

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Structural, Biochemical Characterization, and Homology Based Modeling of Target Protein Interactions with Natural and Synthetic Indolecarbinol Compounds that Control Anti-Proliferative Signaling in Human Melanoma and Breast Cancer Cells

Permalink

<https://escholarship.org/uc/item/0f19g6f0>

Author

Quirit, Jeanne

Publication Date

2016

Peer reviewed|Thesis/dissertation

Structural, Biochemical Characterization, and Homology Based Modeling of Target
Protein Interactions with Natural and Synthetic Indolecarbinol Compounds that Control
Anti-Proliferative Signaling in Human Melanoma and Breast Cancer Cells

By

Jeanne G. Quirit

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

Doctor of Philosophy

in

Endocrinology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Gary L. Firestone, Chair
Professor Jen-Chywan Wang
Professor David Wemmer

Fall 2016

Abstract

Structural, Biochemical Characterization, and Homology Based Modeling of Target Protein Interactions with Natural and Synthetic Indolecarbinol Compounds that Control Anti-Proliferative Signaling in Human Melanoma and Breast Cancer Cells

By

Jeanne G. Quirit

Doctor of Philosophy in Endocrinology

University of California, Berkeley

Professor Gary L. Firestone, Chair

Cancer unforgivingly impinges upon millions of lives daily and is a predominant cause of death worldwide. Conventional cancer treatment is comprised of surgery, radiation, chemotherapy, hormone, immune, and targeted therapy, however, with the deleterious side effects and eventual tumor resistance that invariably ensue, there is an urgency to develop safer, less invasive therapies. Within the past decade, a significant transition in cancer therapeutics has unfurled as molecular targeted therapies have emanated as an alternative treatment for an array of cancers which include breast, colorectal, lung, pancreatic, lymphoma leukemia, and multiple myeloma. Molecular targeted therapies suppress critical biochemical pathways or mutant proteins that are required for tumor cell growth and survival. Within a molecularly defined cluster of patients, these drugs have the capacity to arrest tumor progression and can impel dramatic regressions. Identifying novel classes of highly potent therapeutic agents that specifically act on molecular targets with diminished side effects after continued treatment has been a challenging obstacle to overcome in the treatment of cancer. Among the array of molecular targeted agents employed, indole-3-carbinol (I3C), has emerged as a viable anti-cancer agent. I3C is a bioactive component in cruciferous vegetables that is derived by hydrolysis from glucobrassicin in *Brassica* and exhibits multiple anticarcinogenic and antitumorigenic properties as well as chemo-preventative and strong anti-tumor properties *in vivo* in rodent model systems and in human cancer cell xenograft tumors. A major headway in understanding the I3C anti-proliferative mechanism is our discovery that I3C triggers distinct and overlapping sets of anti-proliferative signaling events by direct interactions with specific target proteins. Through a multi-faceted series of cellular and biochemical experiments, our lab has cemented

three I3C target proteins which include human neutrophil elastase, E3 ubiquitin ligase NEDD4-1, and oncogenic B-RAF V600E serine/threonine kinase.

Because of I3C's off-target and nonspecific cytotoxic effects and its propensity to dimerize into its natural condensation product DIM, which triggers distinctly different antiproliferative cascades, there is a need to generate more potent, target-specific compounds this ultimately lead to a promising new compound, 1-benzyl-I3C, which is substantially more effective in suppressing enzymatic activity, inducing anti-proliferative effects in melanoma and breast cancer cells, and diminishing tumor size in mouse xenografts . Here, we demonstrate that both I3C and 1-benzyl-I3C serve as molecular scaffold for creating a novel, robust enzymatic small molecule inhibitors aimed at disrupting specific target proteins in melanoma and breast cancer cells, particularly NEDD4-1 and elastase respectively. By executing an *in silico* approach, we were able to make predictions about the mechanistic and structural nature of a set of five synthetic I3C and 1-benzyl-I3C derived analogs and employed a combination of *in vitro* protein thermostability assays and enzymatic assays to substantiate our predictions. Notably, compounds 2242 and 2243, the two indolecarbinol analogues with added methyl groups that result in a more nucleophilic benzene ring π system, further inhibited NEDD4-1 enzymatic activity more dramatically with IC₅₀s of 2.7 μ M and 7.6 μ M, respectively. Interestingly, compounds 2242, 2160, and 2243 inhibited elastase activity much more significantly as well with and an IC₅₀ of 30.4 μ M, 25.1 μ M, and 16.9 μ M respectively. In both NEDD4-1 and elastase enzymatic studies, the potency of compound 2242, 2160, and 2243 appeared to be sensitive to structural changes on the phenyl ring, more notably when methyl substituents were added to the para and ortho positions. The activity was abolished with the addition of bulky chemical groups like the thiophene substituent attached to the indole of 2163 and the electron donating methoxy group attached to the phenyl moiety of 2244 in the meta and ortho positions.

