 mL), dried over sodium sulfate, filtered and evaporated to give clear, yellow oil. This oil was purified via automated flash chromatography (EPCLC W-Prep 2XY, Yamazen Corp.) using ethyl acetate/hexanes as the eluents. The fractions containing the desired product were combined and evaporated to give 20.9 g (89.5% yield) of **4** as a white solid. ¹H NMR (300 MHz, CDCl₃) δ 7.34 (d, J = 8.26 Hz, 2H), 7.25 (d, J = 8.17 Hz, 2H), 6.83 (s, br, 1H), 4.74 (s, 2H), 4.49 (d, J = 5.80 Hz, 2H), 0.95 (s, 9H), 0.11 (s, 6H). ¹³C NMR (75 MHz, CDCl₃) δ 157.52, 141.83, 134.51, 128.03, 126.72, 117.90, 64.65, 43.81, 26.04, 18.53, -5.17.

Synthesis of (4-(((tert-butyl)dimethylsilyloxy)methyl)phenyl)methanamine (5; "NBal"). 20.8 g (59.9 mmol) of **4** was dissolved in methanol (100 mL), followed by the addition of a 2 M aqueous solution of potassium carbonate (27.1 g in 100 mL water). The reaction was heated to reflux for seven hours, and then allowed to cool to room temperature. Methanol was evaporated from the reaction mixture in vacuo, the remaining aqueous solution was transferred to a separatory funnel, and was extracted with DCM (2 x 400 mL). The combined organics were washed with water (100 mL), dried over sodium sulfate, filtered and evaporated in vacuo to give a yellow oil. The crude product was purified via automated flash chromatography (EPCLC W-Prep 2XY, Yamazen Corp.) with an increasing gradient of MeOH:DCM w/ 1% TEA. The fractions containing the desired

material were combined and evaporated to give 12.35 g (82.0% yield) of **5** as a clear oil, which solidified upon storage at -20°C . ^1H NMR (500 MHz, CDCl_3) δ 7.28 (d, $J = 8.58$ Hz, 2H), 7.25 (d, $J = 8.14$ Hz, 2H), 4.72 (s, 2H), 4.43 (s, br, 1H), 3.81 (s, 2H), 0.95 (s, 9H), 0.10 (s, 6H). ^{13}C NMR (126 MHz, CDCl_3) δ 141.85, 139.83, 126.87, 126.18, 64.70, 46.16, 29.35, 25.90, -5.29 . HRMS (ESI): calculated mass ($\text{C}_{17}\text{H}_{27}\text{NOSi}$) $[\text{M}+\text{H}]^{1+}$: 252.1784, mass found m/z : 252.1785 $[\text{M}+\text{H}]^{1+}$.



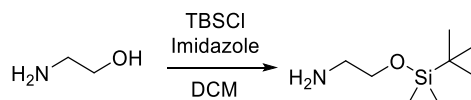
Scheme 2.3
Synthesis of **NBza**

Synthesis of 4-((2,2,2-trifluoroacetamido)methyl)benzoic acid (6). 21.16 g (140 mmol) of 4-(aminomethyl) benzoic acid was suspended in 450 mL of dichloromethane, followed by 42.0 mL (301 mmol) of triethylamine. The reaction was then cooled in an ice bath, and 60.44 g (2.06 eq.) of trifluoroacetic anhydride in 50 mL of dichloromethane was added dropwise over the course of 1 hour. The reaction was stirred for an additional three hours while being allowed to gradually warm to room temperature. 500 mL of aqueous saturated sodium bicarbonate solution was then slowly added to the reaction mixture in portions, and the solution was acidified with 4N HCl ($\text{pH} < 3$). The resultant precipitate was collected via filtration, and the filter cake was washed three times with water and twice with ice-cold ether. The solid was dissolved in ethyl acetate, dried over sodium sulfate, filtered, transferred to a round bottom flask and evaporated to give an off-white solid. The product was recrystallized from ethyl acetate/hexanes three times to give 24.63 g (71.19% yield) of **6** as a white solid. ^1H NMR (300 MHz, DMSO) δ 10.15 (t, $J = 5.91$ Hz, 1H), 8.13 (d, $J = 8.36$ Hz, 2H), 7.51 (d, $J = 8.38$ Hz, 2H), 4.52 (d, $J = 5.97$ Hz, 2H). ^{13}C NMR (75 MHz, DMSO) δ 167.97, 162.96, 145.92, 131.66, 130.49, 128.91, 127.95, 43.27.

Synthesis of tert-butyl 4-((2,2,2-trifluoroacetamido)methyl)benzoate (7). 12.55 g (50.77 mmol) of **6** was dissolved in anhydrous THF (150 mL), and 135 mL of t-butanol and 6.20 g (50.75 mmol) DMAP were added. The reaction was cooled to 0°C under argon, and 29.2 g (152 mmol) of EDCI was added followed by an additional 50 mL of anhydrous THF to wash down the insides of the reaction flask. The reaction was sealed and stirred under argon overnight, allowing it to gradually reach room temperature. After 16 hours of reaction time, 200 mL of water was added to the reaction

and the organics were evaporated in vacuo. The aqueous solution was extracted with DCM (2 x 200 mL), and the subsequent combined organics were washed with 5% HCl (2 x 200 mL), 200 mL of water and 200 mL of brine. The organic layer was dried over sodium sulfate and evaporated to give a yellow oil. The product was purified using normal phase flash chromatography (EPCLC W-Prep 2XY, Yamazen Corp.) with an increasing gradient of ethyl acetate/hexanes (10:90 to 100:0 over 60 minutes). The fractions containing the desired product were combined and evaporated to give 10.78 g (70.0% yield) of **7** as a white solid. ^1H NMR (300 MHz, DMSO) δ 1.53 (s, 9H), 4.46 (d, J = 5.94 Hz, 2H), 7.39 (dt, J = 1.78, 8.33 Hz, 2H), 7.89 (dt, J = 1.83, 8.33 Hz, 2H), 10.09 (t, J = 6.11 Hz, 1H). ^{13}C NMR (75 MHz, DMSO) δ 27.75, 42.36, 80.63, 116.22, 127.40, 129.31, 130.45, 142.54, 156.56, 164.70.

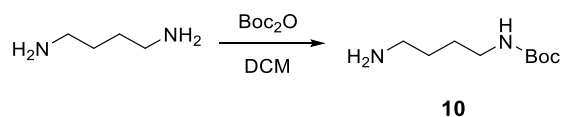
Synthesis of tert-butyl 4-(aminomethyl)benzoate (8; "NBza"). 10.75 g (35.5 mmol) of **7** was dissolved in methanol (45 mL), and 12.25 g (2.5 eq.) of potassium carbonate dissolved in water (45 mL) was added in one portion. The reaction was sealed and stirred overnight at room temperature. After 18 hours of reaction time, methanol was evaporated from the reaction mixture and the remaining aqueous solution was adjusted to pH > 10 with 4N NaOH. The aqueous layer was extracted with DCM (3 x 200 mL), the combined organics washed with water (50 mL) and brine (50 mL), and the organic layer dried over sodium sulfate. The organics were evaporated to give 7.10 g (96.7% yield) of **8** as a clear oil. ^1H NMR (300 MHz, CDCl_3) δ 7.91 (d, J = 8.30 Hz, 2H), 7.31 (d, J = 8.32 Hz, 2H), 3.86 (s, 2H), 1.69 (s, 1H), 1.55 (s, 9H). ^{13}C NMR (75 MHz, CDCl_3) δ 165.57, 147.58, 130.47, 129.60, 126.71, 80.76, 46.02, 28.12. HRMS (ESI): calculated mass ($\text{C}_{12}\text{H}_{18}\text{NO}_2$) $[\text{M}+\text{H}]^{1+}$: 208.1338, mass found m/z : 208.1334 $[\text{M}+\text{H}]^{1+}$.



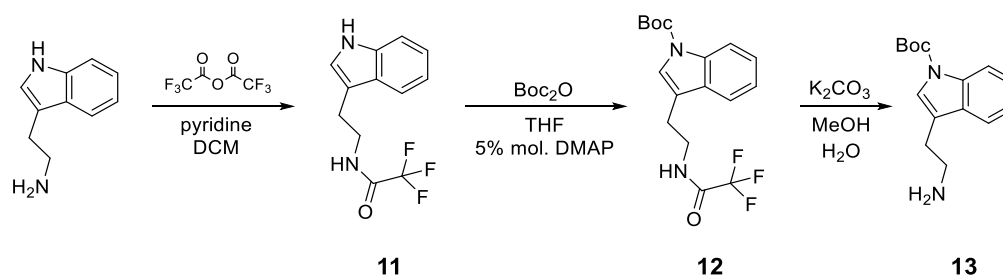
9

Synthesis of 2-((tert-butyldimethylsilyl)oxy)ethan-1-amine (9; "NEal"). 6.11 g (100 mmol) of ethanolamine and 13.62 g (2 equiv.) of imidazole were dissolved in DCM (100 mL) in a 500 mL round bottom flask. 15.83 g (105 mmol) of tert-butyldimethylchlorosilane dissolved in DCM (50 mL) was added dropwise over the course of 20 minutes, and the reaction mixture was stirred for one hour at room temperature. At that time, all of the starting material had been consumed as

confirmed by TLC; 100 mL of water was added and the layers were separated. The aqueous layer was extracted twice with DCM (2 x 100 mL), and the combined organics washed with water (50 mL), dried over sodium sulfate and evaporated to give 13.68 g (77.9% yield) of **9** as a clear oil. ¹H NMR (300 MHz, CDCl₃) δ 3.57 (t, *J* = 5.36 Hz, 2H), 2.71 (t, *J* = 5.25 Hz, 2H), 2.08 (s, 2H, NH₂), 0.84 (s, 9H), 0.00 (s, 6H). ¹³C NMR (75 MHz, CDCl₃) δ 65.13, 44.27, 25.96, 18.35, -5.27. HRMS (ESI) calculated mass (C₉H₂₂NOSi) [M+H]¹⁺: 188.1471, mass found *m/z*: 188.1495 [M+H]¹⁺.



Synthesis of tert-butyl (4-aminobutyl)carbamate (10; "NLys"). 50.02 g (567.4 mmol) of 1,4-diaminobutane was dissolved in chloroform (600 mL) and was cooled to 0°C. 13.17 g (6.03 mmol) of di-tert-butyl dicarbonate dissolved in chloroform (300 mL) was added drop-wise over the course of two hours and the reaction was stirred overnight, allowing it to reach room temperature. After 21 hours of reaction time, the entire reaction mixture was transferred to a separatory funnel and was washed with water (8 x 200 mL), dried over sodium sulfate and evaporated in vacuo to give 10.71 g (94.3% yield) of **10** as a clear oil. ¹H NMR (300 MHz, CDCl₃) δ 4.84 (s, 1H), 3.04 (t, 2H), 2.64 (t, *J* = 6.7 Hz, 2H), 1.46 – 1.38 (m, 4H), 1.37 (s, 9H). ¹³C NMR (75 MHz, CDCl₃) δ 155.93, 78.82, 41.70, 40.30, 30.77, 28.32, 27.37. HRMS (ESI) calculated mass (C₉H₂₁N₂O₂) [M+H]¹⁺: 189.1603, mass found *m/z*: 189.1601 [M+H]¹⁺.



Scheme 2.4
Synthesis of **NTrp**

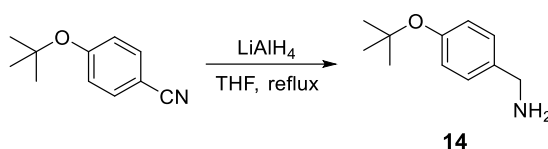
Synthesis of N-(2-(1H-indol-3-yl)ethyl)-2,2,2-trifluoroacetamide (11). 19.99 g (124.8 mmol) of tryptamine was dissolved in DCM (300 mL) followed by the addition of 11.1 mL (1.1 equiv.) of pyridine. This solution was cooled to 0°C and 19.4 mL (1.1 equiv.) of trifluoroacetic anhydride

was added dropwise. After 22 hours of reaction time, all of the starting material had been consumed, as evidenced by TLC; the reaction mixture was washed with 2N HCl (3 x 250 mL), water (100 mL) and brine (100 mL), dried over sodium sulfate and evaporated to give a brown solid. This solid was dissolved in a mixture of acetone and DCM and was absorbed onto silica gel. This was dry loaded into an empty flash column, and the product was purified via normal phase flash chromatography (EPCLC W-Prep 2XY, Yamazen Corp.) using an increasing solvent gradient of ethyl acetate/hexanes (20:80 to 100:0 over 60 minutes). The fractions containing the desired product were combined and evaporated to give 24.1 g (75.3% yield) of **11** as a white solid. ¹H NMR (300 MHz, DMSO) δ 10.86 (s, 1H, NH), 9.55 (t, *J* = 5.56 Hz, 1H), 7.53 (d, *J* = 7.72 Hz, 1H), 7.34 (dd, *J* = 1.05, 7.99 Hz, 1H), 7.16 (d, *J* = 2.35 Hz, 1H), 7.07 (td, *J* = 1.25, 7.55, 8.09 Hz, 1H), 6.98 (ddd, *J* = 1.13, 7.10, 7.90 Hz, 1H), 3.45 (q, *J* = 6.82 Hz, 2H).

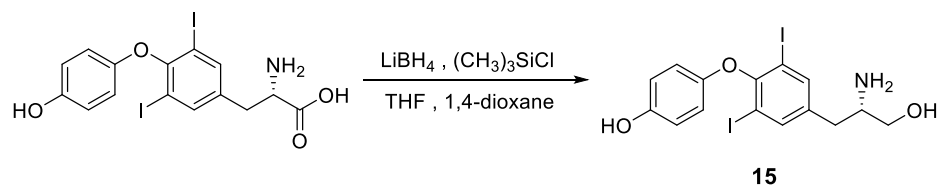
Synthesis of tert-butyl 3-(2-(2,2,2-trifluoroacetamido)ethyl)-1H-indole-1-carboxylate (12). 24.0 g (93.7 mmol) of **11** was dissolved in THF (200 mL), followed by 30.76 g (1.5 equiv.) of di-tert-butyl dicarbonate with an additional 50 mL of THF to wash down the sides of the flask. 0.59 g (0.052 equiv.) of DMAP was then added and the reaction was heated to 40°C for two hours. At that time, TLC showed that all of the starting material had been consumed, so 250 mL of DCM was added to the reaction and the organic solution was washed with water (2 x 100 mL), dried over sodium sulfate and evaporated in vacuo to give a viscous brown oil. The product was purified via normal phase flash chromatography (EPCLC W-Prep 2XY, Yamazen Corp.) using an increasing gradient of ethyl acetate/hexanes (10:90 to 100:0 over 100 minutes). The fractions containing the desired product were combined and evaporated in vacuo to give 17.27 g (51.7% yield) of **12** as a white solid. ¹H NMR (300 MHz, DMSO) δ 9.57 (t, *J* = 5.78 Hz, 1H), 8.05 (d, *J* = 8.20 Hz, 1H), 7.62 (d, *J* = 7.48 Hz, 1H), 7.50 (s, 1H), 7.33 (td, *J* = 1.39, 7.77, 8.28 Hz, 1H), 7.25 (td, *J* = 1.17, 7.44 Hz, 1H), 3.49 (q, *J* = 6.72 Hz, 2H), 2.91 (t, *J* = 6.98 Hz, 2H), 1.61 (s, 9H). ¹³C NMR (75 MHz, DMSO) δ 156.51, 156.03, 149.02, 134.78, 130.13, 124.42, 123.18, 122.52, 119.07, 117.40, 114.74, 83.48, 39.07, 27.67, 23.53.

Synthesis of tert-butyl 3-(2-aminoethyl)-1H-indole-1-carboxylate (13; "NTrp"). 17.20 g (48.27 mmol) of **12** was dissolved in methanol (60 mL) followed by the addition of 16.75 g (2.511 equiv.) of potassium carbonate dissolved in water (60 mL). The reaction flask was covered and the reaction was allowed to stir overnight at room temperature. After 16 hours of reaction time, little progress was seen with the reaction, so it was heated to reflux for seven hours. At that time, TLC showed

that all starting material had been consumed. Methanol was evaporated in vacuo, and the remaining aqueous solution was adjusted to pH >10 with 4N NaOH. The product was extracted with DCM (3 x 200 mL), and the combined organics were washed with water (100 mL) and brine (100 mL), dried over sodium sulfate, filtered and evaporated in vacuo to give 7.05 g (56.1% yield) of **13** as a yellow oil. ¹H NMR (300 MHz, CDCl₃) δ 7.60 – 7.46 (m, 1H), 7.46 – 7.36 (m, 1H), 7.34 – 7.17 (m, 3H), 3.54 (t, *J* = 7.74 Hz, 2H), 3.03 (t, *J* = 7.11 Hz, 2H), 1.66 (s, 9H), 1.64 (s, 2H, NH₂). ¹³C NMR (75 MHz, CDCl₃) δ 167.81, 135.55, 130.74, 124.17, 123.05, 122.72, 122.26, 118.98, 115.15, 83.27, 51.28, 28.17, 26.50. LRMS (ESI) calculated mass (C₁₅H₂₁N₂O₂) [M+H]¹⁺: 261.1, mass found *m/z*: 261.1.



Synthesis of 4-(tert-butoxy)phenylmethanamine (14; “NTyr”). 175 mL of 1 M (175 mmol; 2.6 eq.) lithium aluminum hydride in THF was added to a round bottom flask with a stir bar and was cooled to 0°C. A solution of 11.83 g (67.5 mmol) of 4-(tert-butoxy)benzonitrile in 50 mL of anhydrous THF was added to the stirring solution dropwise over the course of 30 minutes. The reaction was then fitted with a reflux condenser and was heated to reflux for six hours, followed by stirring overnight under argon, allowing the reaction to cool to room temperature. After 22 hours of total reaction time, the reaction mixture was cooled to 0°C and was quenched with 7 mL of water, followed by 6 mL of 15% NaOH (aq) and an additional 17 mL of water. The resulting emulsion was filtered over celite, with the filter cake being washed with methanol (2 x 50 mL) and DCM (2 x 50 mL). The filtrate was evaporated, and the resulting dark yellow oil was dissolved into 75 mL of water. The solution was transferred to a separatory funnel and was extracted with ethyl acetate (4 x 150 mL). The combined extractions were washed with water (100 mL) and brine (100 mL), dried over sodium sulfate, and evaporated in vacuo to give a dark yellow oil. The oil was separated using basic alumina chromatography and a solvent system of ethyl acetate/hexanes (20:80 to 50:50). The fractions containing the product (as evidenced by TLC) were combined and evaporated to give 5.71 g (47.2% yield) of **14** as a clear oil. ¹H NMR (300 MHz, CDCl₃) δ 7.19 (dd, *J* = 2.26, 6.46 Hz, 2H), 6.94 (dd, *J* = 2.32, 6.46 Hz, 2H), 4.38 (s, 2H), 1.31 (s, 9H). ¹³C NMR (75 MHz, CDCl₃) δ 153.80, 135.32, 128.26, 124.27, 78.17, 54.93, 28.75. LRMS (EI) calculated mass (C₁₁H₁₇NO): 179.26, mass found *m/z*: 179.



Synthesis of (S)-4-(4-(2-amino-3-hydroxypropyl)-2,6-diiodophenoxy)phenol (15; “T2AA”). 7.55 mL of 1 M (15.1 mmol) lithium borohydride in THF was added to a 50 mL round bottom flask followed by 5.2 mL of anhydrous THF and 5 mL of anhydrous dioxane. The solution was cooled to 0°C under argon, and 3.85 mL (30.3 mmol) of chlorotrimethylsilane was slowly added. The resulting solution was stirred for 15 minutes at 0°C, and 0.90 g (1.71 mmol) of 3,5-diiodo-L-tyrosine (Combi-Blocks) was added in one portion with the aid of an additional 5.2 mL of anhydrous THF and 5 mL of anhydrous dioxane. The flask was sealed and the reaction was stirred overnight under argon while being allowed to slowly warm to room temperature. After 18 hours of reaction time, the reaction was poured into 25 mL of ice-water, adjusted to pH >9 with 4 N NaOH and was extracted with ethyl acetate (3 x 50 mL). The combined organic extractions were dried over sodium sulfate, filtered and evaporated to give a light brown solid. This solid was dissolved in a mixture of ACN/H₂O (75:25) and was purified on an Agilent ZORBAX SB-C18 reverse phase semi-preparative column on a System Gold 166 (Beckman Coulter) HPLC system, using a gradient of ACN (0.1% TFA)/H₂O (0.1% TFA) 0:100 to 100:0 over 30 minutes with detection at 254 nm. The fractions containing the purified product were combined, frozen and lyophilized to give 460 mg (52.5% yield) of **15** as a fluffy white solid. ¹H NMR (300 MHz, DMSO) δ 9.12 (s, 1H), 7.84 (s, 5H), 6.68 (dt, *J* = 2.30, 3.84, 9.08 Hz, 2H), 6.54 (dt, *J* = 2.30, 3.61, 8.98 Hz, 2H), 5.39 (t, *J* = 4.49 Hz, 1H), 3.56 (dd, *J* = 4.86, 7.97 Hz, 1H), 3.40 (dq, *J* = 5.48, 9.94 Hz, 2H), 2.87 – 2.69 (m, 2H), 2.07 (s, 1H). ¹³C NMR (126 MHz, DMSO) δ 152.65, 152.26, 148.83, 140.67, 137.48, 115.86, 115.82, 92.48, 60.07, 53.27, 40.02, 39.85, 39.78, 39.69, 39.52, 39.35, 39.19, 39.02, 33.06. HRMS (ESI) calculated mass (C₁₅H₁₆I₂NO₃) [M+H]⁺: 511.9220, mass found *m/z*: 511.9678 [M+H]⁺.

2.6 Results

Various applications for trimeric peptoids have already been developed [302,305,318–322,333]. Three key regions at the PIP Box binding site are important for the recognition and binding of PIP Box-containing peptides/proteins [177]. It was predicted that tripeptoids would be sufficient in size to block these key areas. Furthermore, they fall within the general mass range of other inhibitors of

PPIs, and they are small enough to enhance the ease of synthesis while lowering the cost and time to completion. The general approach taken in this project represents a situation in which resources are limited in both materials and time, and should be applicable to other situations where rapid reduction of the size of a library to be synthesized is desired. In this way, this approach provides a unique pathway to fragment-based drug discovery as multiple fragments are “tied” together in the peptoid backbone, and are thus simultaneously screened together.

2.6.1 Preferences for Fragments along the Peptoid Backbone

As outlined in section 2.5.1, computational screens were used to drive the synthetic approach for peptoid library creation. From the results of the screens, as expected, slightly different results were observed for each crystal structure of PCNA (Appendix A, Figure A.1), and different results were seen for each substitution position along the peptoid backbone (Figure 2.5). The 1st (N-terminal) position generally showed a relatively limited preference for a set of several peptoid side chains including NLys, NArg, NTyr, NGln, NEba and NBal. Six fragments were present in the 1st position in at least 7% of the top hit list, which itself consisted of 200 compounds that were compiled from the top 50 hits for each of the four PCNA crystal structures that were screened. The two fragments

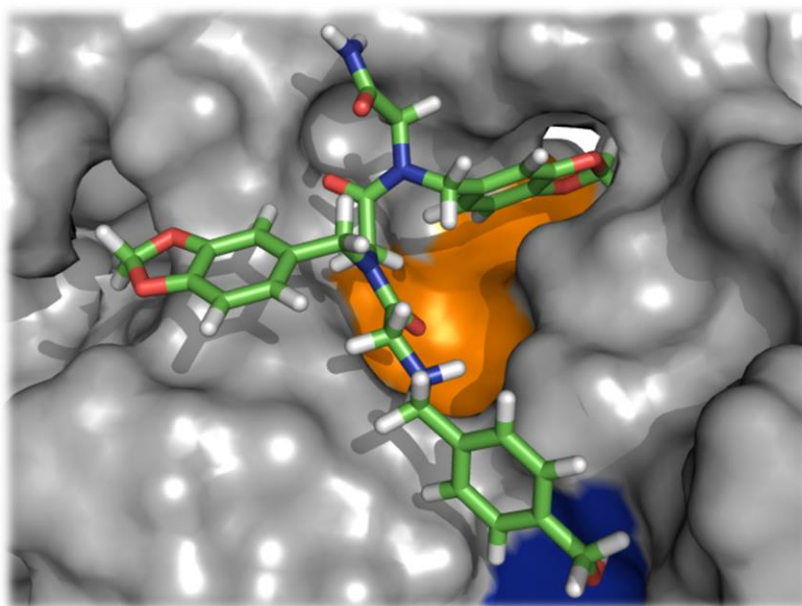


Figure 2.4. Tripeptoids Dock at the PCNA PIP Box Binding Site. The combinatorial tripeptoid library was screened against four crystal structures of PCNA. The top hit of the screen against structure 3VKX (PCNA bound to T3) is shown above. The hydrophobic site is displayed in orange, and the PIP Box glutamine binding site is displayed in blue.

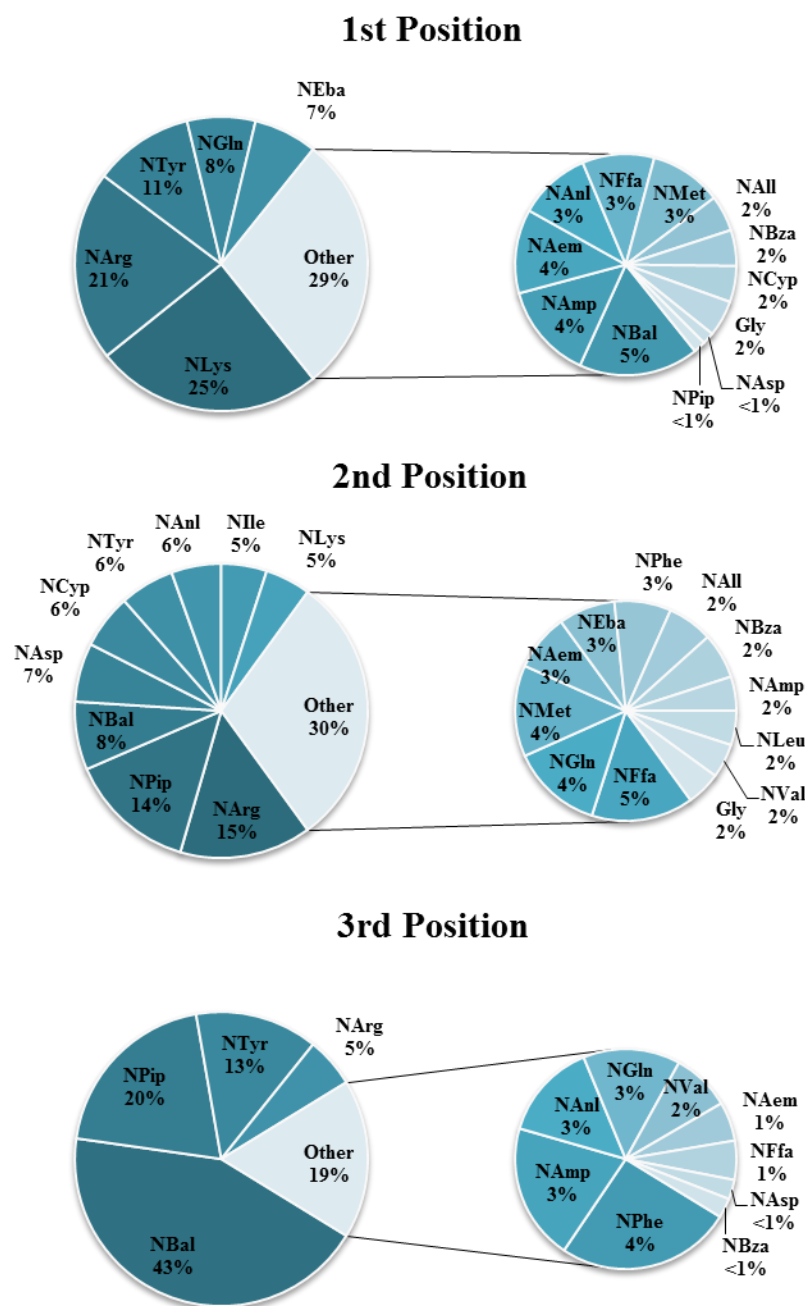


Figure 2.5. Percentage that Amine Fragments in the Initial Library Appeared in the Tripeptide Backbone. The top 50 hits from docking with each of the four PCNA crystal structures were selected and compiled into a total group of 200 compounds; the number of times a specific amine fragment was present at a particular position along the peptoid backbone was tallied. The fragments are represented above as the percentage of the number of times that fragment was present in that position, among the full list of 200 compounds. “1st”, “2nd” and “3rd” positions are as defined in Figure 2.2.

that stood out most significantly were Nlys and NArg, each of which were present in over 20% of the top hit list. Similar to the 1st position, the 2nd position did not show strong preference for a limited number of fragments—rather, ten different fragments were present in the 2nd position in at least 5% of the top hit list, though NArg did stand out with it being present in 15% of the top hits. In contrast to the first two positions, the 3rd position in the peptoid backbone showed significant preference for particular side chains, notably aromatic groups that contained functionalities that allow for hydrogen bonding with protein amino acids. NBal especially stood out as 43% of all of the top hit compounds contained that fragment in the 3rd position.

Given that there were 21 possible substituents at each of the three backbone positions, there would theoretically be 9,261 total possible tripeptoids. With limited resources, this number of compounds may be impossible to synthesize, and may present a non-trivial cost. If a cutoff was set to where only fragments that appeared in 5% or more of the top hits at a given position were considered, for example, the total list of compounds would be narrowed to just 240 (6 in the 1st position x 10 in the 2nd position x 4 in the 3rd position). In principle, an even shorter total list of compounds could be produced if the percentage cutoff for significance is raised (e.g. considering only the fragments that appear in 10% or more of the top hits at a given position), and that was the general strategy pursued here where less than 100 ligands were actually synthesized.

2.6.2 Virtual Incorporation of Second Generation Tripeptoid Ligands

As previously discussed, the small molecules T3 and T2AA have been demonstrated to physically bind in the hydrophobic pocket of the PIP Box binding site [178,231]. Both compounds contain a primary amine, as well as a carboxylic acid, potentially allowing them to be incorporated into a peptoid backbone using submonomer peptoid synthesis. In doing this, either could serve as an anchoring residue, helping to direct small peptoid molecules to the PIP Box binding site on PCNA. Due to the enhanced thyroid activity and low solubility of T3, T2AA was selected as a better candidate for investigation. Using T2AA as a fragment, a virtual combinatorial library was created and virtually screened using similar methodology as described in section 2.5.1, with a larger set of 37 primary amine fragments and two forms of T2AA, itself having defined attachment points at either the primary amine (NT2AA) or carboxylic acid group (CT2AA). Due to the significant conformational difference in the PIP Box binding site between the co-crystal structures of PCNA-T3 and PCNA-peptides, only the structure of PCNA bound to T3 was used (PDB ID: 3VKX) since the hydrophobic pocket on PCNA would be able to accommodate the potential binding of T2AA.

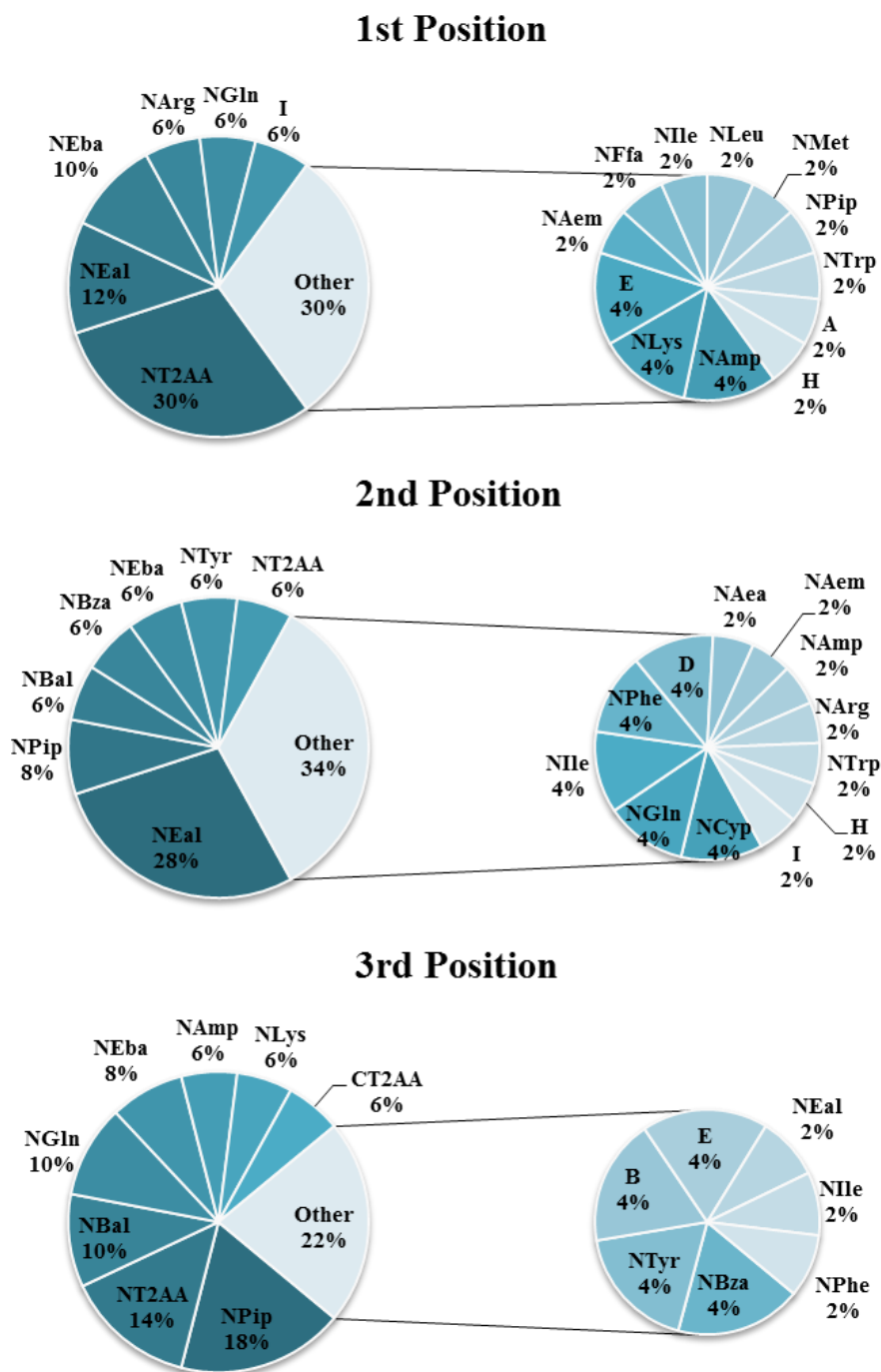


Figure 2.6. Percentage that Fragments Including T2AA Appeared in the Tripeptide Backbone. The top 50 hits from docking were selected and the number of times a specific amine fragment was present at a particular position along the peptoid backbone was tallied. The fragments are represented above as the percentage of the number of times that fragment was present in that position, among the full list of 200 compounds. “1st”, “2nd” and “3rd” positions are as defined in Figure 2.2.

From the results (Figure 2.6), the 1st and 2nd positions along the peptoid backbone showed strong preference for a small set of possible side chains, with T2AA, containing the amine attachment point in the first position, and ethyl alcohol being by far the most favored in the second position. It must be noted that overall, tripeptoids containing T2AA as a side chain docked with much more favorable scores than peptoids that did not contain T2AA; though, based on the crystal structure used for docking (PDB ID: 3VKX), PCNA was conformationally biased towards ligands that took the shape of T3. In contrast, the third position did not show much preference for particular side chains. This was presumably because, according to the docked structures, fragments in the third position tended to be quite flexible and picked up a variety of interactions in the region proximal to the PIP Box glutamine binding site. As a result, none of the third position fragments individually picked up substantial stabilizing interactions with PCNA. Since NT2AA was the most frequent fragment at the first position, it was decided that the most technically straightforward approach would be for all second generation peptoids to have T2AA in the first position. This would substantially increase the ease of synthesis since T2AA could be attached at the N-terminal end of a resin-bound synthetic peptoid without the need for chemical protecting groups.

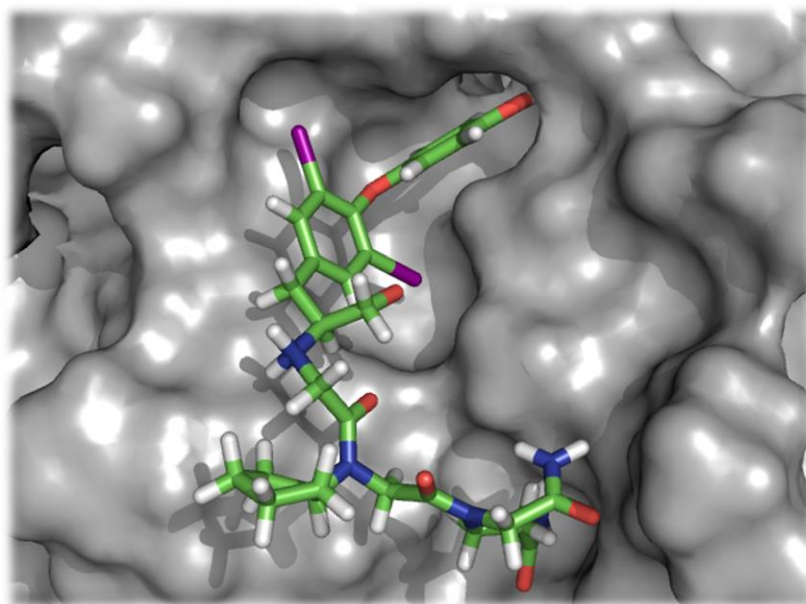


Figure 2.7. Second Generation Peptoid Ligands Dock at the PCNA PIP Box Binding Site. The combinatorial tripeptoid library containing T2AA as a fragment was screened against the crystal structure of PCNA bound to T3 (PDB ID: 3VKX). The top hit from the screen is shown above with the T2AA fragment binding in the hydrophobic pocket site on the surface of the protein.

2.6.3 *In Vitro* Screening of Synthesized Tripeptoids

Fragments that appeared in approximately at least 5% of each position on the tripeptoid backbone were selected for incorporation. Among the top hit side chains were multiple that were not commercially available, including NArg, NBal, NBza, NEba, NEal, NGLn, NLys and NTyr. Fragments that required protecting groups are shown in Appendix A, Figure A.5. Synthesis of NEba and NGLn with acid-labile protecting groups proved to be unexpectedly technically difficult, and to avoid significant costs associated with further attempts at re-synthesis, focus was placed on synthesizing tripeptoids containing the other top hit fragments. The total list of synthesized peptoids can be found in Table A.2 of Appendix A. Some ligands were also synthesized that were not predicted to be good biochemical hits for purposes of comparison to compounds that were predicted to be PCNA-PIP Box antagonists.

Upon synthesis and purification, peptoids were screened in a fluorescence polarization assay to find ligands that disrupted the interaction between His-tagged PCNA and FAM-PL. In the design of the FP assay, the conditions from Pedley *et al.* [177] were used as a starting point. However, customized conditions were needed given that the dynamic range between bound and unbound FAM-PL using 10 nM peptide and 100 nM PCNA was small, though a Z' -factor analysis [334] indicated the variability was small enough to quantitatively determine differences between bound and unbound peptide (Appendix A, Figure A.2). To find a balance of PCNA to FAM-PL that would produce a substantial dynamic range, and to determine the peptide's K_d value, increasing amounts of recombinant PCNA was added to a fixed concentration of the FAM-PL peptide (5 nM) in a two-fold dose-response fashion (Appendix A, Figure A.3). Statistical analysis by non-linear regression indicated that the dissociation constant for the peptide was 107 nM, which is similar to what has been previously reported [335]. Because more than 80% of the peptide was bound at a PCNA concentration of 1 μ M, that concentration of protein was selected for use in subsequent displacement assays. Ultimately, the protein being in that level of excess should greatly enhance the dynamic range of the assay, and would thus increase the robustness in determining which compounds are actually hits in disrupting the binding between PCNA and the PL peptide. After developing conditions under which synthesized peptoids would be screened, 69 of the initially synthesized compounds were screened by FP at concentrations of 1 mM and 250 μ M to find any general hits (Figure 2.8). T2AA was selected as the positive control since it is known to disrupt the interaction between FAM-PL and PCNA, and would be the basis for comparison against the ligands in this study.

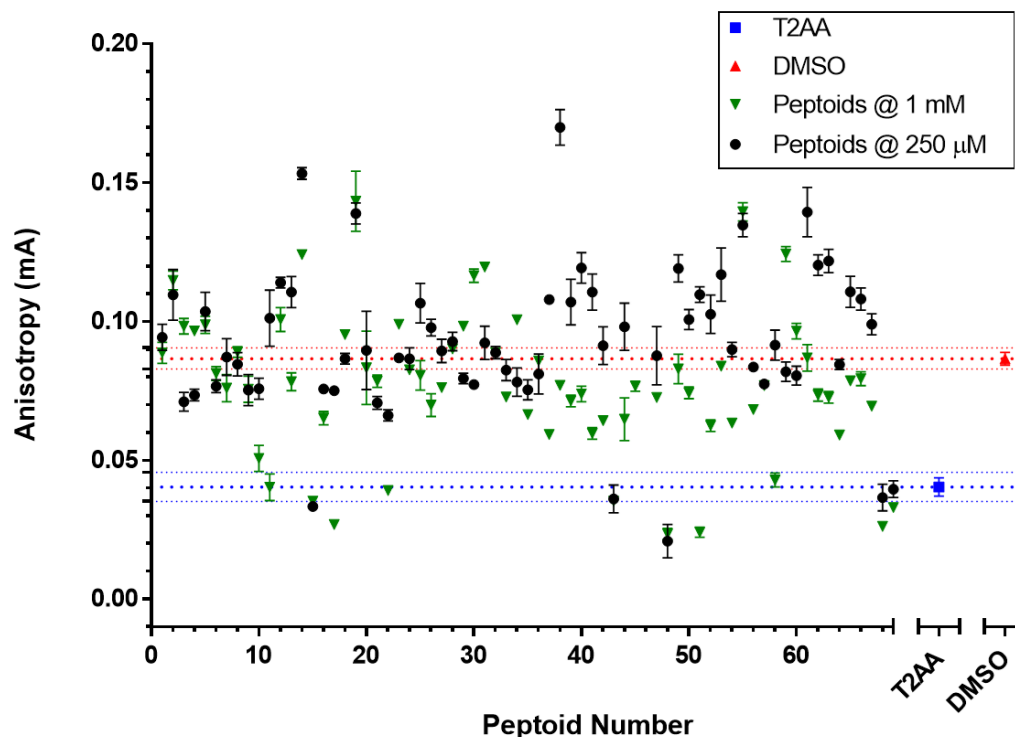


Figure 2.8. Initial Tripeptoid FP Screen. Many of the synthesized tripeptoids were screened using fluorescence polarization at initial concentrations of 1 mM and 250 μ M for their ability to disrupt the binding between PCNA and FAM-PL. 1 mM T2AA serves as the positive control, and DMSO (at the same concentration ratio of DMSO:binding buffer as the screened ligands, but with no compound) serves as the negative control for binding disruption. FAM-PL and PCNA were at 5 nM and 1 μ M, respectively. Error bars represent standard error of the mean, and dotted lines around the controls represent the 95% confidence interval ($\alpha=0.05$) for each. Lower anisotropy values indicate disrupted binding of PCNA-PL. The full list of peptoids screened in this assay, listed by peptoid number, can be found in Appendix A, Table A.3.

From the initial screen, five ligands—NLys-NPip-NBal, NLys-NTyr-NBal, NBal-NLys-NTyr, T2AA-NEal-NPip and T2AA-Gly-NPip—were identified that were able to disrupt the binding between PCNA and FAM-PL at 250 μ M with comparable activity to the positive control, T2AA. Series of two-fold dilutions of each hit (Figure 2.9) were performed to generate dose response curves. Following the screen of the initial set of compounds, additional peptoid-based ligands were synthesized that contained variations on the structure of the hits to see if inhibitory nature of these ligands could be improved. One new fragment, NMba (see Appendix A, Figure A.4), was selected to try in place of NBal due to its similarity in size and shape, but with the ability to donate an additional hydrogen bond. Dose response curves were generated for these new ligands as well (Figure 2.9). IC_{50} values were determined for each peptoid hit by performing non-linear regression fits of each dose response curve using equation 2.4. K_i values were calculated

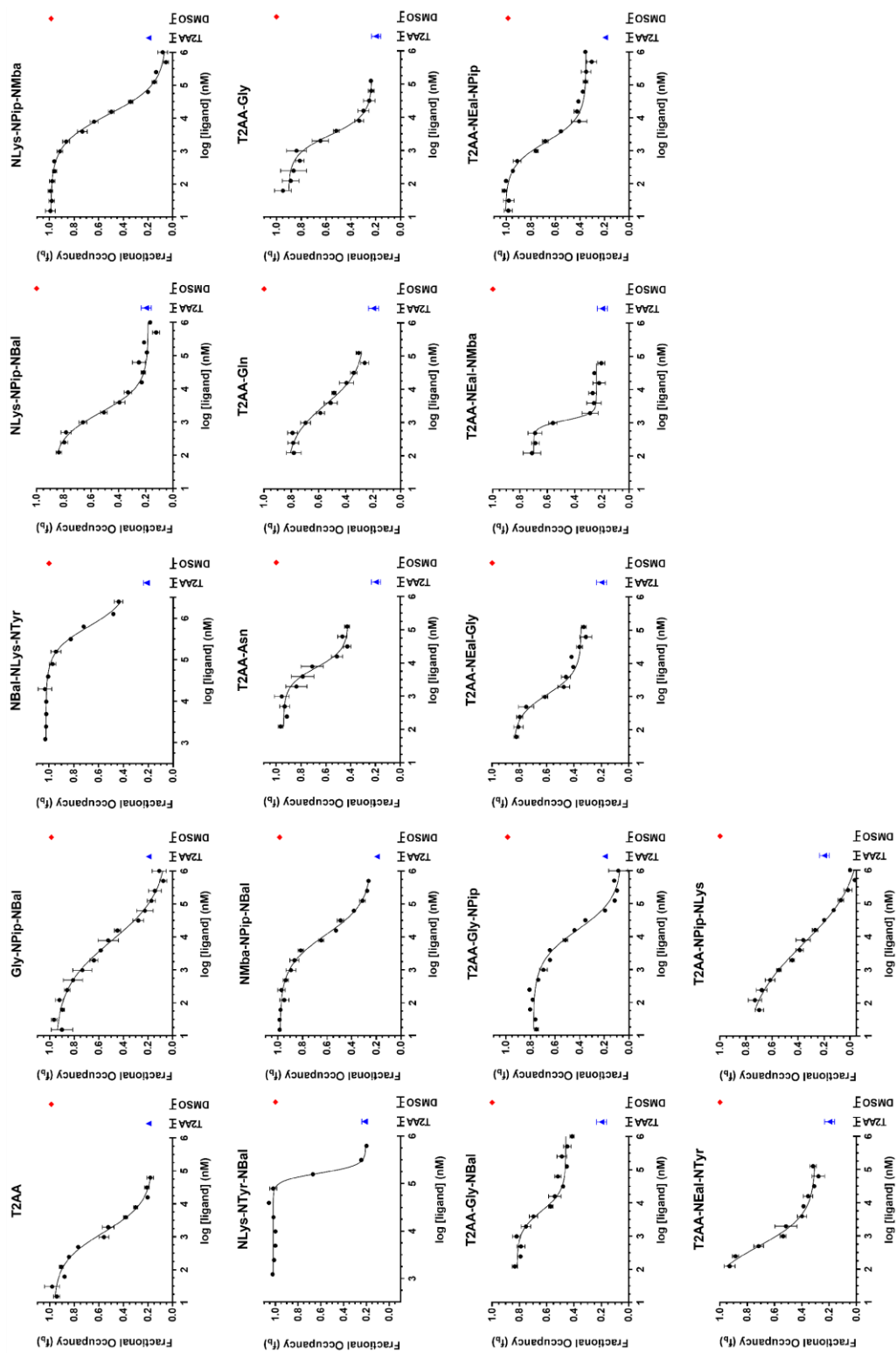


Figure 2.9. Dose Response Curves of Peptoid-Ligand Hits. Two-fold dilution series of each hit compound, in addition to T2AA, were performed to generate dose response curves. 1 mM T2AA serves as the positive control, and DMSO (at the same concentration ratio of DMSO:binding buffer as the screened ligands, but with no ligand) serves as the negative control. Error bars represent standard error of the mean.

Table 2.4
Hit Compound IC₅₀ and K_i Values Measured by Fluorescence Polarization

Compound Name	IC ₅₀ (μM) *	K _i (μM) *†
T2AA	1.34 ± 0.33	1.27 ± 0.29
Gly-NPip-NBal	7.74 ± 3.41	6.84 ± 2.20
NBal-NLys-NTyr	> 600	> 900
NLys-NPip-NBal	1.94 ± 0.51	1.78 ± 0.44
NLys-NPip-NMba	12.93 ± 1.97	12.16 ± 1.71
NLys-NTyr-NBal	~ 165	~ 200
NMba-NPip-NBal	11.69 ± 2.55	10.49 ± 1.93
T2AA-Asn	7.20 ± 2.74	7.23 ± 3.69
T2AA-Gln	3.52 ± 1.65	3.32 ± 1.37
T2AA-Gly	2.91 ± 0.91	2.70 ± 1.08
T2AA-Gly-NBal	5.66 ± 1.67	5.38 ± 1.98
T2AA-Gly-NPip	16.17 ± 3.71	15.10 ± 3.04
T2AA-NEal-Gly	1.17 ± 0.37	1.08 ± 0.39
T2AA-NEal-NMba	1.18 ± 0.24	0.833 ± 0.439
T2AA-NEal-NPip	1.82 ± 0.37	1.73 ± 0.38
T2AA-NEal-NTyr	0.482 ± 0.328	0.521 ± 0.209
T2AA-NPip-NLys	6.13 ± 2.84	6.53 ± 1.92

* Values are represented as the 95% confidence interval around the mean

† Calculated using equation 2.5

(Structures of T2AA-containing ligands can be found in Appendix A, Figure A.7)

from the resultant IC_{50} values using equation 2.5 (Table 2.4). As a control experiment, individual fragments that make up the top hits, including N_GIn, N_Lys, N_Pip, N_Tyr, N_Bal, N_Bza and N_Eal, were screened in the FP assay to determine whether they were individually capable of disrupting PCNA-PL peptide binding (Appendix A, Figure A.6). However, none of the fragments showed any evidence of inhibition.