The quest for additional I3C target proteins has been facilitated through our understanding of homology modeling and identification of patterns in protein folds and ligand binding sites. In order to make predictions about additional indolecarbinol target proteins, we obtained the crystal structures of the four I3C target proteins to date (human neutrophil elastase, ubiquitin E3 ligase NEDD4-1, oncogenic BRAF V600E serine/threonine kinase, WNT) and used them as the starting template structure for creating homologous protein models. Homology models for various homologues of the target proteins were generated to ascertain whether or not similarities in potential indolecarbinol binding sites could be visibly discerned after performing a computational docking analysis of indolecarbinol compounds into their most feasible predicted binding sites. After examining the indolecarbinol binding sites in the homology models, it is evident that the homologous proteins that have greater than 50% sequence identity share a distinct structural architecture that confers binding to the indolecarbinol compounds, but generally, homologous proteins that share 30% sequence identity or below tend to lose essential I3C contact residues.

Examining the binding modes of I3C, 1-benzyl-I3C, and their corresponding synthetic derivatives in complex with their target proteins, namely, elastase, NEDD4-1, oncogenic

BRAF V600E serine/threonine kinase, and wnt, can illuminate various patterns in binding sites and can hence provide valuable information for locating additional indolecarbinol target proteins. Conceivably, information garnered from these studies will be useful in making predictions allowing identification of additional indolecarbinol compound target proteins.

The results presented here provide the fundamental framework for understanding the mode of inhibition of natural and synthetic indolecarbinol compounds and the proteins they effectively target.

Acknowledgements

It was actually the swift and heavy slam of one door that serendipitously opened another, blowing a vibrant and colorful wind of promise from Ithaca, NY towards Berkeley, CA. Second chances granted me the opportunity to pursue and wrestle with my love affair with science, but it was really the love and interminable “you got this” cheers from my parents, family, and friends that allowed me to resolutely move forward and continue being.

First, I would like to thank my advisor, Gary Firestone. With one word, my life forever changed when you expeditiously said, “yes.” By saying “yes,” you not only provided me with a bench to work on and a desk with a window and view, but you gave me another home. As my advisor, you gave me the independence and intellectual freedom to explore the depths of my own scientific creativity and allowed me to embark on projects that challenged and stiffened me, ultimately making me a better scientist. But apart from the science, thanks for parenting us and making sure we were all comfortable sitting in the back seat of your car, listening to country rap jams you specifically selected on our trips to Granlibakken retreats. Thank you for Thanksgiving dinners at your house and invitations to musicals and more significantly, thanks for making me feel like I belonged.

I am very grateful to Kathy Durkin for allowing me to use the Computational Chemistry facility to perform the molecular dynamics and *in silico* studies that were an integral part of this thesis. Moreover, thank you to Kathy, Yinka, and Kelley, for letting me be a part of the garlic pepper pork noodles crew on Thai House Thursdays.

Thank you to Dale Leitman for letting me be a part of the synergistic studies involving the co-ligand binding of E2 and CC7 to the ER- α receptor. But more importantly, thank you for your incessant readiness to help, for your advice about academia, industry, and life in general.

The decision to go to grad school was heavily influenced by my first lab home in Florida. “Two roads diverged in a wood” and I’m grateful that I did eventually take the “one less traveled by” or in my case, the “one” being grad school over med school. Rob and Mavis, not just one day, one month, or one year, but the entirety of my time in your lab could be summarized simply: happy. Rob and Mavis, you were a glimpse into the window of what my life in science ought to be like. There was always a hunger that propelled people in your lab to ask more questions and it was this curiosity and perpetual fervor that inspired me to become a scientist. Most of all, Rob and Mavis, you have shown me that the fundamental core of scientific discovery is still about the love for it all—love for understanding how and why and love for answering how and why.

When I first moved to Berkeley, my only expectation was to improve my technical skills at cloning and protein purification and perhaps learn a new biochemical technique. I never anticipated that I would be living the best four years of my life here.

Bhumika, you welcomed me into the lab as a sister. In the car, when you are driving, before coming to an abrupt stop, you have this instinctive tendency to extend your arm out, almost like an extra seatbelt to protect me. This small action in many ways represents our relationship. There is a certain comfort and security in your friendship that gives me peace and stability. Thank you for making me hot tea on cold nights when I was “studying” for quals. Thank you for always making sure I arrived

home safely after having “Bhumi chicken” dinners at your apartment. Home never felt like more than a 30-minute drive away with you in my life.