2.6.4 Identification of Molecular Recognition Features of PCNA for Peptoid Ligands

In an effort to understand better the molecular features that may be driving the affinity of the peptoids for PCNA, molecular dynamics simulations (MDs) were performed with each of the top five initial hits—N_Lys-N_Pip-N_Bal, N_Lys-N_Tyr-N_Bal, N_Bal-N_Lys-N_Tyr, T2AA-N_Eal-N_Pip and T2AA-Gly-N_Pip—in complex with PCNA. As discussed in section 2.5.2, each ligand was first docked into the co-crystal structure of PCNA and PL-peptide (PDB ID: 1VYJ; Appendix A, Figure A.8), with the peptide removed, using the Glide induced-fit model in Maestro. This was done to give approximate starting points for the MDs, as well as to ensure that there were no conflicts on an atomic scale due to Van der Waals clashing or unfavorable ionic contacts. Each simulation was run on a 5.0 ns scale, or until it converged, as judged by the change in protein C α and side chain RMSD over time (see Appendix A, Figures A.9-A.13). Upon completion, simulation trajectories were first aligned to the first frame of their own simulation, and then the simulations were aligned to one another based on the position of their C α atoms using VMD. The final fifty frames for each MD were averaged to give an average final structure for each PCNA-ligand complex (Figure 2.10). Upon analysis, it was clear that each of the resulting complexes differed significantly from the co-crystal structure of PCNA-PL. It has been previously reported that PCNA is quite flexible and can adopt a variety of conformations to optimize ligand binding [177]. The results here suggest that this principle is conserved given the large differences in the PIP Box binding region on the protein between each MD. Regions on PCNA that appeared to drive the conformational difference between each structure most substantially were between residues 80-86, 93-97, 104-111, 117-136, 162-166, 172-177, 181-194 and 251-257 (Appendix A, Figure A.14). Perhaps unsurprisingly, each of these regions was found in either a β -turn or unordered loop structure, owing to the natural flexibility of these secondary structure protein moieties. The PIP Box binding site itself is surrounded by four distinct flexible regions comprised of β -turn residues 40-46, the disordered interdomain connecting loop residues 117-136, β -turn residues 229-235 and disordered loop residues 251-257. It would make sense that some of the most significant drivers of PCNA conformational change would be

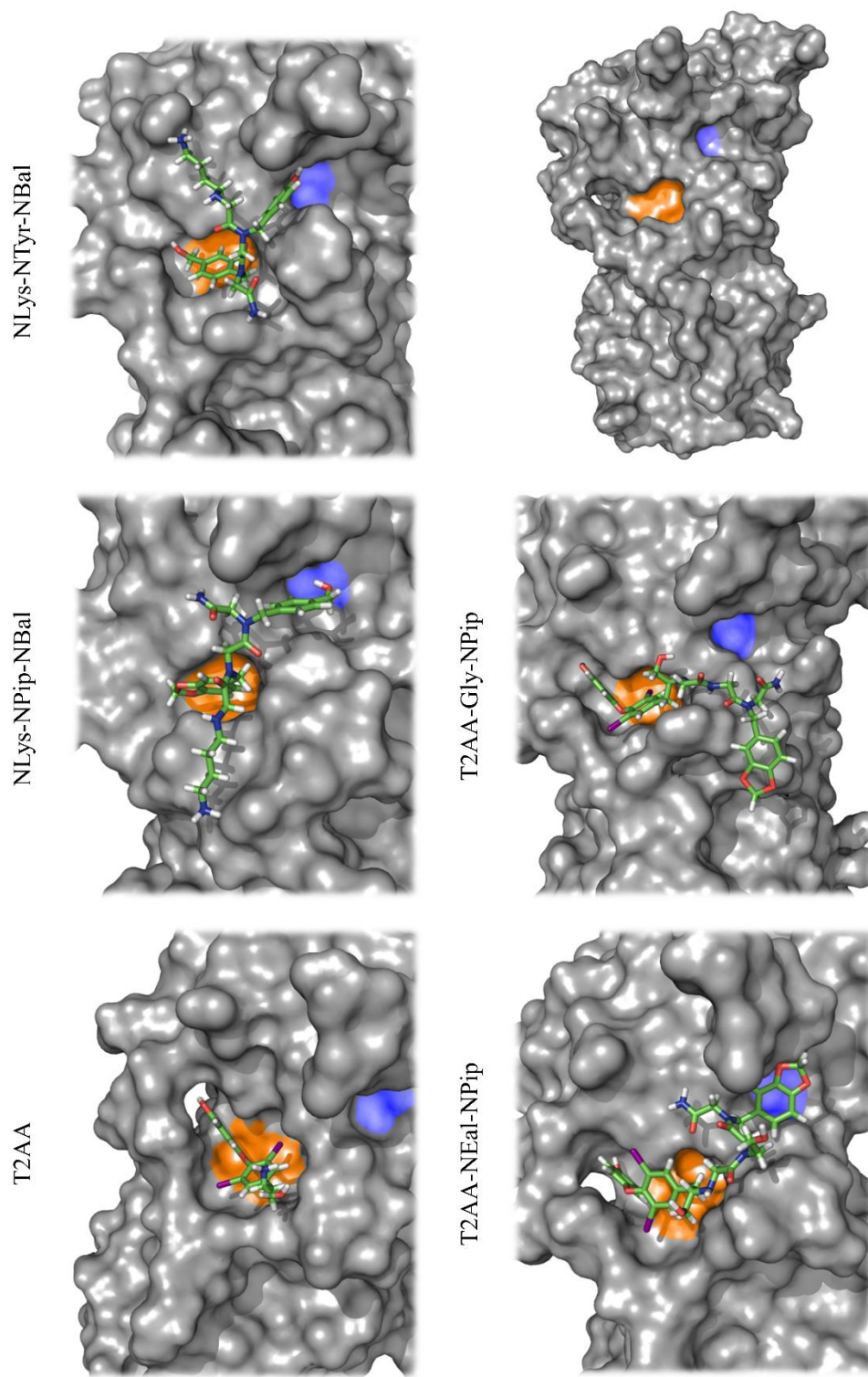


Figure 2.10. Average Structures of the Final 50 Frames of Molecular Dynamic Simulations. The final 50 simulation trajectory frames for each analyzed peptoid ligand were averaged using VMD, and the resulting structures visualized using Pymol. The hydrophobic pocket (*orange*) and PIP Box glutamine binding site (*blue*) are highlighted in each.

found in these regions.. Of the protein residues that looked to be the most important for direct interaction with each peptoid ligand, His44, Pro129, Pro234, Ala252, Pro253 and Ile255 each interacted with each peptoid ligand to various significant degrees in each of the MDs (Appendix A, Figure A.15). Almost all of the significant contacts were shared with T2AA and the PL peptide, indicating that the inhibitors occupy many of the ‘anchoring’ contacts between PCNA and PL, enhancing their antagonistic activity.

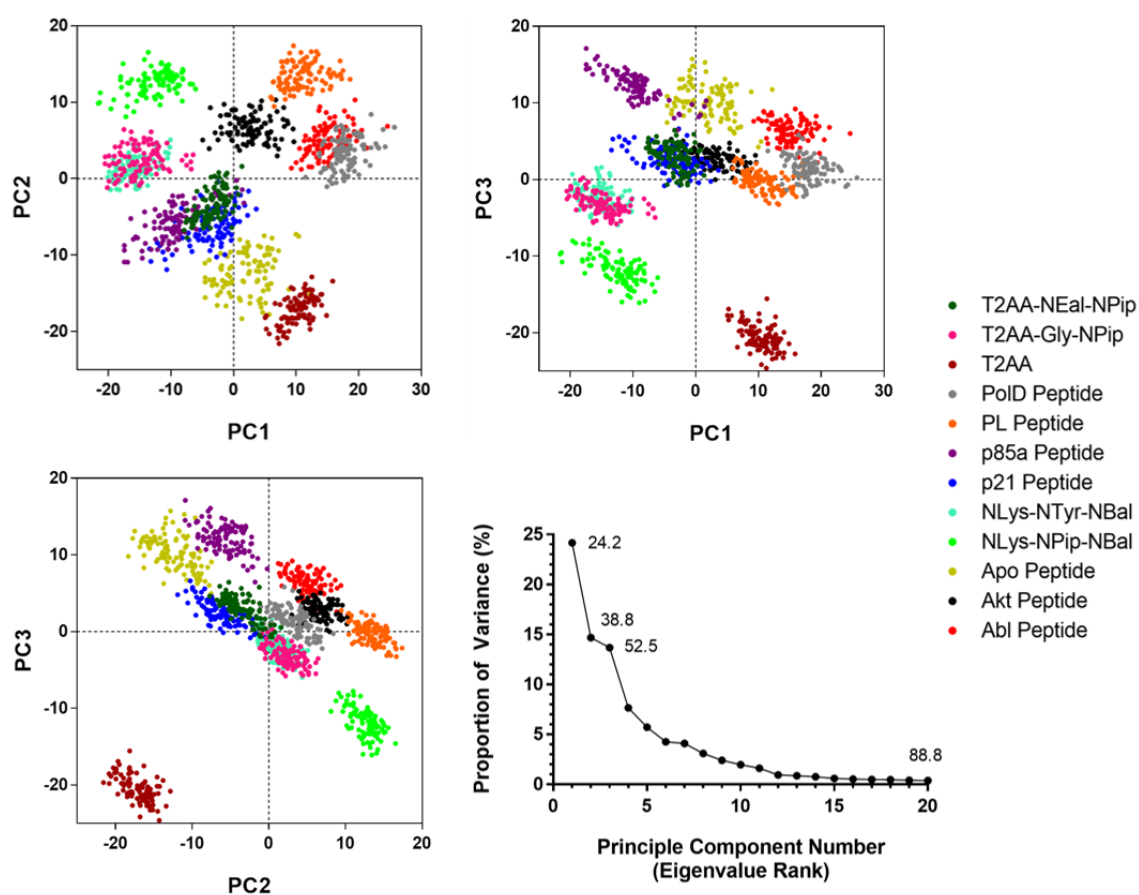


Figure 2.11. Principle Component Analysis of PCNA Topology Variance. The final 100 frames of each PCNA-ligand MD trajectory were aligned based on the position of the PCNA C α backbone atoms. A principle component analysis of the aligned trajectories shows differential clustering of PCNA conformations (residues 1-257) when in complex with either a peptoid-based ligand or a PIP Box-containing peptide. Principle components 1-3 (PC1, PC2 and PC3) for each structure were clustered and plotted along with the proportion of variance for each principle component.

To further visualize the differences between each MD output structure, a principal component analysis (PCA) of the trajectory snapshots for the C α atoms of each PCNA-ligand complex was performed (Figure 2.11). In addition to the outputs from the MDs performed here, trajectories of PCNA bound to various known binding peptides, as used in the study by Pedley, *et al.* 2014 [177], were included in the PCA for comparison. In this additional set were trajectories of PCNA in complex with the DNA polymerase δ , PL, p85- α , p21, Apo, Akt or Abl peptide. In a PCA, the principle components, which are themselves orthogonal eigenvectors, describe the axes of maximal variance of the distribution of structures. The percentage of variance of the fluctuation of protein atom position in each dimension is characterized by a corresponding eigenvalue. By clustering structures in principle component space, one can focus on the relationships between different structures in terms of their major structural displacements. In the context of this work, clustering along principle components 1, 2 and 3 allows for the comparison of the significant structural differences between each conformation of PCNA that covers more than 50% of their conformational variance (Figure 2.11, bottom right panel).

From the results of the clustering analysis, as predicted, the PCA indicated distinct differences in the topology of the PCNA-ligand interaction sites. Of all the structures, PCNA-T2AA was the most different from the rest, likely due to the fact that the inhibitor is much smaller than the peptoid-based compounds and does not project outside of the hydrophobic pocket. Although the structures were mostly separated from one another, there were some similarities in the eigenvectors. For example, though PCNA-PL and PCNA-NLys-NPip-NBal were well separated in clustering space, they both had nearly equivalent second principle components. Likewise, the population distribution of PCNA-PL and PCNA-T2AA had nearly equivalent first principle components. This would indicate that these conformations are very similar in certain dimensions. The potential implications for predicting inhibitory efficiency from this PCA are not yet clear, but this information could be useful for understanding how a ligand interacts with PCNA and how it could be grouped with other similar compounds.

Analysis of the molecular dynamic simulations indicates that the peptoid inhibitors are active due to their ability to disrupt key interactions between PCNA and the PL peptide (Figure 2.12). While in the most general sense they provide a geometric hindrance to PIP Box binding, it is also significant that these compounds prevent PCNA from forming important contacts with PIP Box residues. Computational results from Pedley, *et al.* (2014) suggest that the conserved amino acids of the PIP Box—glutamine in position 1, a hydrophobic residue in position 4 and aromatic residues in positions 7 and 8—act as anchoring residues that drive conformational stability of the

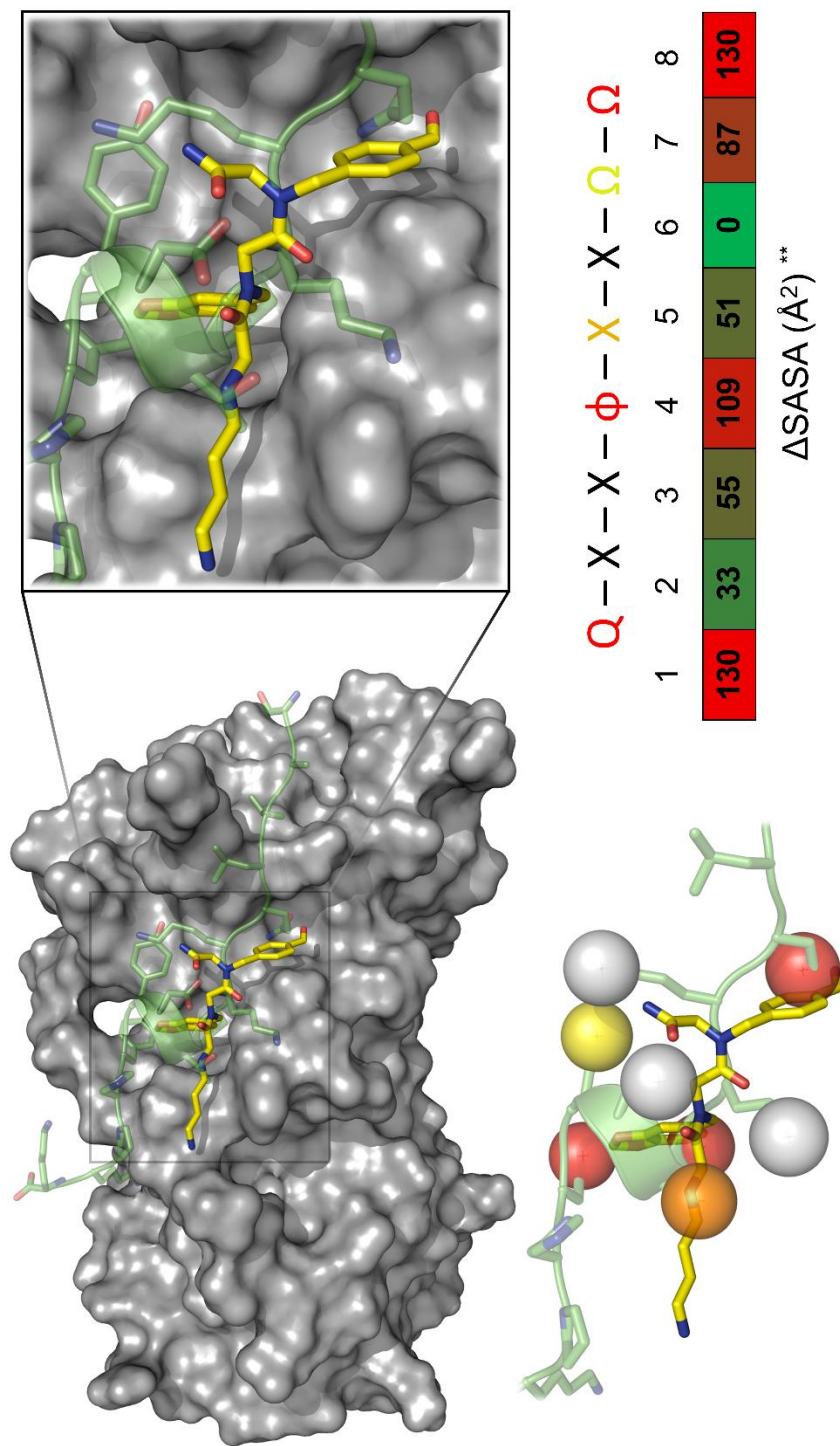


Figure 2.12. Peptoid Inhibitors Disrupt Key PCNA-PIP Box Interactions. (top) The competitive peptoid inhibitors, such as Nlys-NPip-NBal (yellow sticks), are projected to bind at the PIP Box binding site on PCNA (gray surface; PDB ID: 1VYJ), which overlaps with the PL peptide (green cartoon and sticks). (bottom left) Nlys-NPip-NBal overlays key contact points of the PL peptide's PIP Box. Spheres represent PIP Box amino acid residues. Colors indicate direct disruption of key (red) or non-key (orange) residues, and non-disruption of key (yellow) or non-key (white) residues. (bottom right) Results from Pedley, *et al* (2014) demonstrating that residues 1, 4, 7 and 8 of the PL peptide's PIP Box act as anchoring residues. ** Changes in surface accessible surface area (SASA) were calculated with ANCHOR, measuring the differences between bound and unbound forms of the PL peptide.

complex between PCNA and PIP Box containing peptides/proteins. Disrupting these points of contact would substantially weaken their interaction, and would effectively abolish binding.

2.6.5 Analysis of the Diversity and Chemical Classification of Peptoid Inhibitors

The original goal of the design of the tripeptoid library was to incorporate a set of fragments that was reasonably chemically diverse so as to cover a sufficient screening space. To measure the degree of dissimilarity, a clustering analysis was performed using Canvas [336] in Maestro (Figure 2.13) with the molecules shown in Figure 2.3, with the exception of the two variants of T2AA. Fragments were first minimized using the OPLS2005 force field, and similarity was determined with 64-bit precision in a linear fashion based on ring size (if present), aromaticity, hydrogen bond donor/acceptor, ionization potential, whether terminal (in reference to functional group placement),

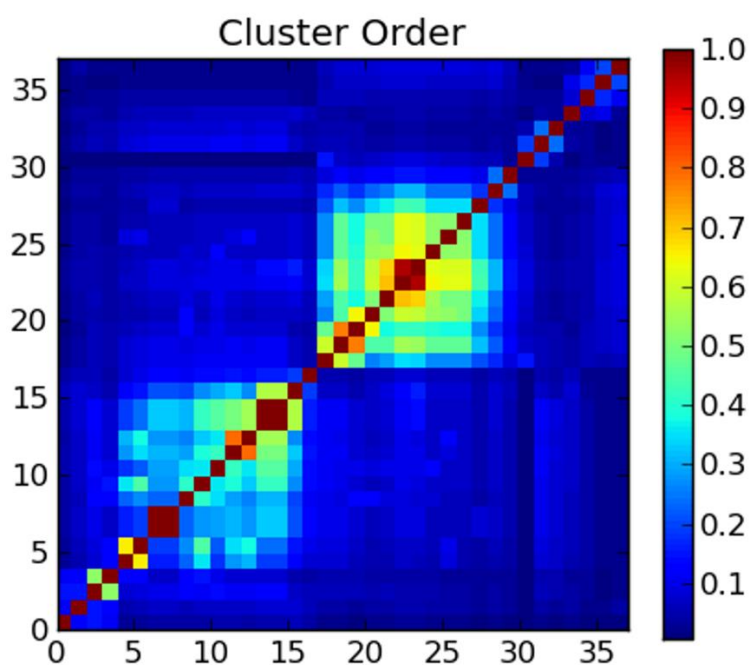


Figure 2.13. Similarity Clustering Analysis of Peptoid Ligands. A similarity clustering analysis was performed for primary amine peptoid fragments using Canvas in the Schrödinger software suite. Fragments were sorted and compared based on chemical features such as aromaticity, ionization potential and the ability to form hydrogen bonds. Similarity is represented by a color gradient with dark red being most similar, and dark blue being least.

whether halogen and bond order. The Tanimoto similarity metric was used with an average linkage method to calculate clustering. The clustering analysis indicated that there was a reasonable separation in the chemical space, with there only being two small groupings of relative similarity. This was due to the fact that a number of fragments contained a single aromatic ring, and a grouping of others were linear. While there was a larger degree of diversity that could have been achieved, this set of peptoid starting materials looked to be sufficient as a proof of concept for this study.

As discussed, inhibitors of protein-protein interactions have characteristics that distinguish themselves from other traditional inhibitors in that they display such features as higher molecular weight, higher hydrophobicity and a larger number of aromatic rings [239,240]. iPPIs also demonstrate higher degrees of globularity, lower distribution of hydrophilic regions, smaller proportions of exposed hydrophilic regions and stronger capacities to bind hydrophobic patches at the core of protein-protein interfaces as compared to inhibitors of classical targets such as enzymes [210]. Though peptoid-like molecules have been demonstrated to disrupt PPIs (Table 2.3), it was not clear whether the library of compounds generated for this study would be predicted to be classified as iPPIs *prima facie*. All of the fragments shown in Figure 2.3, including both variants of T2AA, but not NVal due to its accidental omission, were used to create a combinatorial set of tripeptoids in Schrödinger as before to give 54,862 total ligands.

The peptoids were characterized by implementing a Bayesian classifier method based on the original findings by Morelli, X. *et al* (2011) [240]. The Bayesian categorization model is a simple probabilistic classification model that is based on Bayes' theorem:

$$P(h|d) = \frac{P(d|h)P(h)}{P(d)} \quad (2.6)$$

where h is the hypothesis or model, d is the observed data, $P(h)$ is the prior belief (probability of hypothesis h before observing any data), $P(d)$ is the data evidence (marginal probability of the data), $P(d|h)$ is the likelihood (probability of data d if hypothesis h is true) and $P(h|d)$ is the posterior probability (probability of hypothesis h being true given the observed data d).

Bayesian statistics not only considers the likelihood of a model, it also takes into consideration the complexity of the model. As a result, it automatically picks the simplest model that can explain the observed data, and thus prevents overfitting. In this implementation of the Bayesian modeling, the

learned models are created with a learn-by-example paradigm: the sample data that is of interest (good) is marked, and then the system learns to distinguish them from background data. No tuning parameters are required beyond the selection of the input descriptors from which to learn. The learning process generates a large set of Boolean features from the input descriptors. It then collects the frequency of occurrence of each feature in the good subset and in all data samples. To apply the model to a particular sample, the features of the sample are generated, and a weight is calculated for each feature using a Laplacian-adjusted probability estimate. The weights are summed to provide a probability estimate, which is a relative predictor of the likelihood of that sample being from the good subset. Ultimately, Bayesian categorization was used because it can process large amounts of data, learns fast, and is tolerant of random noise [337–339]. The Laplacian-corrected estimator is used to adjust the uncorrected probability estimate of a feature to account for the different sampling frequencies of different features. The derivation is given below:

- Assume that N samples are available for training, of which, M are active. An estimate of the baseline probability of a randomly chosen sample being active, $P(\text{Active})$, is M/N .
- Next, assume that feature F is contained in B samples, and that A of those samples are active. The uncorrected estimate of activity $P(\text{Active}/F)$ is A/B . As B becomes small, this estimator tends to be less reliable. For example, if $A = 1$ and $B = 1$, $P(\text{Active}/F)$ would be 1 (active), which seems overconfident for a feature that has been seen once. Most likely, the estimator is poor because the feature is not sampled adequately, and further sampling of that feature would improve the estimate.
- It is possible to estimate the effect of further sampling by assuming that the vast majority of features have no relationship with activity. That is, if, for most features, F_i , $P(\text{Active}/F_i)$ is expected to be equal to the baseline probability $P(\text{Active})$.
- If feature K is sampled additional times, $P(\text{Active}) * K$ of those new samples are expected to be active. This provides the information needed to estimate the corrective effect of K additional samples: $P_{\text{corr}}(\text{Active}/F) = (A + P(\text{Active}) * K) / (B + K)$.
- For $K = 1/P(\text{Active})$, this is the Laplacian correction. This stabilizes the estimator: as the number of samples, B , containing a feature approaches zero, the feature's probability contribution converges to $P(\text{Active})$, which is the expected value for most features.
- The final step is to make the estimator a relative estimate by dividing by $P(\text{Active})$:

$$P_{\text{final}}(\text{Active}/F) = P_{\text{corr}}(\text{Active}/F) / P(\text{Active})$$

- For most features, $\log P_{final} \sim 0$. For features more common in actives, $\log P_{final} > 0$. For features less common in actives, $\log P_{final} < 0$. The completed estimate for a particular sample is derived by adding together the $\log P_{final}$ values for all the features present in that sample.

To perform the Bayesian classification, the 2P2I Hunter data set [240,340,341], which is a library of molecules that contains 40 known iPPIs and 1018 small molecules that are not inhibitors of PPIs, was first obtained. All 1058 compounds in that set were converted from 2D to 3D and minimized using the Accelrys Discovery Studio 4.1 Visualizer [342]. Next, four descriptors were calculated for each compound—globularity, CW2, EDmin3 and IW4 (for viewing the data, see the file “2P2I Hunter Active Decoys” on the attached DVD at the end of the thesis). These descriptors measure the following factors in a respective manner: 1) three-dimensional shape globularity; 2) ratio between the surface of the hydrophilic regions calculated at -0.5 kcal/mol and the total molecular surface (it is proportional to the concentration of hydrophilic regions [involved in weak potential polar interactions] compared to the total surface area); 3) third lowest local minimum of the interaction energy (in kcal/mol) of a dry probe (it measures the potential interaction energy of the ligand with a hydrophobic object); 4) unbalance between the center of mass of a molecule and the barycenter of its hydrophilic (IW) interacting regions (a high integrity moment is a clear concentration of hydrophilic interacting regions at one extremity of the compound).

In calculating the Bayesian cutoff, each sample was left out one at a time, and a model built using the results of the samples, with that model used to predict the left-out sample. Once all the samples had predictions, a ROC plot was generated, and the area under the curve (ROC AUC) calculated (Appendix A, Figure A.16). Best Split was determined by picking the split that minimized the sum of the percent misclassified for category members and for category nonmembers, using the cross-validated score for each sample. Using that split, a contingency table was constructed, containing the number of true positives, false negatives, false positives and true negatives. Based on the resulting calculated cutoff of -0.188 , 38 out of the 40 iPPIs as well as 244 out of the 1018 non-iPPIs were predicted to be true iPPIs (file “Bayesian Result on Decoy Set” on attached DVD). With a more than 80% success rate at identifying true positives and negatives, it was determined that the model was satisfactorily accurate.

Next, this same model was applied to the library of tripeptoid ligands, and the same four descriptors were calculated for each molecule—globularity, CW2, EDmin3 and IW4. When

Table 2.5
Peptoid Descriptor Statistics from Bayesian Model

Statistic	Globularity	CW2	EDmin3	IW4
Mean	0.123096	2.19173	-2.62519	1.99723
Median	0.11273	2.204385	-2.59537	1.91726
Minimum	0.027005	1.58044	-4.37756	0.059068
Maximum	0.654219	2.76792	-1.89029	6.11639
Skew	1.691743	-0.33117	-0.6282	0.488549

Table 2.6
iPPI Descriptor Statistics from Bayesian Model

Statistic	Globularity	CW2	EDmin3	IW4
Mean	0.11331	1.953635	-2.84006	2.699039
Median	0.079983	1.92631	-2.81911	2.54485
Minimum	0.013987	1.54225	-3.30688	0.525021
Maximum	0.45662	2.37198	-2.4452	5.5549
Skew	1.613415	0.248895	-0.48162	0.282107

Table 2.7
Non-iPPI Descriptor Statistics from Bayesian Model

Statistic	Globularity	CW2	EDmin3	IW4
Mean	0.055457	2.082103	-2.48144	2.579951
Median	0.037349	2.05015	-2.45079	2.440295
Minimum	0	1.26745	-3.51407	0
Maximum	0.356383	3.04432	-1.70217	8.53636
Skew	1.820263	0.377792	-0.55035	0.525307

considering the entire set of the tripeptoids, both those that are predicted to be iPPIs and those that are not, the statistics for the distribution of descriptor scores (Table 2.5; see Appendix A, Figure A.17 for population distributions for each descriptor) indicate that those molecules are much more similar in globularity to other known iPPIs (Table 2.6) than non-iPPIs (Table 2.7). For CW2, the tripeptoids had a higher score than either iPPIs or non-iPPIs, likely due to the fact that these peptoids, on average, have more exposed hydrophilic regions than what would be expected for classical drugs. The peptoids were not able to be strongly associated with either iPPIs or non-iPPIs

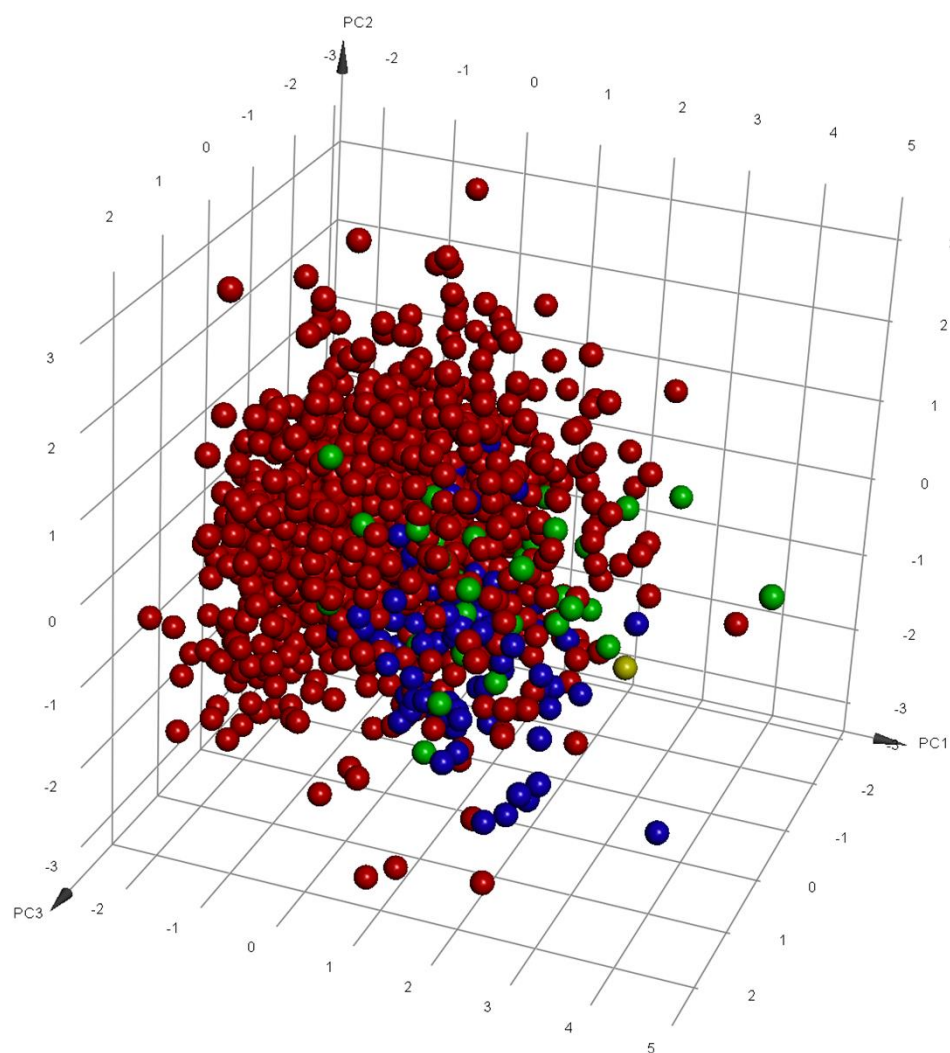


Figure 2.14. Ligand Classification Based On Principle Component Analysis of Descriptors. A principle component analysis of the four descriptors, globularity, CW2, EDmin3 and IW4, was performed on the combined set of iPPIs and non-iPPIs from the 2P2I Hunter collection and 100 randomly selected tripeptoids that were predicted to be iPPIs. Red spheres indicate known non-iPPIs, green spheres indicate known iPPIs, and blue spheres indicate peptoids.

based on the EDmin3 descriptor, though there was very little separation between all three populations. As a result, EDmin3 was not determined to be a good distinguishing factor in this analysis. As might be anticipated based on the necessary number of hydrogen bond donors and acceptors in the tripeptoid backbone, the average score of the final descriptor, IW4, was significantly lower than both iPPIs and non-iPPIs since hydrophilic regions are generally more spread out over an entire ligand, rather than being concentrated at one end.

Though these tripeptoids as an entire set may not be predicted to be classified as iPPIs based on all factors taken together, aside from simply globularity, many of the ligands in the set did adhere to the characteristics that would define them as inhibitors of protein-protein interactions. From the analysis, 20,697 of the 54,862 peptoids were predicted to be iPPIs (results shown in the file “Bayesian Classifier on Tripeptoid Set” on the attached DVD). One hundred of the peptoids predicted as iPPIs by the Bayesian classifier were randomly selected from the set of 20,697, and were combined with all 1058 compounds in the 2P2I Hunter set (file “2P2I and 100 Predicted Peptoid iPPIs_combined” on the attached DVD). A principle component analysis of the descriptors was performed in order to visualize the similarity/dissimilarity between tripeptoids and known iPPIs/non-iPPIs. As can be seen in Figure 2.14, the differences between the three populations are not obvious at face value, and in fact the principle component analysis did not reveal a substantial separation between each of the three populations. As a result, the descriptors—globularity, CW2, EDmin3 and IW4—were determined to be the better gauge of assessing whether a queried molecule can be considered an iPPI.

Table 2.8
Descriptor Statistics for Experimentally Identified Peptoid Inhibitors

Peptoid Inhibitor	Globularity	CW2	EDmin3	IW4	Bayesian Value *
NLys-NPip-NBal	0.195329	2.31062	-3.01837	1.61068	0.207156
T2AA-NEal-NPip	0.130552	2.29504	-2.8844	3.57786	0.471927

* Calculated output of the Bayesian classification model; values above -0.188 are predicted to be iPPIs

Two of the experimentally confirmed inhibitors of PCNA-PIP Box interactions, NLys-NPip-NBal and T2AA-NEal-NPip, were selected and assessed using the Bayesian classification model (file “Bayesian Classifier on Active Peptoids” on attached DVD) for the four main descriptors (Table

2.8). From the results, both peptoid-based inhibitors were predicted to be iPPIs, which is promising for their potential use in the development of next-generation lead compounds for disrupting PCNA-protein interactions.

2.7 Discussion

Fragment-based drug discovery (FBDD) is a popular method of hit discovery and/or lead development in both academia and industry due to its distinct advantages over traditional higher molecular weight chemical libraries [53]. Multiple fragments can be screened simultaneously against a single protein target, which greatly increases the throughput of a screening assay. This would be useful in situations where one would be searching for fragments that could enhance the inhibitory effect of a single fragment that has already been identified for a given site. Findings ligands that enhance the inhibition could provide clues as to what chemical features may be combined to form a single, larger molecule that binds with high affinity. However, when little is known about the target site of interest, including more than one fragment at a time in a screen can be problematic since there is typically no information on the fragments' connectivity to one another. It would be necessary to obtain an NMR or crystal structure of them in complex with their target in order to elucidate how the fragments might be interacting with the protein at the binding site, as well as their spatial proximity/orientation to one another. This is not always easy or even feasible to perform based on the nature of the protein itself and the actual affinity of each fragment. Therefore, having a way to know the relative positions of each small ligand to one another ahead of time would be advantageous since favorable chemical features could theoretically be incorporated into a larger molecule much more quickly.

As an alternative to traditional FBDD methodology, this study demonstrated that multiple individual fragments could be tied together in a single peptoid-based backbone and be screened both *in silico* and *in vitro* to find inhibitors of PCNA-PIP Box interactions. Though much attention in the field of drug discovery is being paid to developing new inhibitors of protein-protein interactions, PCNA itself would not be considered an easy drug target, and most research efforts would shy away from it. PCNA does not have a traditional small molecule binding site, nor does it have any visible deep binding clefts that would make targeting that site with a ligand obvious. Even when compared to other protein-protein interfaces, such as the contact surface between MDM2 and p53, there is not a clear binding groove—instead there is a shallow, relatively small surface pocket where only a couple amino acids bind from PIP Box-containing proteins. However, other studies

[178,231], in addition to this one, have been able to identify small molecules that are able to bind at that site with enough affinity to disrupt the binding between PCNA and a PIP Box-containing peptide. The next step in developing high affinity ligands for PCNA would be to expand upon the chemical information gained from the tripeptoids that showed up as hits in experimental screens, to create next generation ligands that take advantage of the molecular contacts identified in the molecular dynamic simulations. This is discussed further in Chapter 4, section 4.1.

In addition to identifying inhibitors of PCNA, the other major goal of this research was to develop a platform where a combinatorial screening library composed of ligands containing multiple fragments could be quickly narrowed using computational analysis, ultimately saving time and money in subsequent physical synthesis and experimental screening. Previous studies have utilized peptoid-based libraries for identifying inhibitors, including tripeptoids [302,305,318–322,333]. However, in many of those cases there was information already available on potential ligands to fit the respective site of interest, which naturally narrows the number of molecular variants to be made, and in other cases entire expansive chemical libraries were synthesized and screened against their target(s). While the latter method can be effective for finding good hits, it can be quite time consuming and expensive to synthesize tens of thousands or more compounds.

The approach developed here has the potential to be applied to numerous protein targets where an ideal ligand would need a higher degree of globularity and would have to cover a relatively large surface area. Although this method offers some potential versatility in screening against different proteins, there are some natural drawbacks to the ligands. Due to the nature of the peptoid backbone, and the way in which these compounds are cleaved from resin, there are amine and carbonyl groups that are necessarily present, which may result in unfavorable interactions with the target binding site, depending on the orientation of the ligand. Additionally, and perhaps most significantly, because the individual peptoid side chains are tied together into a single backbone, their degrees of freedom are substantially restricted, and this can mean that an individual fragment may not be able to orient itself in the proper way to fully optimize a binding interaction. Furthermore, because the entire library here consisted of peptoid trimers, a particular ligand may maintain favorable interactions between only two of its side chains and the protein; forcing the presence of the third side chain may curtail the overall binding affinity of the total ligand if the individual interaction between that third fragment and the protein is unfavorable. There potentially are also situations where to disrupt a particular protein-protein interaction, three peptoid residues would be insufficient—four or more may be needed. These are all common dilemmas that are faced with FBDD and/or drug screening efforts in that one must always be concerned with the chemical

content and diversity of the compounds in the screening library. In that regard, this approach is not unusual, but it does place limits on how this method can be used. Promisingly, the classification analysis of peptoids that were experimental hits predicted them to be iPPIs. It is possible that ligand selection can be further narrowed in the future by applying an additional filter based on whether each compound is predicted to be an iPPI, but it may not be necessary for a peptoid to be classified as an iPPI for it to efficiently inhibit a protein-protein interaction.

It is not yet clear whether in the context of PCNA-protein complex formation, directly targeting PCNA alone would be sufficient to result in an efficacious treatment option. There is some evidence that PCNA is implicated in some way with every pathway of DNA damage repair given that key proteins in each of those pathways have PIP Boxes. However, many mechanisms in cells are redundant and are capable of compensating for the loss of a single pathway regulator. It may be that targeting PCNA will have utility only when other synthetically lethal [343] drugs are present that target compensatory proteins for the processes of DNA replication and damage repair. Additionally, it is not yet understood how the flexibility of PCNA would influence the efficacy of a single drug targeting the PIP Box binding site. In addition to the study by Pedley, *et al.* [177], the MD simulations performed in this work demonstrate that PCNA can adopt very different conformations, depending on the ligand that is bound to it. The question remains as to whether a given ligand would better inhibit the interaction between one PCNA-protein complex or another. Due to the diversity in the binding partners of PCNA, it is possible that there are multiple binding sites on PCNA; this is supported by the discovery of two different conserved PCNA-binding sequence motifs in addition to the PIP Box: the KA Box and APIM [215,344]. Each of these different motifs has been implicated in the ultimate function of complexes and the regulation of a particular biological response. It may be that a ligand binding to PCNA at one site produces an allosteric effect that would affect PCNA-protein complexes at other sites. The functional consequences of this are not clear, but it could mean that attempting to stabilize a specific conformation of PCNA to achieve a singular response is much more complicated than anticipated.

3. NUCLEAR TARGETING OF KNOWN CANCER THERAPEUTICS

More than 500 genes coding for protein kinases have already been identified in the human genome, and over 150 of these have been shown to be involved in the onset and progression of many cardiovascular, neurodegenerative and inflammatory diseases, as well as cancer [345,346]. As one of four members of the HER receptor tyrosine kinase family [347], epidermal growth factor receptor (EGFR) serves as one of the most studied receptor targets in the fields of oncology and drug development. It exists on the cell surface and becomes activated by the binding of one of its various specific ligands. The activation of EGFR leads to the initiation of growth-promoting signaling cascades that result in cellular activities such as DNA synthesis and proliferation.

A number of studies have focused on the link between EGFR expression and tumorigenesis. Evidence strongly suggests that dysregulated EGFR expression and signaling plays a major role in the onset and progression of various human cancers including head and neck squamous cell carcinoma, lung, breast, colon, anal, pancreatic, ovarian, bladder and oesophageal [348–353], while levels of the protein can be used as a prognosticator of patient outcome in certain cancers [354]. This knowledge has resulted in a significant effort over the last few decades to target EGFR as a therapeutic strategy for cancer. The tyrosine kinase inhibitors (TKIs) that have emerged from this body of work have seen some success in the treatment of disease, but a number of significant problems have limited their range of use.

Ample evidence exists for the utility of TKIs in both experimental and clinical practice. The majority of small molecule TKIs explored to date act as competitive antagonists at the ATP binding site and use a substituted quinoxaline scaffold [355]. Targeted therapies intended to inhibit kinases have suffered from a lack of long-term effectiveness in patients [356], and achieving high levels of specificity with kinase inhibitors has been difficult [159,357–362]. While the implications for non-specificity of these drugs is not yet completely understood, this can ultimately result in complications including a plethora of negative side effects and cardiotoxicity [363]. At the same time, clinical benefit is often limited to only a fraction of treated patients, with factors related to the genome of individual tumors contributing significantly to the observed responses [364], and

acquired resistance to these drugs is a common occurrence in patients undergoing cancer therapy [48,49,365–370]. In fact, all patients with colorectal, pancreatic, metastatic lung or head and neck cancers that initially benefit from therapies that target EGFR eventually develop resistance to those treatments, with a median progression-free survival time of less than one year [49].

Various intracellular kinases have been seen to translocate to the nucleus of cells through endocytosis [371–382]. While the reasons for this are still unclear, EGFR/ErbB1 translocation to the nucleus has been shown to promote cellular proliferation [377], and the phosphorylation state of nuclear-located c-Src is correlated with patient outcome in ER-positive breast cancer [374]. Their overall role in resistant or poorly responsive tumor phenotypes is not yet completely understood, but it is possible that targeting the nuclear-localized forms of these kinases could have a pronounced and/or alternative effect relative to their normally targeted state. In any case, these issues demonstrate the need for developing new therapeutics to overcome many limitations associated with TKIs.

3.1 Targeting Mechanisms of Developed Resistance to Drug Therapies

There is evidence that an increase in the activity of ATP-dependent efflux pumps can reduce the intracellular concentration of drugs, and this is correlated with acquired resistance of various therapeutics such as imatinib (Gleevec), gefitinib (Iressa), doxorubicin (Adriamycin), daunorubicin (Cerubidine), vinblastine, vincristine (Oncovin) and paclitaxel (Taxol) [365,383,384]. Additionally, a reduction in drug importation can lead to drug resistance [365,385,386]. More than 100 different mutations have been described that affect more than 70 amino acids, resulting in different molecular mechanisms of resistance acquisition [387]. The most common mechanism of developed resistance to TKI therapy is the presence of point mutations within the kinase domain. Some of these mutations result in significant conformational changes at the binding site, inhibiting the binding of TKIs due to steric hindrance. Other mutations may cause ATP to preferentially bind the receptor, resulting in it outcompeting the inhibitors. In either case, these mutations impair the apparent affinity of TKIs to these kinases, rendering this type of treatment ultimately ineffective.

There are examples abound of these types of mutations in various types of cancers. Some examples are LYN D189Y in ER+ metastatic breast cancer [388], BCR-ABL T315I in chronic myeloid leukemia [389], and KIT T670I in gastrointestinal stromal tumors [390]. In the case of EGFR-mutant lung cancer, about 70-80% of these tumors contain a somatic mutation in the kinase domain that responds to gefitinib. It has been seen that the drug binds EGFR 20-fold more tightly,

compared to the wild-type protein, when it contains the L858R mutation—this produces a substantial initial patient response [391]. However, all of these patients develop resistance to gefitinib after a median period of 10-16 months. The most commonly observed causative mechanism is the T790M “gatekeeper” mutation, which is seen in approximately 60% of patients that develop TKI resistance [392]. Yet, secondary kinase mutations are a common occurrence in various other cancers that display oncogenic addiction, and targeting this single mutation in and of itself may not prove to be effective in staving off further development of resistance.

In reality, resistance to TKI therapies likely occurs through multiple interacting pathways. The explosion of knowledge surrounding gene networks has demonstrated that central nodes of specific gene interaction clusters undergo significant amounts of cross-talk with other nodes. This is especially true of intracellular signaling cascades which rely on phosphorylative transfer, for example. Disrupting the activity of a single enzyme or protein may not have a significant impact on cellular function/viability since there can be compensatory parallel networks that can make up for that loss. This can be accomplished by activating shared downstream targets (demonstrating biochemical redundancy) [393], or by having other cellular functions cover for the initial loss (e.g. multiple mechanisms of DNA damage repair). As an example, in the case of PARP1, the protein typically has an important functional role in repair of single-stranded DNA breaks (SSBs), but can also function in non-homologous end joining of double stranded breaks (DSBs) when Ku proteins are lost [394], and homologous repair through interaction with MRE11 and ATM [395,396]. In the absence of PARP1, SSB repair is impaired and DSBs subsequently tend to accumulate, but PARP1^{-/-} mice remain viable and fertile, likely due to the compensating effect of PARP2 [397]. Only when BRCA1/2 function is impaired does the organism lose viability. The BRCA1/2 proteins are important for homologous recombination DSB repair, and studies have demonstrated the synthetically lethal relationship between PARP1 and BRCA1/2 [398]. Compensatory relationships between genes have long been studied [399], and the strategy of targeting synthetically lethal proteins has become an interesting approach to treating specific types of cancer [292,400–402].

However, while pursuing synthetic lethality as a therapeutic strategy shows promise, these new drugs show variable clinical benefit in patient populations in both single-agent and combination therapies, and often are associated with dose limited toxicities such as myelosuppression [398]. Furthermore, there still exist the problems of downstream effector targets being activated by alternative mechanisms that become upregulated in response to an intracellular drug-induced insult, as well as the promotion of protein isoform development, which itself can render a particular drug ineffective. These are contributing factors as to why no anti-RAS

therapeutic has been clinically approved, for example [403,404]. The field has recognized these issues, and ongoing research aims to improve our understanding of targeting downstream effectors. What actually defines a “downstream” target is very context dependent, and not all effectors may serve as efficient drug targets themselves, but the role of these proteins in the development of resistance to targeted therapies is being explored for proteins such as EGFR [405–409].

In addition to EGFR’s role as an initiator of phosphorylation signaling cascades, it functions as a transcription factor and enhancer of cellular proliferation in the nucleus [260,373,410–412], where it can directly phosphorylate targets such as proliferating cell nuclear antigen (PCNA), enhancing cellular growth [260,276]. Levels of nuclear EGFR (nEGFR) are also used as a prognostic factor in various cancers [413–415], and its subcellular nuclear distribution has been associated with acquired resistance to cetuximab in non-small cell lung cancer [416], and gefitinib in triple negative breast cancer and epithelial carcinoma [384,412,417]. The total mechanistic role of nEGFR in the progression of disease and development of drug resistance is not completely understood, but it has been proposed that nuclear translocation of the protein allows it to bypass typical downstream effectors, acting directly on the eventual downstream signal recipients [286,418]. This would have obvious implications for the development of resistance since TKI therapies may depend on their targets being accessible at the cell surface, or in the cytoplasm, and cells that upregulate the nuclear import of receptors may be removing these proteins from the targetable pool.

3.1.1 Subcellular Targeting as a Therapeutic Strategy

Traditionally, the pharmaceutical industry has focused on optimizing general molecular parameters such as molecular weight, logP, and capacity to participate in hydrogen bonding as a way to balance target specificity and bioavailability [419]. Upon reaching the organ or tissue of interest, a drug will bind to its target molecule, given that this target is located at the cell surface. However, if the target is located in an intracellular compartment, the drug may not be able to efficiently reach its target, impeding the activity of the drug itself. Practitioners have turned to addressing this issue by designing more membrane-permeable variants, but these may also have their own set of problems related to non-specificity if the drug is able to freely diffuse throughout the cell. To address these crucial obstacles, various methods have been developed to subcellularly target drugs to specific regions within the cell where the drug target of interest exists [420].

A diverse number of drug types such as pepducins, antitumor drugs, antioxidants, Shiga holotoxin and cholesterol drugs have been targeted to different subcellular compartments including the cell membrane, early/late endosomes, endoplasmic reticulum, Golgi, cytosol, mitochondria and nucleus [420]. There are multiple methods through which these have been achieved, ranging from things such as myristoylation and palmitoylation, to the use of lipophilic cations, drug-encapsulating nanoparticles and cell-penetrating peptides [420]. The nucleus itself is an interesting target given that it contains many cancer-related proteins involved in processes such as DNA replication and damage repair. Strategies for targeting molecules to the nucleus arose from observations of the mechanisms of action of various DNA viruses; in fact, viruses themselves have been used as delivery vehicles for gene therapy [421,422], but issues with lack of specificity, toxicity and immune response have limited their range of use. Instead, alternative delivery methods are being explored that lack pathogenic components, but make use of specific viral machinery such as peptide-based sequences.

3.1.1.1 Peptides and Peptoids as Cellular Uptake or Nuclear Targeting Sequences

Nuclear localization sequences (NLSs), often consisting of, but not limited to, short peptides derived from viruses such as Human Immunodeficiency Virus 1 or Simian vacuolating virus 40 [423–428], have been used to target various types of molecules to the nucleus of cells. These peptides can be attached to other molecules that range in size from small, drug-like molecules (MW<500) to mesoporous silica nanoparticles that act as high-payload drug delivery transporters; this has been shown to increase the efficiency of drug delivery and improves the respective desired effect [420,429–431]. These peptidic regions can also be more broadly characterized as “cell penetrating peptides” (CPPs). CPPs tend to be either amphipathic or contain arginine-rich stretches that enhance cellular uptake through proposed mechanisms of cytoskeletal remodeling and actin-encapsulated vesicle endocytosis [432], or through clathrin and dynamin in the case of polyarginine,[433] though this has been disputed [434]. Other various forms of molecules exist that enhance cellular uptake such as β -peptides [435], homochiral cyclic peptides [436] and peptoids.

As discussed in Chapter 2, advantages that peptoids have over peptides are their generally better inherent ability to penetrate cell membranes, higher resistance to proteolysis, resistance to solvent, temperature and chemical denaturants, and generally cheaper cost to produce [296–299]. The two most commonly used kinds of peptoids for enhancing cellular uptake are amino- and guanidinium-based, containing between five and nine residues [437–439]. Guanidinium-peptoids

Table 3.1
Examples of Cell Penetrating or Nuclear Targeting Peptides

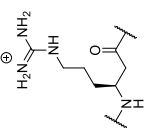
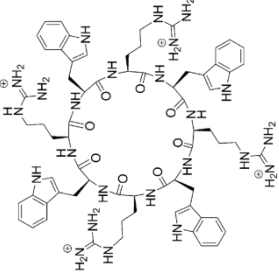
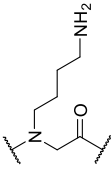
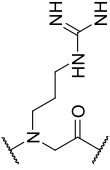
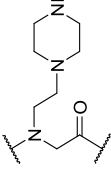
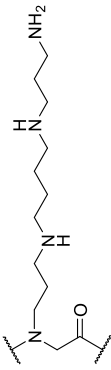
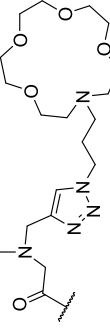
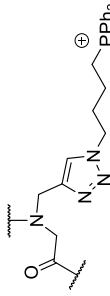
Name	Sequence or Structure (N- to C-Terminus)	Description	Example References
SV40	PKKKRKV	derived from Simian vacuolating virus 40; used as a nuclear targeting sequence	[428,438]
TAT	GRKKRRQRRRPQ	derived from Human Immunodeficiency Virus 1; has been used in numerous applications to enhance nuclear uptake	[426,427,432]
VP-22	NAATATGRSAASRPTQRPRAPA RSASRRRPVN	derived from the herpes simplex virus; enhances cellular uptake	[440]
β -Peptides		analogous to normal peptides, but are more resistant to proteolysis and can adopt well-defined protein-like secondary structures	[435,441]
Homochiral Cyclic Peptides		peptide-based carriers that are more stable to hydrolysis by peptidases	[436]
Proline-Rich	RRIRPRPPRLPRPRRPLPFFPRPG	amphipathic sequence derived from the antimicrobial peptide, Bac 7; enhances cellular uptake	[442]

Table 3.2
Examples of Peptoid Sequences that Enhance Cellular or Intracellular Uptake

Name	Structure	Description	Example References
Amino-Based		enhances cellular uptake and cytosolic distribution	[299,437–439]
Guanidinium-Based		enhances nuclear uptake	[299,437]
PiperazinyI		enhances cellular uptake and is mostly localized in vesicles	[443]
Polyamine		enhances cellular uptake and cytosolic distribution	[443]
Crown Ether		enhances cellular uptake and is mostly localized in vesicles	[443]
Triphenylphosphine		enhances cellular uptake and is mostly localized in vesicles	[443]

are often taken into cells at a higher rate than amino-peptoids, and amino-peptoids are typically localized to the cytoplasm, whereas guanidinium-peptoids are trafficked to the nucleus [437]. Both types are characterized by multiple positively charged residues, but the fact that they are sorted differently indicates that their uptake kinetics and subcellular destination are dependent on the chemical nature of the side chains themselves. Similar to these molecules, polyarginine, -lysine and -histidine peptides have been used to enhance cellular uptake as well. Among these, polyarginine shows a much higher efficiency for uptake than polylysine or -histidine [299,444]. Furthermore, it has been observed that a D-amino acid variant of polyarginine is more effective than the L- form at being taken into cells [299]. Ultimately, however, none of these peptidic polycationic molecules are as effective at enhancing cellular uptake as polyguanidinium-based peptoids [299].

3.2 Rationale

The use of peptide and peptoid-based carriers has long been proposed to enhance the uptake and subcellular targeting of a host of molecular agents [420]. Studies have also shown that peptides tagged with NLSs are able to more effectively act on their target. The utility of this strategy has been established for peptides, proteins, oligonucleotides and nanoparticles alike, and is currently being further explored for use in disease therapeutics. Interestingly, this approach has also been validated for the targeting of small drug-like molecules. Hodoniczky, J. *et al.* demonstrated that the antiproliferative effect of the antiandrogen drug, bicalutamide, was greatly enhanced when the drug was conjugated to a peptide consisting of the cell permeable penetratin and nuclear localizing SV40 sequences [429]. Albeit, not all drugs may be amenable to incorporation through attachment via disulfide or thio-ether linkages, as with that compound, but it provides an important example of the effect of this strategy.

Subcellular drug targeting could potentially have broad applications for various disease types, particularly in cases where the target of interest resides in cellular compartments such as the nucleus. Because upregulated nuclear distribution of EGFR is associated with resistance to cancer therapies [48,49,384,412,416–418], it may serve as a prime subject for this approach [445]. A recent study demonstrated that a peptide derived from PCNA, which was tagged with TAT and binds to EGFR, was able to disrupt phosphorylation at PCNA residue Tyr 211, and subsequently inhibited the growth of human breast cancers cells both *in vitro* and *in vivo* [286]. The implicated reason for this is that nEGFR has been shown to bind to and directly phosphorylate PCNA to control its stability on chromatin [260,261]. Targeting it to the nucleus with TAT enables it to compete

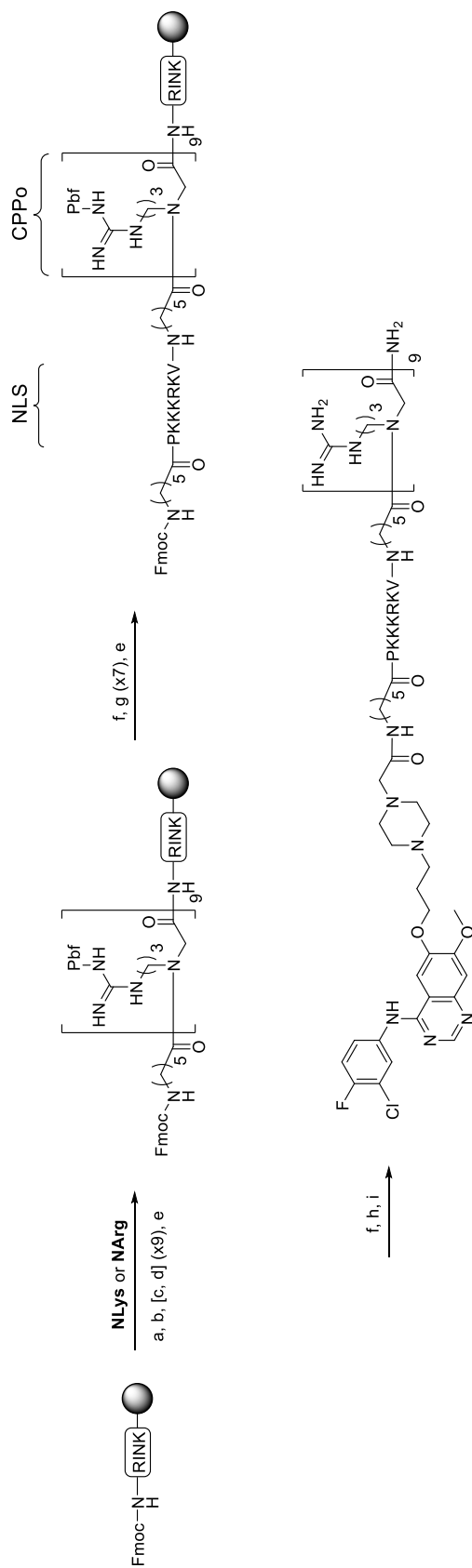
with PCNA for binding to nEGFR, suppressing the phosphorylation of PCNA and resulting in a loss of PCNA stability on chromatin—potentially restoring DNA damage mismatch repair [276,446]. Used in a combination therapeutic regimen, this method of targeting nuclear proteins that are implicated in vital cell survival functions may prove to be very effective in the treatment of disease.

It is important to note, as was demonstrated by the bicalutamide [429] and peptoid-based [438] studies, the presence of a cell penetrating sequence alone may be insufficient in enhancing the specific localization of a compound into the cell, and may instead lead to a general cytosolic or vesicular distribution. The addition of a NLS instead upregulates nuclear uptake and can substantially improve a drug's activity. However, the question of whether simply enhancing cellular uptake or specifically targeting a molecule to the nucleus is more effective is very mechanism dependent, and will completely depend on the therapeutic context. Regardless, being able to direct the nuclear uptake of an inhibitor may prove to be a solution to acquired drug resistance, given that it would bypass “downstream” effector proteins, and would avoid issues associated with receptors being taken in to the nucleus [429]. In this study, a strategy was explored to determine whether, through the incorporation of either a NLS, CPPo or both, gefitinib's activity or mechanism of action could be modulated in triple negative breast cancer cell models and murine tumor cells that irregularly express EGFR. To our knowledge, this study reports for the first time the targeting of a commercially available TKI to the nucleus of cells, and the subsequent enhanced effects, demonstrating a useful strategy for the targeting of traditionally drug-resistant cancer types.

3.3 Experimental

3.3.1 Methodology

A series of peptoid-based conjugates were synthesized using the methodology outlined in Scheme 3.1. These molecules consist of either amino- (NLys) or guanidinium- (NArg) based polycationic cell-penetrating peptoid (CPPo) sequences, with half of the total set also containing a peptide-based SV40 nuclear localization sequence. With the exception of negative controls that contained no drug, the conjugates were N-terminally tagged with either piperazinyl gefitinib (synthesis outlined in Scheme 3.2) or 5-FAM. The CPPo, NLS and drug/dye were each separated by a 6-carbon spacer. The CPPo sequences were synthesized using submonomer peptoid synthesis [300] due to the



Reaction conditions: (a) swell resin in DMF (5-fold volume), 30 min. at RT (b) 20% piperidine in DMF (2 x 15 min.); (c) bromoacetic acid (2 mL of 1 M solution in DMF, 20 equiv.), DIC (19 equiv.), 1 hr. at 35°C; (d) monomer **Nlys** or **NArg** (2 mL of 1 M solution in DMF), 2 hr. at RT (e) Fmoc-6-aminohexanoic acid (2.5 equiv.), HCTU (2.4 equiv.), DIEA (5 equiv.), 0.1 M in DMF, overnight at RT; (f) 20% piperidine in DMF (2 x 15 min.); (g) Fmoc-protected amino acid (10 equiv.), HCTU (9.5 equiv.), DIEA in NMP (500 μ L of 4 M solution, 20 equiv.), 0.4 M in DMF, 1 hr at RT, repeat step (f); (h) piperazine] gefitinib (2.5 equiv.), HCTU (2.49 equiv.), DIEA (5 equiv.), 50 mM in DMF, overnight at RT; (i) TFA/TIS/H₂O (95 : 2.5 : 2.5), 3 hr at RT. "NLS" represents the peptide nuclear localization sequence, and "CPPo" represents the cell-penetrating peptide sequence. NArg (shown above) and Nlys-containing conjugates are synthesized using the same methodology

Scheme 3.1
Synthesis of Peptide-Peptoid
Gefitinib Conjugates

relatively low cost of reagents, and ease of synthesis. The NLS was synthesized, and piperazinyl gefitinib/5-FAM coupled, using standard peptide coupling procedures.

3.3.2 Experimental Materials and Methods

All chemical reagents and solvents were purchased from Sigma Aldrich, Combi-Blocks Inc. or Fisher Scientific, unless noted otherwise. Materials were repurified via recrystallization or distillation as necessary before use. NMR experiments were performed on Bruker (Bruker Corp., Billerica, MA) ARX300 (300 MHz), ARX400 (400 MHz) or DRX500 (500 MHz) instruments. All ^1H chemical shifts (δ) are relative to residual protic solvent (CHCl_3 : δ 7.26, DMSO-d_6 : δ 2.50, CD_3OD : δ 3.31 ppm), and all ^{13}C chemical shifts (δ) are relative to the solvent (CDCl_3 : δ 77.23, DMSO-d_6 : δ 39.52, CD_3OD : δ 49.00 ppm). Low resolution electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI) studies were carried out on an Agilent 6320 Ion Trap (Agilent Labs, Santa Clara, CA) mass spectrometer. High resolution mass measurements were obtained on a LTQ Orbitrap XL mass spectrometer (ThermoScientific Corp.) utilizing electrospray ionization (ESI). Molecular masses and sequences of peptides or peptoids were validated on an Applied Biosystems (Framingham, MA) MALDI-TOF/TOF 4800 mass analyzer, or Applied Biosystems Voyager DE PRO mass spectrometer using either 2,5-dihydroxy benzoic acid or α -cyano-4-hydroxy cinnamic acid matrices. TLC analyses were performed on Merck aluminum-backed F254 silica gel plates. Stock solutions of each non-dye-containing compound were made by measuring the dry mass of each in pre-dried, pre-weighed screw-cap vials, and adding the volume of DMSO necessary to give 10 mM solutions. For mass measurements, molecular weights were calculated based on their TFA salt form. Stock solutions of compounds containing N-terminal 5-carboxyfluorescein were made by measuring the absorbance at 494 nm, using an extinction coefficient of $79,000 \text{ L mol}^{-1} \text{ cm}^{-1}$ and Beer's Law ($A=\epsilon bc$) to calculate concentration. Data analyses and graphical representations were performed in Microsoft Excel, GraphPad Prism 6 or OriginPro 2015. Characterization of synthesized materials is outlined in Appendix B, Table B.2.

General Method for the Synthesis of Conjugate Peptoid Segments. The peptoid portion of each molecule was synthesized using an adapted procedure for submonomer [300] peptoid synthesis. Briefly, 0.100 mmol of H-Rink Amide-ChemMatrix resin (PCAS Biomatrix Inc., Quebec, Canada; loading: 0.51 mmol/g) was transferred to a 25 mL glass fritted peptide reaction vessel and was swelled with DMF for 30 minutes. The resin was then deprotected using two 2.5 mL portions of

20% piperidine in DMF with incubation times of 15 minutes for each addition at room temperature. Following washing of the resin with DMF (6x) and DCM (3x), deprotection was confirmed by a ninhydrin (Kaiser's) test for primary amines. A solution of 1.5 mL of 1M bromoacetic acid (30 equiv.) in DMF and 230 μ L (29.4 equiv.) of N,N'-diisopropylcarbodiimide (DIC) was added, and the resin was placed on an orbital shaker for 1 hour at 37°C. At that time, the resin was washed with DMF (6x) and DCM (3x), and a solution of 1M respective primary amine in DMF was added, with incubation on an orbital shaker for 2 hours at 37°C. These steps were repeated with washing steps in-between to produce the desired peptoid sequence. Following the addition of the final primary amine, a six-carbon linker (aminohexanoic acid) was coupled to the N-terminus as a spacer between the peptoid and peptide portions of each conjugate. Fmoc-6-aminohexanoic acid was coupled using standard peptide coupling conditions (89 mg (2.5 equiv.) Fmoc-6-Ahx-OH, 103 mg (2.49 equiv.) HCTU and 65 mg (5 equiv.) DIEA dissolved up to 2.5 mL with DMF), with incubation on an orbital shaker at room temperature for 16 hours. The resin was Fmoc-deprotected using two 2.5 mL portions of 20% piperidine in DMF with incubation times of 15 minutes for each addition at room temperature.

General Method for the Synthesis of Conjugate Peptide Segments. The peptide segment of each hybrid was constructed using standard solid phase peptide synthesis conditions and Fmoc-protected amino acids (Anaspec Inc.). Amino acids were coupled by incubating the resin with a solution of 2.1 mL 0.45M HCTU (O-(1H-6-Chlorobenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate) in DMF (9.45 equiv.), 500 μ L 4M N,N-diisopropylethylamine in NMP (20 equiv.), and 1 mmol (10 equiv.) of each respective amino acid on an orbital shaker for one hour (incubation time for arginine was extended to two hours) at room temperature. Fmoc deprotection, subsequent to each amino acid coupling, was performed using two portions of 20% piperidine in DMF with incubations on an orbital shaker at room temperature, 20 minutes for each addition. For difficult deprotections, in-between the two additions of piperidine was added 2.5 mL of 3% w/v 1,8-Diazabicyclo[5.4.0]undec-7-ene in 20% piperidine/DMF. This was heated to 45°C for three minutes, and the resin was promptly washed with DMF (6x) and DCM (3x). Fmoc-amino acid couplings and Fmoc-deprotection steps were alternated with washing steps (DMF (6x) and DCM (3x)) in-between to produce the desired peptide sequence. Following the addition of the final amino acid, and subsequent Fmoc-deprotection, a six-carbon linker (aminohexanoic acid) was coupled to the N-terminus as a spacer between the peptide and piperazinyl gefitinib or 5-carboxyfluorescein. Fmoc-6-aminohexanoic acid was coupled using standard peptide coupling conditions (89 mg (2.5

equiv.) Fmoc-6-Ahx-OH, 103 mg (2.49 equiv.) HCTU and 65 mg (5 equiv.) DIEA dissolved up to 2.5 mL with DMF), with incubation on an orbital shaker at room temperature for 16 hours. The resin was Fmoc-deprotected using two 2.5 mL portions of 20% piperidine in DMF with incubation times of 15 minutes for each addition at room temperature. FAM-TAT was synthesized using standard peptide synthesis conditions outlined in Chapter 2.

Coupling of Piperazinyl Gefitinib or 5-Carboxyfluorescein to Conjugates. Piperazinyl gefitinib and 5-carboxyfluorescein were coupled using standard solid phase peptide synthesis conditions. 63 mg (0.125 mmol ; 1.25 equiv.) piperazinyl gefitinib and 51 mg (0.1249 mmol) of HCTU were dissolved in 2.5 mL of DMF, followed by 32 mg (0.25 mmol; 2.5 equiv.) of DIEA. The resin was then incubated with this solution on an orbital shaker overnight at room temperature. For the coupling of 5-carboxyfluorescein, a solution of 75.3 mg of 5-FAM, 80.7 mg of HCTU and 46 mg of DIEA in 2 mL of DMF was added, and the resin was then placed on an orbital shaker overnight, in the dark, at room temperature. After the coupling of either compound, the resin was washed with DMF (6x) and DCM (3x), dried under high vacuum (< 1 mm Hg) and stored at -20°C until cleavage.

General Method for the Cleavage and Purification of Drug Conjugates. Resin-bound compounds were cleaved from resin using a solution of trifluoroacetic acid (TFA)/triisopropylsilane (TIS)/water (95:2.5:2.5), incubating the resin at room temperature for 3 hours. They were then precipitated into ice cold diethyl ether and collected by centrifugation at 4,000 x g for 10 minutes at 4°C. The peptides were purified via HPLC (Beckman Coulter System Gold 166 or 168) using an increasing gradient of acetonitrile (ACN)/water with 0.1% TFA (5:95) to (100:0) over 30 minutes on an Agilent ZORBAX SB-C18 reverse phase semi-preparative column. Molecular masses and sequences were validated via MALDI-TOF/TOF (Applied Biosystems 4700 Proteomics Analyzer 196). Purities were determined by HPLC using absorbencies at 219 or 280 nm.

Methods Performed by Others (See Acknowledgements for Exact Roles)

Probing Phosphorylation Status by Western Blot. Cells were seeded in a 24-well plate, and grown overnight in complete media (DMEM + 10% FBS, penicillin/streptomycin, insulin). The next morning, the media was removed and replaced with serum-free media with inhibitors (1 μ M in DMSO). The cells were incubated for 6 hours, and EGF (25 ng/mL) was added. The cells were incubated for 30 minutes, then the media was removed and lysis buffer was added (RIPA + PPIs,

NaVO₄, NaF, beta-glycerophosphate). After lysis, the insoluble fraction was removed, and the protein concentration was normalized via a bicinchoninic acid assay. The lysates were separated by PAGE and transferred to a PVDF membrane. The membranes were incubated with primary antibody (pERK, pEGFR, pSTAT3, beta-tubulin) overnight at 4°C, then incubated with HRP-conjugated secondary antibody for one hour at room temperature. The membranes were stripped and re-probed if necessary (total-STAT3).

Fluorescent Imaging of Peptoid Conjugates. NME cells were seeded on a 12-well plate, and grown overnight in complete media (DMEM + 10% FBS, penicillin/streptomycin, insulin). FAM-NArg or FAM-SV40-NArg were added at a 10 μM concentration in PBS, and the cells were incubated with the compounds for one hour at room temperature. Cells were then washed with PBS, fixed with 4% formaldehyde in PBS, washed again with PBS, and the nuclei stained with DAPI in PBS. Cells were imaged using the Nikon A1R-MP confocal microscope in the Purdue Bioscience Imaging Facility.

MTT Proliferation Assay. MDA-MB-231 and -468 cell proliferation was measured using the MTT (3-(4, 5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) cell proliferation reagent. Cells were grown in DMEM (phenol-red free) supplemented with 10% FBS, 1% penicillin/streptomycin at 37°C with 5% CO₂. They were then plated on flat bottom 96-well plates at a density of 1 x 10⁴ cells per 100 μL per well. Cells were allowed to attach for 4 hours and were then treated with 5 concentrations of each compound for 72 hours. 10 μL of MTT (5 mg/mL in DMEM phenol-red free media) was added to each well at a final concentration of (0.5 mg/mL) for 4 hours at 37°C. After the period of incubation, 100 μL of solubilization solution (10% Triton-X 100, acidic isopropanol (0.1N HCl)) was added and the plates were sealed and stored from light for 3 days. Absorbance was read at 570 nm and the percent cell growth was normalized by comparison to a day zero control plate with no drug for each respective cell line. The day zero plate standardized 0% and 100% cell viability, with columns plated in triplicates with either 100 μL of DMEM alone, or 1 x 10⁴ cells in 100 μL of DMEM. No inhibitors were added; instead, 10 μL of MTT reagent was added to each well after cells were allowed to attach for 4 hours. At the end 4 hours, 100 μL of solubilization solution was added to the plate, and was it sealed and stored in the dark for 3 days.

Three-Dimensional (3D) Organotypic Growth Assay. 96-well plates were coated with Cultrex (50 μL/well) and cells were resuspended in DMEM supplemented with 10% FBS and 4% Cultrex (150

$\mu\text{L}/\text{well}$). Luciferase-expressing cells were seeded at a density of 1×10^3 cells/well. Media containing the indicated inhibitors and/or EGF was replaced every 4 days and organoid outgrowth was detected by the addition of D-luciferin potassium salt (Gold Biotechnology) to induce bioluminescence, which was quantified using a GloMax-Multi detection system (Promega).

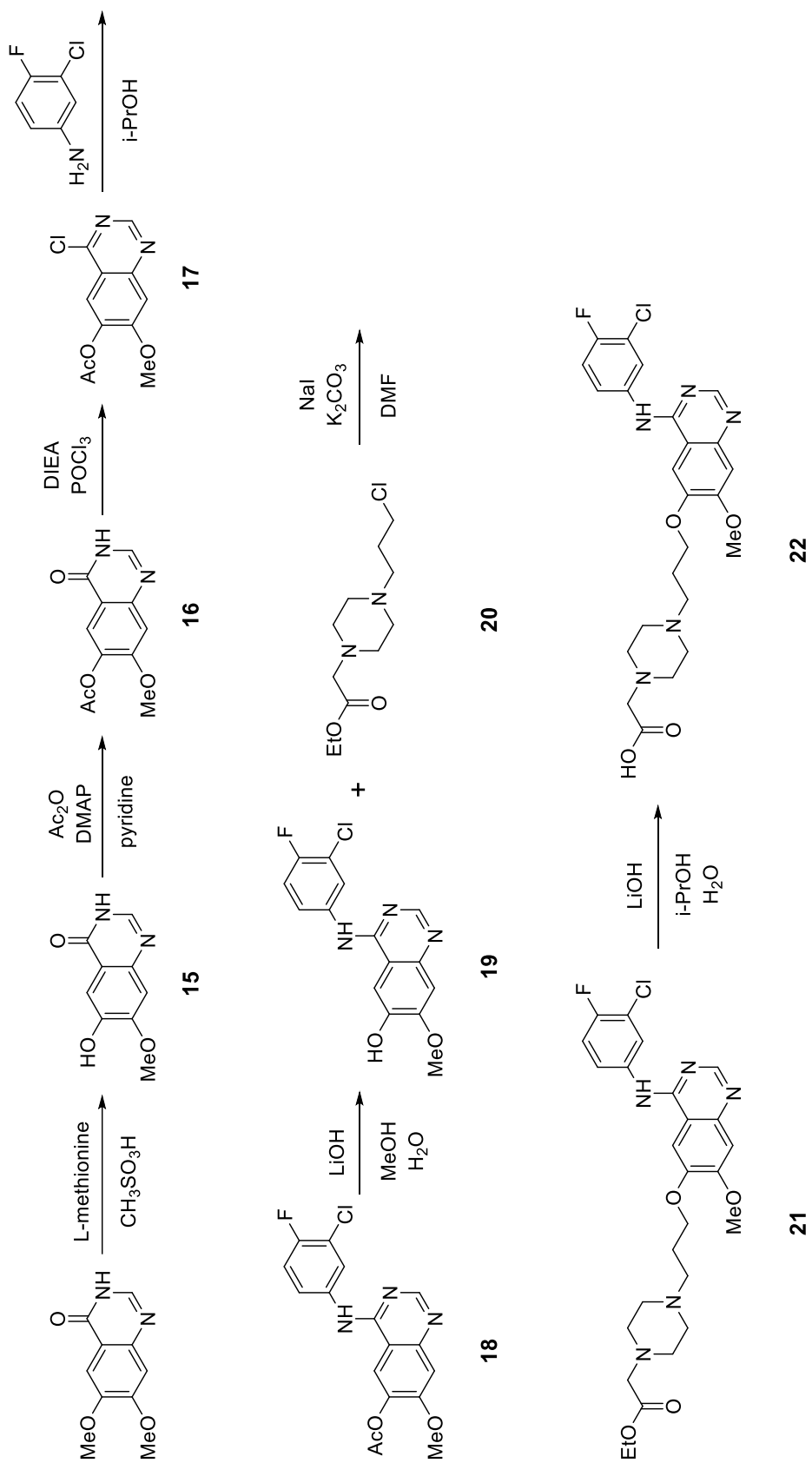
Murine Cell Lines. Normal mammary epithelial (NME) cells were constructed from NMuMG cells as previously described to overexpress EGFR [447]. These cells were then treated with transforming growth factor beta (TGF- β) to induce metastasis, and were grafted into a murine mammary fat pad. TGF- β -treated NME tumor cells that underwent metastasis were subcultured from the lungs of mice—these cells displayed significantly enhanced primary tumor formation, postsurgical recurrence and spontaneous pulmonary metastasis upon secondary mammary fat pad engraftment. These subcultured cells were termed NME lung metastatic (LM1) [448]. Both cell lines were cultured in DMEM + 10% FBS, penicillin/streptomycin, and insulin.

3.3.3 Synthesis of Primary Amines and Piperazinyl Gefinitib

Synthesis of “NArg” and “NLys” were performed in the same way as described previously in Chapter two (see compounds **1**, **2** and **10**).

Synthesis of 6-Hydroxy-7-methoxyquinazolin-4(3H)-one (15): 6.23 g (30.2 mmol) of 6,7-dimethoxyquinazolin-4(3H)-one, 5.40 g (36.2 mmol) of L-methionine and 42.6 mL (21.7 equiv.) of methanesulfonic acid were added to a round bottom flask with stirring. The solution was heated to 115°C and was stirred for 6 hours. The reaction mixture was poured onto crushed ice, and the pH was slowly adjusted upwards with 40% NaOH, until a pH of 7. The white precipitate that formed was filtered off and was washed with water, cold MeOH and diethyl ether. The solid was dried overnight under high vacuum (< 1 mm Hg) to produce 5.54 g (95.4% yield) of **15** as a white solid, which was used in the next step without any further purification. ¹H NMR (300 MHz, DMSO) δ 11.94 (s, 1H), 9.80 (s, 1H), 7.91 (s, 1H), 7.36 (s, 1H), 7.09 (s, 1H), 3.90 (s, 3H).

Synthesis of 7-Methoxy-4-oxo-3,4-dihydroquinazolin-6-yl acetate (16): 4.50 g (23.4 mmol) of **15**, 28 mL (12.7 equiv.) of acetic anhydride and 28 mg (0.01 equiv.) of DMAP were added to a round bottom flask with stirring. Upon filling the flask with argon, 5.6 mL (2.97 equiv.) of anhydrous pyridine was added to the reaction mixture. The reaction became very viscous to the point of not



Scheme 3.2
Synthesis of Piperazinyloxy
Gefinitib

stirring, so another 10 mL of acetic anhydride was added to the flask. The reaction was heated to 100-110°C and was stirred for 4 hours, at which point it was then poured onto crushed ice and stirred vigorously for 30 minutes. The resulting white solid was collected via filtration and washed with water, cold MeOH and small amounts of diethyl ether. The solid was dried under high vacuum (< 1 mm Hg) overnight to give 3.07 g (56% yield) of **16** as a white solid, which was used in the next reaction without further purification. ¹H NMR (300 MHz, DMSO) δ 12.21 (s, 1 H), 8.09 (s, 1H), 7.75 (s, 1H), 7.27 (s, 1H), 3.91 (s, 3H), 2.30 (s, 3H).

Synthesis of 4-Chloro-7-methoxyquinazolin-6-yl acetate (17): 3.80 g (16.2 mmol) of **16** and 8.6 mL (49.4 mmol) of DIEA were combined in a round bottom flask with stirring, and 2.27 mL (24.4 mmol) of phosphoryl chloride was added slowly. The reaction was heated to 80°C and stirred for 3.5 hours, after which the solvent and excess reagent were removed by roto-vap. The resulting brown residue was used directly in the next reaction without purification or further characterization.

Synthesis of 4-[(3-Chloro-4-fluorophenyl)amino]-7-methoxyquinazolin-6-yl acetate (18): The crude product **17** was immediately resuspended in isopropanol (32 mL), and 2.60 g (17.9 mmol) of 3-chloro-4-fluoroaniline was added with stirring under argon. The reaction flask was fitted with a reflux condenser, and the reaction was heated to reflux for 6 hours under argon. The reaction mixture was cooled to room temperature, and the precipitated product was filtered and washed with water, MeOH and cold diethyl ether. The product was dried overnight at reduced pressure to afford 4.57 g (77.9% yield for both steps) of **18** as an off-white powder. ¹H NMR (300 MHz, DMSO) δ 11.51 (s, br, 1H), 8.94 (s, 1H), 8.77 (s, 1H), 8.01-8.09 (m, 1H), 7.70-7.78 (m, 1H), 7.48-7.55 (m, 2H), 4.00 (s, 3 H), 2.38 (s, 3H).

Synthesis of 4-[(3-Chloro-4-fluorophenyl)amino]-7-methoxyquinazolin-6-ol (19). 4.57 g (12.6 mmol) of **18** was dispersed in a mixture of MeOH (300 mL) and H₂O (300 mL) at room temperature with stirring. 1.01 g (42.2 mmol) of LiOH was added and the reaction was stirred at room temperature for 45 minutes. The reaction mixture was neutralized to a pH of 7 with the addition of 10% acetic acid, and the resulting precipitate was collected by filtration, washed with water, MeOH and cold diethyl ether. The product was placed under high vacuum (< 1 mm Hg) to remove all traces of solvent, giving 3.38 g (83.7% yield) of **19** as a white solid. ¹H NMR (300 MHz, DMSO)

δ 9.79 (s, 1H), 9.47 (s, 1H), 8.45 (s, 1H), 8.13-8.20 (m, 1H), 7.70-7.85 (m, 2H), 7.38 (t, $J = 10.2$ Hz, 1H), 7.19 (s, 1H), 3.96 (s, 3H).

Synthesis of Ethyl 2-(4-(3-chloropropyl)piperazin-1-yl)acetate (20). 3.44 g (21.9 mmol; 1.5 equiv.) of 1-bromo-3-chloropropane was added to 2.51 g (14.6 mmol) of ethyl 2-(piperazin-1-yl)acetate in acetonitrile (90 mL) at room temperature with stirring. 4.1 mL (2 equiv.) of triethylamine was added dropwise, and the reaction mixture was stirred at room temperature for 48 hours. At that time, TLC showed that all of the starting material had been consumed, so the solvent was evaporated in vacuo and the crude material purified by normal phase flash chromatography (EPCLC W-Prep 2XY, Yamazen Corp.) using DCM/MeOH as the eluents. The fractions containing isolated product were combined and evaporated to give 1.9 g (52.4% yield) of **20** as a light yellow oil. ^1H NMR (300 MHz, CDCl_3) $\delta = 4.14$ (q, 2 H, $J = 7.2$ Hz), 3.55 (t, 2 H, $J = 6.6$ Hz), 3.17 (s, 2 H), 2.40-2.69 (m, 10 H), 1.88-1.97 (m, 2 H), 1.23 (t, 3 H, $J = 7.2$ Hz); ^{13}C NMR (75 MHz, CDCl_3) $\delta = 170.35, 60.76, 59.56, 55.47, 53.01, 43.27, 29.83, 14.38$.

Synthesis of Ethyl 2-(4-(3-((4-((3-chloro-4-fluorophenyl)amino)-7-methoxyquinazolin-6-yl)oxy)propyl)piperazin-1-yl)acetate (21). To a solution of 1.22 g (3.8 mmol) of **19** in 30 mL of DMF were added 1.89 g (2 equiv.) of **20**, 1.06 g (2 equiv.) of K_2CO_3 and 1.72 g (3 equiv.) of sodium iodide. The reaction mixture was warmed to 60°C and was stirred for 96 hours. The solution was then transferred to a separatory funnel, and the product was extracted with DCM (4 x 100 mL). The combined organic layers were washed with water (2 x 100 mL) and brine (100 mL), dried over anhydrous Na_2SO_4 and then evaporated to dryness. The crude product was separated using normal phase flash chromatography (EPCLC W-Prep 2XY, Yamazen Corp.) with DCM/MeOH as eluents. The fractions containing the desired compound were combined and evaporated to give 1.30 g (64% yield) of **21** as a white solid. mp $> 206^\circ\text{C}$ (decomposes). ^1H NMR (300 MHz, CDCl_3) $\delta = 8.74$ (s, 1 H), 8.66 (s, 1 H), 7.89 (dd, 1 H, $J = 6.3, 2.4$ Hz), 7.55-7.60 (m, 1 H), 7.45 (s, 1 H), 7.16 (s, 1 H), 7.08 (t, $J = 8.7$ Hz, 1 H), 4.17 (q, 2 H, $J = 6.9$ Hz), 3.98 (t, 2 H, $J = 6.6$ Hz), 3.87 (s, 3 H), 3.18 (s, 2 H), 2.40-2.69 (m, 10 H), 1.93-2.00 (m, 2 H), 1.27 (t, 3 H, $J = 6.9$ Hz); ^{13}C NMR (75 MHz, CDCl_3) $\delta = 170.56, 156.70, 155.18, 154.63$ (d, $J = 247.83$ Hz), 153.46, 148.92, 147.31, 135.71, 124.31, 121.97 (d, $J = 6.2$ Hz), 120.81 (d, $J = 18.4$ Hz), 116.48 (d, $J = 21.7$ Hz), 109.32, 107.38, 101.84, 67.64, 60.83, 59.30, 56.17, 54.77, 52.82, 52.72, 26.08, 14.31. ^{19}F NMR (282 MHz, CDCl_3) $\delta -94.94$. HRMS (ESI) Calculated mass ($\text{C}_{26}\text{H}_{32}\text{ClFN}_5\text{O}_4$) $[\text{M}+\text{H}]^{1+}$: 532.2127, mass found m/z : 532.2128 $[\text{M}+\text{H}]^{1+}$. IR: $\tilde{\nu} = 3289, 2940, 2821, 1743, 1624, 1579, 1526, 1500, 1472, 1428, 1397,$

1354, 1337, 1216, 1172, 1141, 1069, 1032, 1011, 931, 858, 821, 794 cm^{-1} . UV/Vis: $\lambda_{\text{max}} = 246, 331$ nm.

Synthesis of 2-(4-(3-((4-((3-chloro-4-fluorophenyl)amino)-7-methoxyquinazolin-6-yl)oxy)propyl)piperazin-1-yl)acetic acid (22; “piperazinyl gefitinib”). 1.30 g (2.44 mmol) of **21** was dispersed in 200 mL of 0.2 M LiOH (i-PrOH/H₂O (1.2/1)), and the reaction mixture was stirred at room temperature in the dark. After 7.5 hours, the i-PrOH was evaporated in vacuo, and 10% HOAc was added dropwise to adjust the pH of the water layer to 5. A pale-white solid precipitated during that time, and the solid was filtered, washed with water, MeOH and diethyl ether. The solid was dried overnight under high vacuum (< 1 mm Hg) to remove residual water to afford 1.13 g (92% yield) of **22** as a white solid. mp > 201 °C (decomposes). ¹H NMR (300 MHz, DMSO) δ 9.61 (s, 1H), 8.50 (s, 1H), 8.12 (dd, $J = 2.57, 6.84$ Hz, 1H), 7.88 – 7.71 (m, 2H), 7.44 (t, $J = 9.11$ Hz, 1H), 7.20 (s, 1H), 4.17 (t, $J = 6.12$ Hz, 2H), 3.94 (s, 3H), 3.16 (s, 2H), 2.81 – 2.58 (m, 4H), 2.08 – 1.88 (m, 2H). ¹³C NMR (75 MHz, DMSO) δ 169.90, 160.71, 156.04, 154.51, 152.63, 148.35, 146.99, 136.88, 123.51, 122.40, 118.89, 116.67, 116.38, 108.81, 107.30, 102.52, 67.11, 58.72, 55.89, 54.30, 51.90, 26.16. ¹⁹F NMR (282 MHz, CDCl₃) δ -120.13. HRMS (ESI) Calculated mass (C₂₄H₂₇ClFN₅O₄): 504.1814, mass found m/z : 504.1838 [M+H]¹⁺. IR: $\tilde{\nu} = 3368, 2952, 2838, 1625, 1581, 1500, 1473, 1429, 1400, 1247, 1234, 1218, 1144, 1004, 855$ cm^{-1} . UV/Vis: $\lambda_{\text{max}} = 249, 331$ nm.

3.4 Results

Peptoid monomers, NLys and NArg, were prepared from 1,4-diaminobutane and 1,3-diaminopropane, respectively, as described in Chapter 2, section 2.4.2 Experimental Materials and Methods. These peptoids were selected based on previous evidence that they enhance cellular and/or nuclear uptake, and were superior to L- or D-peptide sequences. Previous work indicates that the number of residues in polycationic peptoids affects their uptake, with longer sequences being taken into cells more rapidly [438]. To confirm this visually, two poly-lysine peptoids were synthesized that contained either seven or nine residues. These were each N-terminally tagged with FAM, separated by an aminohexanoic acid linker. MBA-MB-231 triple negative breast cancer cells were then exposed to the 7-mer, 9-mer, FAM-tagged TAT peptide (sequence: GRKKRRQRRRPQ) or FAM alone for three hours, fixed with paraformaldehyde and analyzed by confocal microscopy (Figure 3.1). From the results, there was little indication that FAM alone was able to enter cells.

Both the TAT sequence and peptoid sequences demonstrated enhanced cellular uptake, and increasing the peptoid length from seven to nine residues appeared to increase the amount of fluorescent compound taken up, as anticipated. Interestingly, the TAT sequence did not display

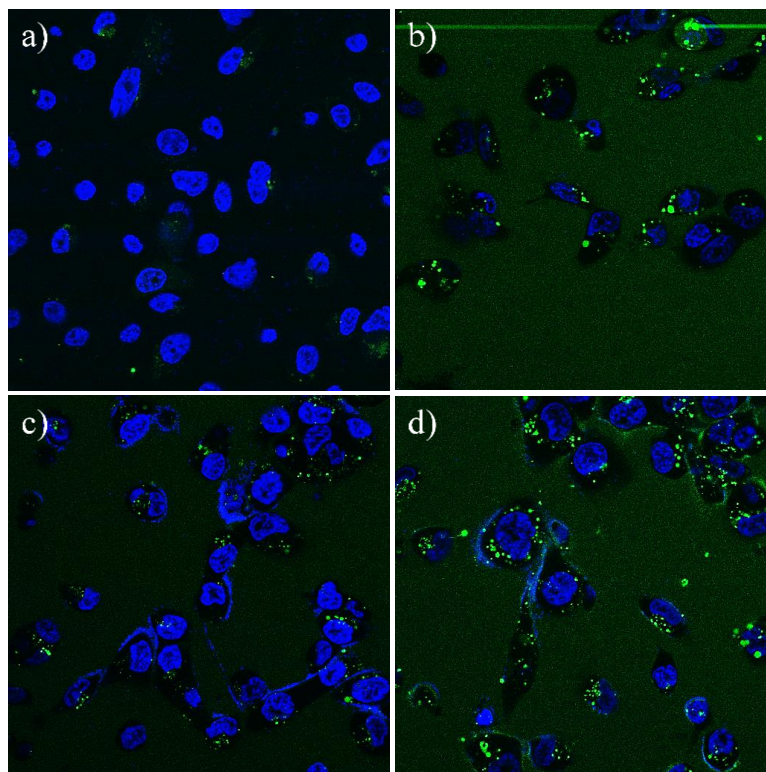


Figure 3.1. Enhanced Uptake of 5-Carboxyfluorescein Using a NLS or CPPo. MDA-MB-231 cells were incubated with 10 μ M (a) 5-carboxyfluorescein (FAM), (b) FAM-labeled TAT peptide, (c) FAM-labeled Nlys₇ peptoid or (d) FAM-labeled Nlys₉ peptoid for 3 hours. Cells were then fixed, stained with Hoechst 33342 and fluorescently imaged using confocal microscopy at 407 nm and 488 nm. Images generated by Jennifer Sturgis, Purdue University College of Veterinary Medicine.

much evidence of enhancing the nuclear localization of the dye, and the nine-member Nlys peptoid showed an equal or better proclivity for inducing cellular entry when compared to TAT. This may not be surprising, though, given that the efficiency of uptake of CPPs can differ between cell lines [449]. Concordant with previous studies, Nlys peptoid entry was characterized by a punctate intracellular distribution, being present inside vesicles. Because the amino-based peptoids did not show any evidence of nuclear uptake, it was decided to pursue the use of poly-arginine peptoids,

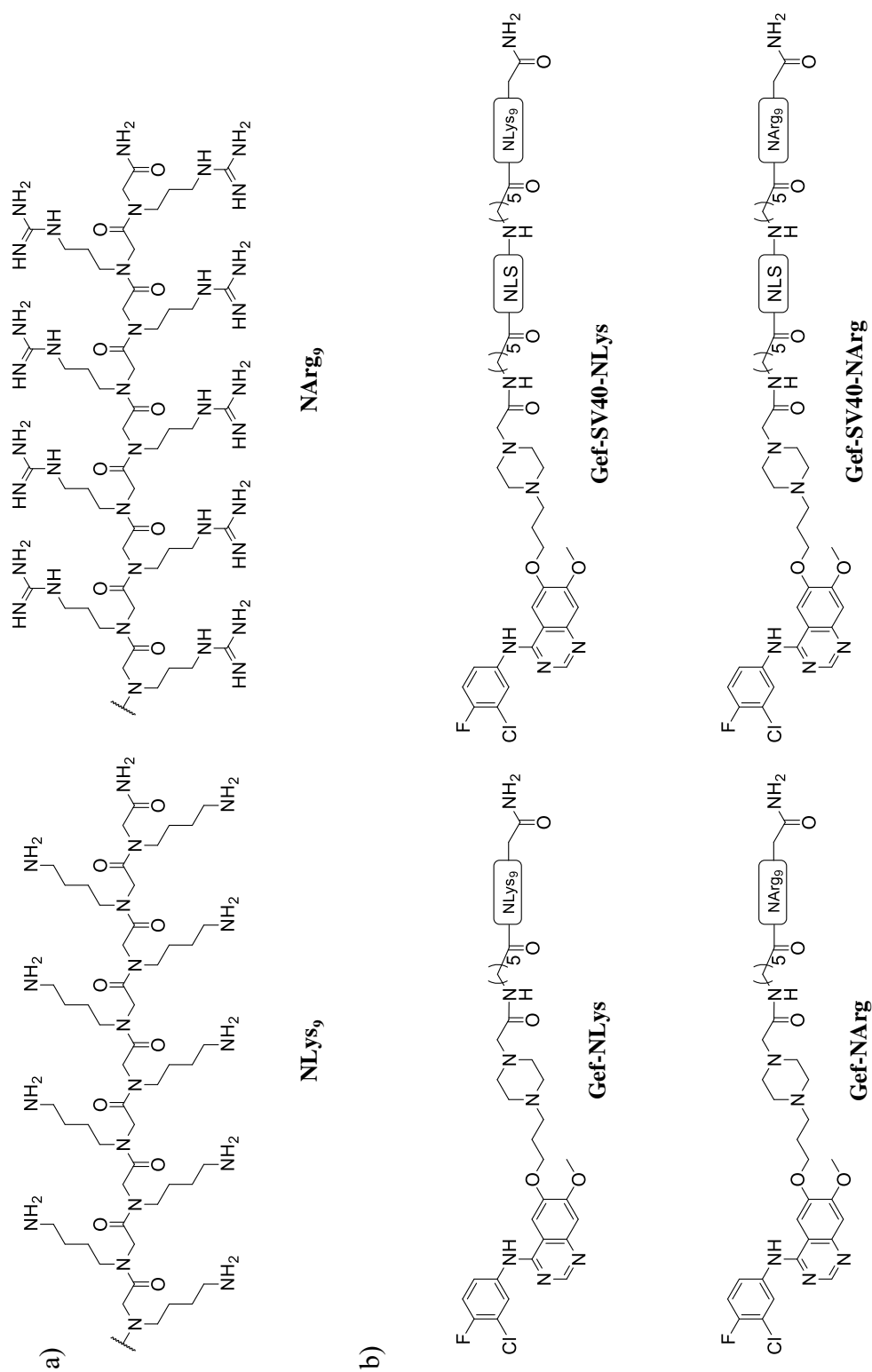


Figure 3.2. Cell Penetrating Peptide Sequences and Peptide-Peptoid Drug Conjugates Used in This Study. (a) Each peptoid-based compound consists of a 9-member amino- (NLys) or guanidinium- (NArg) based polycationic sequence. (b) For compounds containing a conjugated drug, piperazinyl gefitinib was coupled to either to a peptide-peptoid hybrid consisting of SV40 (NLS) and NLys₉ or NArg₉, or simply an NLys₉ or NArg₉ peptoid sequence.

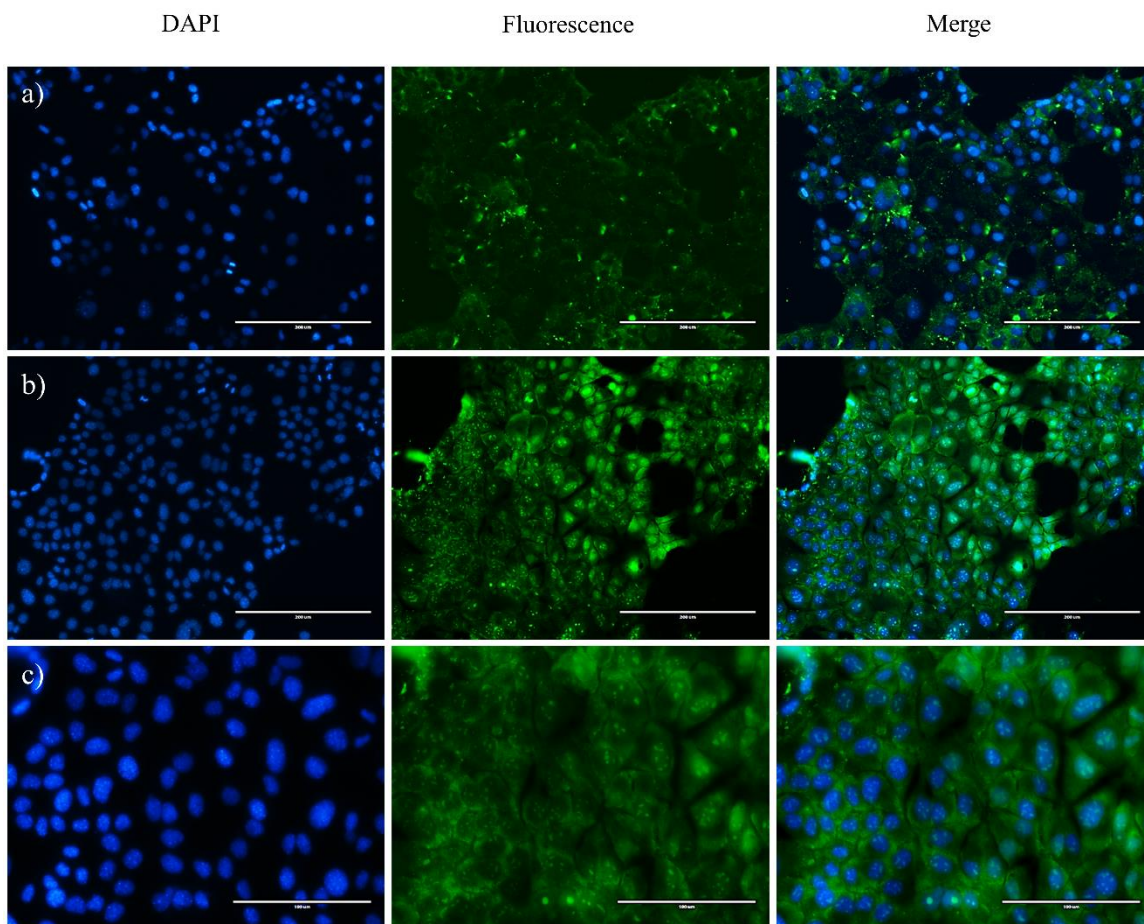


Figure 3.3. NArg-Peptoid Conjugates Enhance Nuclear Uptake. Normal Mammary Epithelial (NME) cells were incubated with either 10 μM FAM-NArg or FAM-SV40-NArg for one hour, fixed and stained with DAPI. The cells were then imaged by confocal microscopy. (a) FAM-NArg at 20X magnification (bars = 200 μm); (b) FAM-SV40-NArg at 20X magnification (bars = 200 μm); (c) FAM-SV40-NArg at 40X magnification (bars = 100 μm). Images generated by Wells Brown, Purdue University College of Pharmacy.

given that they should theoretically increase the nuclear distribution of a tagged molecule [437]. Based on the observations of peptoid length-dependence on the efficiency of cellular entry, guanidinium-based compounds were also synthesized with a length of nine residues (Figure 3.2a).

With the goal of further enhancing nuclear uptake, peptoids were also synthesized containing the NLS from the simian virus 40 large T antigen (sequence: PKKKRKV). This sequence, which is naturally found on the surface of certain proteins, promotes nuclear transport through its recognition by importins, taking it through the nuclear pore complex [450]. The SV40 sequence was selected due to its relative ease of synthesis when compared to TAT. Following the

submonomer synthesis [300] of each respective CPPo on Rink-amide resin, the NLS was assembled using basic Fmoc-based peptide synthesis conditions. The NLS and CPPo were separated by a six-carbon linker; an additional six-carbon spacer was placed on the N-terminus of the NLS, to which either piperazinyl gefitinib (Pip Gef) or FAM would be attached (Figure 3.2b). To examine the cellular distribution of the FAM-conjugated NArg₉-containing molecules, normal mammary epithelial (NME) cells, which were engineered to express elevated levels of EGFR, were exposed to either FAM-NArg or FAM-SV40-NArg (Figure 3.3). Both FAM-NArg and FAM-SV40-NArg were efficiently taken into the epithelial cells. From the results of imaging, FAM-NArg was distributed inside the cytoplasm (Figure 3.3a), but the addition of the SV40 sequence seemed to enhance the nuclear, and overall, uptake of these compounds, as expected (Figures 3.3b & c).

3.4.1 NLys-Based Conjugates Enhance the Anti-Proliferative Effect of Gefitinib in Resistant Cells

Although Pip Gef is nearly identical to commercial gefitinib, with the exception of a piperazinyl acetate moiety in place of the morpholino group, minor changes to its structure could have substantial consequences for its affinity for EGFR, or could cause other issues due to changes in its cellular uptake or solubility. To assess the two compounds for any difference in activity, commercial gefitinib and Pip Gef were compared in 3D cell cultures of NME and LM1 cells (Figure 3.4). In both cells lines, gefitinib efficiently downregulated cell growth. In NME cells, PIP Gef did not show as much of an antiproliferative effect, though at 10 μ M the effect was similar to that of gefitinib. In LM1 cells, however, the difference between gefitinib and Pip Gef was substantial, with Pip Gef having comparable activity to the DMSO control. Because of the reduced potency of Pip Gef, the next step was to assess whether incorporating the drug into a peptide-peptoid conjugate could improve the observed activity. Due to the relative ease of synthesis and reduced cost of NLys-based peptoids as opposed to NArg, Gef-NLys and Gef-SV40-NLys were assessed for their ability to negatively impact the growth of the tumor-based cell lines MDA-MB-231 and MDA-MB-468.

Both the 231 and 468 cell lines are derived from mammary gland/breast tissue that has metastasized, and are classified as models for triple negative breast cancer (do not express estrogen receptor, progesterone receptor or Her2/neu). Each line overexpresses EGFR [451,452], and although 231 and 468 cells are relatively insensitive to gefitinib monotherapy when compared to other breast cancer lines, MDA-MB-231 is especially resistant to the drug, displaying two-fold or lower sensitivity than other triple negative cell types [453]. Furthermore, they show enhanced resistance to other anti-tumor drugs including erlotinib, cetuximab, carboplatin, doxorubicin and docetaxel. As anticipated, gefitinib had little effect on the growth of MDA-MB-231 cells, while

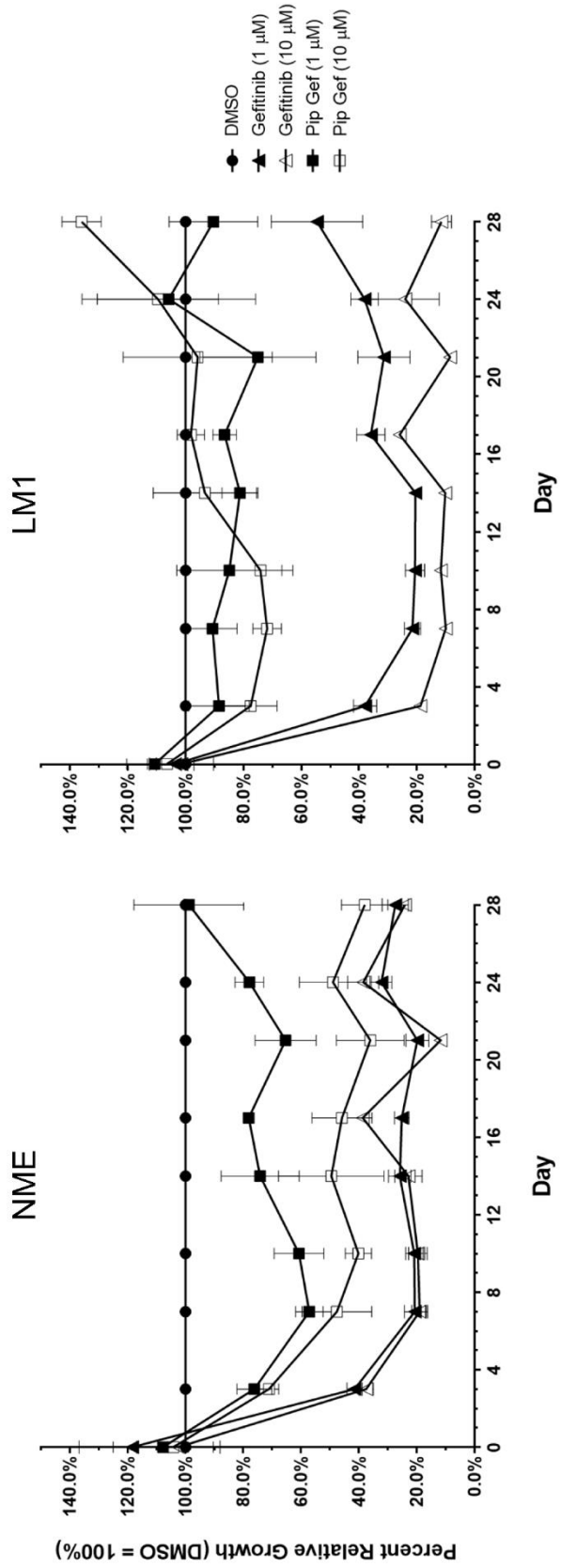


Figure 3.4. Activity of Commercial Gefitinib and Pip Gef in 3D Cell Cultures of NME and LM1. Luciferase-expressing cells were seeded on 96-well plates coated in Cultrex. Media containing gefitinib or Pip Gef, and/or EGF, was replaced every 4 days and organoid outgrowth was detected by the addition of D-luciferin potassium salt to induce bioluminescence, which was quantified as percent relative growth. Data was generated by Wells Brown, Purdue University College of Pharmacy.

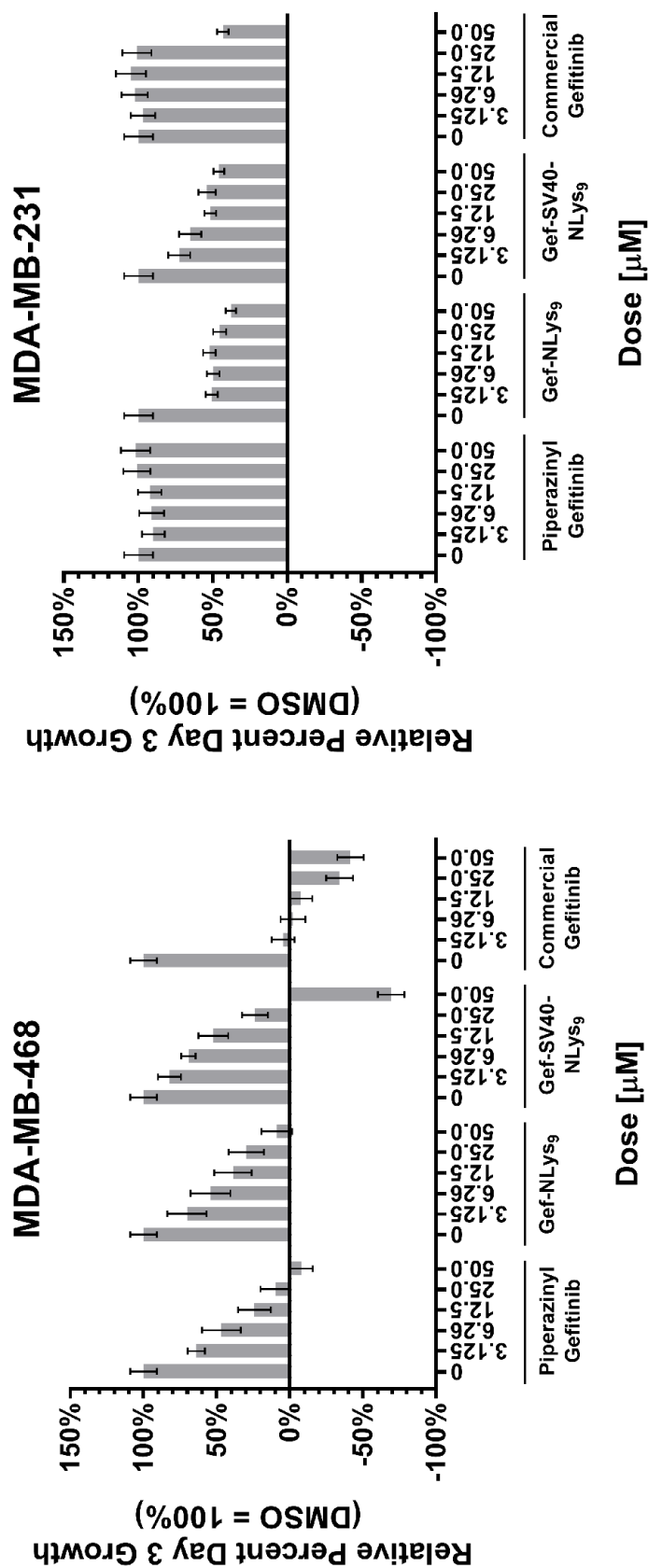


Figure 3.5. NLys-Based Drug Conjugates Negatively Impact Growth of Gefitinib-Resistant Cell Lines. MDA-MB-231 and -468 cell proliferation was assessed using the MTT (3-(4, 5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) cell proliferation reagent. Cells were plated on 96-well plates and were allowed to attach for 4 hours, followed by treatment with 5 concentrations of each compound and subsequent incubation at 37°C for 72 hours. 10 µL of MTT (5 mg/mL in DMEM phenol-red free media) was added to each well at a final concentration of (0.5 mg/mL) for 4 hours at 37°C. After the period of incubation, 100 µL of solubilization solution (10% Triton-X 100, acidic isopropanol (0.1N HCL)) was added and the plates were sealed and stored from light for 3 days. Absorbance was read at 570 nm and the percent cell growth was normalized by comparison to a control plate with no drug for each respective cell line. Data was generated by Ray Fatig, Purdue University College of Pharmacy.

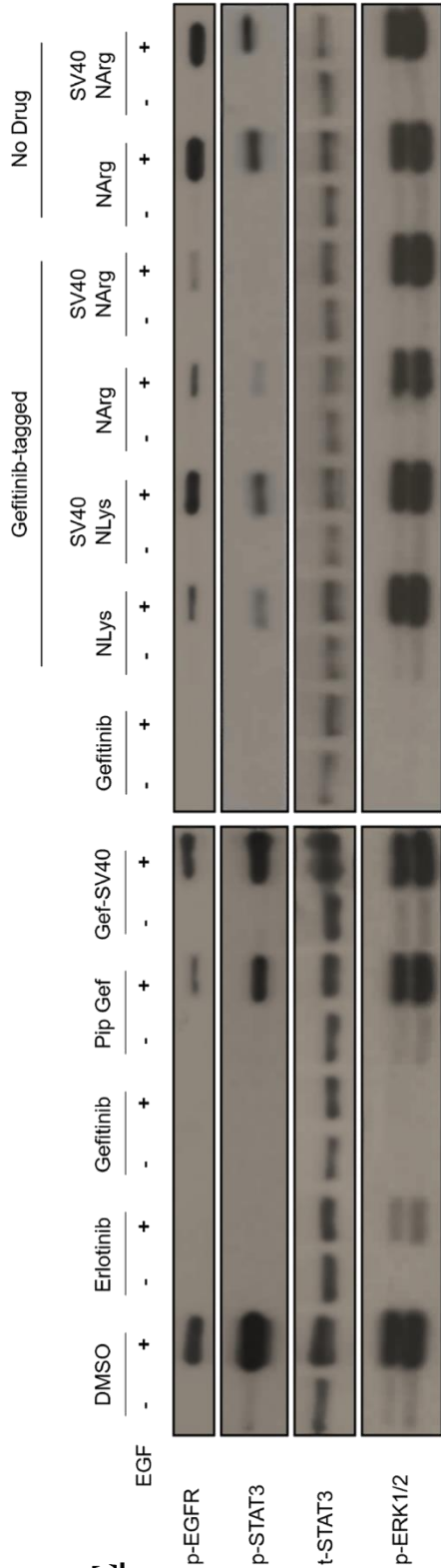
PIP Gef had no measurable activity after three days, even at concentrations up to 50 μM (Figure 3.5). However, when PIP Gef was coupled to either NLys or SV40-NLys, a significant decrease in cell growth was observed. At the same time, NLys and SV40-NLys peptoid sequences with no drug conjugate (Ahx-NLys and Ahx-SV40-NLys, respectively) did not show any antiproliferative activity (data not shown). In MDA-MB-468 cells, Gef-NLys and Gef-SV40-NLys also negatively impacted growth—Gef-NLys had comparable activity to Pip Gef, though all of the test compounds were inferior to gefitinib, with the apparent exception of Gef-SV40-NLys at a concentration of 50 μM . Given that NLys-based compounds do not appear to get taken into the nucleus (Figure 3.3), the ability of the peptoids to improve the potency of gefitinib in resistant cells may be due to enhanced cellular uptake, though it is not yet mechanistically clear why some cell lines would be more sensitive to this effect than others.

3.4.2 Drug-Peptoid Conjugates Alter the Phosphorylation Status of STAT3

Because the peptoid conjugates demonstrated that they are taken into cells, and that they produce a phenotypic growth effect, the next step was to examine the effect of gefitinib-tagged conjugates on EGFR as well as downstream intracellular proteins. The phosphorylation statuses of EGFR, extracellular-signal-regulated kinases 1/2 (ERK1/2) and signal transducer and activator of transcription 3 (STAT3) were probed to observe the effect of the conjugates on signaling pathways known to be activated in part by EGFR. ERK proteins exist within the family of classical MAP kinases, and are involved in various intracellular signaling processes that regulate mitosis, meiosis and postmitotic events in cells. They are activated by a plethora of upstream stimuli, including growth factors, ligands for heterotrimeric GPCRs, cytokines, mitogens and carcinogens [454]. ERK is naturally located in the cytoplasm, where it can translocate to the nucleus upon phosphorylative activation. STAT3, on the other hand, which is a transcription factor that is activated in response to various cytokines and growth factors, is able to translocate into and out of the nucleus independently of phosphorylation, and maintains a prominent nuclear presence [455,456]. Furthermore, evidence exists that STAT3 physically interacts with nuclear EGFR, leading to the activation of its transcriptional control [457,458].

To observe the effect of the conjugates on EGFR, ERK1/2 and STAT3, two murine tumor cells lines, termed NME and LM1, were exposed to either gefitinib or erlotinib (Tarceva), or one of the test compounds. As previously discussed, NME cells overexpress EGFR in the whole cell; in LM1 cells, however, the level of EGFR is near normal, but the protein is disproportionately

NME



LMI

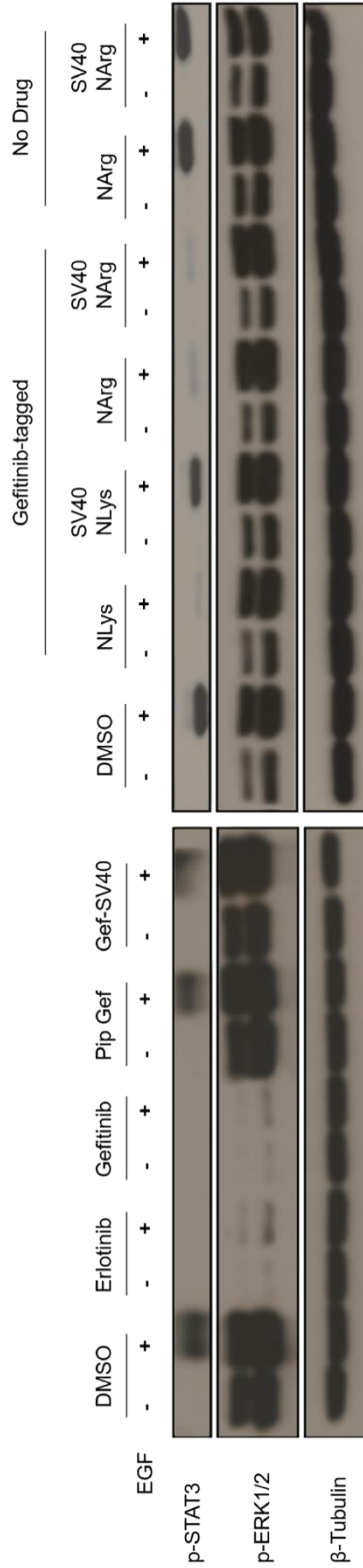


Figure 3.6. Effect of Nuclear-Targeted Conjugates on Phosphorylation Status of EGFR, ERK1/2 and STAT3. NME (top) and LMI (bottom) cells were incubated with erlotinib, gefitinib or the NLys- and NArg-based drug conjugates at concentrations of 1 μM in the presence or absence of EGF (25 ng/mL). Cells were lysed and probed for p-EGFR, p-STAT3 (residue Y705), p-ERK1/2, total STAT3 (t-STAT3) and/or β-tubulin. Data generated by Wells Brown, Purdue University College of Pharmacy.

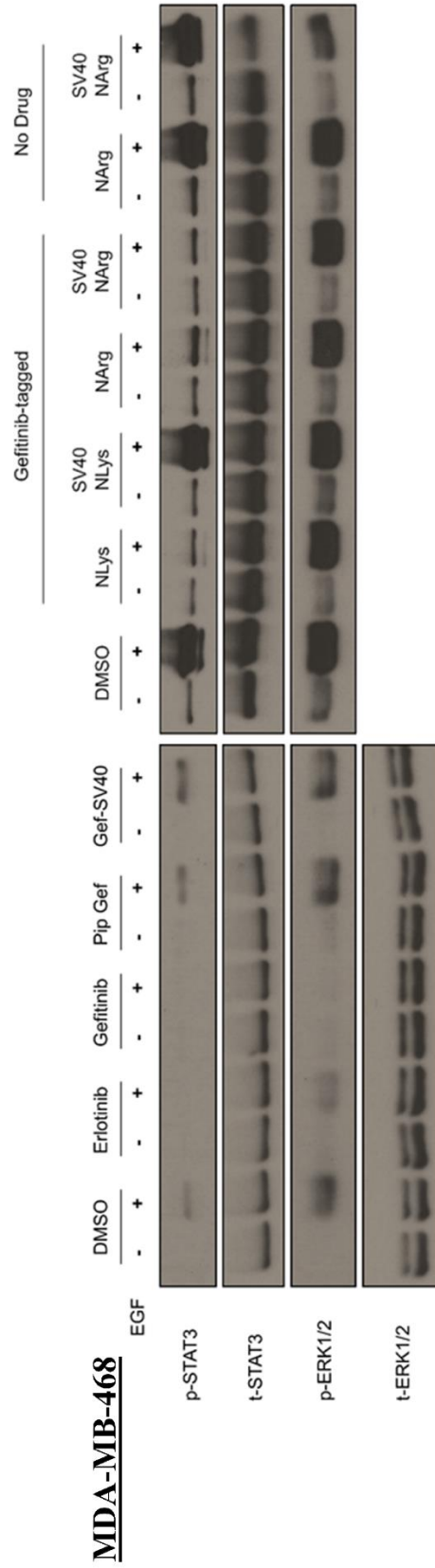


Figure 3.7. Analysis of the Phosphorylation Status of ERK1/2 and STAT3 in MDA-MB-468 Cells. As performed previously with the NME and LM1 lines, cells were incubated with erlotinib, gefitinib or the NLys- and NArg-based drug conjugates at concentrations of 1 μ M in the presence or absence of EGF (25 ng/mL). Cells were lysed and probed for p-STAT3 (Y705), total STAT3 (t-STAT3), p-ERK1/2 and total ERK1/2 (t-ERK1/2). Data generated by Wells Brown, Purdue University College of Pharmacy.

located in the nucleus, as opposed to the rest of the cell (see Appendix B, Figure B.1). In the NME line, cell lysates were probed for the phosphorylation status of EGFR (p-EGFR), STAT3 (p-STAT3) and ERK1/2 (p-ERK1/2), with the total level of STAT3 (t-STAT3) used as both a loading and internal control (Figure 3.6 (*top*)). Erlotinib and gefitinib, both clinically established inhibitors of EGFR, abrogated the phosphorylation of EGFR and STAT3; both antagonists had a negative effect on ERK1/2 phosphorylation, with gefitinib nearly completely abolishing it. Pip Gef had relatively little effect on the phosphorylation of EGFR, and no noticeable effect on STAT3 or ERK1/2. It was also observed that attaching the SV40 nuclear targeting sequence to Pip Gef seemed to hurt the activity of the drug. Interestingly, the NLys- and NArg-based drug conjugates had a more profound effect on STAT3 phosphorylation than on EGFR, and each of the conjugates did not have a measurable effect on ERK1/2. Of the compounds, Gef-SV40-NArg had the overall most significant impact. LM1 cell lysates were also probed for p-STAT3 and p-ERK1/2 (Figure 3.6 (*bottom*)), and as with the other cell line, NLys- and NArg-based drug conjugates disrupted phosphorylation of STAT3, but had little effect on ERK1/2.

Because murine cells lines showed a differential effect of the compounds on STAT3 and ERK1/2, the human cell line MDA-MB-468, derived from a mammary gland adenocarcinoma, was investigated (Figure 3.7). As was seen with the NME and LM1 lines, the NLys- and NArg-based compounds decreased the levels of p-STAT3, but did not affect p-ERK1/2. Moreover, Gef-SV40-NArg was again the most potent compound. To assess the distribution of phosphorylated STAT3 in the cell, MDA-MB-468 cells were cultured in the presence of EGF or bovine serum albumin (BSA), and were imaged using immunofluorescence (Figure 3.8). t-STAT3 showed distribution throughout the entire cell, both inside and outside the nucleus. When the cells were treated with DMSO, p-STAT3 appeared to be restricted to the nucleus; but on exposure to Gef-SV40-NArg, the cells showed little to no trace of STAT3 phosphorylation. Overall, it was not clear from the results of the previous assays why Gef-SV40-NLys had little to no effect on EGFR, STAT3 and ERK1/2. A potential explanation could be the aggregation or formation of unanticipated secondary/tertiary structures in solution, but this issue was not pursued any further.

3.4.3 NArg-Based Compounds Disrupt STAT3 Phosphorylation in a Dose-Dependent Manner

Based on the results from the Western blots, Gef-NArg and Gef-SV40-NArg were the most effective compounds at disrupting STAT3 phosphorylation. Given that commercial gefitinib effectively downregulated p-STAT3 as well as p-ERK1/2, it was investigated whether the NArg-

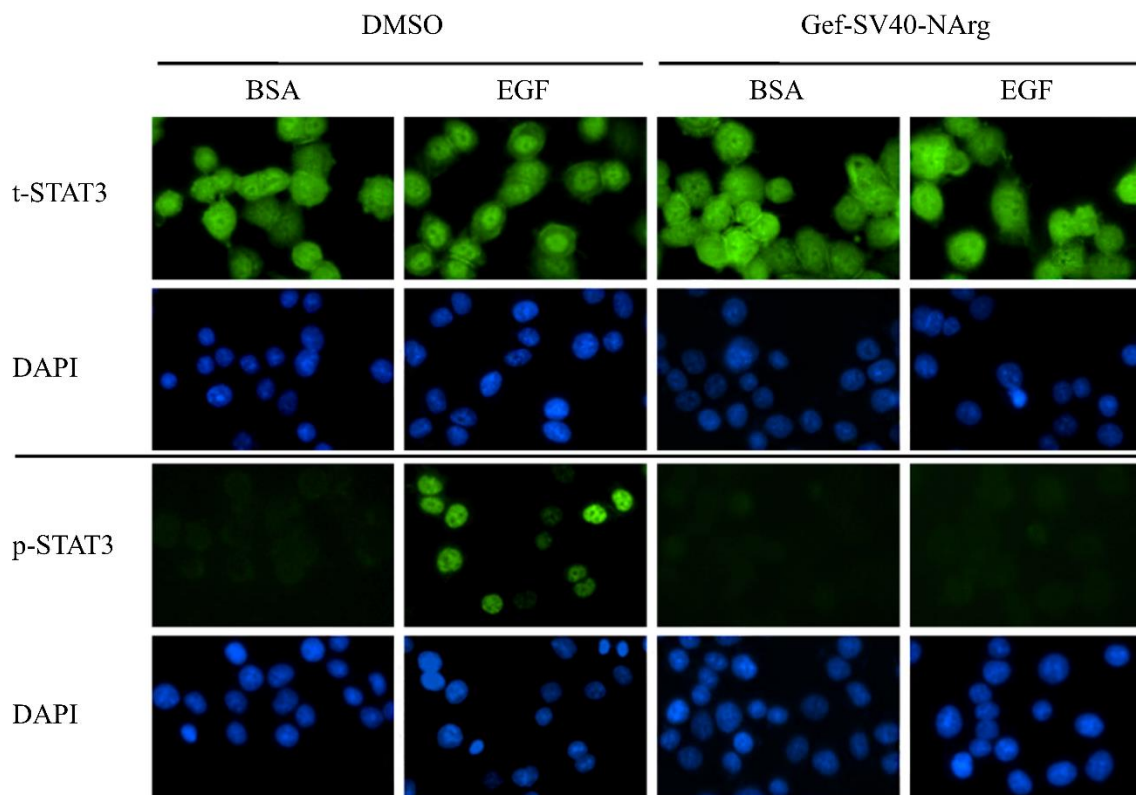


Figure 3.8. Gef-SV40-NArg Disrupts Nuclear Accumulation of Phosphorylated STAT3. MDA-MB-468 cells were seeded onto glass coverslips and cultured overnight in the presence of DMSO or Gef-SV40-NArg (1 μ M), and BSA or EGF. The coverslips were washed with PBS, the cells fixed with PBS plus 4% formaldehyde for 20 minutes at room temperature, and then the cells permeabilized with PBS plus 0.1% Triton X-100 for 20 minutes at room temperature. The cells were washed with PBS, followed by PBS plus 2% BSA, and were incubated overnight at 4°C with primary antibody (anti-STAT3 or -p-STAT3) in PBS plus 2% BSA. After incubation, the cells were washed and again incubated for one hour at room temperature with secondary antibody (donkey anti-mouse Alexa Fluor® 488) in PBS plus 2% BSA. The cells were washed with PBS plus 2% BSA, PBS alone, followed by incubation with DAPI in PBS for 5 minutes at room temp. The cells were washed in PBS three more times, then mounted to a glass slide using Prolong Gold anti-fade reagent. Images were acquired using a Nikon A1R-MP confocal microscope. Images generated by Wells Brown, Purdue University College of Pharmacy.

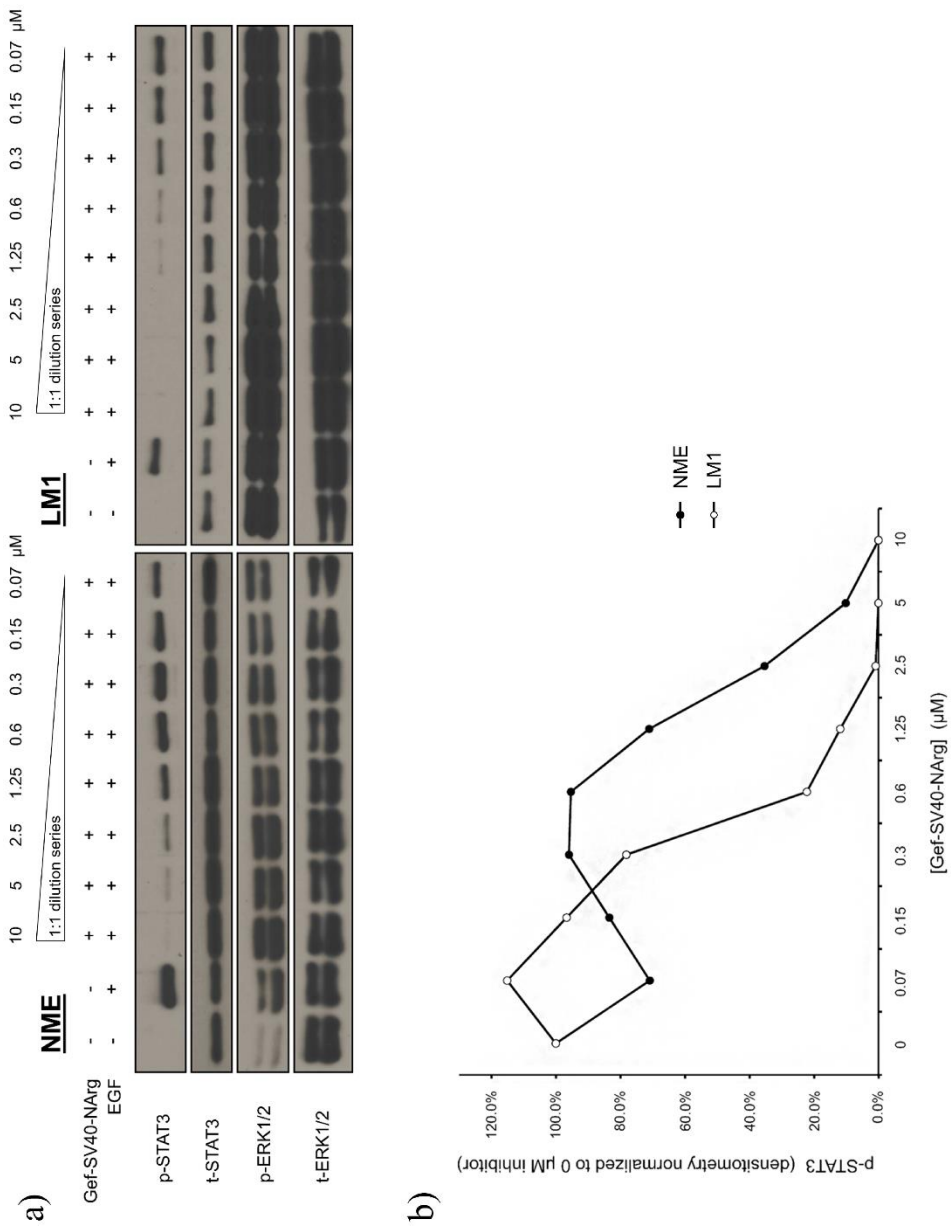


Figure 3.9. Dose Response of Gef-SV40-NArg against STAT3 and ERK1/2 in NME and LM1 Cells. NME and LM1 Cells. NME and LM1 cells were exposed to increasing concentrations of Gef-SV40-NArg in the presence of EGF (25 ng/mL). (a) Cell lysates were probed for p-STAT3 (Y705), t-STAT3, p-ERK1/2 and t-ERK1/2. (b) Densitometry analysis was performed for each Western blot using the Biorad Chemidoc touch imaging system to determine relative levels of p-STAT3 in each cell line at given concentrations of the drug conjugate. Data generated by Wells Brown, Purdue University College of Pharmacy.

based drug conjugates would effect ERK1/2 phosphorylation when the concentration of each compound was increased. NME and LM1 cells were exposed to two-fold increasing concentrations of Gef-SV40-NArg up to 10 μ M (Figure 3.9a). In each line, the level of p-STAT3 decreased in a dose-dependent manner, while p-ERK1/2 was not affected. The results also indicated that in LM1 cells, which have upregulated levels of nuclear EGFR, p-STAT3 was more sensitive to modulation upon exposure to the drug conjugate than in NME cells. A densitometry analysis of the amount of p-STAT3, normalized to the amount of total STAT3 in each cell line, further demonstrated the higher sensitivity of LM1 cells to Gef-SV40-NArg (Figure 3.9b).

In the previous set of experiments, at 1 μ M gefitinib eliminated the phosphorylation of STAT3 and ERK1/2. However, it was unknown whether p-STAT3 would be fully abrogated before p-ERK1/2 levels were affected at lower concentrations. To investigate this possibility, MDA-MB-468 cells were exposed to Gef-NArg, Gef-SV40-NArg or gefitinib in a two-fold series of concentrations up to 1 μ M for gefitinib, or 5 μ M for each peptoid conjugate (Figure 3.10). From the results, it was seen that gefitinib nearly eliminated p-STAT3 at a concentration of approximately 125 nM, and increasing gefitinib four-fold to 0.5 μ M abrogated p-ERK1/2. Gef-NArg and Gef-SV40-NArg both disrupted p-STAT3, with the latter abolishing the protein's phosphorylation at between 150 and 300 nM. However, even when the concentration of Gef-SV40-NArg was increased more than ten-fold to 5 μ M, there was no evidence of p-ERK1/2 being affected. This could suggest that gefitinib and the peptoid-drug conjugates are acting through different mechanisms. If the difference in STAT3 phosphorylation between gefitinib and Gef-SV40-NArg was simply due to the differential affinity of the two compounds for EGFR, it would be expected that as the concentration of the NArg-based ligand continued to be increased, levels of p-ERK1/2 would eventually begin to decrease. In effect, a dose response curve of the peptoid compounds would be shifted to the right, relative to gefitinib. However, this was not observed, as gefitinib inhibited p-ERK1/2 at four times the concentration of p-STAT3 inhibition, but Gef-SV40-NArg, the most potent drug conjugate, did not inhibit p-ERK1/2 at even more than ten times the concentration of p-STAT3 abrogation.

3.4.4 Drug-Peptoid Conjugates Downregulate EGFR-Dependent, but not JAK-Dependent STAT3 Phosphorylation

As discussed, STAT3 is a transcriptional activator that can translocate into and out of the nucleus, where it mediates the expression of a variety of genes. In a canonical sense, STAT3 is most often associated with activation by Janus Kinase 1 (JAK1), but it can also be activated through either

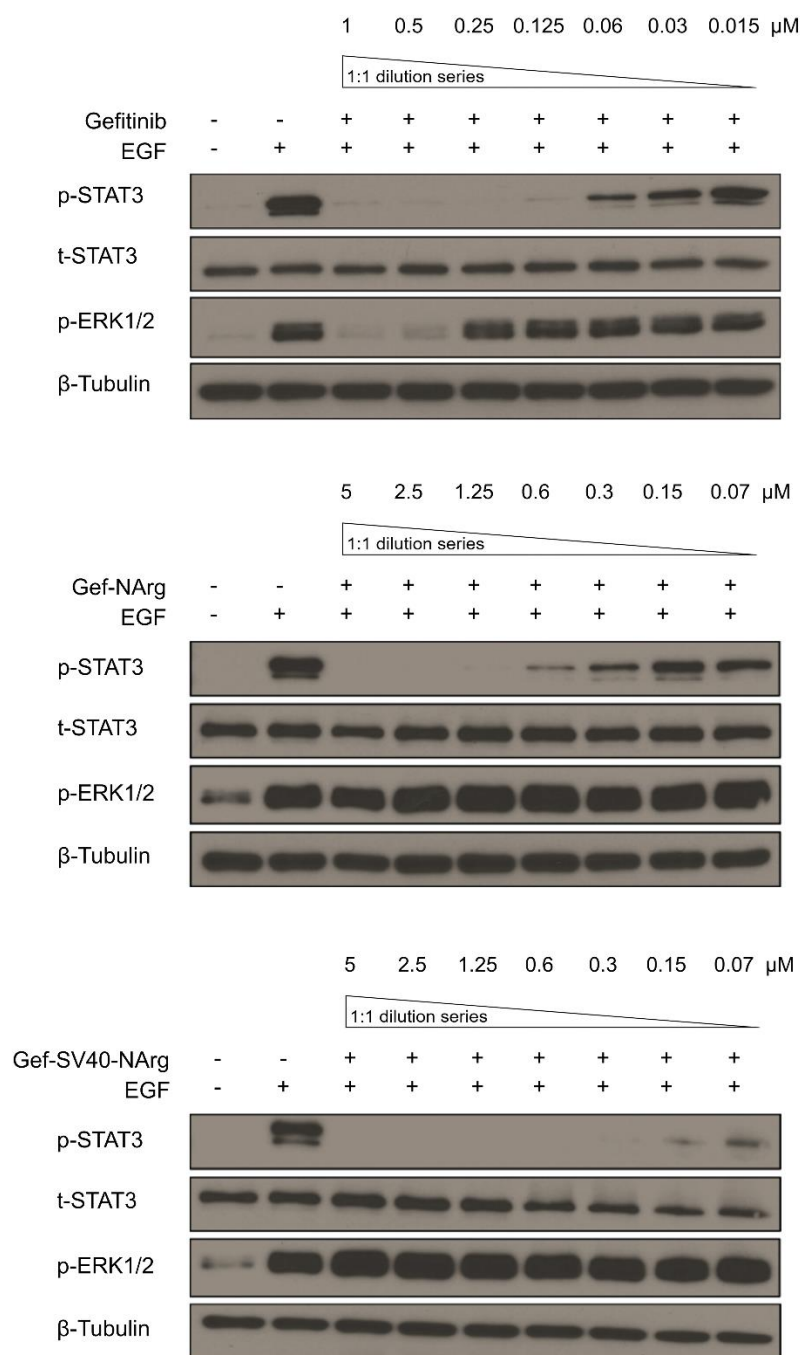


Figure 3.10. Dose Response of NArg-Based Drug Conjugates in MDA-MB-468 Cells. 468 cells were exposed to increasing concentrations of gefitinib, Gef-NArg or Gef-SV40-NArg in the presence of EGF (25 ng/mL). Cell lysates were probed for p-STAT3 (Y705), t-STAT3, p-ERK1/2 and β -tubulin. Data generated by Wells Brown, Purdue University College of Pharmacy.

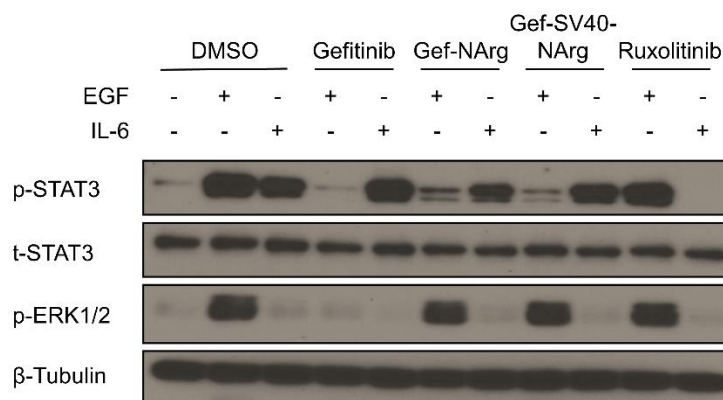


Figure 3.11. Drug-Peptoid Conjugates Downregulate EGFR-Dependent Phosphorylation of STAT3. MDA-MB-468 Cells were cultured in the presence of either EGF (25 ng/mL) or IL-6 (20 ng/mL), followed by exposure to 1 μ M Gef-NArg, Gef-SV40-NArg, gefitinib or the JAK1 inhibitor ruxolitinib. Cell lysates were probed for p-STAT3 (Y705), t-STAT3, p-ERK1/2 and β -tubulin. Data generated by Wells Brown, Purdue University College of Pharmacy.

direct interaction with EGFR or downstream pathways of EGFR [457–460]. To investigate whether the drug-peptoid conjugates were affecting JAK-dependent STAT3 phosphorylation, MDA-MB-468 cells were exposed to either EGF or IL-6 to activate EGFR- or JAK-dependent signaling cascades, respectively. Gef-NArg, Gef-SV40-NArg, gefitinib and the JAK1 inhibitor ruxolitinib were added to determine the effect on levels of p-STAT3 and p-ERK1/2 (Figure 3.11). As before, in the presence of EGF Gef-NArg, Gef-SV40-NArg and gefitinib disrupted the phosphorylation of STAT3, while only gefitinib affected the levels of p-ERK1/2. Conversely, in the presence of IL-6, none of the compounds had an effect on p-STAT3, with the exception of ruxolitinib. Additionally, IL-6 exposure did not result in ERK1/2 phosphorylation, indicating that STAT3 and ERK1/2 do not associate with one another under the conditions of interleukin-induced JAK activation.

3.4.5 Gef-SV40-NArg Has Strong Potential for Oral Absorption

To investigate whether the drug conjugates might make good drugs in and of themselves, a permeability study was performed with Gef-SV40-NArg to see if it was able to be absorbed through the gastrointestinal tract. Caco-2 epithelial colorectal adenocarcinoma cells were cultured in DMEM and plated as a single layer in Corning 12-well 0.4 μ m polyester Transwell filter supports (see Appendix B for experimental methodology). In order to ensure an adequate acceptor

compartment concentration, the ligand concentration was kept well over 100-fold higher than the limit of detection. Gef-SV40-NArg permeability was measured with an initial concentration of 100 μM in 1% DMSO/HBSS in triplicate samples. All samples gave an area under the curve well above background, and future studies could likely be conducted over a shorter time scale or up to 10-fold lower concentration. A significant lag period of between 30 and 60 minutes was seen; however, linearity was stable after 60 minutes. This lag could be caused by slow permeation and equilibration across the cell monolayer, but a jump in dC/dt slope can also be indicative of cell toxicity. It should be noted that while Caco-2 cells are immortalized, they are contact-inhibited, and monolayers are fully differentiated and no longer dividing at the time of study. Permeability results for Gef-SV40-NArg are shown below:

Table 3.3
Permeability of Gef-SV40-NArg in
Caco-2 Cells

Triplicate 1	$6.72 \times 10^{-6} *$
Triplicate 2	$6.29 \times 10^{-6} *$
Triplicate 3	$5.98 \times 10^{-6} *$
Average	6.33×10^{-6}
Deviation	3.68×10^{-7}

* Values are reported as the permeability across the cell monolayer in cm/s

(Experiment performed by Christopher Kulczar, Purdue University College of Pharmacy)

In the assay, the mass balance for the drug conjugate was nearly 100%, indicating that the compound remained stable throughout the experiment. Based on the results, it was concluded that the permeability rates were considered to be in the middle range, suggesting that the drug conjugate would have reasonable absorption through the gastrointestinal tract when ingested orally.

3.5 Discussion

This study establishes a strategy for subcellularly targeting a small molecule that aims to inhibit receptors at the level of the nucleus. The direction of molecules to intracellular compartments has been studied for at least a couple decades, but this work reports, to our knowledge, the first instance of incorporating a tyrosine kinase inhibitor into a NLS or CPPo. This same approach is versatile in that it could theoretically be used to target numerous small molecules that are agonists/antagonists of various protein targets. Both the peptide and peptoid portions of these compounds are synthesized without much difficulty, and chemical modification of the small molecules would allow for facile conjugation to the sequences using basic conditions for solid-phase peptide synthesis.

The NLS and CPPo sequences used in this work were selected based on their success in trafficking small molecule fluorescent dyes into the cell [437–439]. The design of these compounds to simultaneously contain both the peptide-based NLS and peptoid-based CPPo proved to enhance their cellular uptake and activity, but the physical structure of the drug conjugates, as a whole, may not be ideal for every situation. While the six-methylene linkers provide spatial separation of approximately 10 Å between the drug and peptide/peptoid portion of these molecules, it is possible that the targeting sequence still destabilizes the binding of the small molecule to its target due to steric clashing between the sequence and the protein. Other methods of attachment could be explored such as either disulfide or thioether-based covalent linkages [429], which could become reduced upon cellular uptake, releasing the drug. This, however, may not be an ideal alternative for targeting the nucleus since the constrained timeframe of cellular machinery transporting the ligands to the nucleus would necessitate a non-labile linker.

A major finding of this work is the observation that the drug conjugates downregulate phosphorylation of STAT3, but do not affect ERK1/2, whereas commercial gefitinib abrogates p-ERK1/2 as cells are exposed to increasing concentrations of the drug. Both ERK1/2 and STAT3 can be found downstream of EGFR, but the involvement of each in cellular processes is quite different (see Figure 3.12). ERK1 and ERK2, also referred to as MAPK3 and MAPK1, respectively, are both part of the well-studied MAPK/ERK pathway that acts as a mediator of signal transduction from the surface of cells to the nucleus. ERK1/2 are found in the cytoplasm until they become activated, upon which they dimerize and pass into the nucleus. STAT3, on the other hand, can be found in both the cytoplasm and nucleus, as it can translocate back and forth independent of its phosphorylation [455,456]. In order to perform its role as a transcriptional factor, it becomes

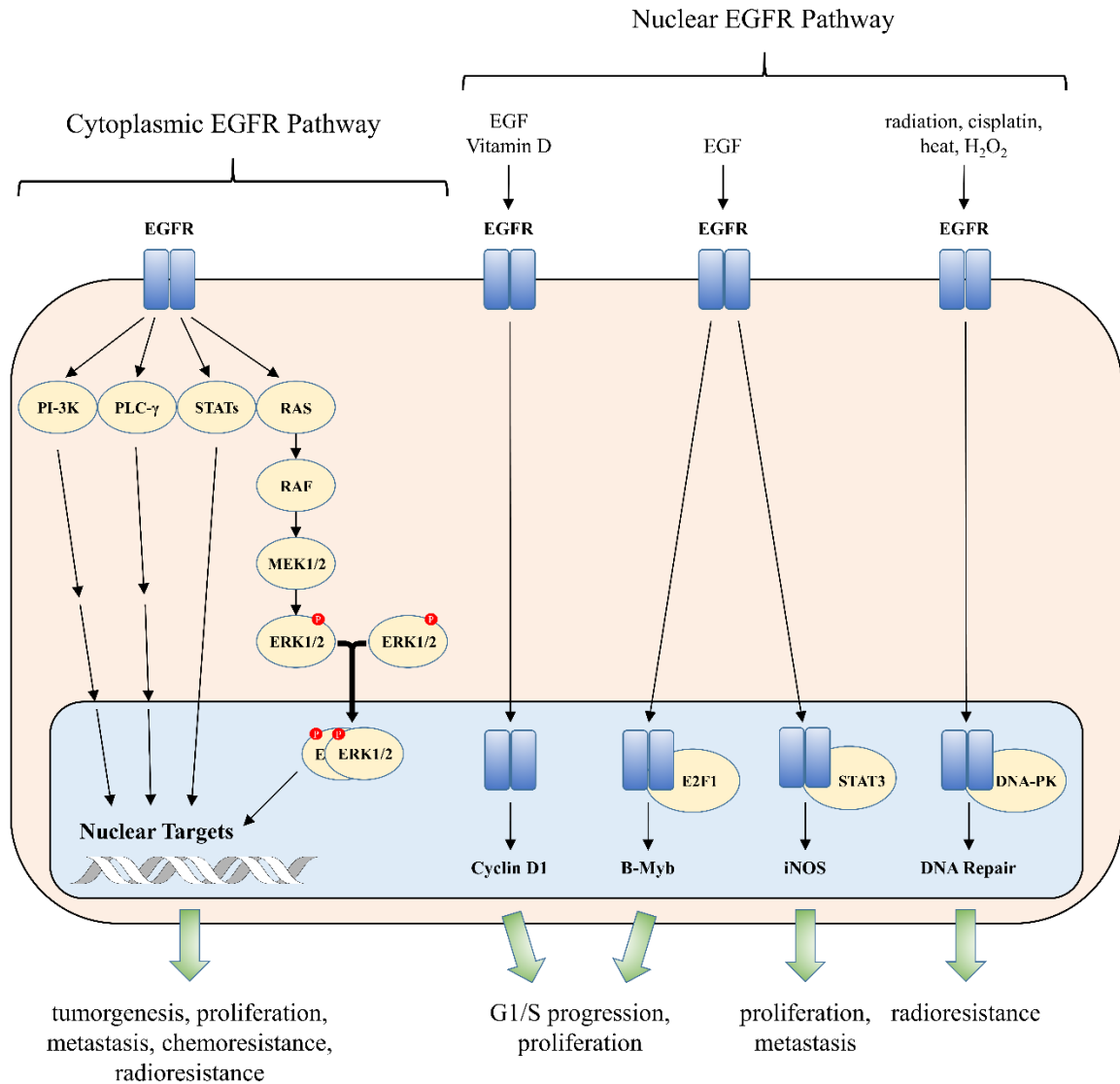


Figure 3.12. Cytoplasmic and Nuclear Modes of EGFR Signaling. Signaling originating from EGFR propagates through the cell via two major pathways—one occurring in the cytoplasm, and the other in the nucleus. Upon EGFR activation at the cell surface, EGFR either: 1) activates downstream signaling proteins such as PI-3K, PLC-γ, one of the STATs, or RAS; these subsequently activate further downstream proteins, causing a signal cascade that eventually reaches targets in the nucleus, activating various cellular growth- or survival-related processes. 2) translocates to the nucleus where it binds with and activates proteins that then lead to the initiation of cellular responses such as checkpoint regulation, proliferation and DNA repair.

activated through phosphorylation at residue Y705 in response to a number of cytokines and growth factors, and has been identified as one of the two most important members of the STAT protein family for cancer progression [461–463]. Unlike ERK1/2, there is evidence that STAT3 physically associates and colocalizes with EGFR in the cell nucleus following EGF stimulation [457]. This is a compelling observation since it would imply that it is possible that the drug conjugates downregulate STAT3 in the nucleus of each of the cell lines.

Interestingly, in the context of downregulating STAT3 phosphorylation, LM1 cells displayed a higher sensitivity to Gef-SV40-NArg than did NME cells (Figure 3.9). As previously mentioned, as opposed to what is seen in NME cells, EGFR is upregulated in the nucleus of LM1 cells (Appendix B, Figure B.1), though the overall whole-cell expression level of the receptor is comparable to wild-type cells. This could provide an explanation for the enhanced sensitivity to downregulation of p-STAT3, but some investigation is still needed. As of yet, there is still no proof of direct inhibition of STAT3 activation by EGFR in the nucleus, though the evidence presented here could suggest that is occurring. Future experiments involving alternative subcellular tags or nuclear separation could help to answer that question.

Ultimately, the discovery of ligands that preferentially downregulate signaling proteins downstream of EGFR could have important implications for the treatment of certain cancers. STAT3 is an important signaling mediator in malignant disease, and it has been seen to be persistently activated in 22-65% of all non-small cell lung cancers [464–466]. Previous studies have demonstrated that suppressing signaling from EGFR-STAT3 [467], as well as STAT3 signaling in general [468–472], can result in tumor cell apoptosis in breast, melanoma, leukemia, myeloma and lung cancer types. Because constitutive IL-6R/JAK1/STAT3 signaling has been proposed as a major mechanism of resistance for EGFR-TKIs [473], using the nuclear-targeted conjugates in combination with a IL-6R/JAK1 pathway inhibitor could be an effective treatment for patients with TKI-resistant cancer [474]. Together with the observations that the gefitinib-conjugated molecules curtail the growth of gefitinib-resistant cells more efficiently than commercial gefitinib, and that they hold the potential for good absorptivity through oral administration, demonstrating that these compounds are able to efficiently target specific segments of signaling pathways could substantially increase their value as therapeutics for disease.

4. FUTURE DIRECTIONS

The search for new therapeutics to treat human disease is constant. Although there is a steady stream of new drug applications in the United States every year, the total number of these applications pales in comparison to that seen just a decade and a half ago. Due to the ballooning costs associated with the research and development of a new drug, which are further exacerbated by the tremendous subsequent investment necessary to take the drug to market, many entities have scaled back drug discovery efforts and generally focus on ‘safer’ targets. This unfortunately results in the vast majority of research focusing on a relatively small set of proteins [206], and the chemotypes present in many libraries used for high-throughput screening tend to be biased toward traditional drug targets such as receptor tyrosine kinases and enzymes [207–209]. The work presented here represents two unique strategies that diverge from prototypical drug discovery efforts in that they aim to provide new avenues for the discovery of new drug chemotypes. Ultimately, the initial discovery of compounds that inhibit PCNA-protein interactions or target subcellularly-located protein receptors lays a foundation for future work that could result in second or third generation compounds that are much improved, and could see broader application in the research & development or clinical spaces.

4.1 Using Chemical Information from Peptoid Hits to Develop More Drug-Like Inhibitors

In the analysis of the peptoids that inhibited the interaction between PCNA and the PL peptide, there were similar chemical features that were shared between the top hits. A general weakness of these compounds is that they each have many rotational degrees of freedom, which, among other features, does not make them ideal lead drug candidates. It would be advantageous if the features that contributed to the ligands’ individual affinities could be conserved, but incorporated into a more rigid drug-like structure (Figure 4.1). Chemistry currently exists that allows for the rigidification of peptoid structures on resin by causing ring closure of a portion of the peptoid’s backbone. This could be used, as an example, in the pursuit of next-generation lead molecules.

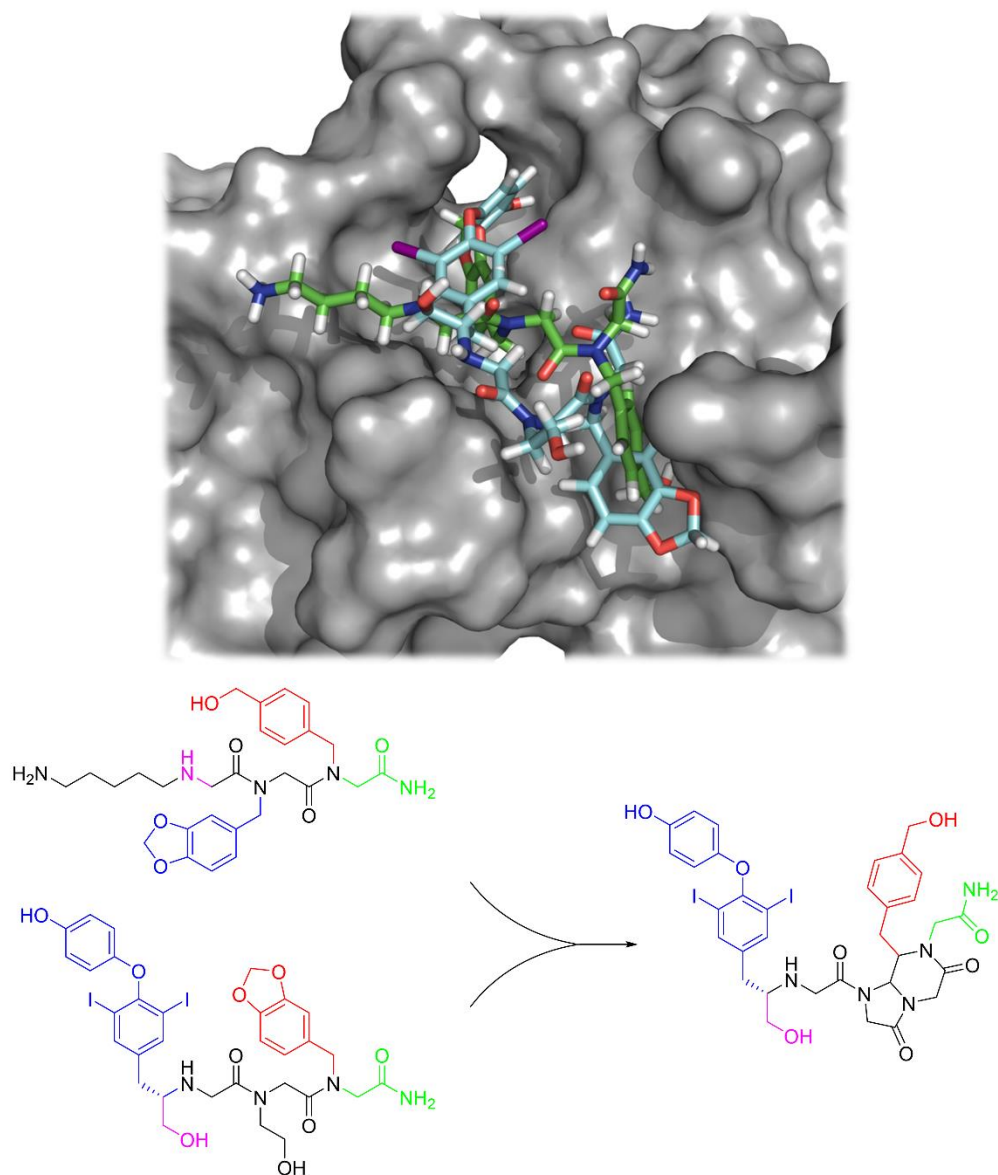


Figure 4.1. Shared Chemical Features Between Peptoid Hits Can Be Incorporated into Lead Compounds. (*top*) The average output positions of Nlys-NPip-NBal (green sticks) and T2AA-NEal-NPip (blue sticks) from their respective MD simulations with PCNA are shown overlaid. (*bottom*) Overlapping chemical features between the two ligands are color coded in blue, pink, red or green, and these can be incorporated into a more rigid, drug-like scaffold to create next-generation lead molecules.

Along with combining chemical features to make better lead antagonists, the peptoid side chain library could be expanded to include fragments that have been used in other various studies [302,305,312,318,320,475], as well as different chemotypes in general. With the additional 46 fragments shown in Figure 4.2, the total combinatorial library of tripeptoids would increase to more than 570,000 compounds, and more than 585,000 when T2AA is included as a potential substituent

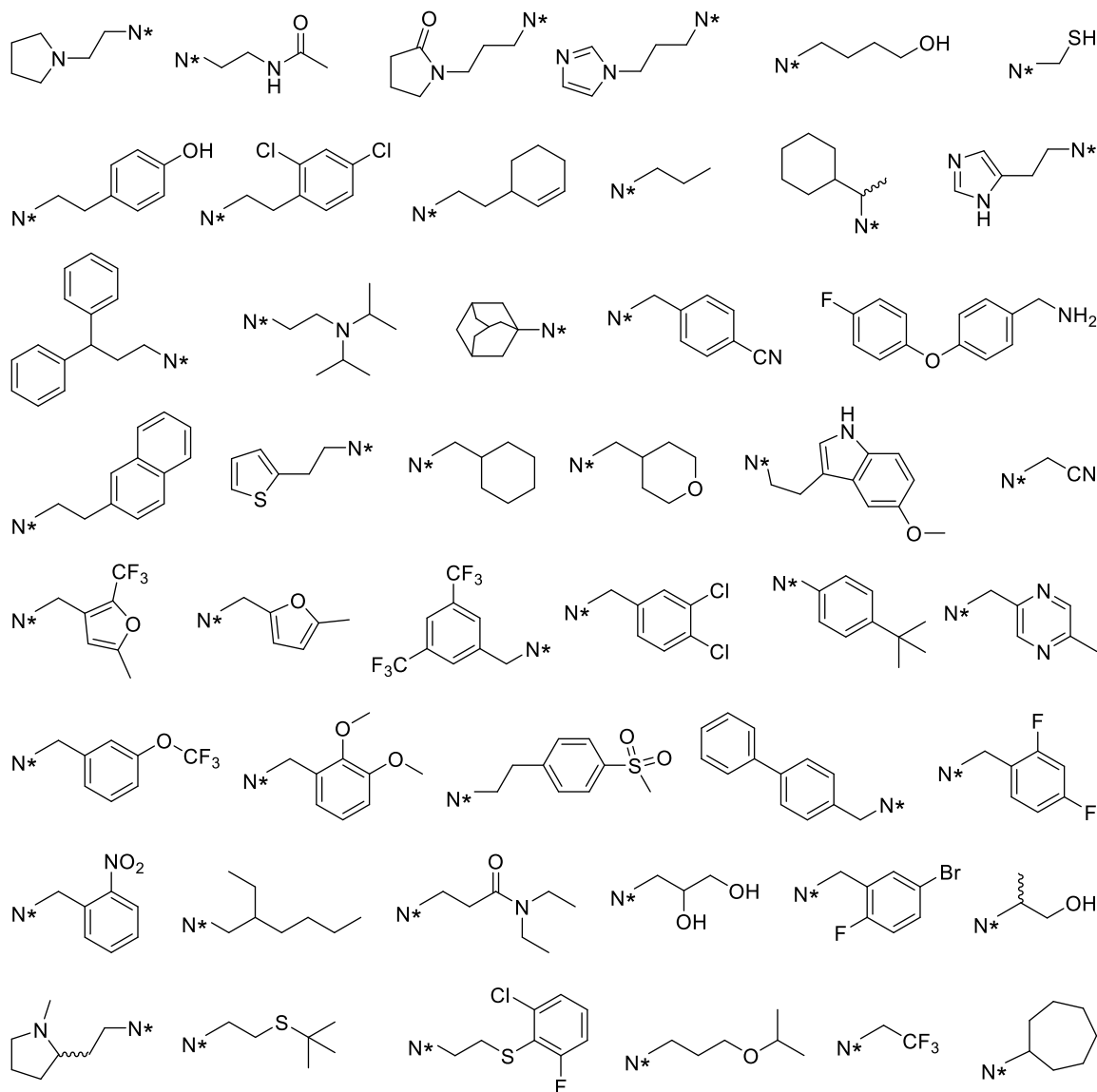


Figure 4.2. Set of Fragments for Incorporation into Future Virtual Libraries. In order to expand the chemical space covered by tripeptoid-based screening, peptoid fragments previously used in other studies can be incorporated into the virtual screening library.

in the first position. Given that their features are distinct from inhibitors of classical targets, and that they typically share certain properties with other members of their own class of compounds, it is likely that inhibitors of protein-protein interactions will not cover extensive chemical space. However, protein interaction surfaces can be quite diverse from one another, with substantial differences in polar and nonpolar surface area as well as size and depth of surface pockets. Covering a much larger fragment space for the purposes of using tripeptoids as a base for screening would be beneficial in this case giving that it would, of course, greatly increase the probability of discovering hit ligands.

As a logical next step following biochemical validation of tripeptoids, the best hits would be tested in human cell lines to determine whether they elicit a quantifiable growth phenotype. Abnormal regulation of PCNA levels, as well as post-translational modification of the protein complex, has been observed in a number of malignant cell types, particularly breast and prostate cancers [260,261]. As a result, cell lines such as MDA-MB-231 (breast), MDA-MB-468 (breast), LNCaP (prostate), PC3 (prostate) and DU145 (prostate) would likely be used as the models for testing. Demonstrating that the tripeptoid ligands can inhibit the growth of tumor-derived cells would be an important validation of this study's general approach toward discovering inhibitors of protein-protein interactions. However, a major factor that is not taken into consideration in the computational or biochemical testing processes is the ability for these compounds to efficiently cross a cell membrane and/or be taken up via pinocytosis. It may be that the tripeptoids synthesized for this study do not have favorable features for allowing them to efficiently be taken into cells, but this has not yet been fully assessed.

Other future studies would aim to further investigate the mechanism of action of the tripeptoid antagonists, as well as gauge their *in vivo* activity. The ability of these compounds to affect DNA translesion synthesis (TLS) could be assessed using published methodology [178], where, in the presence of PCNA antagonist, nuclear excision repair-resistant cells would be transfected with plasmid DNA containing an intrastrand cisplatin cross-link in a coding region of the *lacZ* gene, and another plasmid lacking the cross-link would serve as a control. Following a period of incubation, replicated plasmid would be recovered and analyzed for TLS via *E. Coli* transformation/colony selection. DNA damage response would be further monitored through the use of tumor cell lines incubated with antagonist, cisplatin, or both. Cells would be stained using γ H2AX, a biomarker for double-stranded DNA breaks [476], and analyzed with flow cytometry. In parallel to those studies, questions of whether blocking the PIP Box binding site could have an effect on PCNA post-translational modification (PTM), and whether levels of PTM have an effect

on binding affinity at the PIP Box site could also be pursued. These are interesting questions since the plasticity of PCNA [177] may accommodate for a larger variety ligands that bind at the PIP Box site, and this could have ‘long-range’ effects on the opposing face of the protein complex that directly affects levels of PTM—thus potentially having an effect on overall levels of DNA damage repair. Furthermore, demonstrating that PCNA ubiquitination has an effect on the affinity of ligands to bind at the PIP Box binding site could have huge benefits for drug-target specificity.

In further validation of tripeptoid-based inhibitors as potential drug candidates, top lead molecules could be further evaluated for their pharmacological properties in early ADME/tox profile assays. Prior to any *in vivo* efficacy testing, acute toxicity studies can be executed on all compounds deemed as leads for *in vivo* testing. Pharmacokinetic studies can also be pursued in rodent models. Ultimately, experimental data obtained from the tripeptoid molecules could be used in the design of compound formulations that enable the drug properties and hypothesis testing to be conducted with new tripeptoid molecular entities. Additionally, it is possible that due to the homotrimeric nature of PCNA, a nanotechnology-esque approach could be utilized where three PIP Box antagonists are linked together to inhibit all three binding sites on PCNA simultaneously (Figure 4.3)—thus greatly enhancing the specificity and potency of these compounds.

4.2. Further Assessment of the Nuclear Effect of Subcellularly-Targeted Gefitinib, and Exploring the Localization of Alternative Small Molecule Inhibitors

The study outlined in Chapter 3 showed the feasibility of incorporating the known tyrosine kinase inhibitor, gefitinib, into a peptoid-based sequence in an effort to target it to the nucleus of cancer cells. The results indicated that as opposed to commercial gefitinib, these peptoid-based conjugates were able to downregulate the phosphorylation of STAT3 while not affecting the phosphorylation of ERK1/2. Fluorescently-tagged variants of the conjugates were able to be taken up into cells, and FAM-SV40-NArg in particular was taken into the nucleus; however, direct evidence of the gefitinib-tagged molecules in the nucleus would be important to establish. The next major piece of evidence for the justification of the use of these compounds would be showing that they have a strong growth phenotype. A two-dimensional MTT assay was performed using Gef-NLys and Gef-SV40-NLys, but the experiment would need to be performed with the NArg-based ligands, especially given that there was evidence of them being more effective in the disruption of p-STAT3, as well as having enhanced cellular and nuclear uptake. In three-dimensional cell growth studies using NME and LM1 cell lines, the peptoid-based conjugates did not show much evidence of growth inhibition over a four week study (data not shown). However, when T2AA, which has been

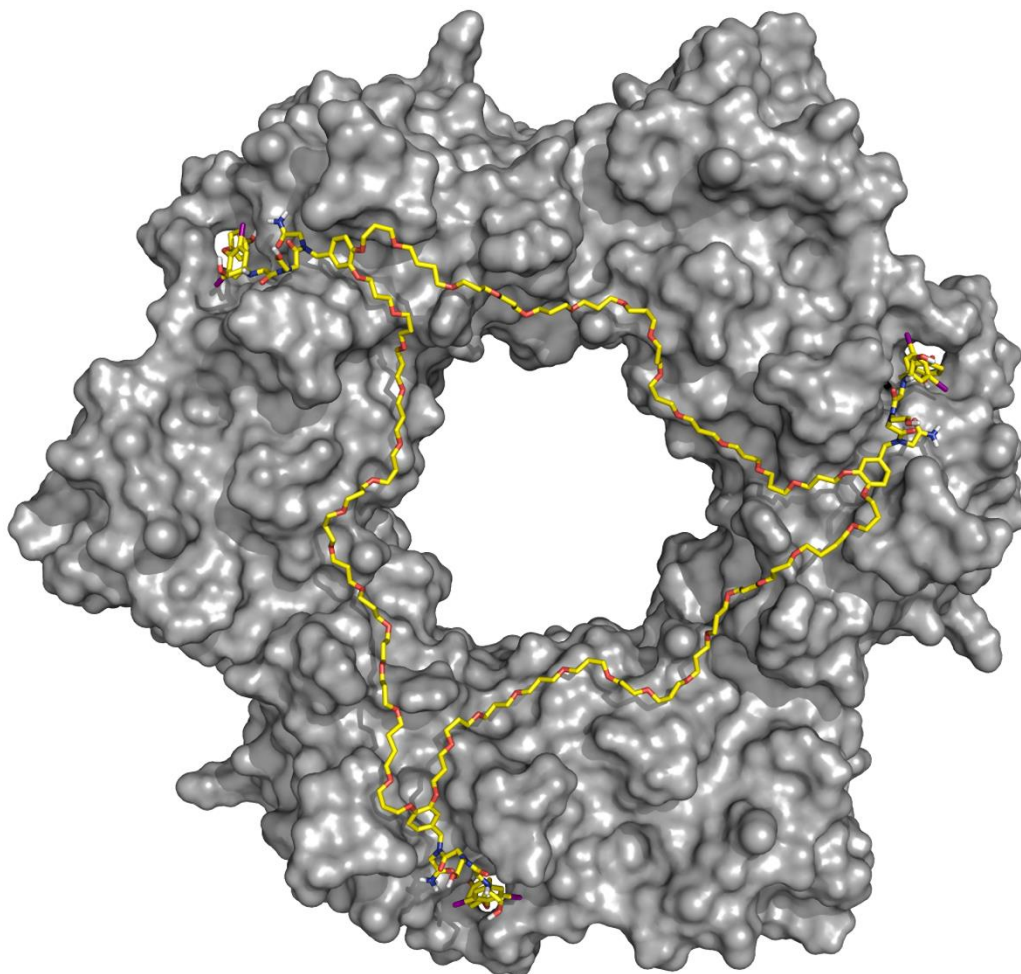


Figure 4.3. Simultaneous Targeting of the Three PIP Box Binding Sites on PCNA. Attaching three of the same PIP Box antagonists, such as T2AA-NEal-NPip (shown here), together with a polypropylene glycol linker could allow for the simultaneous targeting of all three PIP Box binding sites of PCNA upon the formation of the protein homotrimer. PCNA is shown as a gray surface, while the trimeric antagonist is shown as yellow sticks.

seen to disrupt cell proliferation in the U2OS bone osteocarcinoma cell line, was introduced to both NMEs and LM1s in the same 3D assay, an *enhancement* of growth was observed. As a result, there are likely multiple factors that affect the sensitivity of these cell lines to different types of inhibitors, and so a proper model system would need to be developed.

To further validate the activity of the inhibitors in the nucleus, a number of different experiments could be performed. In contrast to the nuclear localization sequence that was used in this study to target the molecules to the nucleus of cells, a nuclear exclusion signal could be incorporated, which would instead target the compounds for export from the nucleus into the cytoplasm [477–479]. The phosphorylation statuses of STAT3 and ERK1/2, as well as potential growth phenotype data, could be compared between the two classes of compounds to see if the targeting sequence has a true effect. Separately, it was established here that inhibition of STAT3 phosphorylation by the peptoid antagonists was dependent on EGF stimulation. While this would make sense at face value given that the ligands should be inhibitors of EGFR, the downstream mechanisms of this inhibition are not yet completely understood. As of yet, it has not been proven whether the observed effect on p-STAT3 in NME, LM1 and MDA-MB-468 cells is due to the direct interaction of STAT3 with nuclear EGFR, or if there is simply inhibition occurring at the cell surface, and the downstream signal originating from EGFR gets disrupted. This would be a point of emphasis in elucidating the mechanism of action of the gefitinib-peptoid conjugates in the disruption of p-STAT3. Furthermore, the functional consequence of abrogating STAT3 phosphorylation would need to be assessed. The protein naturally acts as transcription activator upon its phosphorylation, and in cancer types dependent on altered EGFR-STAT3 signaling, abolishing STAT3 phosphorylation may prove to be an effective strategy for inhibiting tumorigenesis [467].

It is not yet known whether the peptoid/peptide sequences have a negative impact on the drug binding to its target. While a six-member methylene linker was used to separate the drug from the peptide or peptoid portion of these conjugate molecules, if the ligands are not subjected to intracellular proteolysis, which may subsequently cause the drug to be released from the targeting sequence, the drug binding to its protein receptor would necessitate the remainder of the molecule (including the peptide and peptoid portions) to either wrap around the surface of the protein, or project freely out into solution. It is likely that there are non-zero penalties to solvation and desolvation energies, as well as potentially added steric clashes between these molecules and the receptor, as opposed to the commercial drug alone. Therefore, it may be beneficial to design a method for the drug to be released from the peptide/peptoid targeting sequence within the cell. One

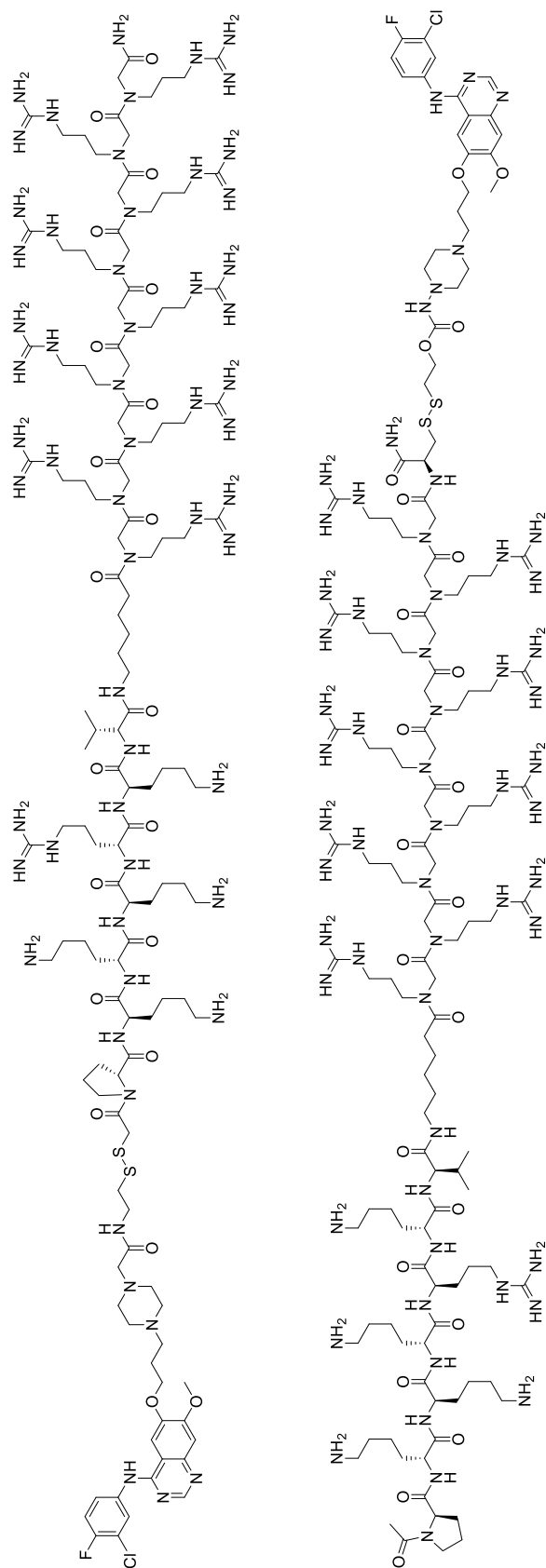


Figure 4.4. Alternative Reducible Linkers for Drug-Peptidic Conjugates. It is possible that coupling drugs, such as gefitinib shown here, to alternative reducible linkers could provide a way for the drug to be cleaved from accompanying peptid-peptide sequences inside the cell. Two versions of a simple reducible disulfide linker is shown here which allows gefitinib to be attached to either the N-terminus (*top*) or C-terminus (*bottom*) of the targeting sequence.

idea is to use a reducible linker that involves a disulfide bond (Figure 4.4). This is a strategy currently used in the design of vintafolide, an anticancer drug currently in Phase IIb clinical trials for the treatment of non-small-cell lung carcinoma. Variants of this spacer could be designed so as to be able to attach a desired drug to either the N- or C-terminus of a targeting sequence.

In addition to gefitinib, other protein kinase inhibitors could be explored for their potential for incorporation into peptide/peptoid-conjugates. In this way, the subcellular targeting strategy can be pursued against alternative receptors. For example, Figure 4.5 shows other potentially modifiable kinase inhibitors such as erlotinib, imatinib, dasatinib, crizotinib and lapatinib, which are antagonists of EGFR, BCR-Abl, BCR-Abl, ALK/ROS1 or HER2/neu/EGFR, respectively. In the original design of piperazinyl gefitinib, the morpholino group was chosen as the site for modification since it is in a portion of the molecule that is easily substituted, and it projects out into solution, away from the interior of the binding pocket on EGFR (see PDB ID: 2ITY). The modifications proposed in Figure 4.5 were selected for the same reason, where in each case the location of substitution projects outward from the interior of the binding pocket. In addition to exploring other kinase inhibitors, alternative targeting sequences could be pursued. For example, instead of the SV40 nuclear localization sequence, TAT may be used. Other subcellular compartments could be targeted in the future as well, where drug-conjugates could be directed to the nucleolus [480], mitochondria [481] or peroxisomes [482], or could be targeted for secretion or retention in the endoplasmic reticulum [483,484]. This could prove to be a versatile strategy for increasing the intracellular concentration of a drug molecule in a desired compartment of interest.

In further investigation of the significance of the nuclear-targeted ligands, future work could be done to determine whether a possible connection to the processes of DNA damage repair exists. nEGFR plays a very important role in DNA repair pathway activation, and previous studies have shown increased localization of EGFR to the nucleus after ionizing radiation (IR) [485–487]. Additionally, nEGFR's direct interactions with PCNA and DNA-PK are both heavily implicated in DNA repair [276,446,485,487,488]. Along with the previously discussed importance of PCNA and its relation to the MMR repair pathway, DNA-PK is an essential protein to both the HR and NHEJ repair pathways that address double strand breaks, and it becomes activated by phosphorylation by EGFR allowing it to carry out its function in either pathway. The increased levels of phosphorylation of PCNA and DNA-PK in various cancer types could explain in part the high resistance to genotoxic agents, such as cisplatin, and IR. The loss of phosphorylation of PCNA would decrease both the amount of active PCNA and the amount that is bound to chromatin; both of these effects have been implicated in cell cycle progression. The loss of active DNA-PK would

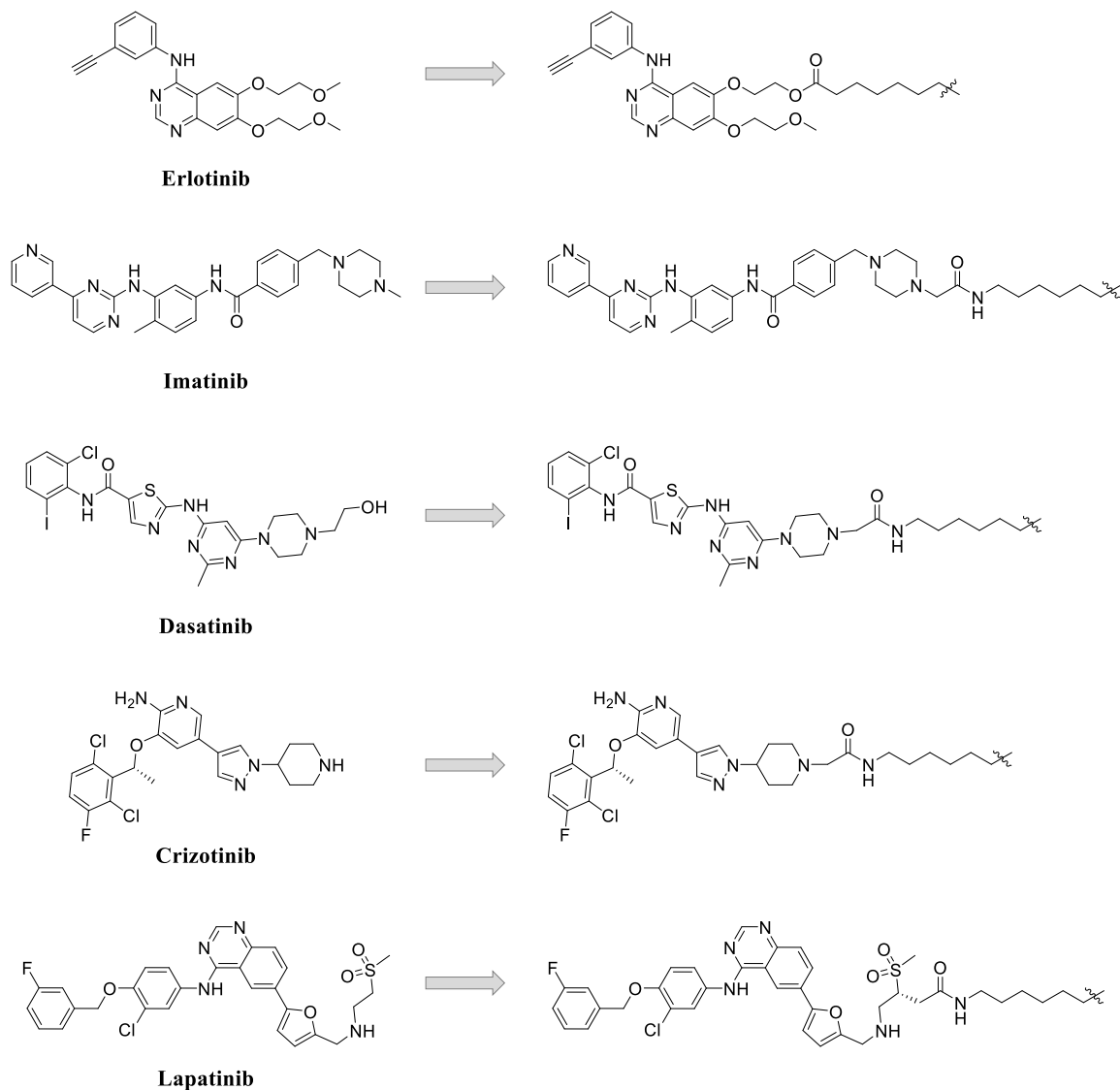


Figure 4.5. Variants of Protein Kinase Inhibitors Can Be Incorporated into a Peptide/Peptoid. Similar to gefitinib in this study, other protein kinase inhibitors can be modified so as to be able to conjugate them to a peptide or peptoid. Here, the linkages are analogous to the one used for gefitinib, with a six-member methylene spacer separating the drug from the targeting sequence.

see wide reaching effects given that it has many interactors, but much of the DNA repair aspects of DNA-PK can be supplemented by PARP1, which can also activate NHEJ pathway independent of DNA-PK [489]. It is possible that the addition of a PARP1 inhibitor along with gefitinib could cause a synthetic lethal relationship by inhibiting the alternate NHEJ pathway [490,491] in addition to downregulating the phosphorylation of PCNA, which itself has been strongly correlated with enhanced tumorigenesis [260,261]. In effect, both double strand break repair pathways could be inhibited, and this could be further supplemented by the inhibition of PCNA-dependent [175,176,256] repair mechanisms. Ultimately, this could prove to be an effective treatment for multiple forms of cancer.

LIST OF REFERENCES

LIST OF REFERENCES

- [1] Lodish H., Berk A., Matsudaira P., Kaiser C. A., Krieger M., Scott M. P., *et al.* (2004). *Molecular Biology of the Cell*. 5th ed. New York, NY: WH Freeman.
- [2] Lindahl T. and Barnes D. E. (2000). Repair of endogenous DNA damage. *Cold Spring Harb Symp Quant Biol.* 65:127–33.
- [3] De Bont R. and van Larebeke N. (2004). Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis.* 19(3):169–85.
- [4] Dizdaroglu M. (2012). Oxidatively induced DNA damage: mechanisms, repair and disease. *Cancer Lett.* 327(1-2):26–47.
- [5] Helleday T., Petermann E., Lundin C., Hodgson B. and Sharma R. A. (2008). DNA repair pathways as targets for cancer therapy. *Nat Rev Cancer.* 8(3):193–204.
- [6] Jackson S. P. and Bartek J. (2009). The DNA-damage response in human biology and disease. *Nature.* 461(7267):1071–8.
- [7] Ljungman M. and Lane D. P. (2004). Transcription - guarding the genome by sensing DNA damage. *Nat Rev Cancer.* 4(9):727–37.
- [8] Mandal P. K., Blanpain C. and Rossi D. J. (2011). DNA damage response in adult stem cells: pathways and consequences. *Nat Rev Mol Cell Biol.* 12(3):198–202.
- [9] Rass U., Ahel I. and West S. C. (2007). Defective DNA repair and neurodegenerative disease. *Cell.* 130(6):991–1004.
- [10] d’Adda di Fagagna F. (2008). Living on a break: cellular senescence as a DNA-damage response. *Nat Rev Cancer.* 8(7):512–22.
- [11] Schumacher B., Garinis G. A. and Hoeijmakers J. H. J. (2008). Age to survive: DNA damage and aging. *Trends Genet TIG.* 24(2):77–85.
- [12] Khanna K. K. and Jackson S. P. (2001). DNA double-strand breaks: signaling, repair and the cancer connection. *Nat Genet.* 27(3):247–54.
- [13] Harper J. W. and Elledge S. J. (2007). The DNA damage response: ten years after. *Mol Cell.* 28(5):739–45.

- [14] Polo S. E. and Jackson S. P. (2011). Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications. *Genes Dev.* 25(5):409–33.
- [15] Sancar A., Lindsey-Boltz L. A., Unsal-Kaçmaz K. and Linn S. (2004). Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu Rev Biochem.* 73:39–85.
- [16] Kulkarni A. and Wilson D. M. (2008). The Involvement of DNA-Damage and -Repair Defects in Neurological Dysfunction. *Am J Hum Genet.* 82(3):539–66.
- [17] Bartkova J., Rajpert-De Meyts E., Skakkebaek N. E., Lukas J. and Bartek J. (2007). DNA damage response in human testes and testicular germ cell tumours: biology and implications for therapy. *Int J Androl.* 30(4):282–91.
- [18] Matzuk M. M. and Lamb D. J. (2008). The biology of infertility: research advances and clinical challenges. *Nat Med.* 14(11):1197–213.
- [19] Schneider J. G., Finck B. N., Ren J., Standley K. N., Takagi M., Maclean K. H., *et al.* (2006). ATM-dependent suppression of stress signaling reduces vascular disease in metabolic syndrome. *Cell Metab.* 4(5):377–89.
- [20] Kastan M. B. (2008). DNA damage responses: mechanisms and roles in human disease: 2007 G.H.A. Clowes Memorial Award Lecture. *Mol Cancer Res MCR.* 6(4):517–24.
- [21] Mirkin S. M. (2007). Expandable DNA repeats and human disease. *Nature.* 447(7147):932–40.
- [22] Kovtun I. V. and McMurray C. T. (2008). Features of trinucleotide repeat instability in vivo. *Cell Res.* 18(1):198–213.
- [23] Yang J.-L., Weissman L., Bohr V. A. and Mattson M. P. (2008). Mitochondrial DNA damage and repair in neurodegenerative disorders. *DNA Repair.* 7(7):1110–20.
- [24] Hoeijmakers J. H. (2001). Genome maintenance mechanisms for preventing cancer. *Nature.* 411(6835):366–74.
- [25] Bartkova J., Horejsí Z., Koed K., Krämer A., Tort F., Zieger K., *et al.* (2005). DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature.* 434(7035):864–70.
- [26] Gorgoulis V. G., Vassiliou L.-V. F., Karakaidos P., Zacharatos P., Kotsinas A., Liloglou T., *et al.* (2005). Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature.* 434(7035):907–13.
- [27] Kaina B., Christmann M., Naumann S. and Roos W. P. (2007). MGMT: Key node in the battle against genotoxicity, carcinogenicity and apoptosis induced by alkylating agents. *DNA Repair.* 6(8):1079–99.
- [28] Ciccia A. and Elledge S. J. (2010). The DNA Damage Response: Making It Safe to Play with Knives. *Mol Cell.* 40(2):179–204.

- [29] Caldecott K. W. (2008). Single-strand break repair and genetic disease. *Nat Rev Genet.* 9(8):619–31.
- [30] Liang L., Deng L., Chen Y., Li G. C., Shao C. and Tischfield J. A. (2005). Modulation of DNA End Joining by Nuclear Proteins. *J Biol Chem.* 280(36):31442–9.
- [31] Lieber M. R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem.* 79:181–211.
- [32] Waters L. S., Minesinger B. K., Wiltrout M. E., D'Souza S., Woodruff R. V. and Walker G. C. (2009). Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance. *Microbiol Mol Biol Rev MMBR.* 73(1):134–54.
- [33] Hendel A., Krijger P. H. L., Diamant N., Goren Z., Langerak P., Kim J., *et al.* (2011). PCNA Ubiquitination Is Important, But Not Essential for Translesion DNA Synthesis in Mammalian Cells. *PLoS Genet.* 7(9):e1002262.
- [34] Shaheen M., Shanmugam I. and Hromas R. (2010). The Role of PCNA Posttranslational Modifications in Translesion Synthesis. *J Nucleic Acids.* 2010:e761217.
- [35] Knobel P. A. and Marti T. M. (2011). Translesion DNA synthesis in the context of cancer research. *Cancer Cell Int.* 11(1):39.
- [36] Ciriello G., Miller M. L., Aksoy B. A., Senbabaoglu Y., Schultz N. and Sander C. (2013). Emerging landscape of oncogenic signatures across human cancers. *Nat Genet.* 45(10):1127–33.
- [37] Bartkova J., Rezaei N., Liontos M., Karakaidos P., Kletsas D., Issaeva N., *et al.* (2006). Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. *Nature.* 444(7119):633–7.
- [38] Di Micco R., Fumagalli M., Cicalese A., Piccinin S., Gasparini P., Luise C., *et al.* (2006). Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. *Nature.* 444(7119):638–42.
- [39] Halazonetis T. D., Gorgoulis V. G. and Bartek J. (2008). An Oncogene-Induced DNA Damage Model for Cancer Development. *Science.* 319(5868):1352–5.
- [40] Negrini S., Gorgoulis V. G. and Halazonetis T. D. (2010). Genomic instability--an evolving hallmark of cancer. *Nat Rev Mol Cell Biol.* 11(3):220–8.
- [41] Stratton M. R., Campbell P. J. and Futreal P. A. (2009). The cancer genome. *Nature.* 458(7239):719–24.
- [42] Holohan C., Van Schaeybroeck S., Longley D. B. and Johnston P. G. (2013). Cancer drug resistance: an evolving paradigm. *Nat Rev Cancer.* 13(10):714–26.
- [43] Damia G. and D'Incalci M. (2007). Targeting DNA repair as a promising approach in cancer therapy. *Eur J Cancer Oxf Engl 1990.* 43(12):1791–801.

- [44] Hosoya N. and Miyagawa K. (2014). Targeting DNA damage response in cancer therapy. *Cancer Sci.* 105(4):370–88.
- [45] Persidis A. (2000). Cancer multidrug resistance. *Nat Biotechnol.* 18:IT18–20.
- [46] Fojo T. (2001). Cancer, DNA Repair Mechanisms, and Resistance to Chemotherapy. *J Natl Cancer Inst.* 93(19):1434–6.
- [47] Luqmani Y. A. (2005). Mechanisms of drug resistance in cancer chemotherapy. *Med Princ Pract Int J Kuwait Univ Health Sci Cent.* 14 Suppl 1:35–48.
- [48] Chen Y. and Fu L. (2011). Mechanisms of acquired resistance to tyrosine kinase inhibitors. *Acta Pharm Sin B.* 1(4):197–207.
- [49] Chong C. R. and Jänne P. A. (2013). The quest to overcome resistance to EGFR-targeted therapies in cancer. *Nat Med.* 19(11):1389–400.
- [50] Srinivasan A. and Gold B. (2012). Small-molecule inhibitors of DNA damage-repair pathways: an approach to overcome tumor resistance to alkylating anticancer drugs. *Future Med Chem.* 4(9):1093–111.
- [51] Bartolowits M. and Davisson V. J. (2015). Considerations of Protein Subpockets in Fragment-Based Drug Design. *Chem Biol Drug Des [Internet]*. Available from: <http://onlinelibrary.wiley.com/doi/10.1111/cbdd.12631/abstract>
- [52] Erlanson D. A. (2012). Introduction to fragment-based drug discovery. *Top Curr Chem.* 317:1–32.
- [53] Erlanson D. A., McDowell R. S. and O'Brien T. (2004). Fragment-Based Drug Discovery. *J Med Chem.* 47(14):3463–82.
- [54] Henen M. A., Coudevylle N., Geist L. and Konrat R. (2012). Toward Rational Fragment-Based Lead Design without 3D Structures. *J Med Chem.* 55(17):7909–19.
- [55] Sheng C. and Zhang W. (2013). Fragment Informatics and Computational Fragment-Based Drug Design: An Overview and Update. *Med Res Rev.* 33(3):554–98.
- [56] Hu Y. and Bajorath J. (2013). What is the Likelihood of an Active Compound to Be Promiscuous? Systematic Assessment of Compound Promiscuity on the Basis of PubChem Confirmatory Bioassay Data. *AAPS J.* 15(3):808–15.
- [57] Babaoglu K. and Shoichet B. K. (2006). Deconstructing fragment-based inhibitor discovery. *Nat Chem Biol.* 2(12):720–3.
- [58] Schneider G. (2014). Future De Novo Drug Design. *Mol Inform.* 33(6-7):397–402.
- [59] Konec J. and Janežič D. (2014). Binding site comparison for function prediction and pharmaceutical discovery. *Curr Opin Struct Biol.* 25:34–9.

- [60] Joseph-McCarthy D., Campbell A. J., Kern G. and Moustakas D. (2014). Fragment-Based Lead Discovery and Design. *J Chem Inf Model.* 54(3):693–704.
- [61] Feixas F., Lindert S., Sinko W. and McCammon J. A. (2014). Exploring the role of receptor flexibility in structure-based drug discovery. *Biophys Chem.* 186:31–45.
- [62] Rognan D. (2013). Towards the Next Generation of Computational Chemogenomics Tools. *Mol Inform.* 32(11-12):1029–34.
- [63] Jalencas X. and Mestres J. (2013). Identification of Similar Binding Sites to Detect Distant Polypharmacology. *Mol Inform.* 32(11-12):976–90.
- [64] Scott D. E., Coyne A. G., Hudson S. A. and Abell C. (2012). Fragment-Based Approaches in Drug Discovery and Chemical Biology. *Biochemistry (Mosc).* 51(25):4990–5003.
- [65] Kumar A., Voet A. and Zhang K. Y. J. (2012). Fragment based drug design: from experimental to computational approaches. *Curr Med Chem.* 19(30):5128–47.
- [66] Mestres J., Gregori-Puigjané E., Valverde S. and Solé R. V. (2009). The topology of drug–target interaction networks: implicit dependence on drug properties and target families. *Mol Biosyst.* 5(9):1051–7.
- [67] Vogt I. and Mestres J. (2010). Drug-Target Networks. *Mol Inform.* 29(1-2):10–4.
- [68] Brianso F., Carrascosa M. C., Oprea T. I. and Mestres J. (2011). Cross-Pharmacology Analysis of G Protein-Coupled Receptors. *Curr Top Med Chem.* 11(15):1956–63.
- [69] Jalencas X. and Mestres J. (2012). On the origins of drug polypharmacology. *MedChemComm.* 4(1):80–7.
- [70] Schneider G. (2010). Virtual screening: an endless staircase? *Nat Rev Drug Discov.* 9(4):273–6.
- [71] Eder J., Sedrani R. and Wiesmann C. (2014). The discovery of first-in-class drugs: origins and evolution. *Nat Rev Drug Discov.* 13(8):577–87.
- [72] Swinney D. C. and Anthony J. (2011). How were new medicines discovered? *Nat Rev Drug Discov.* 10(7):507–19.
- [73] Berman H. M., Westbrook J., Feng Z., Gilliland G., Bhat T. N., Weissig H., *et al.* (2000). The Protein Data Bank. *Nucleic Acids Res.* 28(1):235–42.
- [74] Kinnings S. L. and Jackson R. M. (2009). Binding Site Similarity Analysis for the Functional Classification of the Protein Kinase Family. *J Chem Inf Model.* 49(2):318–29.
- [75] Weber A., Casini A., Heine A., Kuhn D., Supuran C. T., Scozzafava A., *et al.* (2004). Unexpected nanomolar inhibition of carbonic anhydrase by COX-2-selective celecoxib: new pharmacological opportunities due to related binding site recognition. *J Med Chem.* 47(3):550–7.

- [76] Gherardini P. F., Wass M. N., Helmer-Citterich M. and Sternberg M. J. E. (2007). Convergent evolution of enzyme active sites is not a rare phenomenon. *J Mol Biol.* 372(3):817–45.
- [77] Xie L. and Bourne P. E. (2008). Detecting evolutionary relationships across existing fold space, using sequence order-independent profile–profile alignments. *Proc Natl Acad Sci.* 105(14):5441–6.
- [78] Baskin I. and Varnek A. (2008). Building a chemical space based on fragment descriptors. *Comb Chem High Throughput Screen.* 11(8):661–8.
- [79] Zuegg J. and Cooper M. A. (2012). Drug-likeness and increased hydrophobicity of commercially available compound libraries for drug screening. *Curr Top Med Chem.* 12(14):1500–13.
- [80] Congreve M., Carr R., Murray C. and Jhoti H. (2003). A “rule of three” for fragment-based lead discovery? *Drug Discov Today.* 8(19):876–7.
- [81] Köster H., Craan T., Brass S., Herhaus C., Zentgraf M., Neumann L., *et al.* (2011). A Small Nonrule of 3 Compatible Fragment Library Provides High Hit Rate of Endothiapepsin Crystal Structures with Various Fragment Chemotypes. *J Med Chem.* 54(22):7784–96.
- [82] Chan A. W. E., Laskowski R. A. and Selwood D. L. (2010). Chemical Fragments that Hydrogen Bond to Asp, Glu, Arg, and His Side Chains in Protein Binding Sites. *J Med Chem.* 53(8):3086–94.
- [83] Wang L., Xie Z., Wipf P. and Xie X.-Q. (2011). Residue Preference Mapping of Ligand Fragments in the Protein Data Bank. *J Chem Inf Model.* 51(4):807–15.
- [84] Cox B. D., Prosser A. R., Katzman B. M., Alcaraz A. A., Liotta D. C., Wilson L. J., *et al.* (2014). Anti-HIV small-molecule binding in the peptide subpocket of the CXCR4:CVX15 crystal structure. *Chembiochem Eur J Chem Biol.* 15(11):1614–20.
- [85] Wu B., Chien E. Y. T., Mol C. D., Fenalti G., Liu W., Katritch V., *et al.* (2010). Structures of the CXCR4 chemokine receptor in complex with small molecule and cyclic peptide antagonists. *Science.* 330(6007):1066–71.
- [86] Kawatkar S. P., Yan M., Gevariya H., Lim M. Y., Eisold S., Zhu X., *et al.* (2011). Computational analysis of the structural mechanism of inhibition of chemokine receptor CXCR4 by small molecule antagonists. *Exp Biol Med.* 236(7):844–50.
- [87] Zhang C., Hou T., Feng Z. and Li Y. (2013). Structure-based development of antagonists for chemokine receptor CXCR4. *Curr Comput Aided Drug Des.* 9(1):60–75.
- [88] Baell J. and Walters M. A. (2014). Chemistry: Chemical con artists foil drug discovery. *Nature.* 513(7519):481–3.
- [89] Pohjala L. and Tammela P. (2012). Aggregating Behavior of Phenolic Compounds — A Source of False Bioassay Results? *Molecules.* 17(9):10774–90.

- [90] Mendgen T., Steuer C. and Klein C. D. (2012). Privileged Scaffolds or Promiscuous Binders: A Comparative Study on Rhodanines and Related Heterocycles in Medicinal Chemistry. *J Med Chem.* 55(2):743–53.
- [91] Baell J. B. and Holloway G. A. (2010). New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. *J Med Chem.* 53(7):2719–40.
- [92] Baell J. B., Ferrins L., Falk H. and Nikolakopoulos G. (2013). PAINS: Relevance to Tool Compound Discovery and Fragment-Based Screening. *Aust J Chem.* 66(12):1483–94.
- [93] Priyadarsini K. I. (2013). Chemical and structural features influencing the biological activity of curcumin. *Curr Pharm Des.* 19(11):2093–100.
- [94] Ingólfsson H. I., Thakur P., Herold K. F., Hobart E. A., Ramsey N. B., Periole X., *et al.* (2014). Phytochemicals perturb membranes and promiscuously alter protein function. *ACS Chem Biol.* 9(8):1788–98.
- [95] Haupt V. J. and Schroeder M. (2011). Old friends in new guise: repositioning of known drugs with structural bioinformatics. *Brief Bioinform.* 12:312–26.
- [96] Henrich S., Salo-Ahen O. M. H., Huang B., Rippmann F. F., Cruciani G. and Wade R. C. (2010). Computational approaches to identifying and characterizing protein binding sites for ligand design. *J Mol Recognit JMR.* 23(2):209–19.
- [97] Kellenberger E., Schalón C. and Rognan D. (2008). How to measure the similarity between protein ligand-binding sites? *Curr Comput - Aided Drug Des.* 4(3):209–20.
- [98] Nisius B., Sha F. and Gohlke H. (2012). Structure-based computational analysis of protein binding sites for function and druggability prediction. *J Biotechnol.* 159(3):123–34.
- [99] Salentin S., Haupt V. J., Daminelli S. and Schroeder M. (2014). Polypharmacology rescored: Protein–ligand interaction profiles for remote binding site similarity assessment. *Prog Biophys Mol Biol.* 116(2–3):174–86.
- [100] Volkamer A. and Rarey M. (2014). Exploiting structural information for drug-target assessment. *Future Med Chem.* 6(3):319–31.
- [101] Vulpetti A., Kalliokoski T. and Milletti F. (2012). Chemogenomics in drug discovery: computational methods based on the comparison of binding sites. *Future Med Chem.* 4(15):1971–9.
- [102] Walter P., Metzger J., Thiel C. and Helms V. (2013). Predicting where Small Molecules Bind at Protein-Protein Interfaces. *PLoS ONE.* 8(3):e58583.
- [103] Wirth M., Volkamer A., Zoete V., Rippmann F., Michelin O., Rarey M., *et al.* (2013). Protein pocket and ligand shape comparison and its application in virtual screening. *J Comput Aided Mol Des.* 27(6):511–24.

- [104] Wood D. J., de Vlieg J., Wagener M. and Ritschel T. (2012). Pharmacophore fingerprint-based approach to binding site subpocket similarity and its application to bioisostere replacement. *J Chem Inf Model.* 52(8):2031–43.
- [105] Zheng X., Gan L., Wang E. and Wang J. (2013). Pocket-Based Drug Design: Exploring Pocket Space. *AAPS J.* 15(1):228–41.
- [106] Jordan R. A., EL-Manzalawy Y., Dobbs D. and Honavar V. (2012). Predicting protein-protein interface residues using local surface structural similarity. *BMC Bioinformatics.* 13(1):41.
- [107] Xie L. and Bourne P. E. (2007). A robust and efficient algorithm for the shape description of protein structures and its application in predicting ligand binding sites. *BMC Bioinformatics.* 8(Suppl 4):S9.
- [108] Konc J. and Janezic D. (2012). ProBiS-2012: web server and web services for detection of structurally similar binding sites in proteins. *Nucleic Acids Res.* 40(Web Server issue):W214–21.
- [109] Roy A., Yang J. and Zhang Y. (2012). COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Res.* 40(Web Server issue):W471–7.
- [110] Yang J., Roy A. and Zhang Y. (2013). BioLiP: a semi-manually curated database for biologically relevant ligand-protein interactions. *Nucleic Acids Res.* 41(Database issue):D1096–103.
- [111] Nadzirin N., Willett P., Artymiuk P. J. and Firdaus-Raih M. (2013). IMAAAGINE: a webserver for searching hypothetical 3D amino acid side chain arrangements in the Protein Data Bank. *Nucleic Acids Res.* 41(Web Server issue):W432–40.
- [112] Nilmeier J. P., Kirshner D. A., Wong S. E. and Lightstone F. C. (2013). Rapid Catalytic Template Searching as an Enzyme Function Prediction Procedure. *PLoS ONE.* 8(5):e62535.
- [113] Kirshner D. A., Nilmeier J. P. and Lightstone F. C. (2013). Catalytic site identification—a web server to identify catalytic site structural matches throughout PDB. *Nucleic Acids Res.* 41(W1):W256–65.
- [114] Hendlich M., Rippmann F. and Barnickel G. (1997). LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J Mol Graph Model.* 15(6):359–63, 389.
- [115] Schmitt S., Hendlich M. and Klebe G. (2001). From Structure to Function: A New Approach to Detect Functional Similarity among Proteins Independent from Sequence and Fold Homology. *Angew Chem Int Ed.* 40(17):3141–4.
- [116] Schmitt S., Kuhn D. and Klebe G. (2002). A New Method to Detect Related Function Among Proteins Independent of Sequence and Fold Homology. *J Mol Biol.* 323(2):387–406.

- [117] Bergner A. and Günther J. (2005). Structural Aspects of Binding Site Similarity: A 3D Upgrade for Chemogenomics. In: *Chemogenomics in Drug Discovery: A Medicinal Chemistry Perspective*. Wiley-VCH Verlag GmbH & Co. KGaA. p. 97–135.
- [118] Kuhn D., Weskamp N., Hüllermeier E. and Klebe G. (2007). Functional classification of protein kinase binding sites using Cavbase. *ChemMedChem*. 2(10):1432–47.
- [119] Durrant J. D., Friedman A. J. and McCammon J. A. (2011). CrystalDock: A Novel Approach to Fragment-Based Drug Design. *J Chem Inf Model*. 51(10):2573–80.
- [120] Volkamer A., Griewel A., Grombacher T. and Rarey M. (2010). Analyzing the topology of active sites: on the prediction of pockets and subpockets. *J Chem Inf Model*. 50(11):2041–52.
- [121] Schmidtke P. and Barril X. (2010). Understanding and Predicting Druggability. A High-Throughput Method for Detection of Drug Binding Sites. *J Med Chem*. 53(15):5858–67.
- [122] Volkamer A., Kuhn D., Grombacher T., Rippmann F. and Rarey M. (2012). Combining Global and Local Measures for Structure-Based Druggability Predictions. *J Chem Inf Model*. 52(2):360–72.
- [123] Vulpetti A., Schiering N. and Dalvit C. (2010). Combined use of computational chemistry, NMR screening, and X-ray crystallography for identification and characterization of fluorophilic protein environments. *Proteins Struct Funct Bioinforma*. 78(16):3281–91.
- [124] Brylinski M. and Skolnick J. (2009). FINDSITE_LHM: A Threading-Based Approach to Ligand Homology Modeling. *PLoS Comput Biol*. 5(6):e1000405.
- [125] Tang G. W. and Altman R. B. (2014). Knowledge-based Fragment Binding Prediction. *PLoS Comput Biol*. 10(4):e1003589.
- [126] Kinjo A. R. and Nakamura H. (2007). Similarity search for local protein structures at atomic resolution by exploiting a database management system. *Biophysics*. 3:75–84.
- [127] Kinjo A. R. and Nakamura H. (2009). Comprehensive Structural Classification of Ligand-Binding Motifs in Proteins. *Structure*. 17(2):234–46.
- [128] Lee H. S. and Im W. (2012). Identification of Ligand Templates using Local Structure Alignment for Structure-Based Drug Design. *J Chem Inf Model*. 52(10):2784–95.
- [129] Davis F. P. (2011). Proteome-wide prediction of overlapping small molecule and protein binding sites using structure. *Mol Biosyst*. 7(2):545.
- [130] MEDIT-SA. (Accessed 2014, October 1). [Internet]. Available from: <http://medit-pharma.com/>
- [131] Parca L., Ferré F., Ausiello G. and Helmer-Citterich M. (2013). Nucleos: a web server for the identification of nucleotide-binding sites in protein structures. *Nucleic Acids Res*. 41(Web Server issue):W281–5.

- [132] Zhu X., Xiong Y. and Kihara D. (2015). Large-scale binding ligand prediction by improved patch-based method Patch-Surfer2.0. *Bioinformatics*. 31(5):707–13.
- [133] Parca L., Mangone I., Gherardini P. F., Ausiello G. and Helmer-Citterich M. (2011). Phosfinder: a web server for the identification of phosphate-binding sites on protein structures. *Nucleic Acids Res.* 39(Web Server issue):W278–82.
- [134] Liu T. and Altman R. B. (2011). Using Multiple Microenvironments to Find Similar Ligand-Binding Sites: Application to Kinase Inhibitor Binding. *PLoS Comput Biol.* 7(12):e1002326.
- [135] Yeturu K. and Chandra N. (2011). PocketAlign a novel algorithm for aligning binding sites in protein structures. *J Chem Inf Model.* 51(7):1725–36.
- [136] Ito J.-I., Tabei Y., Shimizu K., Tsuda K. and Tomii K. (2012). PoSSuM: a database of similar protein–ligand binding and putative pockets. *Nucleic Acids Res.* 40(D1):D541–8.
- [137] Weisel M., Bitter H.-M., Diederich F., So W. V. and Kondru R. (2012). PROLIX: rapid mining of protein-ligand interactions in large crystal structure databases. *J Chem Inf Model.* 52(6):1450–61.
- [138] Regad L., Martin J., Nuel G. and Camproux A.-C. (2010). Mining protein loops using a structural alphabet and statistical exceptionality. *BMC Bioinformatics.* 11(1):75.
- [139] Nuel G., Regad L., Martin J. and Camproux A.-C. (2010). Exact distribution of a pattern in a set of random sequences generated by a Markov source: applications to biological data. *Algorithms Mol Biol.* 5(1):15.
- [140] Teyra J., Doms A., Schroeder M. and Pisabarro M. T. (2006). SCOWLP: a web-based database for detailed characterization and visualization of protein interfaces. *BMC Bioinformatics.* 7:104.
- [141] Sehnal D., Vařeková R. S., Huber H. J., Geidl S., Ionescu C.-M., Wimmerová M., *et al.* (2012). SiteBinder: An Improved Approach for Comparing Multiple Protein Structural Motifs. *J Chem Inf Model.* 52(2):343–59.
- [142] Lin Y., Yoo S. and Sanchez R. (2012). SiteComp: a server for ligand binding site analysis in protein structures. *Bioinforma Oxf Engl.* 28(8):1172–3.
- [143] Kalliokoski T., Olsson T. S. G. and Vulpetti A. (2013). Subpocket Analysis Method for Fragment-Based Drug Discovery. *J Chem Inf Model.* 53(1):131–41.
- [144] Wang L., Ma C., Wipf P., Liu H., Su W. and Xie X.-Q. (2013). TargetHunter: an in silico target identification tool for predicting therapeutic potential of small organic molecules based on chemogenomic database. *AAPS J.* 15(2):395–406.
- [145] Kokh D. B., Richter S., Henrich S., Czodrowski P., Rippmann F. and Wade R. C. (2013). TRAPP: A Tool for Analysis of Transient Binding Pockets in Proteins. *J Chem Inf Model.* 53(5):1235–52.

- [146] von Behren M. M., Volkamer A., Henzler A. M., Schomburg K. T., Urbaczek S. and Rarey M. (2013). Fast protein binding site comparison via an index-based screening technology. *J Chem Inf Model.* 53(2):411–22.
- [147] Wallach I. and Lilien R. H. (2009). Prediction of sub-cavity binding preferences using an adaptive physicochemical structure representation. *Bioinformatics.* 25(12):i296–304.
- [148] Feldman H. J. and Labute P. (2010). Pocket similarity: are alpha carbons enough? *J Chem Inf Model.* 50(8):1466–75.
- [149] Yeturu K. and Chandra N. (2008). PocketMatch: A new algorithm to compare binding sites in protein structures. *BMC Bioinformatics.* 9(1):543.
- [150] Villoutreix B. O., Lagorce D., Labbé C. M., Sperandio O. and Miteva M. A. (2013). One hundred thousand mouse clicks down the road: selected online resources supporting drug discovery collected over a decade. *Drug Discov Today.* 18(21–22):1081–9.
- [151] Bergner A., Günther J., Hendlich M., Klebe G. and Verdonk M. (2001). Use of Relibase for retrieving complex three-dimensional interaction patterns including crystallographic packing effects. *Biopolymers.* 61(2):99–110.
- [152] Krotzky T., Grunwald C., Egerland U. and Klebe G. (2015). Large-Scale Mining for Similar Protein Binding Pockets: With RAPMAD Retrieval on the Fly Becomes Real. *J Chem Inf Model.* 55(1):165–79.
- [153] Wei L., Altman R. B. and Chang J. T. (1997). Using the radial distributions of physical features to compare amino acid environments and align amino acid sequences. *Pac Symp Biocomput.* :465–76.
- [154] Tabei Y., Uno T., Sugiyama M. and Tsuda K. (2010). Single versus multiple sorting for all pairs similarity search. In Tokyo, Japan: the 2nd Asian Conference on Machine Learning (ACML2010). p. 145–60.
- [155] Chikhi R., Sael L. and Kihara D. (2010). Real-time ligand binding pocket database search using local surface descriptors. *Proteins Struct Funct Bioinforma.* 78(9):2007–28.
- [156] Kabsch W. (1976). A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr Sect A.* 32(5):922–3.
- [157] Cock P. J. A., Antao T., Chang J. T., Chapman B. A., Cox C. J., Dalke A., *et al.* (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinforma Oxf Engl.* 25(11):1422–3.
- [158] Traxler P. and Furet P. (1999). Strategies toward the Design of Novel and Selective Protein Tyrosine Kinase Inhibitors. *Pharmacol Ther.* 82(2–3):195–206.
- [159] Anastassiadis T., Deacon S. W., Devarajan K., Ma H. and Peterson J. R. (2011). Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. *Nat Biotechnol.* 29(11):1039–45.

- [160] Pargellis C., Tong L., Churchill L., Cirillo P. F., Gilmore T., Graham A. G., *et al.* (2002). Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nat Struct Biol.* 9(4):268–72.
- [161] Liu Y. and Gray N. S. (2006). Rational design of inhibitors that bind to inactive kinase conformations. *Nat Chem Biol.* 2(7):358–64.
- [162] Bogoyevitch M. A. and Fairlie D. P. (2007). A new paradigm for protein kinase inhibition: blocking phosphorylation without directly targeting ATP binding. *Drug Discov Today.* 12(15-16):622–33.
- [163] Han B. G., Shim M. B. and Ahn H. C. (2013). The crystal structure of AMP-bound JNK3 [Internet]. *RCSB PDB Protein Data Bank*. Available from: <http://www.pdb.org/pdb/explore/explore.do?structureId=4KKE>
- [164] Scapin G., Patel S. B., Lisnock J., Becker J. W. and LoGrasso P. V. (2003). The Structure of JNK3 in Complex with Small Molecule Inhibitors: Structural Basis for Potency and Selectivity. *Chem Biol.* 10(8):705–12.
- [165] Tanramluk D., Schreyer A., Pitt W. R. and Blundell T. L. (2009). On the Origins of Enzyme Inhibitor Selectivity and Promiscuity: A Case Study of Protein Kinase Binding to Staurosporine. *Chem Biol Drug Des.* 74(1):16–24.
- [166] Tian-Jun Zhou L.-G. S. (2006). Crystal structure of the MAP3K TAO2 kinase domain bound by an inhibitor staurosporine. *Acta Biochim Biophys Sin.* 38(6):385–92.
- [167] Debreczeni J. E., Eswaran J., Bullock A., Filippakopoulos P., Kavanagh K., Amos A., *et al.* (2005). Crystal Structure of the Human Serine-Threonine Kinase 16 in Complex with Staurosporine [Internet]. *RCSB PDB Protein Data Bank*. Available from: <http://www.pdb.org/pdb/explore/explore.do?structureId=2buj>
- [168] Knowles P. P., Murray-Rust J., Kjær S., Scott R. P., Hanrahan S., Santoro M., *et al.* (2006). Structure and Chemical Inhibition of the RET Tyrosine Kinase Domain. *J Biol Chem.* 281(44):33577–87.
- [169] Hwang H., Vreven T., Janin J. and Weng Z. (2010). Protein-protein docking benchmark version 4.0. *Proteins.* 78(15):3111–4.
- [170] Arkin M. R. and Wells J. A. (2004). Small-molecule inhibitors of protein–protein interactions: progressing towards the dream. *Nat Rev Drug Discov.* 3(4):301–17.
- [171] Nero T. L., Morton C. J., Holien J. K., Wielens J. and Parker M. W. (2014). Oncogenic protein interfaces: small molecules, big challenges. *Nat Rev Cancer.* 14(4):248–62.
- [172] Jones S. and Thornton J. M. (1996). Principles of protein-protein interactions. *Proc Natl Acad Sci U S A.* 93(1):13–20.
- [173] Koes D., Khoury K., Huang Y., Wang W., Bista M., Popowicz G. M., *et al.* (2012). Enabling Large-Scale Design, Synthesis and Validation of Small Molecule Protein-Protein Antagonists. *PLoS ONE.* 7(3):e32839.

- [174] Rajamani D., Thiel S., Vajda S. and Camacho C. J. (2004). Anchor residues in protein–protein interactions. *Proc Natl Acad Sci U S A*. 101(31):11287–92.
- [175] Maga G. and Hübscher U. (2003). Proliferating cell nuclear antigen (PCNA): a dancer with many partners. *J Cell Sci*. 116(15):3051–60.
- [176] Moldovan G.-L., Pfander B. and Jentsch S. (2007). PCNA, the maestro of the replication fork. *Cell*. 129(4):665–79.
- [177] Pedley A. M., Lill M. A. and Davisson V. J. (2014). Flexibility of PCNA-Protein Interface Accommodates Differential Binding Partners. *PLoS ONE*. 9(7):e102481.
- [178] Punchihewa C., Inoue A., Hishiki A., Fujikawa Y., Connelly M., Evison B., *et al.* (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):14289–300.
- [179] Gulbis J. M., Kelman Z., Hurwitz J., O’Donnell M. and Kuriyan J. (1996). Structure of the C-terminal region of p21(WAF1/CIP1) complexed with human PCNA. *Cell*. 87(2):297–306.
- [180] Sanz G., Schlegel C., Pernollet J.-C. and Briand L. (2005). Comparison of Odorant Specificity of Two Human Olfactory Receptors from Different Phylogenetic Classes and Evidence for Antagonism. *Chem Senses*. 30(1):69–80.
- [181] Buck L. B. (2005). Unraveling the Sense of Smell (Nobel Lecture). *Angew Chem Int Ed*. 44(38):6128–40.
- [182] DeMaria S. and Ngai J. (2010). The cell biology of smell. *J Cell Biol*. 191(3):443–52.
- [183] Figueroa X. A., Cooksey G. A., Votaw S. V., Horowitz L. F. and Folch A. (2010). Large-scale investigation of the olfactory receptor space using a microfluidic microwell array. *Lab Chip*. 10(9):1120–7.
- [184] Hsieh J.-H., Yin S., Wang X. S., Liu S., Dokholyan N. V. and Tropsha A. (2012). Cheminformatics meets molecular mechanics: a combined application of knowledge-based pose scoring and physical force field-based hit scoring functions improves the accuracy of structure-based virtual screening. *J Chem Inf Model*. 52(1):16–28.
- [185] Nichols S. E., Baron R. and McCammon J. A. (2012). On the use of molecular dynamics receptor conformations for virtual screening. *Methods Mol Biol Clifton NJ*. 819:93–103.
- [186] Wereszczynski J. and McCammon J. A. (2012). Accelerated molecular dynamics in computational drug design. *Methods Mol Biol Clifton NJ*. 819:515–24.
- [187] Baron R. and McCammon J. A. (2013). Molecular recognition and ligand association. *Annu Rev Phys Chem*. 64:151–75.

- [188] Lindert S., Meiler J. and McCammon J. A. (2013). Iterative Molecular Dynamics-Rosetta Protein Structure Refinement Protocol to Improve Model Quality. *J Chem Theory Comput.* 9(8):3843–7.
- [189] Lindert S., Maslennikov I., Chiu E. J. C., Pierce L. C., McCammon J. A. and Choe S. (2014). Drug screening strategy for human membrane proteins: from NMR protein backbone structure to in silica- and NMR-screened hits. *Biochem Biophys Res Commun.* 445(4):724–33.
- [190] Sinko W., Lindert S. and McCammon J. A. (2013). Accounting for receptor flexibility and enhanced sampling methods in computer-aided drug design. *Chem Biol Drug Des.* 81(1):41–9.
- [191] Ivetac A. and McCammon J. A. (2011). Molecular recognition in the case of flexible targets. *Curr Pharm Des.* 17(17):1663–71.
- [192] Bohnuud T., Kozakov D. and Vajda S. (2014). Evidence of Conformational Selection Driving the Formation of Ligand Binding Sites in Protein-Protein Interfaces. *PLoS Comput Biol.* 10(10):e1003872.
- [193] Basom E. J., Spearman J. W. and Thielges M. C. (2015). Conformational Landscape and the Selectivity of Cytochrome P450cam. *J Phys Chem B.* 119(22):6620–7.
- [194] Johnson D. K. and Karanicolas J. (2015). Selectivity by Small-Molecule Inhibitors of Protein Interactions Can Be Driven by Protein Surface Fluctuations. *PLoS Comput Biol.* 11(2):e1004081.
- [195] Mahasenan K. V. and Li C. (2012). Novel Inhibitor Discovery through Virtual Screening against Multiple Protein Conformations Generated via Ligand-Directed Modeling: A Maternal Embryonic Leucine Zipper Kinase Example. *J Chem Inf Model.* 52(5):1345–55.
- [196] Nair P. C., Malde A. K., Drinkwater N. and Mark A. E. (2012). Missing Fragments: Detecting Cooperative Binding in Fragment-Based Drug Design. *ACS Med Chem Lett.* 3(4):322–6.
- [197] Kuhn B., Fuchs J. E., Reutlinger M., Stahl M. and Taylor N. R. (2011). Rationalizing Tight Ligand Binding through Cooperative Interaction Networks. *J Chem Inf Model.* 51(12):3180–98.
- [198] Li H. and Li C. (2010). Multiple ligand simultaneous docking: Orchestrated dancing of ligands in binding sites of protein. *J Comput Chem.* 31(10):2014–22.
- [199] Vass M., Tarcsay Á. and Keserü G. M. (2012). Multiple ligand docking by Glide: implications for virtual second-site screening. *J Comput Aided Mol Des.* 26(7):821–34.
- [200] Yu W., Xiao H., Lin J. and Li C. (2013). Discovery of Novel STAT3 Small Molecule Inhibitors via in Silico Site-Directed Fragment-Based Drug Design. *J Med Chem.* 56(11):4402–12.

- [201] Li H., Xiao H., Lin L., Jou D., Kumari V., Lin J., *et al.* (2014). Drug Design Targeting Protein–Protein Interactions (PPIs) Using Multiple Ligand Simultaneous Docking (MLSD) and Drug Repositioning: Discovery of Raloxifene and Bazedoxifene as Novel Inhibitors of IL-6/GP130 Interface. *J Med Chem.* 57(3):632–41.
- [202] Li H., Liu A., Zhao Z., Xu Y., Lin J., Jou D., *et al.* (2011). Fragment-Based Drug Design and Drug Repositioning Using Multiple Ligand Simultaneous Docking (MLSD): Identifying Celecoxib and Template Compounds as Novel Inhibitors of Signal Transducer and Activator of Transcription 3 (STAT3). *J Med Chem.* 54(15):5592–6.
- [203] Reed S., Li H., Li C. and Lin J. (2011). Celecoxib inhibits STAT3 phosphorylation and suppresses cell migration and colony forming ability in rhabdomyosarcoma cells. *Biochem Biophys Res Commun.* 407(3):450–5.
- [204] Johnson C. N., Adelinet C., Berdini V., Beke L., Bonnet P., Brehmer D., *et al.* (2014). Structure-Based Design of Type II Inhibitors Applied to Maternal Embryonic Leucine Zipper Kinase. *ACS Med Chem Lett.* 6(1):31–6.
- [205] Arkin M. R., Tang Y. and Wells J. A. (2014). Small-Molecule Inhibitors of Protein-Protein Interactions: Progressing toward the Reality. *Chem Biol.* 21(9):1102–14.
- [206] Edwards A. M., Isserlin R., Bader G. D., Frye S. V., Willson T. M. and Yu F. H. (2011). Too many roads not taken. *Nature.* 470(7333):163–5.
- [207] Fedorov O., Müller S. and Knapp S. (2010). The (un)targeted cancer kinome. *Nat Chem Biol.* 6(3):166–9.
- [208] Macarron R. (2006). Critical review of the role of HTS in drug discovery. *Drug Discov Today.* 11(7-8):277–9.
- [209] Shoichet B. K. (2004). Virtual screening of chemical libraries. *Nature.* 432(7019):862–5.
- [210] Kuenemann M. A., Bourbon L. M. L., Labbé C. M., Villoutreix B. O. and Sperandio O. (2014). Which Three-Dimensional Characteristics Make Efficient Inhibitors of Protein-Protein Interactions? *J Chem Inf Model.* 54(11):3067–79.
- [211] Ramani A. K., Bunescu R. C., Mooney R. J. and Marcotte E. M. (2005). Consolidating the set of known human protein-protein interactions in preparation for large-scale mapping of the human interactome. *Genome Biol.* 6(5):R40.
- [212] Lage K. (2014). Protein-protein interactions and genetic diseases: The Interactome. *Biochim Biophys Acta.* 1842(10):1971–80.
- [213] Berridge M. J. (2012). Module 6: Spatial and Temporal Aspects of Signaling. *Cell Signal Biol.* 6:6.1–6.52.
- [214] Warbrick E. (1998). PCNA binding through a conserved motif. *BioEssays.* 20(3):195–9.

- [215] Gilljam K. M., Feyzi E., Aas P. A., Sousa M. M. L., Müller R., Vågbø C. B., *et al.* (2009). Identification of a novel, widespread, and functionally important PCNA-binding motif. *J Cell Biol.* 186(5):645–54.
- [216] Mullard A. (2012). Protein–protein interaction inhibitors get into the groove. *Nat Rev Drug Discov.* 11(3):173–5.
- [217] Babine R. E. and Bender S. L. (1997). Molecular Recognition of Protein–Ligand Complexes: Applications to Drug Design. *Chem Rev.* 97(5):1359–472.
- [218] Lo Conte L., Chothia C. and Janin J. (1999). The atomic structure of protein-protein recognition sites. *J Mol Biol.* 285(5):2177–98.
- [219] Clackson T. and Wells J. A. (1995). A hot spot of binding energy in a hormone-receptor interface. *Science.* 267(5196):383–6.
- [220] Laraia L., McKenzie G., Spring D. R., Venkitaraman A. R. and Huggins D. J. (2015). Overcoming Chemical, Biological, and Computational Challenges in the Development of Inhibitors Targeting Protein-Protein Interactions. *Chem Biol.* 22(6):689–703.
- [221] Burgess K. (2001). Solid-Phase Syntheses of β -Turn Analogues To Mimic or Disrupt Protein–Protein Interactions. *Acc Chem Res.* 34(10):826–35.
- [222] Garner J. and Harding M. M. (2007). Design and synthesis of α -helical peptides and mimetics. *Org Biomol Chem.* 5(22):3577–85.
- [223] Kemp D. S., Bowen B. R. and Muendel C. C. (1990). Synthesis and conformational analysis of epindolidione-derived peptide models for β -sheet formation. *J Org Chem.* 55(15):4650–7.
- [224] Arnold U., Hinderaker M. P., Nilsson B. L., Huck B. R., Gellman S. H. and Raines R. T. (2002). Protein Prosthesis: A Semisynthetic Enzyme with a β -Peptide Reverse Turn. *J Am Chem Soc.* 124(29):8522–3.
- [225] Fasan R., Dias R. L. A., Moehle K., Zerbe O., Obrecht D., Mittl P. R. E., *et al.* (2006). Structure–Activity Studies in a Family of β -Hairpin Protein Epitope Mimetic Inhibitors of the p53–HDM2 Protein–Protein Interaction. *ChemBioChem.* 7(3):515–26.
- [226] Sadowsky J. D., Fairlie W. D., Hadley E. B., Lee H.-S., Umezawa N., Nikolovska-Coleska Z., *et al.* (2007). ($\alpha/\beta+\alpha$)-Peptide Antagonists of BH3 Domain/Bcl-xL Recognition: Toward General Strategies for Foldamer-Based Inhibition of Protein–Protein Interactions. *J Am Chem Soc.* 129(1):139–54.
- [227] Na Z., Peng B., Ng S., Pan S., Lee J.-S., Shen H.-M., *et al.* (2015). A small-molecule protein-protein interaction inhibitor of PARP1 that targets its BRCT domain. *Angew Chem Int Ed Engl.* 54(8):2515–9.
- [228] Jordheim L. P., Barakat K. H., Heinrich-Balard L., Matera E.-L., Cros-Perrial E., Bouletradak K., *et al.* (2013). Small Molecule Inhibitors of ERCC1-XPF Protein-Protein Interaction Synergize Alkylating Agents in Cancer Cells. *Mol Pharmacol.* 84(1):12–24.

- [229] Ding Q., Zhang Z., Liu J.-J., Jiang N., Zhang J., Ross T. M., *et al.* (2013). Discovery of RG7388, a Potent and Selective p53–MDM2 Inhibitor in Clinical Development. *J Med Chem.* 56(14):5979–83.
- [230] Zhao Y., Aguilar A., Bernard D. and Wang S. (2015). Small-Molecule Inhibitors of the MDM2–p53 Protein–Protein Interaction (MDM2 Inhibitors) in Clinical Trials for Cancer Treatment. *J Med Chem.* 58(3):1038–52.
- [231] Actis M., Inoue A., Evison B., Perry S., Punchihewa C. and Fujii N. (2013). Small molecule inhibitors of PCNA/PIP-box interaction suppress translesion DNA synthesis. *Bioorg Med Chem.* 21(7):1972–7.
- [232] Inoue A., Kikuchi S., Hishiki A., Shao Y., Heath R., Evison B. J., *et al.* (2014). A Small Molecule Inhibitor of Monoubiquitinated Proliferating Cell Nuclear Antigen (PCNA) inhibits Repair of Interstrand DNA Crosslink, enhances DNA Double-strand Break, and sensitizes Cancer Cells to Cisplatin. *J Biol Chem.* 289(10):7109–20.
- [233] Tan Z., Wortman M., Dillehay K. L., Seibel W. L., Evelyn C. R., Smith S. J., *et al.* (2012). Small-Molecule Targeting of Proliferating Cell Nuclear Antigen Chromatin Association Inhibits Tumor Cell Growth. *Mol Pharmacol.* 81(6):811–9.
- [234] Dillehay K. L., Lu S. and Dong Z. (2014). Anti-tumor effects of a novel small molecule targeting PCNA chromatin association in prostate cancer. *Mol Cancer Ther.* 13(12):2817–26.
- [235] Dillehay K. L., Seibel W. L., Zhao D., Lu S. and Dong Z. (2015). Target validation and structure-activity analysis of a series of novel PCNA inhibitors. *Pharmacol Res Perspect.* 3(2):e00115.
- [236] Guo W., Wisniewski J. A. and Ji H. (2014). Hot spot-based design of small-molecule inhibitors for protein-protein interactions. *Bioorg Med Chem Lett.*
- [237] Lipinski C. A., Lombardo F., Dominy B. W. and Feeney P. J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev.* 23(1–3):3–25.
- [238] Wells J. A. and McClendon C. L. (2007). Reaching for high-hanging fruit in drug discovery at protein–protein interfaces. *Nature.* 450(7172):1001–9.
- [239] Sperandio O., Reynès C. H., Camproux A.-C. and Villoutreix B. O. (2010). Rationalizing the chemical space of protein–protein interaction inhibitors. *Drug Discov Today.* 15(5–6):220–9.
- [240] Morelli X., Bourgeas R. and Roche P. (2011). Chemical and structural lessons from recent successes in protein–protein interaction inhibition (2P2I). *Curr Opin Chem Biol.* 15(4):475–81.
- [241] Yin H. and Hamilton A. D. (2005). Strategies for targeting protein-protein interactions with synthetic agents. *Angew Chem Int Ed Engl.* 44(27):4130–63.

- [242] Turnbull A., Boyd S. and Walse B. (2014). Fragment-based drug discovery and protein-protein interactions. *Res Rep Biochem.* 4:13–26.
- [243] Spencer R. W. (1998). High-throughput screening of historic collections: observations on file size, biological targets, and file diversity. *Biotechnol Bioeng.* 61(1):61–7.
- [244] Bleicher K. H., Böhm H.-J., Müller K. and Alanine A. I. (2003). Hit and lead generation: beyond high-throughput screening. *Nat Rev Drug Discov.* 2(5):369–78.
- [245] Pritchard J. F., Jurima-Romet M., Reimer M. L. J., Mortimer E., Rolfe B. and Cayen M. N. (2003). Making better drugs: Decision gates in non-clinical drug development. *Nat Rev Drug Discov.* 2(7):542–53.
- [246] Pagliaro L., Felding J., Audouze K., Nielsen S. J., Terry R. B., Krog-Jensen C., *et al.* (2004). Emerging classes of protein-protein interaction inhibitors and new tools for their development. *Curr Opin Chem Biol.* 8(4):442–9.
- [247] Hopkins A. L. and Groom C. R. (2002). The druggable genome. *Nat Rev Drug Discov.* 1(9):727–30.
- [248] Chène P. (2006). Drugs Targeting Protein–Protein Interactions. *ChemMedChem.* 1(4):400–11.
- [249] Teague S. J. (2011). Learning lessons from drugs that have recently entered the market. *Drug Discov Today.* 16(9–10):398–411.
- [250] Overington J. P., Al-Lazikani B. and Hopkins A. L. (2006). How many drug targets are there? *Nat Rev Drug Discov.* 5(12):993–6.
- [251] Kelman Z. and O'Donnell M. (1995). Structural and functional similarities of prokaryotic and eukaryotic DNA polymerase sliding clamps. *Nucleic Acids Res.* 23(18):3613–20.
- [252] Leonardi E., Girlando S., Serio G., Mauri F. A., Perrone G., Scampini S., *et al.* (1992). PCNA and Ki67 expression in breast carcinoma: correlations with clinical and biological variables. *J Clin Pathol.* 45(5):416–9.
- [253] Essers J., Theil A. F., Baldeyron C., van Cappellen W. A., Houtsmuller A. B., Kanaar R., *et al.* (2005). Nuclear Dynamics of PCNA in DNA Replication and Repair. *Mol Cell Biol.* 25(21):9350–9.
- [254] Tsurimoto T. and Stillman B. (1990). Functions of replication factor C and proliferating-cell nuclear antigen: functional similarity of DNA polymerase accessory proteins from human cells and bacteriophage T4. *Proc Natl Acad Sci U S A.* 87(3):1023–7.
- [255] Tsurimoto T. and Stillman B. (1991). Replication factors required for SV40 DNA replication in vitro. II. Switching of DNA polymerase alpha and delta during initiation of leading and lagging strand synthesis. *J Biol Chem.* 266(3):1961–8.
- [256] Stoimenov I. and Helleday T. (2009). PCNA on the crossroad of cancer. *Biochem Soc Trans.* 37(Pt 3):605–13.

- [257] Egelkrout E. M., Mariconti L., Settlege S. B., Cella R., Robertson D. and Hanley-Bowdoin L. (2002). Two E2F Elements Regulate the Proliferating Cell Nuclear Antigen Promoter Differently during Leaf Development. *Plant Cell Online*. 14(12):3225–36.
- [258] Paunesku T., Mittal S., Protić M., Oryhon J., Korolev S. V., Joachimiak A., *et al.* (2001). Proliferating cell nuclear antigen (PCNA): ringmaster of the genome. *Int J Radiat Biol*. 77(10):1007–21.
- [259] Hoege C., Pfander B., Moldovan G.-L., Pyrowolakis G. and Jentsch S. (2002). RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature*. 419(6903):135–41.
- [260] Wang S.-C., Nakajima Y., Yu Y.-L., Xia W., Chen C.-T., Yang C.-C., *et al.* (2006). Tyrosine phosphorylation controls PCNA function through protein stability. *Nat Cell Biol*. 8(12):1359–68.
- [261] Zhao H., Lo Y.-H., Ma L., Waltz S. E., Gray J. K., Hung M.-C., *et al.* (2011). Targeting Tyrosine Phosphorylation of PCNA Inhibits Prostate Cancer Growth. *Mol Cancer Ther*. 10:29–36.
- [262] Zhao H., Ho P.-C., Lo Y.-H., Espejo A., Bedford M. T., Hung M.-C., *et al.* (2012). Interaction of Proliferation Cell Nuclear Antigen (PCNA) with c-Abl in Cell Proliferation and Response to DNA Damages in Breast Cancer. *PLoS ONE*. 7(1):e29416.
- [263] Zhao H., Chen M.-S., Lo Y.-H., Waltz S. E., Wang J., Ho P.-C., *et al.* (2014). The Ron receptor tyrosine kinase activates c-Abl to promote cell proliferation through tyrosine phosphorylation of PCNA in breast cancer. *Oncogene*. 33(11):1429–37.
- [264] Mailand N., Gibbs-Seymour I. and Bekker-Jensen S. (2013). Regulation of PCNA-protein interactions for genome stability. *Nat Rev Mol Cell Biol*. 14(5):269–82.
- [265] Harper M. E., Glynne-Jones E., Goddard L., Wilson D. W., Griffiths K., Conn I. G., *et al.* (1992). Relationship of proliferating cell nuclear antigen (PCNA) in prostatic carcinomas to various clinical parameters. *The Prostate*. 20(3):243–53.
- [266] Tahan S. R., Neuberger D. S., Dieffenbach A. and Yacoub L. (1993). Prediction of early relapse and shortened survival in patients with breast cancer by proliferating cell nuclear antigen score. *Cancer*. 71(11):3552–9.
- [267] Zhao R., Hirano M. and Kurita S. (1996). Expression of proliferating cell nuclear antigen in premalignant lesions of the larynx. *Am J Otolaryngol*. 17(1):36–44.
- [268] Baldi G., Baldi F., Maguire M. and Massaro-Giordano M. (1998). Prognostic factors for survival after enucleation for choroidal melanoma. *Int J Oncol*. 13(6):1185–9.
- [269] Tsuji T., Kitajima S. and Koashi Y. (1998). Expression of proliferating cell nuclear antigen (PCNA) and apoptosis related antigen (LeY) in epithelial skin tumors. *Am J Dermatopathol*. 20(2):164–9.

- [270] Korkolopoulou P., Angelopoulou M. K., Kontopidou F., Tsengas A., Patsouris E., Kittas C., *et al.* (1998). Prognostic implications of proliferating cell nuclear antigen (PCNA), AgNORs and P53 in non-Hodgkin's lymphomas. *Leuk Lymphoma*. 30(5-6):625–36.
- [271] Chu J. S., Huang C. S. and Chang K. J. (1998). Proliferating cell nuclear antigen (PCNA) immunolabeling as a prognostic factor in invasive ductal carcinoma of the breast in Taiwan. *Cancer Lett*. 131(2):145–52.
- [272] Tsurimoto T. (1998). PCNA, a multifunctional ring on DNA. *Biochim Biophys Acta*. 1443(1-2):23–39.
- [273] Aaltomaa S., Lipponen P. and Syrjänen K. (1993). Proliferating cell nuclear antigen (PCNA) immunolabeling as a prognostic factor in axillary lymph node negative breast cancer. *Anticancer Res*. 13(2):533–8.
- [274] Bergh J. (1999). Clinical studies of p53 in treatment and benefit of breast cancer patients. *Endocr Relat Cancer*. 6(1):51–9.
- [275] Lynch H. T., Watson P., Tinley S., Snyder C., Durham C., Lynch J., *et al.* (1999). An update on DNA-based BRCA1/BRCA2 genetic counseling in hereditary breast cancer. *Cancer Genet Cytogenet*. 109(2):91–8.
- [276] Ortega J., Li J. Y., Lee S., Tong D., Gu L. and 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):5667–72.
- [277] Wang S.-C. (2014). PCNA: a silent housekeeper or a potential therapeutic target? *Trends Pharmacol Sci*. 35(4):178–86.
- [278] Smith S. J., Gu L., Phipps E. A., Dobrolecki L. E., Mabrey K. S., Gulley P., *et al.* (2014). A Peptide Mimicking a Region in Proliferating Cell Nuclear Antigen (PCNA) Specific to Key Protein Interactions is Cytotoxic to Breast Cancer. *Mol Pharmacol*. 87(2):263–76.
- [279] Lingeman R. G., Hickey R. J. and Malkas L. H. (2014). Expression of a novel peptide derived from PCNA damages DNA and reverses cisplatin resistance. *Cancer Chemother Pharmacol*. 74(5):981–93.
- [280] Gu L., Smith S., Li C., Hickey R. J., Stark J. M., Fields G. B., *et al.* (2014). A PCNA-Derived Cell Permeable Peptide Selectively Inhibits Neuroblastoma Cell Growth. *PloS One*. 9(4):e94773.
- [281] Warbrick E., Lane D. P., Glover D. M. and 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):275–82.
- [282] Pan Z.-Q., Reardon J. T., Li L., Flores-Rozas H., Legerski R., Sancar A., *et al.* (1995). Inhibition of Nucleotide Excision Repair by the Cyclin-dependent Kinase Inhibitor p21. *J Biol Chem*. 270(37):22008–16.

- [283] Chen J., Peters R., Saha P., Lee P., Theodoras A., Pagano M., *et al.* (1996). A 39 Amino Acid Fragment of the Cell Cycle Regulator p21 Is Sufficient to Bind PCNA and Partially Inhibit DNA Replication in vivo. *Nucleic Acids Res.* 24(9):1727–33.
- [284] Fridman Y., Palgi N., Dovrat D., Ben-Aroya S., Hieter P. and Aharoni A. (2010). Subtle Alterations in PCNA-Partner Interactions Severely Impair DNA Replication and Repair. *PLoS Biol.* 8(10):e1000507.
- [285] Fridman Y., Gur E., Fleishman S. J. and Aharoni A. (2013). Computational protein design suggests that human PCNA-partner interactions are not optimized for affinity. *Proteins Struct Funct Bioinforma.* 81(2):341–8.
- [286] Yu Y.-L., Chou R.-H., Liang J.-H., Chang W.-J., Su K.-J., Tseng Y.-J., *et al.* (2013). Targeting the EGFR/PCNA Signaling Suppresses Tumor Growth of Triple-Negative Breast Cancer Cells with Cell-Penetrating PCNA Peptides. *PLoS ONE.* 8(4):e61362.
- [287] Hishiki A., Hashimoto H., Hanafusa T., Kamei K., Ohashi E., Shimizu T., *et al.* (2009). Structural Basis for Novel Interactions between Human Translesion Synthesis Polymerases and Proliferating Cell Nuclear Antigen. *J Biol Chem.* 284(16):10552–60.
- [288] Gulbis J. M., Kelman Z., Hurwitz J., O'Donnell M. and Kuriyan J. (1996). Structure of the C-Terminal Region of p21WAF1/CIP1 Complexed with Human PCNA. *Cell.* 87(2):297–306.
- [289] Sakurai S., Kitano K., Yamaguchi H., Hamada K., Okada K., Fukuda K., *et al.* (2005). Structural basis for recruitment of human flap endonuclease 1 to PCNA. *EMBO J.* 24(4):683–93.
- [290] Bruning J. B. and Shamoo Y. (2004). Structural and thermodynamic analysis of human PCNA with peptides derived from DNA polymerase-delta p66 subunit and flap endonuclease-1. *Struct Lond Engl* 1993. 12(12):2209–19.
- [291] He X., Wei C., Song T., Yuan J., Zhang Y., Ma Q., *et al.* (2009). Proliferating cell nuclear antigen destabilizes c-Abl tyrosine kinase and regulates cell apoptosis in response to DNA damage. *Apoptosis Int J Program Cell Death.* 14(3):268–75.
- [292] Guo W., Wu S., Liu J. and Fang B. (2008). Identification of a small molecule with synthetic lethality for K-ras and protein kinase C iota. *Cancer Res.* 68(18):7403–8.
- [293] Warbrick E., Lane D. P., Glover D. M. and Cox L. S. (1997). Homologous regions of Fen1 and p21Cip1 compete for binding to the same site on PCNA: a potential mechanism to coordinate DNA replication and repair. *Oncogene.* 14(19):2313–21.
- [294] Zheleva D. I., Zhelev N. Z., Fischer P. M., Duff S. V., Warbrick E., Blake D. G., *et al.* (2000). A Quantitative Study of the in Vitro Binding of the C-Terminal Domain of p21 to PCNA: Affinity, Stoichiometry, and Thermodynamics. *Biochemistry (Mosc).* 39(25):7388–97.
- [295] Otvos L. and Wade J. D. (2014). Current challenges in peptide-based drug discovery. *Front Chem.* 2:62.

- [296] Kwon Y.-U. and Kodadek T. (2007). Quantitative Evaluation of the Relative Cell Permeability of Peptoids and Peptides. *J Am Chem Soc.* 129(6):1508–9.
- [297] Miller S. M., Simon R. J., Ng S., Zuckermann R. N., Kerr J. M. and Moos W. H. (1994). Proteolytic studies of homologous peptide and N-substituted glycine peptoid oligomers. *Bioorg Med Chem Lett.* 4(22):2657–62.
- [298] Sanborn T. J., Wu C. W., Zuckermann R. N. and Barron A. E. (2002). Extreme stability of helices formed by water-soluble poly-N-substituted glycines (polypeptoids) with alpha-chiral side chains. *Biopolymers.* 63(1):12–20.
- [299] Wender P. A., Mitchell D. J., Pattabiraman K., Pelkey E. T., Steinman L. and Rothbard J. B. (2000). The design, synthesis, and evaluation of molecules that enable or enhance cellular uptake: peptoid molecular transporters. *Proc Natl Acad Sci U S A.* 97(24):13003–8.
- [300] Zuckermann R. N., Kerr J. M., Kent S. B. H. and Moos W. H. (1992). Efficient method for the preparation of peptoids [oligo(N-substituted glycines)] by submonomer solid-phase synthesis. *J Am Chem Soc.* 114(26):10646–7.
- [301] Diaz J. P., Chirayil R., Chirayil S., Tom M., Head K. J. and Luebke K. J. (2014). Association of a peptoid ligand with the apical loop of pri-miR-21 inhibits cleavage by Drosha. *RNA.* 20(4):528–39.
- [302] Vendrell-Navarro G., Rúa F., Bujons J., Brockmeyer A., Janning P., Ziegler S., *et al.* (2015). Positional Scanning Synthesis of a Peptoid Library Yields Novel Inducers of Apoptosis Targeting Karyopherins and Tubulin. *ChemBioChem.* 16(11):1580–7.
- [303] Scheper J., Guerra-Rebollo M., Sanclimens G., Moure A., Masip I., González-Ruiz D., *et al.* (2010). Protein-Protein Interaction Antagonists as Novel Inhibitors of Non-Canonical Polyubiquitylation. *PLoS ONE.* 5(6):e11403.
- [304] Goodson B., Ehrhardt A., Ng S., Nuss J., Johnson K., Giedlin M., *et al.* (1999). Characterization of Novel Antimicrobial Peptoids. *Antimicrob Agents Chemother.* 43(6):1429–34.
- [305] Ng S., Goodson B., Ehrhardt A., Moos W. H., Siani M. and Winter J. (1999). Combinatorial discovery process yields antimicrobial peptoids. *Bioorg Med Chem.* 7(9):1781–5.
- [306] Culf A. S. and Ouellette R. J. (2010). Solid-Phase Synthesis of N-Substituted Glycine Oligomers (α -Peptoids) and Derivatives. *Molecules.* 15(8):5282–335.
- [307] Fowler S. A. and Blackwell H. E. (2009). Structure–function relationships in peptoids: Recent advances toward deciphering the structural requirements for biological function. *Org Biomol Chem.* 7(8):1508–24.
- [308] Ronald N Zuckermann T. K. (2009). Peptoids as potential therapeutics. *Curr Opin Mol Ther.* 11(3):299–307.
- [309] Brown N. J., Johansson J. and Barron A. E. (2008). Biomimicry of Surfactant Protein C. *Acc Chem Res.* 41(10):1409–17.

- [310] Tran H., Gael S. L., Connolly M. D. and Zuckermann R. N. (2011). Solid-phase Submonomer Synthesis of Peptoid Polymers and their Self-Assembly into Highly-Ordered Nanosheets. *J Vis Exp JoVE*. (57):e3373.
- [311] Olivier G. K., Cho A., Sanii B., Connolly M. D., Tran H. and Zuckermann R. N. (2013). Antibody-Mimetic Peptoid Nanosheets for Molecular Recognition. *ACS Nano*. 7(10):9276–86.
- [312] Reddy M. M., Wilson R., Wilson J., Connell S., Gocke A., Hynan L., *et al.* (2011). Identification of Candidate IgG Biomarkers for Alzheimer’s Disease via Combinatorial Library Screening. *Cell*. 144(1):132–42.
- [313] Udugamasooriya D. G., Dineen S. P., Brekken R. A. and Kodadek T. (2008). A Peptoid “Antibody Surrogate” That Antagonizes VEGF Receptor 2 Activity. *J Am Chem Soc*. 130(17):5744–52.
- [314] Nguyen J. T., Turck C. W., Cohen F. E., Zuckermann R. N. and Lim W. A. (1998). Exploiting the Basis of Proline Recognition by SH3 and WW Domains: Design of N-Substituted Inhibitors. *Science*. 282(5396):2088–92.
- [315] Shankaramma S. C., Moehle K., James S., Vrijbloed J. W., Obrecht D. and Robinson J. A. (2003). A family of macrocyclic antibiotics with a mixed peptide-peptoid beta-hairpin backbone conformation. *Chem Commun Camb Engl*. (15):1842–3.
- [316] Chongsiriwatana N. P., Patch J. A., Czyzewski A. M., Dohm M. T., Ivankin A., Gidalevitz D., *et al.* (2008). Peptoids that mimic the structure, function, and mechanism of helical antimicrobial peptides. *Proc Natl Acad Sci*. 105(8):2794–9.
- [317] Kapoor R., Wadman M. W., Dohm M. T., Czyzewski A. M., Spormann A. M. and Barron A. E. (2011). Antimicrobial Peptoids Are Effective against *Pseudomonas aeruginosa* Biofilms. *Antimicrob Agents Chemother*. 55(6):3054–7.
- [318] Zuckermann R. N., Martin E. J., Spellmeyer D. C., Stauber G. B., Shoemaker K. R., Kerr J. M., *et al.* (1994). Discovery of nanomolar ligands for 7-transmembrane G-protein-coupled receptors from a diverse N-(substituted)glycine peptoid library. *J Med Chem*. 37(17):2678–85.
- [319] Thompson D. A., Chai B.-X., Rood H. L. E., Siani M. A., Douglas N. R., Gantz I., *et al.* (2003). Peptoid mimics of agouti related protein. *Bioorg Med Chem Lett*. 13(8):1409–13.
- [320] García-Martínez C., Humet M., Planells-Cases R., Gomis A., Caprini M., Viana F., *et al.* (2002). Attenuation of thermal nociception and hyperalgesia by VR1 blockers. *Proc Natl Acad Sci*. 99(4):2374–9.
- [321] William A. Kleschick, L. Navelle Davis, Michael R. Dick, Joseph R. Garlich, Eric J. Martin, Nailah Orr, *et al.* (2000). The Application of Combinatorial Chemistry in Agrochemical Discovery. In: *Agrochemical Discovery [Internet]*. American Chemical Society. p. 205–13. Available from: <http://dx.doi.org/10.1021/bk-2001-0774.ch018>

- [322] Heizmann G., Hildebrand P., Tanner H., Ketterer S., Pansky A., Froidevaux S., *et al.* (1999). A combinatorial peptoid library for the identification of novel MSH and GRP/bombesin receptor ligands. *J Recept Signal Transduct Res.* 19(1-4):449–66.
- [323] Persidis A. (1999). Cancer multidrug resistance. *Nat Biotechnol.* 17(1):94–5.
- [324] Lehmann A. R. and Fuchs R. P. (2006). Gaps and forks in DNA replication: Rediscovering old models. *DNA Repair.* 5(12):1495–8.
- [325] Simon R. J., Kania R. S., Zuckermann R. N., Huebner V. D., Jewell D. A., Banville S., *et al.* (1992). Peptoids: A Modular Approach to Drug Discovery. *Proc Natl Acad Sci.* 89(20):9367–71.
- [326] Schrödinger Release 2013-2: Maestro, version 9.5, Schrödinger, LLC, New York, NY, 2013.
- [327] (2013). Small-Molecule Drug Discovery Suite 2013-3: Schrödinger Suite 2013-3 Induced Fit Docking protocol; Glide version 6.1, Schrödinger, LLC, New York, NY, 2013; Prime version 3.4, Schrödinger, LLC, New York, NY, 2013. New York, NY.
- [328] Small-Molecule Drug Discovery Suite 2013-2: Glide, version 6.0, Schrödinger, LLC, New York, NY, 2013.
- [329] Nikolovska-Coleska Z., Wang R., Fang X., Pan H., Tomita Y., Li P., *et al.* (2004). Development and optimization of a binding assay for the XIAP BIR3 domain using fluorescence polarization. *Anal Biochem.* 332(2):261–73.
- [330] Bowers K. J., Chow E., Xu H., Dror R. O., Eastwood M. P., Gregersen B. A., *et al.* (2006). Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In: *Proceedings of the ACM/IEEE SC 2006 Conference.* p. 43–43.
- [331] Humphrey W., Dalke A. and Schulten K. (1996). VMD: visual molecular dynamics. *J Mol Graph.* 14(1):33–8, 27–8.
- [332] Grant B. J., Rodrigues A. P. C., ElSawy K. M., McCammon J. A. and Caves L. S. D. (2006). Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics.* 22(21):2695–6.
- [333] Bremner J. B., Keller P. A., Pyne S. G., Boyle T. P., Brkic Z., Morgan J., *et al.* (2010). Synthesis and antibacterial studies of binaphthyl-based tripeptoids. Part 2. *Bioorg Med Chem.* 18(13):4793–800.
- [334] Zhang J.-H., Chung T. D. Y. and Oldenburg K. R. (1999). A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J Biomol Screen.* 4(2):67–73.
- [335] Kontopidis G., Wu S.-Y., Zheleva D. I., Taylor P., McInnes C., Lane D. P., *et al.* (2005). Structural and biochemical studies of human proliferating cell nuclear antigen complexes provide a rationale for cyclin association and inhibitor design. *Proc Natl Acad Sci U S A.* 102(6):1871–6.

- [336] (2015). Schrödinger Release 2015-3: Canvas, version 2.5. New York, NY: Schrödinger, LLC.
- [337] Xia X., Maliski E. G., Gallant P. and Rogers D. (2004). Classification of Kinase Inhibitors Using a Bayesian Model. *J Med Chem.* 47(18):4463–70.
- [338] Rogers D., Brown R. D. and Hahn M. (2005). Using Extended-Connectivity Fingerprints with Laplacian-Modified Bayesian Analysis in High-Throughput Screening Follow-Up. *J Biomol Screen.* 10(7):682–6.
- [339] Glick M., Jenkins J. L., Nettles J. H., Hitchings H. and Davies J. W. (2006). Enrichment of High-Throughput Screening Data with Increasing Levels of Noise Using Support Vector Machines, Recursive Partitioning, and Laplacian-Modified Naive Bayesian Classifiers. *J Chem Inf Model.* 46(1):193–200.
- [340] Bourgeas R., Basse M.-J., Morelli X. and Roche P. (2010). Atomic analysis of protein-protein interfaces with known inhibitors: the 2P2I database. *PLoS One.* 5(3):e9598.
- [341] Hamon V., Bourgeas R., Ducrot P., Theret I., Xuereb L., Basse M. J., *et al.* (2014). 2P2I HUNTER: a tool for filtering orthosteric protein-protein interaction modulators via a dedicated support vector machine. *J R Soc Interface R Soc.* 11(90):20130860.
- [342] Accelrys Software Inc. (2007). Discovery Studio Modeling Environment, Release 4.1. San Diego: Accelrys Software Inc.
- [343] Tucker C. L. and Fields S. (2003). Lethal combinations. *Nat Genet.* 35(3):204–5.
- [344] Xu H., Zhang P., Liu L. and Lee M. Y. (2001). A novel PCNA-binding motif identified by the panning of a random peptide display library. *Biochemistry (Mosc).* 40(14):4512–20.
- [345] Cohen P. (2001). The role of protein phosphorylation in human health and disease. *Eur J Biochem.* 268(19):5001–10.
- [346] Manning G., Whyte D. B., Martinez R., Hunter T. and Sudarsanam S. (2002). The protein kinase complement of the human genome. *Science.* 298(5600):1912–34.
- [347] Yarden Y. and Sliwkowski M. X. (2001). Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol.* 2(2):127–37.
- [348] Libermann T. A., Razon N., Bartal A. D., Yarden Y., Schlessinger J. and Soreq H. (1984). Expression of epidermal growth factor receptors in human brain tumors. *Cancer Res.* 44(2):753–60.
- [349] Libermann T. A., Nusbaum H. R., Razon N., Kris R., Lax I., Soreq H., *et al.* (1985). Amplification and overexpression of the EGF receptor gene in primary human glioblastomas. *J Cell Sci Suppl.* 3:161–72.
- [350] Libermann T. A., Nusbaum H. R., Razon N., Kris R., Lax I., Soreq H., *et al.* (1985). Amplification, enhanced expression and possible rearrangement of EGF receptor gene in primary human brain tumours of glial origin. *Nature.* 313(5998):144–7.

- [351] Veale D., Ashcroft T., Marsh C., Gibson G. J. and Harris A. L. (1987). Epidermal growth factor receptors in non-small cell lung cancer. *Br J Cancer*. 55(5):513–6.
- [352] Weichselbaum R. R., Dunphy E. J., Beckett M. A., Tybor A. G., Moran W. J., Goldman M. E., *et al.* (1989). Epidermal growth factor receptor gene amplification and expression in head and neck cancer cell lines. *Head Neck*. 11(5):437–42.
- [353] Walker F., Abramowitz L., Benabderrahmane D., Duval X., Descatoire V., Hénin D., *et al.* (2009). Growth factor receptor expression in anal squamous lesions: modifications associated with oncogenic human papillomavirus and human immunodeficiency virus. *Hum Pathol*. 40(11):1517–27.
- [354] Nicholson R. I., Gee J. M. and Harper M. E. (2001). EGFR and cancer prognosis. *Eur J Cancer Oxf Engl 1990*. 37 Suppl 4:S9–15.
- [355] Levitzki A. (1999). Protein tyrosine kinase inhibitors as novel therapeutic agents. *Pharmacol Ther*. 82(2-3):231–9.
- [356] Wind N. S. and Holen I. (2011). Multidrug Resistance in Breast Cancer: From In Vitro Models to Clinical Studies. *Int J Breast Cancer*. 2011:1–12.
- [357] Fabian M. A., Biggs W. H., Treiber D. K., Atteridge C. E., Azimioara M. D., Benedetti M. G., *et al.* (2005). A small molecule-kinase interaction map for clinical kinase inhibitors. *Nat Biotech*. 23(3):329–36.
- [358] Karaman M. W., Herrgard S., Treiber D. K., Gallant P., Atteridge C. E., Campbell B. T., *et al.* (2008). A quantitative analysis of kinase inhibitor selectivity. *Nat Biotech*. 26(1):127–32.
- [359] Fedorov O., Marsden B., Pogacic V., Rellos P., Müller S., Bullock A. N., *et al.* (2007). A systematic interaction map of validated kinase inhibitors with Ser/Thr kinases. *Proc Natl Acad Sci*. 104(51):20523–8.
- [360] Bain J., Plater L., Elliott M., Shpiro N., Hastie C. J., Mclauchlan H., *et al.* (2007). The selectivity of protein kinase inhibitors: a further update. *Biochem J*. 408(3):297.
- [361] Bain J., McLauchlan H., Elliott M. and Cohen P. (2003). The specificities of protein kinase inhibitors: an update. *Biochem J*. 371(1):199.
- [362] Davies S. P., Reddy H., Caivano M. and Cohen P. (2000). Specificity and mechanism of action of some commonly used protein kinase inhibitors. *Biochem J*. 351(Pt 1):95–105.
- [363] Force T. and Kolaja K. L. (2011). Cardiotoxicity of kinase inhibitors: the prediction and translation of preclinical models to clinical outcomes. *Nat Rev Drug Discov*. 10(2):111–26.
- [364] Jänne P. A., Gray N. and Settleman J. (2009). Factors underlying sensitivity of cancers to small-molecule kinase inhibitors. *Nat Rev Drug Discov*. 8(9):709–23.

- [365] Bixby D. and Talpaz M. (2009). Mechanisms of resistance to tyrosine kinase inhibitors in chronic myeloid leukemia and recent therapeutic strategies to overcome resistance. *ASH Educ Program Book*. 2009(1):461–76.
- [366] Engelman J. A. and Settleman J. (2008). Acquired resistance to tyrosine kinase inhibitors during cancer therapy. *Curr Opin Genet Dev*. 18(1):73–9.
- [367] Chen F. L., Xia W. and Spector N. L. (2008). Acquired resistance to small molecule ErbB2 tyrosine kinase inhibitors. *Clin Cancer Res Off J Am Assoc Cancer Res*. 14(21):6730–4.
- [368] Wilson T. R., Fridlyand J., Yan Y., Penuel E., Burton L., Chan E., *et al.* (2012). Widespread potential for growth-factor-driven resistance to anticancer kinase inhibitors. *Nature*. 487(7408):505–9.
- [369] Ohashi K., Maruvka Y. E., Michor F. and Pao W. (2013). Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitor–Resistant Disease. *J Clin Oncol*. 31(8):1070–80.
- [370] Cortot A. B., Repellin C. E., Shimamura T., Capelletti M., Zejnullahu K., Ercan D., *et al.* (2012). Resistance to Irreversible EGFR Tyrosine Kinase Inhibitors through a Multistep Mechanism Involving the IGF1R Pathway. *Cancer Res*. 73(2):834–43.
- [371] Aoyama K., Fukumoto Y., Ishibashi K., Kubota S., Morinaga T., Horiike Y., *et al.* (2011). Nuclear c-Abl-mediated tyrosine phosphorylation induces chromatin structural changes through histone modifications that include H4K16 hypoacetylation. *Exp Cell Res*. 317(20):2874–903.
- [372] De Angelis Campos A. C., Rodrigues M. A., de Andrade C., de Goes A. M., Nathanson M. H. and Gomes D. A. (2011). Epidermal growth factor receptors destined for the nucleus are internalized via a clathrin-dependent pathway. *Biochem Biophys Res Commun*. 412(2):341–6.
- [373] Wang S.-C. and Hung M.-C. (2009). Nuclear Translocation of the Epidermal Growth Factor Receptor Family Membrane Tyrosine Kinase Receptors. *Clin Cancer Res*. 15:6484–9.
- [374] Campbell E. J., McDuff E., Tatarov O., Tovey S., Brunton V., Cooke T. G., *et al.* (2008). Phosphorylated c-Src in the nucleus is associated with improved patient outcome in ER-positive breast cancer. *Br J Cancer*. 99(11):1769–74.
- [375] Thiel K. W. and Carpenter G. (2006). ErbB-4 and TNF-alpha converting enzyme localization to membrane microdomains. *Biochem Biophys Res Commun*. 350(3):629–33.
- [376] Wang R. and Brattain M. G. (2006). AKT can be activated in the nucleus. *Cell Signal*. 18(10):1722–31.
- [377] Lo H.-W., Hsu S.-C. and Hung M.-C. (2006). EGFR signaling pathway in breast cancers: from traditional signal transduction to direct nuclear translocalization. *Breast Cancer Res Treat*. 95(3):211–8.
- [378] Polo S., Pece S. and Di Fiore P. P. (2004). Endocytosis and cancer. *Curr Opin Cell Biol*. 16(2):156–61.

- [379] Carpenter G. (2003). Nuclear localization and possible functions of receptor tyrosine kinases. *Curr Opin Cell Biol.* 15(2):143–8.
- [380] Ito Y., Pandey P., Mishra N., Kumar S., Narula N., Kharbanda S., *et al.* (2001). Targeting of the c-Abl tyrosine kinase to mitochondria in endoplasmic reticulum stress-induced apoptosis. *Mol Cell Biol.* 21(18):6233–42.
- [381] David-Pfeuty T., Bagrodia S. and Shalloway D. (1993). Differential localization patterns of myristoylated and nonmyristoylated c-Src proteins in interphase and mitotic c-Src overexpresser cells. *J Cell Sci.* 105 (3):613–28.
- [382] Carpenter G. and Liao H.-J. (2013). Receptor tyrosine kinases in the nucleus. *Cold Spring Harb Perspect Biol.* 5(10):a008979.
- [383] Ambudkar S. V., Dey S., Hrycyna C. A., Ramachandra M., Pastan I. and Gottesman M. M. (1999). Biochemical, cellular, and pharmacological aspects of the multidrug transporter. *Annu Rev Pharmacol Toxicol.* 39:361–98.
- [384] Chen Y.-J., Huang W.-C., Wei Y.-L., Hsu S.-C., Yuan P., Lin H. Y., *et al.* (2011). Elevated BCRP/ABCG2 Expression Confers Acquired Resistance to Gefitinib in Wild-Type EGFR-Expressing Cells. *PLoS ONE.* 6(6):e21428.
- [385] Thomas J., Wang L., Clark R. E. and Pirmohamed M. (2004). Active transport of imatinib into and out of cells: implications for drug resistance. *Blood.* 104(12):3739–45.
- [386] An X., Tiwari A. K., Sun Y., Ding P.-R., Ashby C. R. Jr and Chen Z.-S. (2010). BCR-ABL tyrosine kinase inhibitors in the treatment of Philadelphia chromosome positive chronic myeloid leukemia: a review. *Leuk Res.* 34(10):1255–68.
- [387] Hanfstein B., Müller M. C., Kreil S., Ernst T., Schenk T., Lorentz C., *et al.* (2011). Dynamics of mutant BCR-ABL-positive clones after cessation of tyrosine kinase inhibitor therapy. *Haematologica.* 96(3):360–6.
- [388] Stanford J. C., Young C., Hicks D., Owens P., Williams A., Vaught D. B., *et al.* (2014). Efferocytosis produces a prometastatic landscape during postpartum mammary gland involution. *J Clin Invest.* 124(11):4737–52.
- [389] Roche-Lestienne C. and Preudhomme C. (2003). Mutations in the ABL kinase domain pre-exist the onset of imatinib treatment. *Semin Hematol.* 40(2 Suppl 2):80–2.
- [390] Tamborini E., Bonadiman L., Greco A., Albertini V., Negri T., Gronchi A., *et al.* (2004). A new mutation in the KIT ATP pocket causes acquired resistance to imatinib in a gastrointestinal stromal tumor patient. *Gastroenterology.* 127(1):294–9.
- [391] Suda K., Onozato R., Yatabe Y. and Mitsudomi T. (2009). EGFR T790M mutation: a double role in lung cancer cell survival? *J Thorac Oncol Off Publ Int Assoc Study Lung Cancer.* 4(1):1–4.

- [392] Yu H. A., Arcila M. E., Rekhtman N., Sima C. S., Zakowski M. F., Pao W., *et al.* (2013). Analysis of tumor specimens at the time of acquired resistance to EGFR-TKI therapy in 155 patients with EGFR-mutant lung cancers. *Clin Cancer Res Off J Am Assoc Cancer Res.* 19(8):2240–7.
- [393] Sun C. and Bernards R. (2014). Feedback and redundancy in receptor tyrosine kinase signaling: relevance to cancer therapies. *Trends Biochem Sci.* 39(10):465–74.
- [394] Simsek D. and Jasin M. (2010). Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation. *Nat Struct Mol Biol.* 17(4):410–6.
- [395] Haince J.-F., Kozlov S., Dawson V. L., Dawson T. M., Hendzel M. J., Lavin M. F., *et al.* (2007). Ataxia telangiectasia mutated (ATM) signaling network is modulated by a novel poly(ADP-ribose)-dependent pathway in the early response to DNA-damaging agents. *J Biol Chem.* 282(22):16441–53.
- [396] Aguilar-Quesada R., Muñoz-Gómez J. A., Martín-Oliva D., Peralta A., Valenzuela M. T., Matínez-Romero R., *et al.* (2007). Interaction between ATM and PARP-1 in response to DNA damage and sensitization of ATM deficient cells through PARP inhibition. *BMC Mol Biol.* 8(1):29.
- [397] Ménissier de Murcia J., Ricoul M., Tartier L., Niedergang C., Huber A., Dantzer F., *et al.* (2003). Functional interaction between PARP-1 and PARP-2 in chromosome stability and embryonic development in mouse. *EMBO J.* 22(9):2255–63.
- [398] Fang B. (2014). Development of Synthetic Lethality Anticancer Therapeutics. *J Med Chem.* 57(19):7859–73.
- [399] Lucchesi J. C. (1968). Synthetic lethality and semi-lethality among functionally related mutants of *Drosophila melanogaster*. *Genetics.* 59(1):37–44.
- [400] Dolma S., Lessnick S. L., Hahn W. C. and Stockwell B. R. (2003). Identification of genotype-selective antitumor agents using synthetic lethal chemical screening in engineered human tumor cells. *Cancer Cell.* 3(3):285–96.
- [401] Farmer H., McCabe N., Lord C. J., Tutt A. N. J., Johnson D. A., Richardson T. B., *et al.* (2005). Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature.* 434(7035):917–21.
- [402] Kaelin W. G. (2005). The concept of synthetic lethality in the context of anticancer therapy. *Nat Rev Cancer.* 5(9):689–98.
- [403] Heidorn S. J., Milagre C., Whittaker S., Nourry A., Niculescu-Duvas I., Dhomen N., *et al.* (2010). Kinase-Dead BRAF and Oncogenic RAS Cooperate to Drive Tumor Progression through CRAF. *Cell.* 140(2):209–21.
- [404] Callahan M. K., Rampal R., Harding J. J., Klimek V. M., Chung Y. R., Merghoub T., *et al.* (2012). Progression of RAS-Mutant Leukemia during RAF Inhibitor Treatment. *N Engl J Med.* 367(24):2316–21.

- [405] Cappuzzo F., Hirsch F. R., Rossi E., Bartolini S., Ceresoli G. L., Bemis L., *et al.* (2005). Epidermal growth factor receptor gene and protein and gefitinib sensitivity in non-small-cell lung cancer. *J Natl Cancer Inst.* 97(9):643–55.
- [406] Han S.-W., Hwang P. G., Chung D. H., Kim D.-W., Im S.-A., Kim Y. T., *et al.* (2005). Epidermal growth factor receptor (EGFR) downstream molecules as response predictive markers for gefitinib (Iressa, ZD1839) in chemotherapy-resistant non-small cell lung cancer. *Int J Cancer J Int Cancer.* 113(1):109–15.
- [407] Scartozzi M., Bearzi I., Berardi R., Mandolesi A., Pierantoni C. and Cascinu S. (2007). Epidermal growth factor receptor (EGFR) downstream signalling pathway in primary colorectal tumours and related metastatic sites: optimising EGFR-targeted treatment options. *Br J Cancer.* 97(1):92–7.
- [408] Molinari F., Martin V., Saletti P., De Dosso S., Spitale A., Camponovo A., *et al.* (2009). Differing deregulation of EGFR and downstream proteins in primary colorectal cancer and related metastatic sites may be clinically relevant. *Br J Cancer.* 100(7):1087–94.
- [409] Berg M. and Soreide K. (2012). EGFR and Downstream Genetic Alterations in KRAS/BRAF and PI3K/AKT Pathways in Colorectal Cancer — Implications for Targeted Therapy. *Discov Med.* 14(76):207–14.
- [410] Lin S. Y., Makino K., Xia W., Matin A., Wen Y., Kwong K. Y., *et al.* (2001). Nuclear localization of EGF receptor and its potential new role as a transcription factor. *Nat Cell Biol.* 3(9):802–8.
- [411] Huo L., Wang Y.-N., Xia W., Hsu S.-C., Lai C.-C., Li L.-Y., *et al.* (2010). RNA helicase A is a DNA-binding partner for EGFR-mediated transcriptional activation in the nucleus. *Proc Natl Acad Sci U S A.* 107(37):16125–30.
- [412] Wheeler D. L., Dunn E. F. and Harari P. M. (2010). Understanding resistance to EGFR inhibitors-impact on future treatment strategies. *Nat Rev Clin Oncol.* 7(9):493–507.
- [413] Lo H.-W., Xia W., Wei Y., Ali-Seyed M., Huang S.-F. and Hung M.-C. (2005). Novel prognostic value of nuclear epidermal growth factor receptor in breast cancer. *Cancer Res.* 65(1):338–48.
- [414] Psyrrri A., Yu Z., Weinberger P. M., Sasaki C., Haffty B., Camp R., *et al.* (2005). Quantitative determination of nuclear and cytoplasmic epidermal growth factor receptor expression in oropharyngeal squamous cell cancer by using automated quantitative analysis. *Clin Cancer Res Off J Am Assoc Cancer Res.* 11(16):5856–62.
- [415] Xia W., Wei Y., Du Y., Liu J., Chang B., Yu Y.-L., *et al.* (2009). Nuclear expression of epidermal growth factor receptor is a novel prognostic value in patients with ovarian cancer. *Mol Carcinog.* 48(7):610–7.
- [416] Li C., Iida M., Dunn E. F., Ghia A. J. and Wheeler D. L. (2009). Nuclear EGFR contributes to acquired resistance to cetuximab. *Oncogene.* 28(43):3801–13.

- [417] Huang W.-C., Chen Y.-J., Li L.-Y., Wei Y.-L., Hsu S.-C., Tsai S.-L., *et al.* (2011). Nuclear Translocation of Epidermal Growth Factor Receptor by Akt-dependent Phosphorylation Enhances Breast Cancer-resistant Protein Expression in Gefitinib-resistant Cells. *J Biol Chem.* 286(23):20558–68.
- [418] Brand T. M., Iida M., Luthar N., Starr M. M., Huppert E. J. and Wheeler D. L. (2013). Nuclear EGFR as a molecular target in cancer. *Radiother Oncol.* 108(3):370–7.
- [419] Lipinski C. A., Lombardo F., Dominy B. W. and Feeney P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev.* 46(1-3):3–26.
- [420] Rajendran L., Knolker H.-J. and Simons K. (2010). Subcellular targeting strategies for drug design and delivery. *Nat Rev Drug Discov.* 9(1):29–42.
- [421] Kamiya H., Tsuchiya H., Yamazaki J. and Harashima H. (2001). Intracellular trafficking and transgene expression of viral and non-viral gene vectors. *Adv Drug Deliv Rev.* 52(3):153–64.
- [422] Leopold P. L. and Crystal R. G. (2007). Intracellular trafficking of adenovirus: many means to many ends. *Adv Drug Deliv Rev.* 59(8):810–21.
- [423] Mattaj J. W. and Englmeier L. (1998). Nucleocytoplasmic transport: the soluble phase. *Annu Rev Biochem.* 67:265–306.
- [424] Timmers A. C. ., Stuger R., Schaap P. J., van 't Riet J. and Raué H. A. (1999). Nuclear and nucleolar localization of *Saccharomyces cerevisiae* ribosomal proteins S22 and S25. *FEBS Lett.* 452(3):335–40.
- [425] Lee B. J., Cansizoglu A. E., Süel K. E., Louis T. H., Zhang Z. and Chook Y. M. (2006). Rules for Nuclear Localization Sequence Recognition by Karyopherin β 2. *Cell.* 126(3):543–58.
- [426] Kalderon D., Roberts B. L., Richardson W. D. and Smith A. E. (1984). A short amino acid sequence able to specify nuclear location. *Cell.* 39(3 Pt 2):499–509.
- [427] Frankel A. D. and Pabo C. O. (1988). Cellular uptake of the tat protein from human immunodeficiency virus. *Cell.* 55(6):1189–93.
- [428] Lanford R. E. and Butel J. S. (1984). Construction and characterization of an SV40 mutant defective in nuclear transport of T antigen. *Cell.* 37(3):801–13.
- [429] Hodoniczky J., Sims C. G., Best W. M., Bentel J. M. and Wilce J. A. (2008). The intracellular and nuclear-targeted delivery of an antiandrogen drug by carrier peptides. *Pept Sci.* 90(5):595–603.
- [430] Pan L., He Q., Liu J., Chen Y., Ma M., Zhang L., *et al.* (2012). Nuclear-Targeted Drug Delivery of TAT Peptide-Conjugated Monodisperse Mesoporous Silica Nanoparticles. *J Am Chem Soc.* 134(13):5722–5.

- [431] Nitin N., LaConte L., Rhee W. J. and Bao G. (2009). Tat Peptide Is Capable of Importing Large Nanoparticles Across Nuclear Membrane in Digitonin Permeabilized Cells. *Ann Biomed Eng.* 37(10):2018–27.
- [432] Mishra A., Lai G. H., Schmidt N. W., Sun V. Z., Rodriguez A. R., Tong R., *et al.* (2011). Translocation of HIV TAT peptide and analogues induced by multiplexed membrane and cytoskeletal interactions. *Proc Natl Acad Sci.* 108(41):16883–8.
- [433] Khalil I. A., Futaki S., Niwa M., Baba Y., Kaji N., Kamiya H., *et al.* (2004). Mechanism of improved gene transfer by the N-terminal stearylolation of octaarginine: enhanced cellular association by hydrophobic core formation. *Gene Ther.* 11(7):636–44.
- [434] Ter-Avetisyan G., Tünnemann G., Nowak D., Nitschke M., Herrmann A., Drab M., *et al.* (2009). Cell Entry of Arginine-rich Peptides Is Independent of Endocytosis. *J Biol Chem.* 284(6):3370–8.
- [435] Umezawa N., Gelman M. A., Haigis M. C., Raines R. T. and Gellman S. H. (2001). Translocation of a β -Peptide Across Cell Membranes. *J Am Chem Soc.* 124(3):368–9.
- [436] Mandal D., Nasrolahi Shirazi A. and Parang K. (2011). Cell-Penetrating Homochiral Cyclic Peptides as Nuclear-Targeting Molecular Transporters. *Angew Chem Int Ed.* 50(41):9633–7.
- [437] Schröder T., Niemeier N., Afonin S., Ulrich A. S., Krug H. F. and Bräse S. (2008). Peptoidic Amino- and Guanidinium-Carrier Systems: Targeted Drug Delivery into the Cell Cytosol or the Nucleus. *J Med Chem.* 51(3):376–9.
- [438] Unciti-Broceta A., Diezmann F., Ou-Yang C. Y., Fara M. A. and Bradley M. (2009). Synthesis, penetrability and intracellular targeting of fluorescein-tagged peptoids and peptide-peptoid hybrids. *Bioorg Med Chem.* 17(3):959–66.
- [439] Peretto I., Sanchez-Martin R. M., Wang X., Ellard J., Mittoo S. and Bradley M. (2003). Cell penetrable peptoid carrier vehicles: synthesis and evaluation. *Chem Commun.* (18):2312.
- [440] Elliott G. and O'Hare P. (1997). Intercellular trafficking and protein delivery by a herpesvirus structural protein. *Cell.* 88(2):223–33.
- [441] Potocky T. B., Menon A. K. and Gellman S. H. (2003). Cytoplasmic and nuclear delivery of a TAT-derived peptide and a beta-peptide after endocytic uptake into HeLa cells. *J Biol Chem.* 278(50):50188–94.
- [442] Sadler K., Eom K. D., Yang J.-L., Dimitrova Y. and Tam J. P. (2002). Translocating proline-rich peptides from the antimicrobial peptide bactenecin 7. *Biochemistry (Mosc).* 41(48):14150–7.
- [443] Kölmel D. K., Hörner A., Röncke F., Nieger M., Schepers U. and Bräse S. (2014). Cell-penetrating peptoids: Introduction of novel cationic side chains. *Eur J Med Chem.* 79:231–43.

- [444] Mitchell D. J., Kim D. T., Steinman L., Fathman C. G. and Rothbard J. B. (2000). Polyarginine enters cells more efficiently than other polycationic homopolymers. *J Pept Res Off J Am Pept Soc.* 56(5):318–25.
- [445] Brand T. M., Iida M., Dunn E. F., Luthar N., Kostopoulos K. T., Corrigan K. L., *et al.* (2014). Nuclear epidermal growth factor receptor is a functional molecular target in triple-negative breast cancer. *Mol Cancer Ther.* 13(5):1356–68.
- [446] Tong D., Ortega J., Kim C., Huang J., Gu L. and Li G.-M. (2015). Arsenic Inhibits DNA Mismatch Repair by Promoting EGFR Expression and PCNA Phosphorylation. *J Biol Chem.* 290(23):14536–41.
- [447] Wendt M. K., Smith J. A. and Schiemann W. P. (2010). Transforming growth factor- β -induced epithelial-mesenchymal transition facilitates epidermal growth factor-dependent breast cancer progression. *Oncogene.* 29(49):6485–98.
- [448] Wendt M. K., Taylor M. A., Schiemann B. J., Sossey-Alaoui K. and Schiemann W. P. (2014). Fibroblast growth factor receptor splice variants are stable markers of oncogenic transforming growth factor β 1 signaling in metastatic breast cancers. *Breast Cancer Res.* 16(2):R24.
- [449] Madani F., Lindberg S., Langel Ü, lo, Futaki S., Grä, *et al.* (2011). Mechanisms of Cellular Uptake of Cell-Penetrating Peptides. *J Biophys.* 2011:e414729.
- [450] Jans D. A., Xiao C. Y. and Lam M. H. (2000). Nuclear targeting signal recognition: a key control point in nuclear transport? *BioEssays News Rev Mol Cell Dev Biol.* 22(6):532–44.
- [451] Fitzpatrick S. L., LaChance M. P. and Schultz G. S. (1984). Characterization of epidermal growth factor receptor and action on human breast cancer cells in culture. *Cancer Res.* 44(8):3442–7.
- [452] Filmus J., Pollak M. N., Cailleau R. and Buick R. N. (1985). MDA-468, a human breast cancer cell line with a high number of epidermal growth factor (EGF) receptors, has an amplified EGF receptor gene and is growth inhibited by EGF. *Biochem Biophys Res Commun.* 128(2):898–905.
- [453] Corkery B., Crown J., Clynes M. and O'Donovan N. (2009). Epidermal growth factor receptor as a potential therapeutic target in triple-negative breast cancer. *Ann Oncol.* 20(5):862–7.
- [454] Roux P. P. and Blenis J. (2004). ERK and p38 MAPK-Activated Protein Kinases: a Family of Protein Kinases with Diverse Biological Functions. *Microbiol Mol Biol Rev.* 68(2):320–44.
- [455] Liu L., McBride K. M. and Reich N. C. (2005). STAT3 nuclear import is independent of tyrosine phosphorylation and mediated by importin- α 3. *Proc Natl Acad Sci U S A.* 102(23):8150–5.
- [456] Yang J. and Stark G. R. (2008). Roles of unphosphorylated STATs in signaling. *Cell Res.* 18(4):443–51.

- [457] Lo H.-W., Hsu S.-C., Ali-Seyed M., Gunduz M., Xia W., Wei Y., *et al.* (2005). Nuclear interaction of EGFR and STAT3 in the activation of the iNOS/NO pathway. *Cancer Cell*. 7(6):575–89.
- [458] Park O. K., Schaefer T. S. and Nathans D. (1996). In vitro activation of Stat3 by epidermal growth factor receptor kinase. *Proc Natl Acad Sci U S A*. 93(24):13704–8.
- [459] Zhong Z., Wen Z. and Darnell J. E. (1994). Stat3: a STAT family member activated by tyrosine phosphorylation in response to epidermal growth factor and interleukin-6. *Science*. 264(5155):95–8.
- [460] Yuan Z., Guan Y., Wang L., Wei W., Kane A. B. and Chin Y. E. (2004). Central Role of the Threonine Residue within the p+1 Loop of Receptor Tyrosine Kinase in STAT3 Constitutive Phosphorylation in Metastatic Cancer Cells. *Mol Cell Biol*. 24(21):9390–400.
- [461] Buettner R., Mora L. B. and Jove R. (2002). Activated STAT Signaling in Human Tumors Provides Novel Molecular Targets for Therapeutic Intervention. *Clin Cancer Res*. 8(4):945–54.
- [462] Yu H. and Jove R. (2004). The STATs of cancer — new molecular targets come of age. *Nat Rev Cancer*. 4(2):97–105.
- [463] Haura E. B., Turkson J. and Jove R. (2005). Mechanisms of Disease: insights into the emerging role of signal transducers and activators of transcription in cancer. *Nat Rev Clin Oncol*. 2(6):315–24.
- [464] Zimmer S., Kahl P., Buhl T. M., Steiner S., Wardelmann E., Merkelbach-Bruse S., *et al.* (2009). Epidermal growth factor receptor mutations in non-small cell lung cancer influence downstream Akt, MAPK and Stat3 signaling. *J Cancer Res Clin Oncol*. 135(5):723–30.
- [465] Jiang R., Jin Z., Liu Z., Sun L., Wang L. and Li K. (2011). Correlation of activated STAT3 expression with clinicopathologic features in lung adenocarcinoma and squamous cell carcinoma. *Mol Diagn Ther*. 15(6):347–52.
- [466] Looyenga B. D., Hutchings D., Cherni I., Kingsley C., Weiss G. J. and MacKeigan J. P. (2012). STAT3 Is Activated by JAK2 Independent of Key Oncogenic Driver Mutations in Non-Small Cell Lung Carcinoma. *PLoS ONE*. 7(2):e30820.
- [467] Jiang Y.-Q., Zhou Z.-X. and Ji Y.-L. (2014). Suppression of EGFR-STAT3 signaling inhibits tumorigenesis in a lung cancer cell line. *Int J Clin Exp Med*. 7(8):2096–9.
- [468] Catlett-Falcone R., Landowski T. H., Oshiro M. M., Turkson J., Levitzki A., Savino R., *et al.* (1999). Constitutive activation of Stat3 signaling confers resistance to apoptosis in human U266 myeloma cells. *Immunity*. 10(1):105–15.
- [469] Niu G., Bowman T., Huang M., Shivers S., Reintgen D., Daud A., *et al.* (2002). Roles of activated Src and Stat3 signaling in melanoma tumor cell growth. *Oncogene*. 21(46):7001–10.

- [470] Garcia R., Bowman T. L., Niu G., Yu H., Minton S., Muro-Cacho C. A., *et al.* (2001). Constitutive activation of Stat3 by the Src and JAK tyrosine kinases participates in growth regulation of human breast carcinoma cells. *Oncogene*. 20(20):2499–513.
- [471] Huang M., Dorsey J. F., Epling-Burnette P. K., Nimmanapalli R., Landowski T. H., Mora L. B., *et al.* (2002). Inhibition of Bcr-Abl kinase activity by PD180970 blocks constitutive activation of Stat5 and growth of CML cells. *Oncogene*. 21(57):8804–16.
- [472] Levis M., Allebach J., Tse K.-F., Zheng R., Baldwin B. R., Smith B. D., *et al.* (2002). A FLT3-targeted tyrosine kinase inhibitor is cytotoxic to leukemia cells in vitro and in vivo. *Blood*. 99(11):3885–91.
- [473] Harada D., Takigawa N. and Kiura K. (2014). The Role of STAT3 in Non-Small Cell Lung Cancer. *Cancers*. 6(2):708–22.
- [474] Kim S. M., Kwon O.-J., Hong Y. K., Kim J. H., Solca F., Ha S.-J., *et al.* (2012). Activation of IL-6R/JAK1/STAT3 signaling induces de novo resistance to irreversible EGFR inhibitors in non-small cell lung cancer with T790M resistance mutation. *Mol Cancer Ther*. 11(10):2254–64.
- [475] Nalband D. M., Warner B. P., Zahler N. H. and Kirshenbaum K. (2014). Rapid identification of metal-binding peptoid oligomers by on-resin X-ray fluorescence screening. *Pept Sci*. 102(5):407–15.
- [476] Kuo L. J. and Yang L.-X. (2008). Gamma-H2AX - a novel biomarker for DNA double-strand breaks. *Vivo Athens Greece*. 22(3):305–9.
- [477] Cour T. la, Gupta R., Rapacki K., Skriver K., Poulsen F. M. and Brunak S. (2003). NESbase version 1.0: a database of nuclear export signals. *Nucleic Acids Res*. 31(1):393–6.
- [478] Cour T. la, Kiemer L., Mølgaard A., Gupta R., Skriver K. and Brunak S. (2004). Analysis and prediction of leucine-rich nuclear export signals. *Protein Eng Des Sel*. 17(6):527–36.
- [479] Xu D., Farmer A., Collett G., Grishin N. V. and Chook Y. M. (2012). Sequence and structural analyses of nuclear export signals in the NESdb database. *Mol Biol Cell*. 23(18):3677–93.
- [480] Emmott E. and Hiscox J. A. (2009). Nucleolar targeting: the hub of the matter. *EMBO Rep*. 10(3):231–8.
- [481] Omura T. (1998). Mitochondria-targeting sequence, a multi-role sorting sequence recognized at all steps of protein import into mitochondria. *J Biochem (Tokyo)*. 123(6):1010–6.
- [482] Brocard C. and Hartig A. (2006). Peroxisome targeting signal 1: Is it really a simple tripeptide? *Biochim Biophys Acta BBA - Mol Cell Res*. 1763(12):1565–73.

- [483] Gomord V., Denmat L. A., Fichette-Lainé A. C., Satiat-Jeunemaitre B., Hawes C. and Faye L. (1997). The C-terminal HDEL sequence is sufficient for retention of secretory proteins in the endoplasmic reticulum (ER) but promotes vacuolar targeting of proteins that escape the ER. *Plant J Cell Mol Biol.* 11(2):313–25.
- [484] Stornaiuolo M., Lotti L. V., Borgese N., Torrisi M.-R., Mottola G., Martire G., *et al.* (2003). KDEL and KKXX Retrieval Signals Appended to the Same Reporter Protein Determine Different Trafficking between Endoplasmic Reticulum, Intermediate Compartment, and Golgi Complex. *Mol Biol Cell.* 14(3):889–902.
- [485] Dittmann K., Mayer C., Fehrenbacher B., Schaller M., Raju U., Milas L., *et al.* (2005). Radiation-induced Epidermal Growth Factor Receptor Nuclear Import Is Linked to Activation of DNA-dependent Protein Kinase. *J Biol Chem.* 280(35):31182–9.
- [486] Dittmann K., Mayer C., Fehrenbacher B., Schaller M., Kehlbach R. and Rodemann H. P. (2010). Nuclear EGFR shuttling induced by ionizing radiation is regulated by phosphorylation at residue Thr654. *FEBS Lett.* 584(18):3878–84.
- [487] Liccardi G., Hartley J. A. and Hochhauser D. (2011). EGFR nuclear translocation modulates DNA repair following cisplatin and ionizing radiation treatment. *Cancer Res.* 71(3):1103–14.
- [488] Yu Y.-L., Chou R.-H., Wu C.-H., Wang Y.-N., Chang W.-J., Tseng Y.-J., *et al.* (2012). Nuclear EGFR suppresses ribonuclease activity of polynucleotide phosphorylase through DNAPK-mediated phosphorylation at serine 776. *J Biol Chem.* 287(37):31015–26.
- [489] Beck C., Robert I., Reina-San-Martin B., Schreiber V. and Dantzer F. (2014). Poly(ADP-ribose) polymerases in double-strand break repair: Focus on PARP1, PARP2 and PARP3. *Exp Cell Res.* 329(1):18–25.
- [490] Friedmann B. J., Caplin M., Savic B., Shah T., Lord C. J., Ashworth A., *et al.* (2006). Interaction of the epidermal growth factor receptor and the DNA-dependent protein kinase pathway following gefitinib treatment. *Mol Cancer Ther.* 5(2):209–18.
- [491] Al-Ejeh F., Shi W., Miranda M., Simpson P. T., Vargas A. C., Song S., *et al.* (2013). Treatment of triple-negative breast cancer using anti-EGFR-directed radioimmunotherapy combined with radiosensitizing chemotherapy and PARP inhibitor. *J Nucl Med Off Publ Soc Nucl Med.* 54(6):913–21.

APPENDICES

A. Characterization of PCNA Peptoid Inhibitors

Table A.1
PCNA Crystal Structures Used for Computational Screening

PDB ID [†]	Description
1AXC	Structure of the C-terminal region of p21(WAF1/CIP1) complexed with human PCNA
3VKX	PCNA complexed with the small molecule antagonist T3 at the PIP Box binding site
1U7B	Structure of human PCNA in complex with a peptide derived from residues 331-350 of flap endonuclease-1 (FEN1)
2ZVK	Structure of human PCNA in complex with a fragment of DNA polymerase eta

[†] All crystal structures were obtained from the RCSB Protein Data Bank (<http://www.rcsb.org/pdb/home/home.do>)

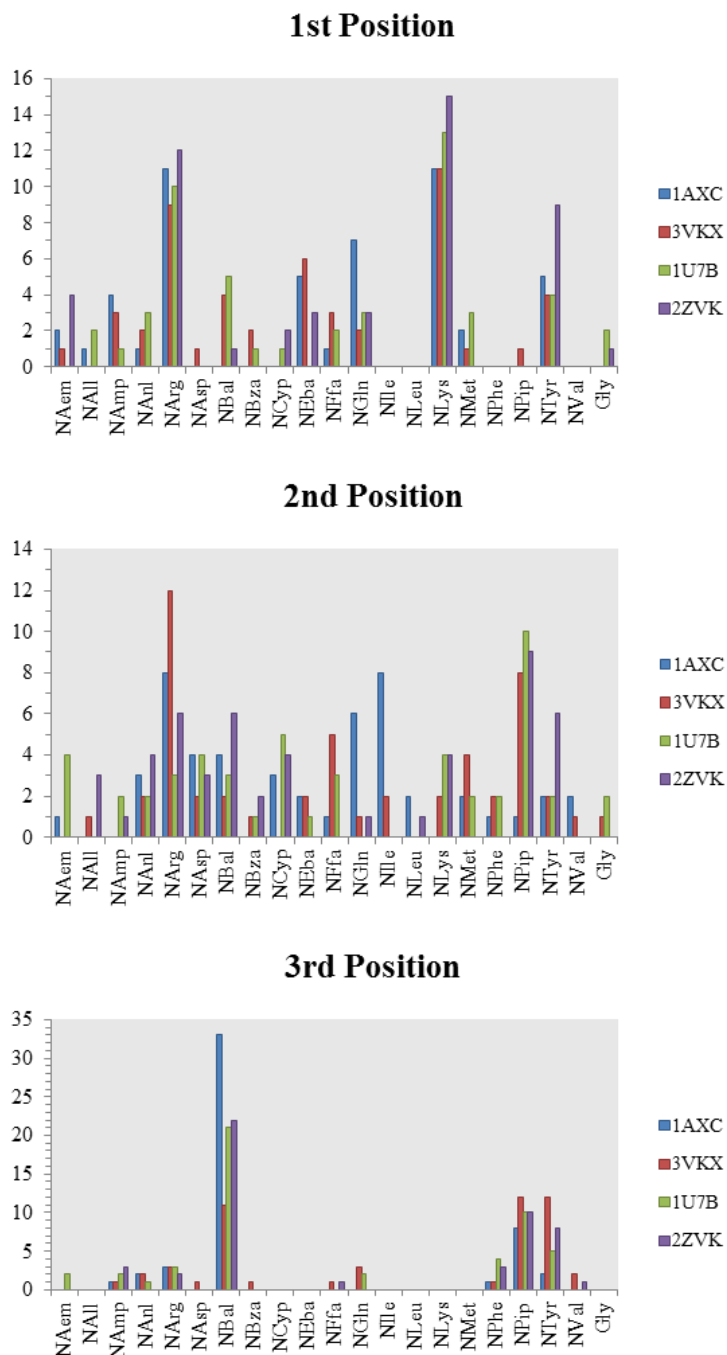


Figure A.1. Initial Peptoid Fragment Tally by PCNA Crystal Structure. The initial amine fragment library consisting of 20 primary amines as well as hydrogen as a substituent (Gly) was screened against four different crystal structures of PCNA—PDB IDs: 1AXC, 3VKX, 1U7B and 2ZVK. The number of times a particular fragment appeared in a specific substitution position on the peptoid backbone, in the top 50 hits for each crystal structure, was tallied. Each structure displayed slightly different results. “1st”, “2nd” and “3rd” positions are as defined in Figure 2.2.

Table A.2
Characterization of Screened Tripeptoids

Peptoid Sequence ^a	Calculated Mass [M+H] ¹⁺	Observed Mass ^b (m/z)	% Yield ^c
Gly-NPip-NBal	443.1931	443.5391	7.19%
NAem-NArg-NPip	535.2948	535.2581	47.49%
NAem-NArg-NTyr	507.2999	507.3044	22.44%
NAem-NEal-NBal	466.2621	466.2658	15.51%
NAem-NLys-NPip	507.2887	507.3019	32.08%
NAem-NLys-NTyr	479.2938	479.2624	33.93%
NAem-NPip-NBal	556.2772	556.2767	22.40%
NAem-NPip-NBza	570.2520	570.2553	34.71%
NAem-NPip-NPip	570.2520	570.3911	73.95%
NAem-NPip-NTyr	542.2570	542.2764	50.28%
NAem-NTyr-NPip	542.2615	542.3094	40.60%
NAem-NTyr-NTyr	514.2621	514.3411	7.16%
NArg-NArg-NPip	521.2904	521.4318	32.30%
NArg-NIle-NBal	464.2986	464.2994	9.40%
NArg-NLys-NBza	493.2843	493.2887	32.32%
NArg-NLys-NPip	493.2843	493.5698	34.09%
NArg-NPip-NBal	542.2683	542.2703	29.86%
NArg-NPip-NBza	556.2475	556.1994	13.39%
NArg-NPip-NPip	556.2475	556.2516	8.29%
NArg-NPip-NTyr	528.2526	528.2560	3.31%
NArg-NTyr-NBal	514.2778	514.2761	10.11%
NArg-NTyr-NTyr	500.2577	500.2616	13.73%
NBal-NArg-NBal	528.2890	528.2910	12.43%
NBal-NEal-NBza	487.2193	487.2186	29.94%
NBal-NLys-NPip	514.2621	514.2656	29.10%
NBal-NLys-NTyr	486.2672	486.3338	10.92%
NBal-NPip-NBal	563.2462	563.1897	5.32%
NBal-NPip-NBza	577.2254	577.2296	19.10%

^a N-substituted side chains listed from N-terminus to most C-terminal

^b as determined by high resolution electrospray ionization or MALDI-TOF

^c based on dry mass recovery after HPLC purification

Table A.2 Continued

NBal-NPip-NTyr	549.2305	549.2338	46.74%
NBal-NTyr-NBal	535.2512	535.2536	12.01%
NBal-NTyr-NPip	549.2350	549.2787	26.26%
NBza-NArg-NBal	542.2683	542.2712	20.25%
NBza-NArg-NPip	556.2475	556.2506	49.91%
NBza-NLys-NBza	528.2414	528.2452	23.80%
NBza-NLys-NPip	528.2414	528.2443	49.19%
NBza-NLys-NTyr	500.2465	500.2509	27.46%
NBza-NPip-NBal	577.2254	577.2283	14.47%
NBza-NPip-NBza	591.2047	591.2075	14.18%
NBza-NTyr-NBal	549.2305	549.2157	10.93%
NBza-NTyr-NPip	563.2098	563.4002	17.49%
NEal-NEal-NBza	411.1836	411.3020	10.34%
NEal-NEal-NPip	411.1836	411.1878	22.09%
NEal-NEal-NTrp	420.2203	420.2237	28.15%
NEal-NLys-NPip	438.2308	438.2353	33.02%
NEal-NPip-NBza	501.1941	501.1979	1.71%
NEal-NPip-NPip	501.1941	501.1972	37.39%
NEal-NPip-NTrp	510.2353	510.2364	11.03%
NLys-NArg-NPip	493.2843	493.4769	29.45%
NLys-NBal-NPip	514.2621	514.2684	35.56%
NLys-NEal-NBza	438.2308	438.2353	7.50%
NLys-NEal-NPip	438.2308	438.2351	34.86%
NLys-NIle-NBal	437.2958	437.4813	0.82%
NLys-NLys-NPip	465.2781	465.5013	28.13%
NLys-NPip-NBal	514.2621	514.2656	8.62%
NLys-NPip-NBza	528.2414	528.3454	36.49%
NLys-NPip-NMba	513.2826	513.6709	6.75%
NLys-NPip-NPip	528.2414	528.3143	27.24%
NLys-NTyr-NBal	486.2672	486.2708	20.16%
NLys-NTyr-NBza	500.2465	500.2403	21.69%
NLys-NTyr-NPip	500.2465	500.2713	10.44%
NMba-NPip-NBal	562.2666	562.6433	23.68%

Table A.2 Continued

NPip-NPip-NPip	591.2047	591.8057	3.97%
NTyr-NArg-NBal	514.2734	514.3009	11.59%
NTyr-NArg-NPip	528.2526	528.3313	3.44%
NTyr-NEal-NBal	459.2199	459.2245	16.40%
NTyr-NIle-NBal	471.2563	471.3852	12.99%
NTyr-NLys-NPip	500.2509	500.2503	47.78%
NTyr-NLys-NTyr	472.2516	472.1906	4.57%
NTyr-NPip-NBal	549.2350	549.6621	0.75%
NTyr-NPip-NBza	563.2098	563.2132	27.46%
NTyr-NPip-NPip	563.2098	563.2128	23.04%
NTyr-NPip-NTyr	535.2193	535.2186	45.90%
NTyr-NTyr-NPip	535.2149	535.2551	23.10%
T2AA-Asn	682.9864	682.9854	5.78%
T2AA-Gln	697.0021	697.0003	8.15%
T2AA-Gly	625.9649	625.9629	0.81%
T2AA-Gly-NBal	803.0439	803.0422	9.83%
T2AA-Gly-NPip	817.0232	817.0266	4.51%
T2AA-NEal-Gly	727.0126	727.0109	3.81%
T2AA-NEal-NBal	847.0701	847.0695	1.67%
T2AA-NEal-NMba	846.0861	846.8161	13.13%
T2AA-NEal-NPip	861.0450	861.0513	5.95%
T2AA-NEal-NTyr	833.0545	833.0562	15.00%
T2AA-NPip-NLys	888.0967	888.0983	9.41%
T2AA-NPip-NPip	951.0560	951.0586	4.95%

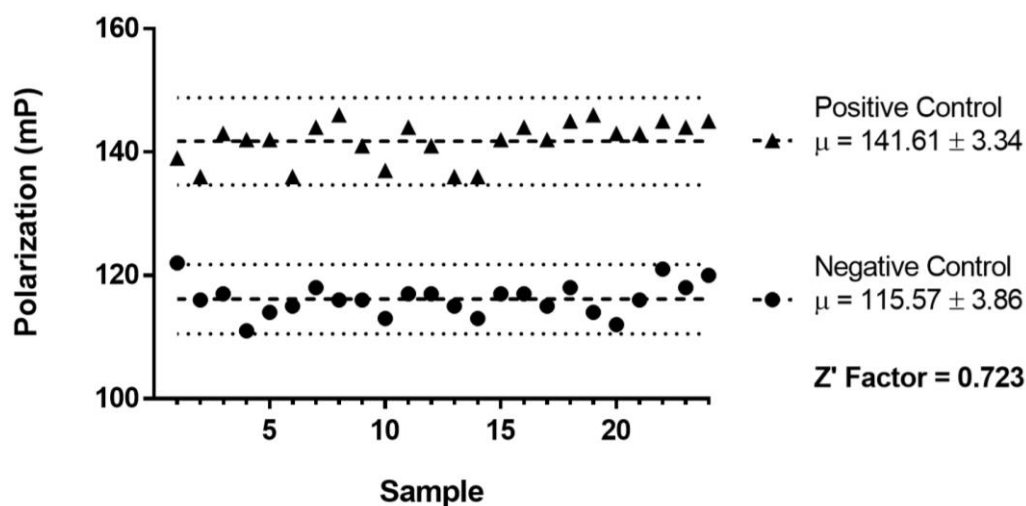


Figure A.2. Z'-Factor Analysis. Polarization values for 24 replicate samples of both positive and negative controls were used to evaluate the quality of the assay platform. 10 nM FAM-PL and 100 nM PCNA served as the positive control, while 10 nM FAM-PL in binding buffer served as the negative control. Dashed lines indicate the mean value of each set of controls, and dotted lines indicate the 95% prediction interval.

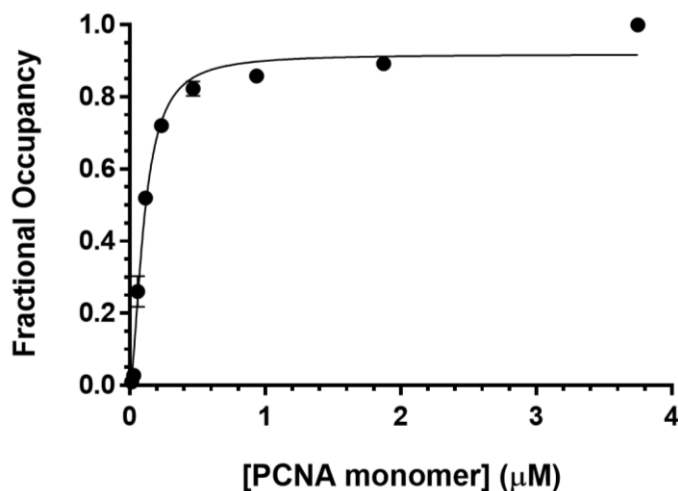


Figure A.3. PCNA Titration for Determining Affinity of FAM-PL. Increasing amounts of recombinant PCNA were added to a fixed concentration of the FAM-PL peptide (5 nM) in binding buffer. The data were fit to Equation 2.3 was used to determine the K_d value for the peptide (here, the calculated affinity was 107 nM).

Table A.3
Peptoid Number from Initial Tripeptoid FP Screen (Figure 2.8)

Peptoid Sequence (1 st – 3 rd Position)	Peptoid Number
NAem-NArg-NPip	1
NAem-NArg-NTyr	2
NAem-NEal-NBal	3
NAem-NLys-NPip	4
NAem-NLys-NTyr	5
NAem-NPip-NBza	6
NAem-NPip-NPip	7
NAem-NPip-NTyr	8
NAem-NTyr-NPip	9
NAem-NTyr-NTyr	10
NArg-NArg-NPip	11
NArg-NLys-NBza	12
NArg-NLys-NPip	13
NArg-NPip-NBal	14
NArg-NPip-NBza	15
NArg-NPip-NPip	16
NArg-NPip-NTyr	17
NArg-NTyr-NTyr	18
NBal-NArg-NBal	19
NBal-NEal-NBza	20
NBal-NLys-NPip	21
NBal-NLys-NTyr	22
NBal-NPip-NBal	23
NBal-NPip-NBza	24
NBal-NPip-NTyr	25
NBal-NTyr-NBal	26
NBal-NTyr-NPip	27
NBza-NArg-NBal	28
NBza-NArg-NPip	29
NBza-NLys-NBza	30

Table A.3 Continued

NBza-NLys-NPip	31
NBza-NLys-NTyr	32
NBza-NPip-NBal	33
NBza-NPip-NBza	34
NBza-NPip-NTyr	35
NBza-NTyr-NBal	36
NBza-NTyr-NPip	37
NEal-NEal-NBza	38
NEal-NEal-NPip	39
NEal-NLys-NPip	40
NEal-NPip-NBza	41
NEal-NPip-NPip	42
NLys-NArg-NPip	43
NLys-NBal-NPip	44
NLys-NEal-NBza	45
NLys-NEal-NPip	46
NLys-NLys-NPip	47
NLys-NPip-NBal	48
NLys-NPip-NBza	49
NLys-NPip-NPip	50
NLys-NTyr-NBal	51
NLys-NTyr-NBza	52
NLys-NTyr-NPip	53
NPip-NPip-NPip	54
NTyr-NArg-NBal	55
NTyr-NArg-NPip	56
NTyr-NEal-NBal	57
NTyr-NIle-NBal	58
NTyr-NLys-NBza	59
NTyr-NLys-NPip	60
NTyr-NLys-NPip	61
NTyr-NLys-NTyr	62
NTyr-NPip-NBza	63

Table A.3 Continued

NTyr-NPip-NPip	64
NTyr-NPip-NTyr	65
NTyr-NTyr-NBza	66
NTyr-NTyr-NPip	67
T2AA-NEal-NPip	68
T2AA-Gly-NPip	69

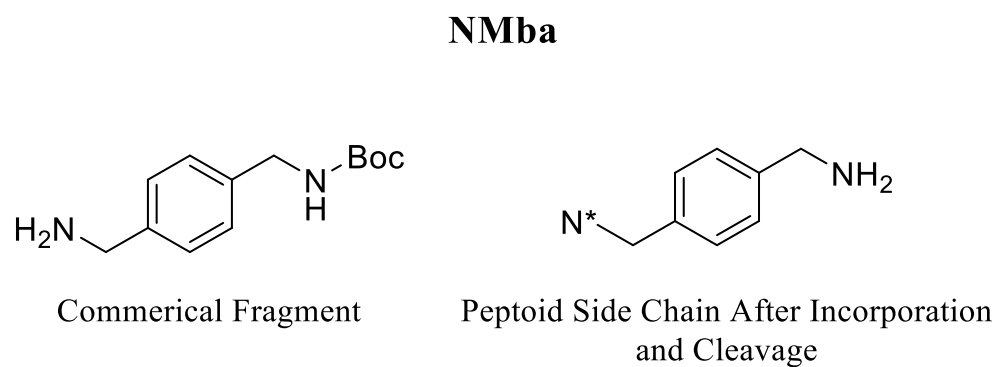


Figure A.4. Chemical Structure of NMba. After initial results from experimental screening, the fragment NMba was selected for incorporation into peptoids as a substitute for NBal due to its structural similarity, but ability to form an additional hydrogen bond. The fragment was purchased as a Boc-protected amine (*left*), but the side chain becomes exposed upon incorporation into a tripeptoid and subsequent cleavage with trifluoroacetic acid (*right*). N* denotes a peptoid backbone nitrogen.

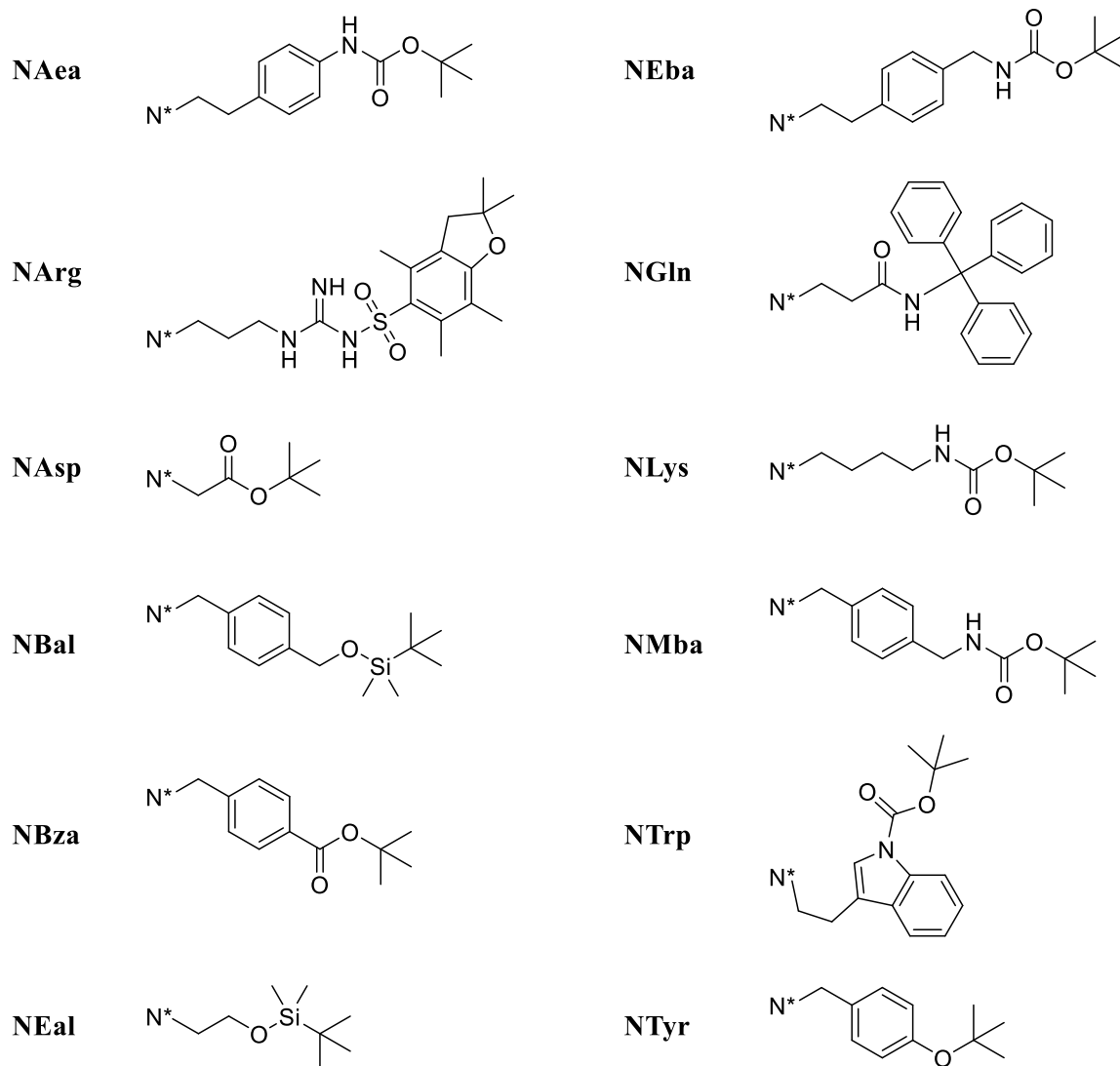


Figure A.5. Necessary Protecting Groups on Peptoid Fragments. Some of the fragments selected for incorporation into either the virtual or synthetic libraries require protecting groups on reactive species. N* indicates the location of the $-NH_2$ group, which is the position of substitution into a peptoid backbone.

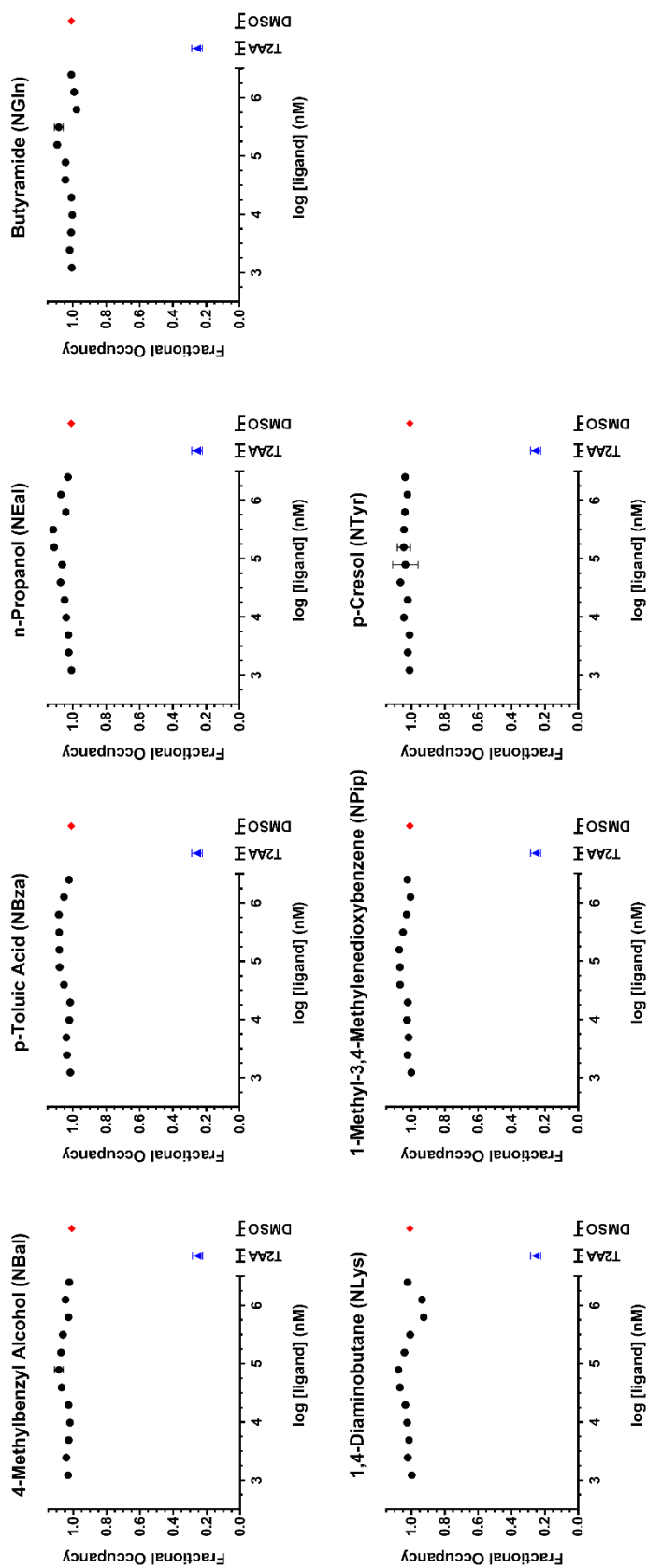


Figure A.6. Dose Response Curves of Individual Peptoid Fragments. The fragments present in hit tripeptoid molecules were screened individually in the FP assay to determine if they could individually disrupt the binding between PCNA and the PL peptide. None of the fragments showed any evidence of inhibition. T2AA was selected as the positive control, and DMSO as the negative control. Error bars represent the standard error of the mean.

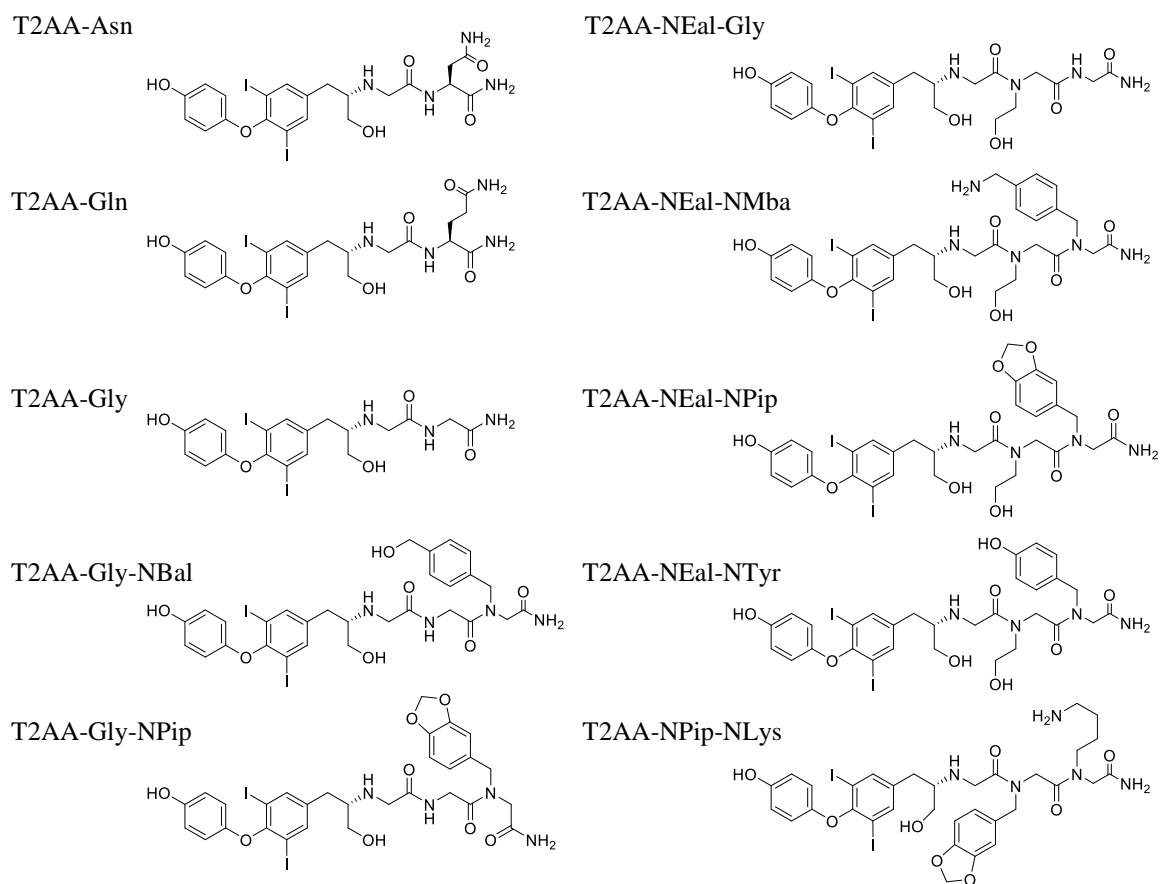


Figure A.7. Structure of T2AA-Conjugates. For second generation peptoid inhibitors of PCNA, T2AA was coupled to the N-terminus of peptoids or peptides anchored on solid phase resin to generate the molecules shown above.

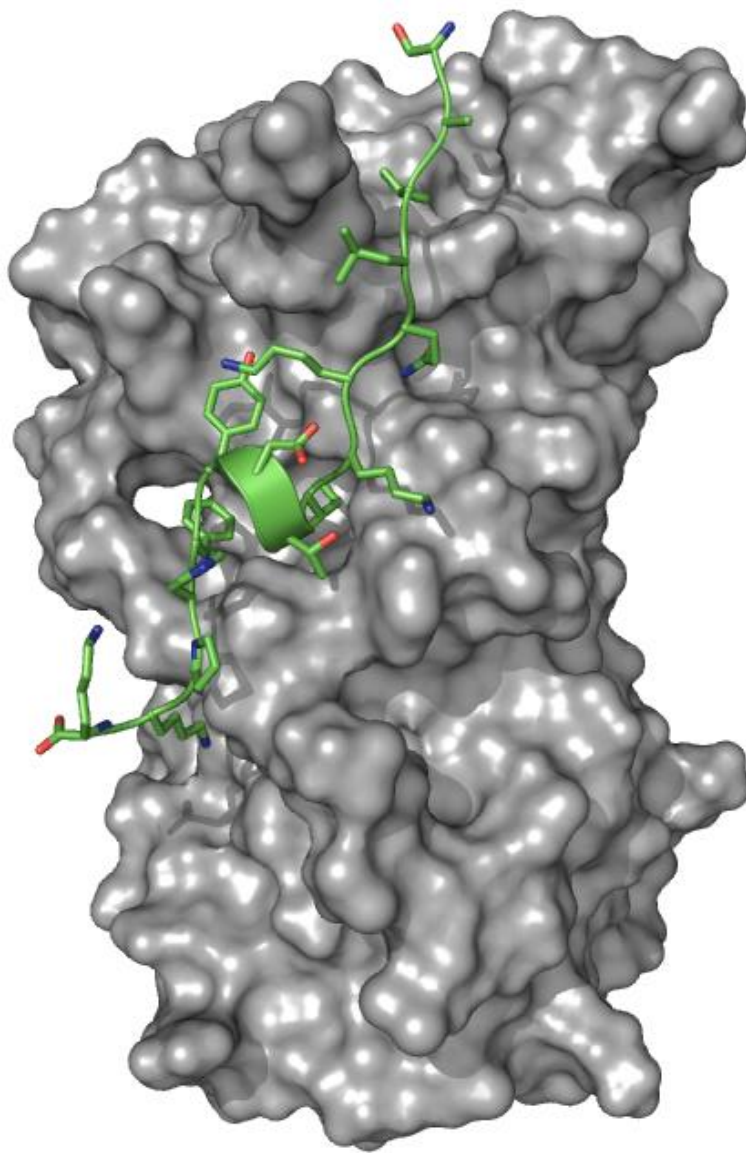
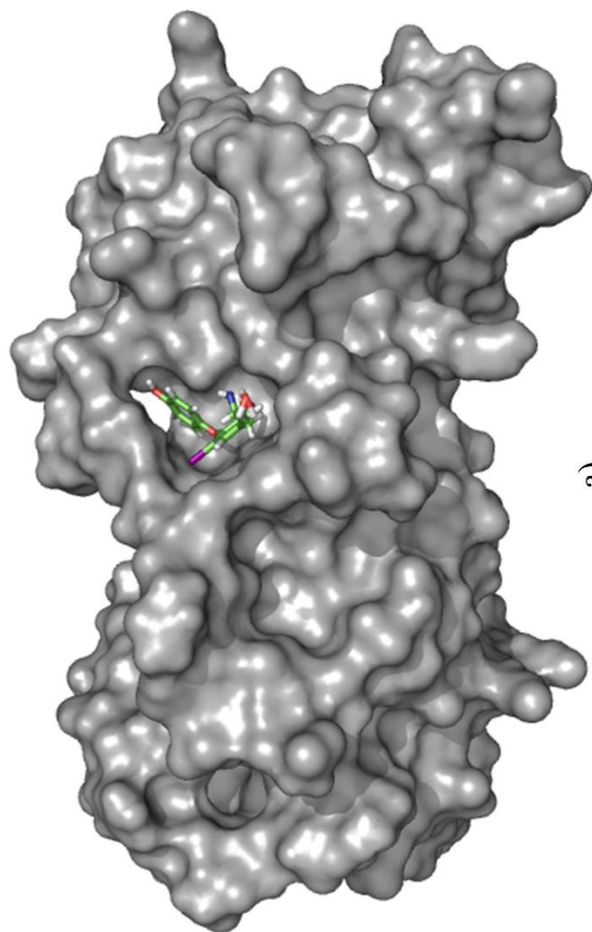
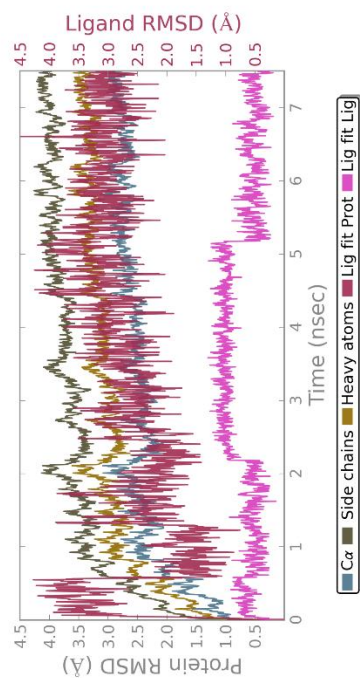


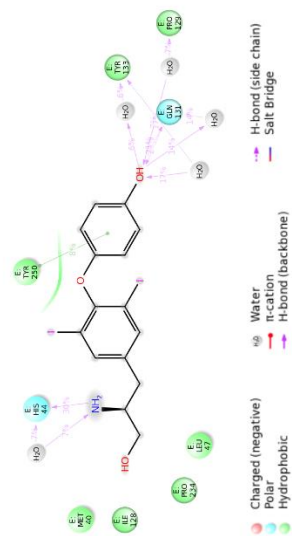
Figure A.8. Co-Crystal Structure of PCNA-PL Peptide (PDB ID: 1VYJ).



a)



b)



c)

Figure A.9. PCNA-T2AA MD Simulation Output. (a) Average structure of the final 50 frames of the PCNA-T2AA MD simulation. (b) RMSD diagram demonstrating simulation convergence. (c) Ligand interaction diagram showing molecular contacts over the course of the simulation.

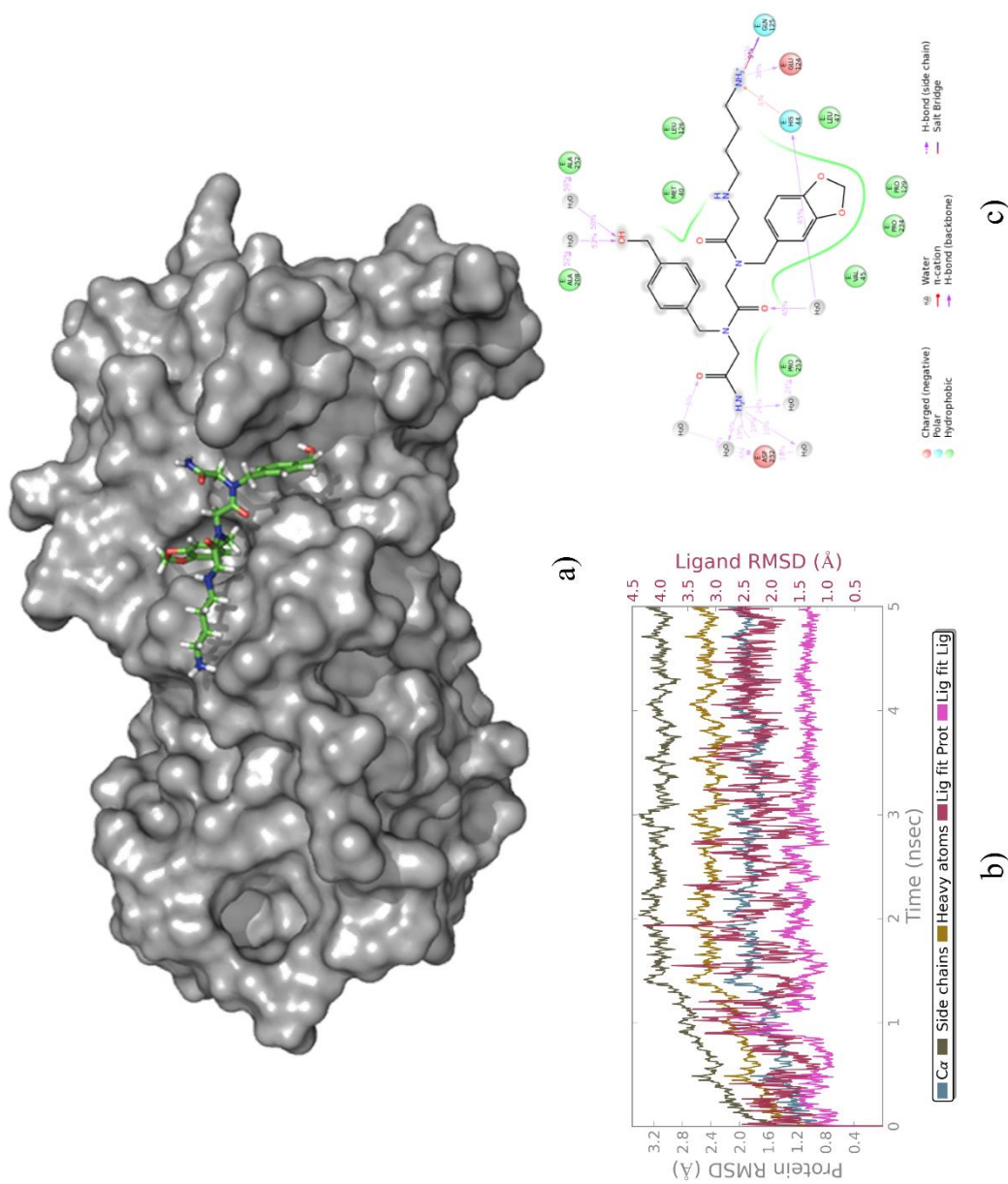


Figure A.10. PCNA-NLys-NPip-NBaI MD Simulation Ouput. (a) Average structure of the PCNA-NLys-NPip-NBaI MD simulation. (b) RMSD diagram demonstrating simulation convergence. (c) Ligand interaction diagram showing molecular contacts over the course of the simulation.

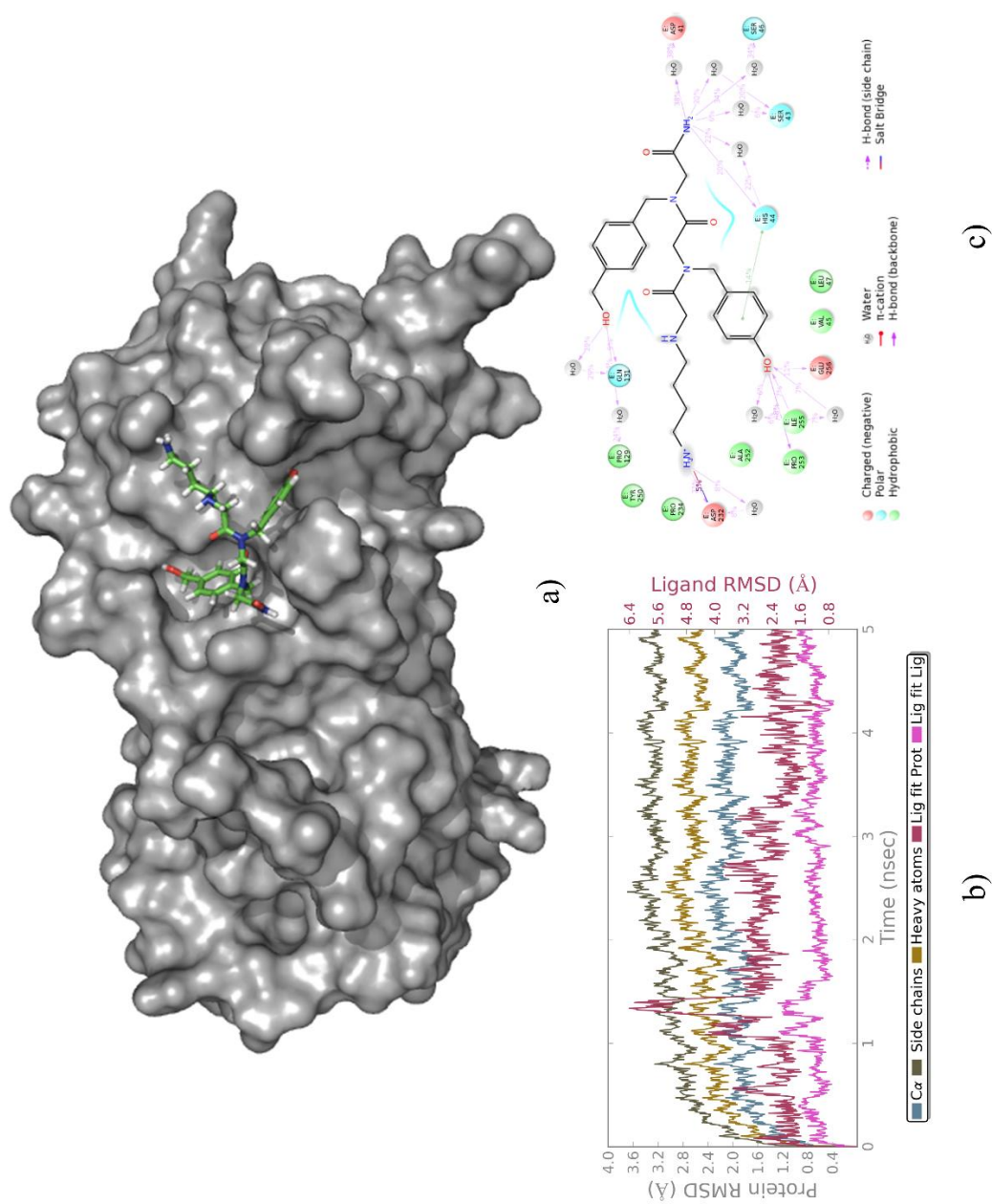


Figure A.1.1. PCNA-NLys-NTyr-NBa1 MD Simulation Output. (a) Average structure of the final 50 frames of the PCNA-NLys-NTyr-NBa1 MD simulation. (b) RMSD diagram demonstrating simulation convergence. (c) Ligand interaction diagram showing molecular contacts over the course of the simulation.

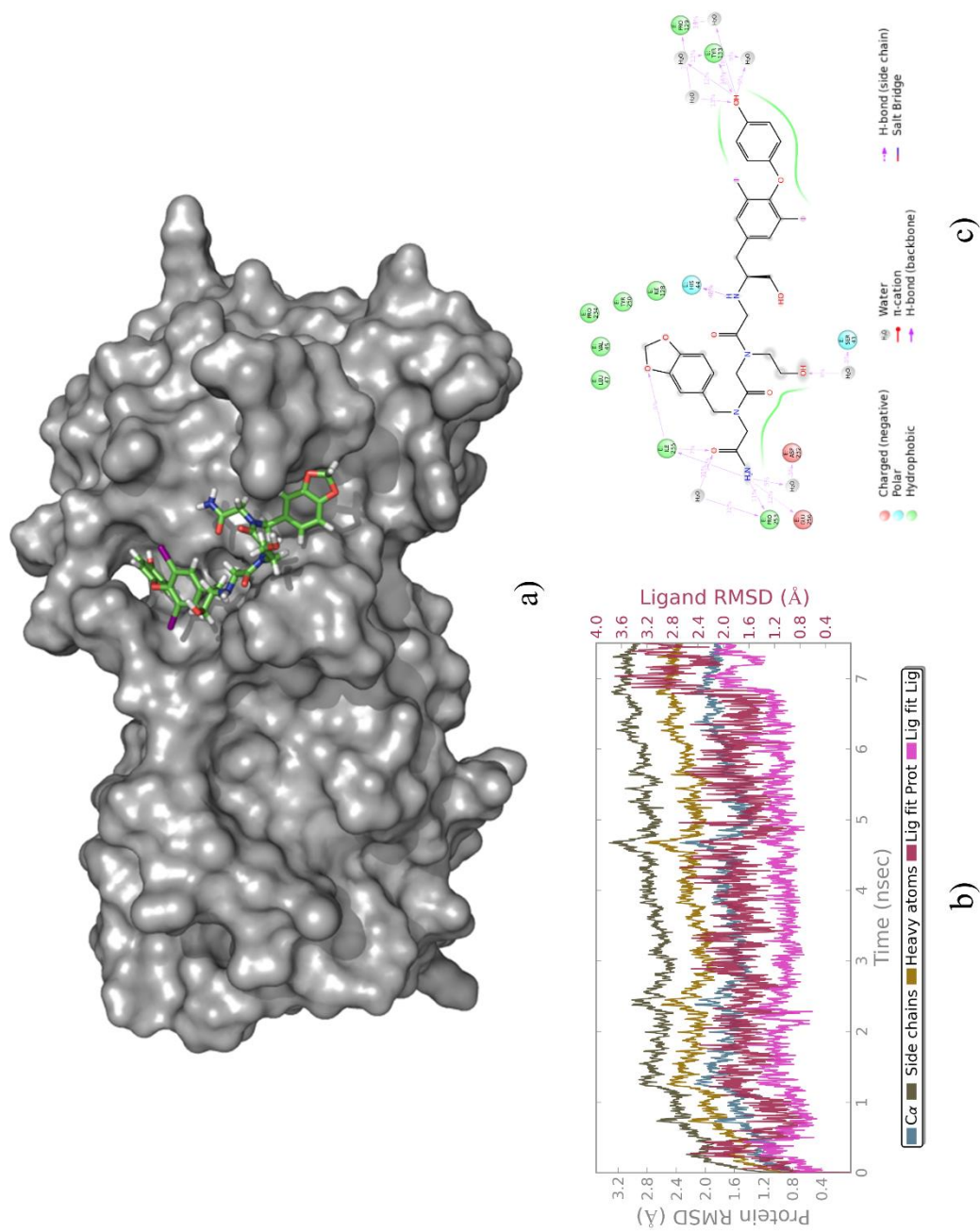


Figure A.12. PCNA-T2AA-NEal-NPip MD Simulation Output. (a) Average structure of the final 50 frames of the PCNA-T2AA-NEal-NPip MD simulation. (b) RMSD diagram demonstrating simulation convergence. (c) Ligand interaction diagram showing molecular contacts over the course of the simulation.

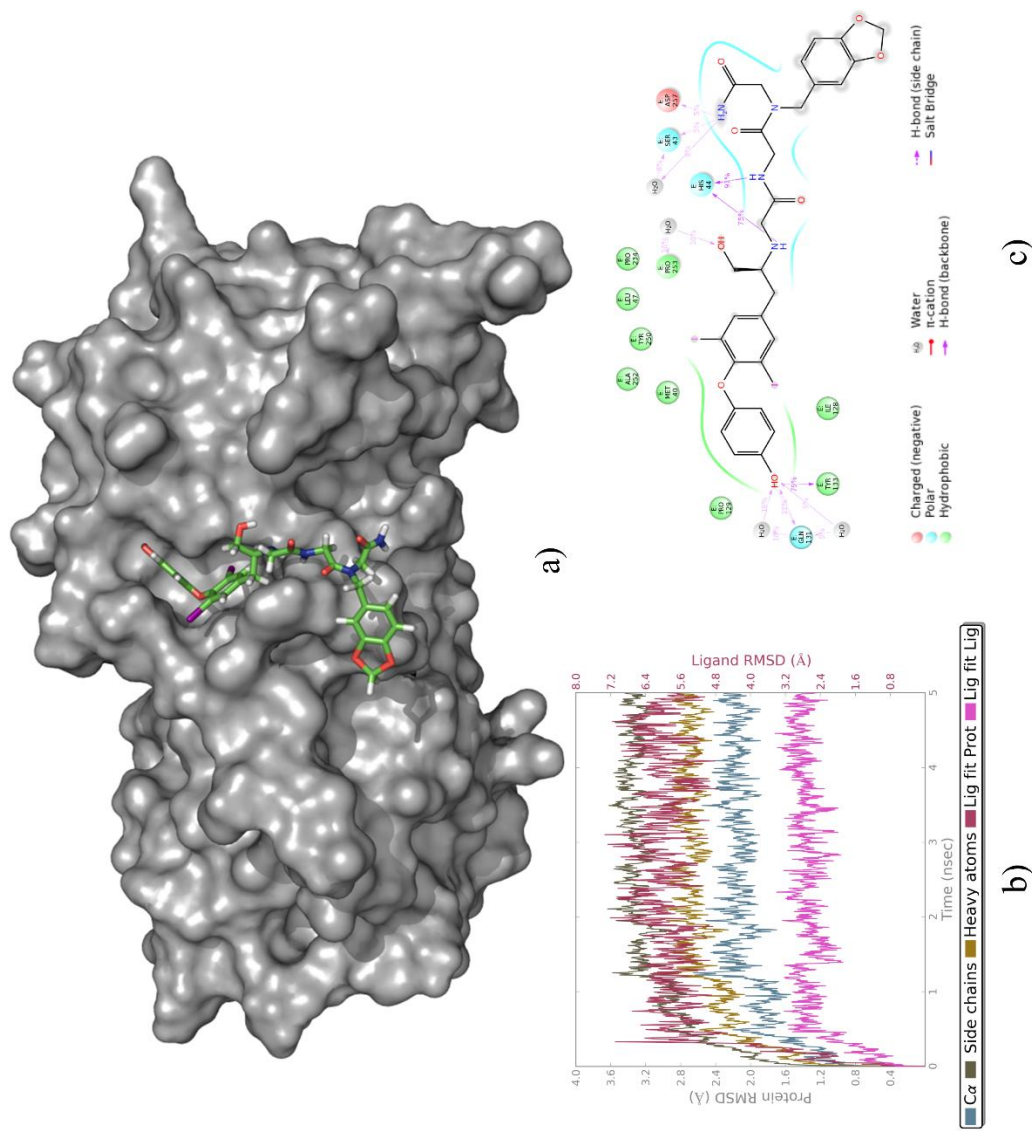


Figure A.13. PCNA-T2AA-Gly-NPip MD Simulation Output. (a) Average structure of the PCNA-T2AA-Gly-NPip MD simulation. (b) RMSD diagram demonstrating simulation convergence. (c) Ligand interaction diagram showing molecular contacts over the course of the simulation.

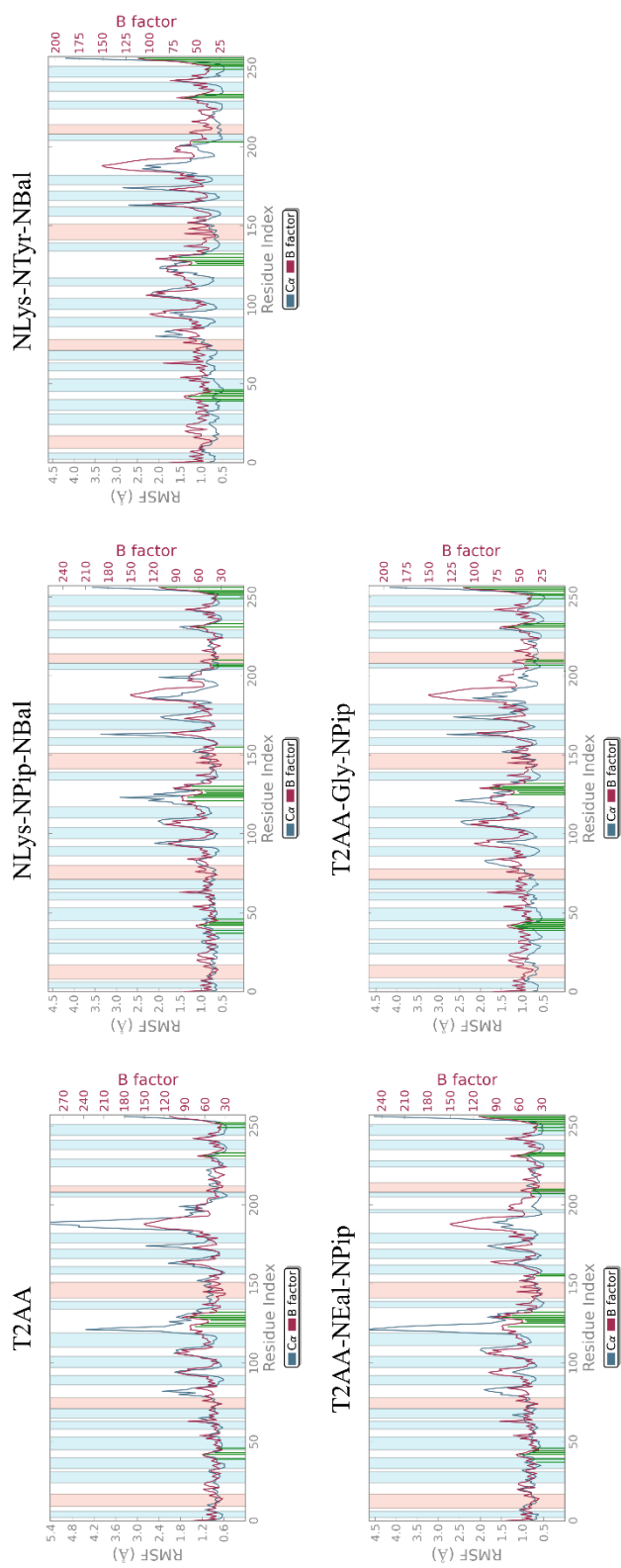
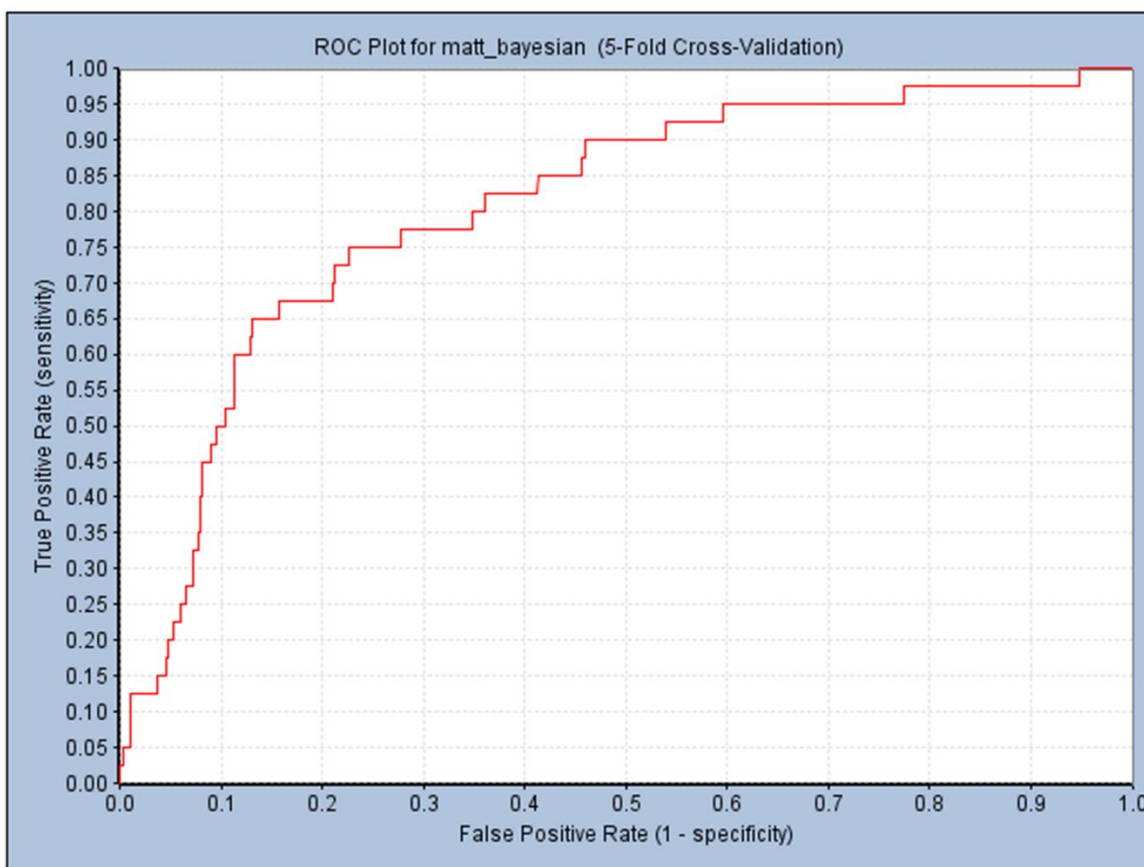


Figure A.14. MD Simulation C α Atom RMSD Fluctuation by Residue Number. RMSD diagrams of each PCNA-peptide MD simulation are shown. Peaks indicate areas of the protein that fluctuate the most during the simulation. Alpha helical and beta-strand regions are highlighted in orange and blue, respectively. Contacts between the ligand and PCNA residues are represented as green lines projecting upward from the x-axis.

XV ROC AUC	Best Split	TP/FN FP/TN	# in Category
0.876	-0.188	38/4 249/769	40

5-Fold Cross-Validation Result									
Model Name	ROC Score	ROC Rating	True Positive	False Negative	False Positive	True Negative	Sensitivity	Specificity	Concordance
matt_bayesian	0.806	Good	38	2	344	674	0.950	0.662	0.673



Category %	1%	5%	10%	25%	50%	75%	90%	95%	99%
3.781%	5%	35%	52.5%	85%	95%	100%	100%	100%	100%

Figure A.16. Enrichment Plot for Bayesian Model Prediction of the 2P2I Hunter Compound Set. A leave-one-out cross-validation of the Bayesian model for the 1058-ligand 2P2I Hunter set was performed, an enrichment plot was generated, and the percentages of true category members captured at particular cutoff percentages were listed. From this, a best split value was calculated as -0.188 for the Bayesian score.

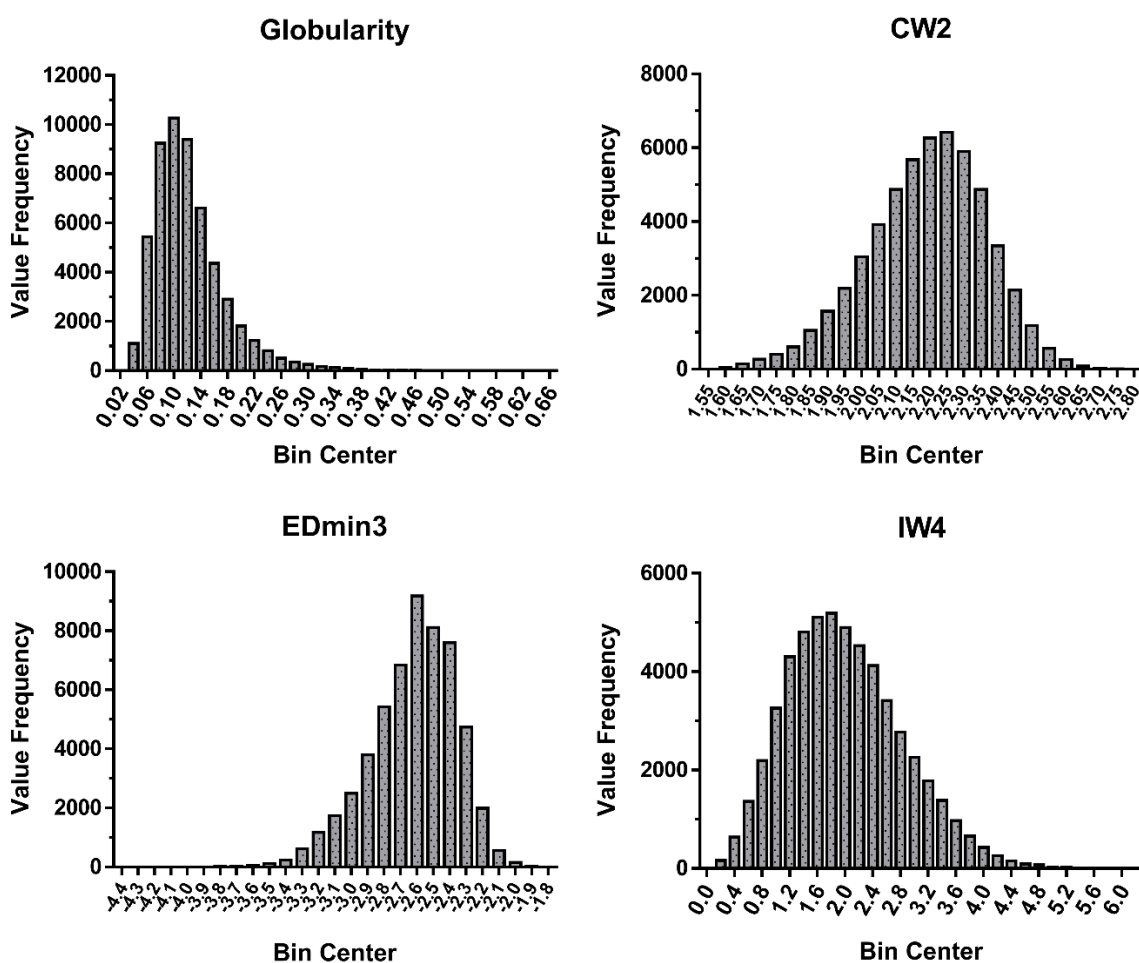


Figure A.17. Distribution of Peptoid Descriptor Statistics. After characterizing each peptoid from the virtual library of 38 fragments (36 side chains and two variants of T2AA that can be incorporated into a peptoid backbone), four descriptors for each ligand were calculated—globularity, CW2, EDmin3 and IW4. Histograms showing the distribution of scores for each descriptor amongst the entire set of peptoids are shown.

B. Supplementary Information for Nuclear-Targeted Therapeutics and Cell Lines

Supplementary Materials and Methods

Imaging of NME and LM1 Cells by Immunofluorescence. Cells were seeded onto glass coverslips and cultured overnight. The coverslips were washed with PBS, the cells fixed with PBS plus 4% formaldehyde for 20 minutes at room temperature, and then the cells permeabilized with PBS plus 0.1% Triton X-100 for 20 minutes at room temperature. The cells were washed with PBS, followed by PBS plus 2% BSA, and were incubated overnight at 4°C with primary antibody (anti-EGFR) in PBS plus 2% BSA. After incubation, the cells were washed and again incubated for one hour at room temperature with secondary antibody (donkey anti-mouse Alexa Fluor® 488) in PBS plus 2% BSA. The cells were washed with PBS plus 2% BSA, PBS alone, followed by incubation with DAPI and phalloidin (Alexa Fluor® 594) in PBS for 5 minutes at room temp. The cells were washed in PBS three more times, then mounted to a glass slide using Prolong Gold anti-fade reagent. Images were acquired using the Nikon A1R-MP confocal microscope in the Purdue Bioscience Imaging Facility.

Assessing Oral Absorptivity of Gef-SV40-NArg with Caco-2 Cells. Caco-2 Cells were obtained from ATCC (Manassas, VA). Dulbecco's Modified Eagle's Medium (DMEM), Fetal Bovine Serum (FBS), Penicillin/Streptomycin (P/S), Non-Essential Amino Acids (NEAA), L-Glutamine, Hank's Balanced Salt Solution (HBSS), Phosphate Buffered Saline (PBS), Trypsin-EDTA, Type I rat tail collagen, culture flasks, Corning 12-well 0.4 µm polyester Transwell filter supports, and all other supplies were purchased from Sigma-Aldrich (St. Louis, MO). Caco-2 cells were cultured in DMEM supplemented with 10% FBS, 1X P/S, 1X NEAA, and 2mM L-Glutamine in T-75 flasks at 37°C in 5% CO₂ and 90% RH. Media was changed every other day. Cells were passaged at a ratio between 1:3 and 1:5 at 80-90% confluence. Cells were grown for at least 2 weeks after removing from cryopreservation before studies were conducted.

Permeability studies were conducted in triplicate on Corning 12-well 0.4 µm polyester Transwell filter supports. Transwells were pre-treated with 65 µL of 1 mg/mL Type I rat tail collagen in 60% ethanol. Transwells were left overnight to evaporate ethanol prior to plating. Passage 41 Caco-2's were plated at a density of 70,000 cells/cm². After plating, media was changed every other day for 28 days. To ensure monolayer integrity, TEER was measured with a WPI Inc.

EVOMX Volt-Ohm meter. All Transwells were above 400 ohm*cm². Prior to studies, cells were washed twice with PBS and then equilibrated for 30 minutes in HBSS, pH 7.4. After equilibration, HBSS was removed and 0.5 mL of drug solutions were added to the apical chamber. Transwells were then moved into well plates with 1.5 mL of pre-warmed HBSS. 200 µL samples were removed from the basolateral compartment at 30, 60, 90, 120, 150, and 180 minute time points. After each sample, 200 µL of HBSS was added to replace sample volume. Apparent permeability coefficients were determined using the following equation:

$$P_{app} = \frac{V_A * (dC/dt)}{SA * C_0 * 60} \quad (\text{B.1})$$

where P_{app} is the apparent permeability, V_A is the volume of the acceptor compartment, dC/dt is the change in basolateral compartment concentration over time, SA is the Transwell surface area, C_0 is the initial donor concentration, and 60 is a unit conversion from minutes to seconds.

The analysis was conducted on an Agilent 1100 HPLC with an Agilent Zorbax Eclipse XBD-C18 Column, 5 µm, 4.6 mm x 150 mm. Samples were kept in the autosampler at 4°C. A gradient method was used and is shown below where A is water with 0.1% Trifluoroacetic acid filtered through 0.22 µm nylon filters, and B is Chromsolv Acetonitrile with 0.1% Trifluoroacetic acid.

Table B.1. Solvent Gradient for HPLC Method

Time (min)	%A	%B
0	90	10
9	10	90
10	10	90
12	90	10
15	90	10

Flow was set to 1 mL/min with a column temperature of 40°C. Gef-SV40-NArg was measured at 345 nm. The retention time for the compound was 5.1 minutes. C_0 and Donor_{t-180} samples were injected at 5 µL, while time point samples were 100 µL injections.

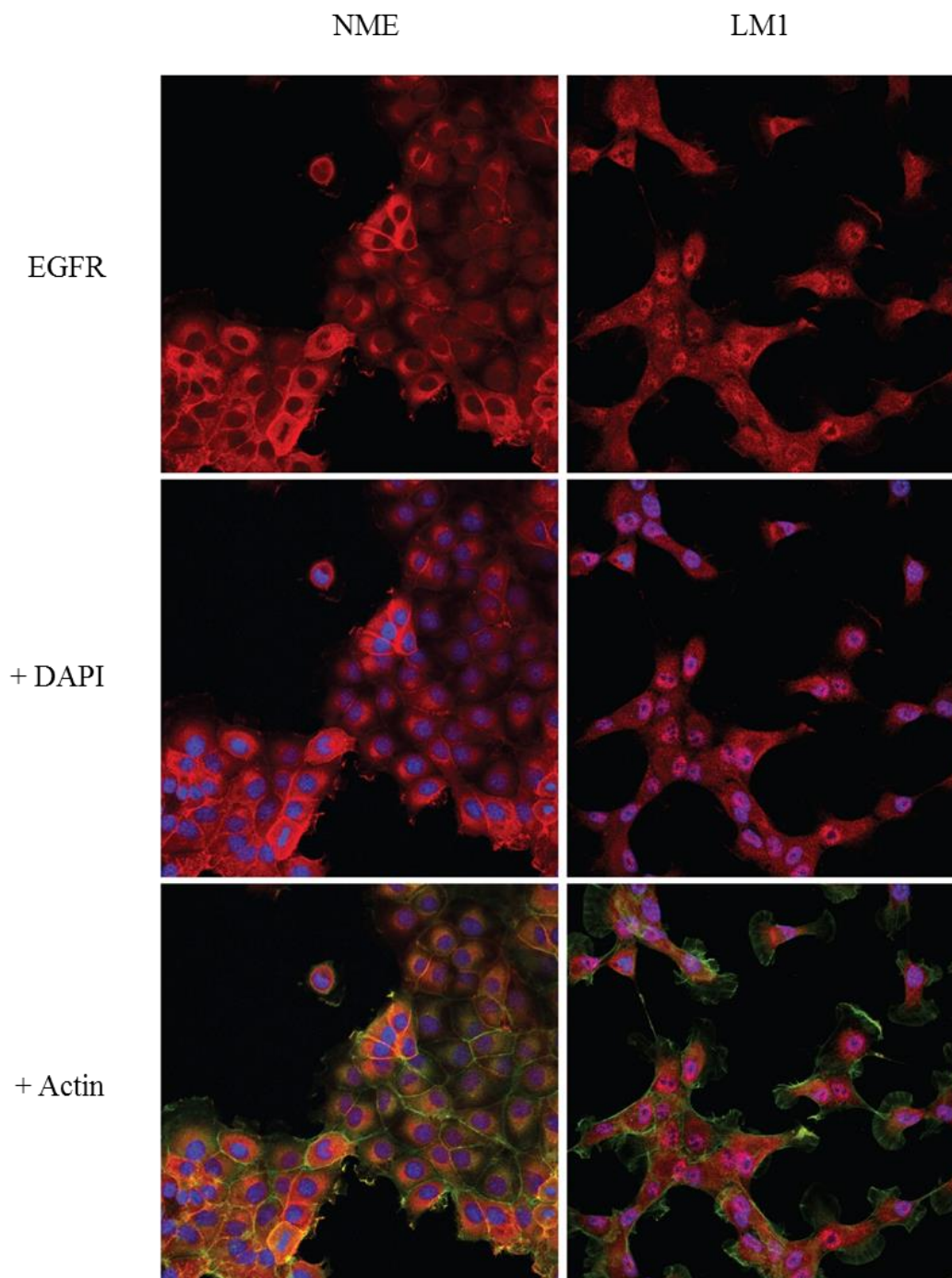


Figure B.1. Localization of EGFR in NME and LM1 Cell Lines. Both cell lines were subjected to analysis by immunofluorescence using a monoclonal antibody to detect EGFR. After immunostaining, cells were stained with both DAPI and phalloidin (Alexa Fluor[®] 594) to detect nuclei and actin, respectively. NME cells display upregulated whole-cell EGFR expression, while LM1 cells show EGFR mostly localized to the nucleus. Images generated by Wells Brown, Purdue University College of Pharmacy.

Table B.2
 Characterization of Subcellularly-Targeted Peptoid- or Peptide-Based Conjugates

Peptoid- or Peptide-Based Conjugate ^a	Calculated Mass [M+H] ¹⁺	Observed Mass ^c (m/z)	% Yield ^d
Ahx-NArg	1534.0128 ^b	1534.4058 ^b	1.11 %
Ahx-SV40-NArg	2513.7174 ^b	2513.8523 ^b	1.29 %
FAM-NArg	1893.0684 ^b	1893.5849 ^b	0.45 %
FAM-NLys(7)	1383.8153 ^b	1383.9733 ^b	2.76 %
FAM-NLys(9)	1640.0053 ^b	1640.2131 ^b	2.11 %
FAM-SV40-NArg	2871.4355 ^b	2871.0960 ^b	1.15 %
FAM-TAT	2091.3680 ^b	2091.5021 ^b	3.46 %
Gef-NArg	2021.8010	2022.7603	2.62 %
Gef-NLys	1770.6882	1770.5853	0.34 %
Gef-SV40	1482.2582	1482.3820	3.61 %
Gef-SV40-NArg	2999.0735	2999.4851	0.47 %
Gef-SV40-NLys	2748.9687	2748.9822	3.20 %

^a naming conventions outlined in Chapter 3

^b calculated/detected as negative ion [M-H]¹⁻

^c as determined by MALDI-TOF

^d based on dry mass recovery after HPLC purification

VITA

VITA

Matthew David Bartolowits was born to parents David and Leslie on February 17, 1987 in Indianapolis, Indiana. After graduating from St. Monica Catholic School in 2001, he attended Pike High School in Indianapolis, and graduated in 2005. He then attended Purdue University in West Lafayette, Indiana and earned a Bachelor of Science in Honors Biochemistry with a minor in Chemistry. During his undergraduate studies, he worked with Dr. Peter Hirst in the Department of Horticulture and Landscape Architecture on the characterization of apple strains based on cell size, density and DNA composition. He later worked in the laboratory of Dr. Mark Hall in the Department of Biochemistry, studying eukaryotic cell cycle control, and earned a Purdue SURF research fellowship to conduct honors-level research, for which a thesis was written. Following his graduation in May 2010, Matthew joined the research team of Dr. George Sandusky at Indiana University School of Medicine where he studied the link between genetic mutations and schizophrenia, and assisted in the development of a monoclonal antibody targeting a cancer-associated isoform of Proliferating Cell Nuclear Antigen (PCNA). He began his graduate studies in August 2010 in the Department of Medicinal Chemistry and Molecular Pharmacology at Purdue University. In December of that year he joined the laboratory of Dr. V. Jo Davisson studying the application of peptoid screening technology toward the discovery of new inhibitors of PCNA. In May 2014 he received a Purdue Research Foundation assistantship, and in April 2015 received a Bilsland Dissertation Fellowship from Purdue. After graduating with his Ph.D. in December 2015, Matt co-founded RamaGLO, a company focused on developing a diagnostic tool for detecting early stage acute kidney injury.