Kevin, you dismantled all my stereotypes of the cold, rigid scientist and sustained my faith that there are genuinely good people who help for the pure sake of helping. More than being a colleague in the lab, you are a friend whose reliability is indisputable and the word “no” to you just doesn’t belong to your vernacular. Thank you for bringing me pizza and coffee when I was preparing for my seminar. Thank you for taking me to In-N-Out so I could enjoy my burger properly for the second time. You were the butt of my every joke and you girl-talked with me when I needed advice. One day I’ll admit that you’re always right, but for now, just accept my thanks.

To my sparkles in the lab—Janice and Shanice. Many of the projects we tackled would not have come to full fruition had it not been for your dedication and tenacity. You both kept me smiling and you rejuvenated my love for science. Thanks for being the bursts of sunshine on all days—good or bad.

To all the Firestone lab members, past and present—you were my extended family and carried me through the years with your grace, your laughter, and your stories.

I would like to extend my deepest gratitude to my Lola Nene and Lolo Pepe who never stopped believing in me. I’m grateful to my Grandma Sol and my cousins, Jennifer and Jackie and my family and cousins in LA. My support group from near and afar kept me grounded: Natasha & Earl, Lily, Bala, Lisa, Alisa & Lenny, Susanne, Christine, Amanda, Calvin, Ash, Sid, Jessica, Hai, Abeer, Tiff & Nuan, Mizue, Harald, Helen, Debbie, Annie, Gil, Anna & Tony. My thesis writing group—Allison, Ioana, and Andrew, imbued me with their sense of serenity and warmth and they kept me sane with our late night dance parties. I’m also grateful to Janice and Isaak at Yali’s café for always saving me a blueberry scone in the morning. I’d like to thank The Shins for creating ‘Simple Song’ and Sylvia Plath for writing ‘The Bell Jar’ which both gave me clarity throughout grad school.

In the end, I owe it all to my parents. I want to thank you for being a constant refuge to which I can always retreat. You nurtured and watered my growth with love and you gave me a blank canvas to which I could paint my masterpiece. After my painful departure from Cornell, your response to my decision to start anew in California was plainly and yet so earnestly, “how can we help?” There was no flicker, no hesitation—just support and a plan for how to proceed. The long, sinuous drive from Ithaca, NY to Palm City, FL, and finally to Berkeley, CA left me less fearful of the future, all because we were undertaking this journey together. From picking me up after tumbling down the slide on the playground to capturing moments of me eating pancakes at IHOP on video when I was 5, you always were and continue to be at my side. Thank you for letting me sing loudly as a child. Thank you for letting me catch frogs. Thank you for sit-down dinners together every night. I know love because of you and the courage and resolve that I have now are only fortified by your example.

If it is true that we are just elements of the people we meet and the relationships we forge, then I am lucky just to know all of you because you all have augmented my world in ways that I could only dream of. Thank you.

To my parents who let me bloom.

Table of Contents

List of Figures.....	vi-vii
----------------------	--------

General Introduction.....	viii-xiv
--	-----------------

Chapter I

Indole-3-carbinol (I3C) analogues are potent small molecule inhibitors of NEDD4-1 ubiquitin ligase activity that disrupt proliferation of human melanoma cells.....	1
Abstract.....	2
Introduction.....	3-4
Materials and Methods.....	5-10
Results.....	11-27
Discussion.....	28-31
References.....	32-37

Chapter II

Indole-3-carbinol (I3C) analogues are potent small molecule inhibitors of human neutrophil elastase activity that disrupt proliferation of human breast cancer cells.....	38
Abstract.....	39
Introduction.....	40-41
Materials and Methods.....	42-44

Results.....	45-56
Discussion.....	57-61
References.....	62-65

Chapter III

Protein homology modeling unveils novel targets for natural and synthetic indolecarbinol compounds.....66

Abstract.....	67
Introduction.....	68-87
Materials and Methods.....	88-89
Results.....	90-160
Discussion.....	161-166
References.....	167-179

Chapter IV

Conclusion and Future Directions

Conclusion and Future Directions.....	180-185
---------------------------------------	---------

List of Figures

Chapter I Figures

Figure 1.1: In silico computational modeling of molecular interactions, direct binding of I3C and 1-Benzyl-I3C to the HECT domain of the ubiquitin ligase NEDD4-1, and effects of I3C and 1-benzyl-I3C on the enzymatic activity of NEDD4-1

Figure 1.2: Effects of I3C and 1-benzyl-I3C on MITF-M protein levels and cell cycle in G361 human melanoma cells.

Figure 1.3: Structures of I3C, 1-benzyl-I3C, and indolecarbinol analogues.

Figure 1.4: *In silico* computational models of predicted molecular interactions of indolecarbinol synthetic compounds with the human NEDD4-1 HECT domain x-ray crystal structure

Figure 1.5: Binding of indolecarbinol compounds to purified NEDD4-1 HECT domain was evaluated via protein thermal shift assays

Figure 1.6: Effects of indolecarbinol compounds on the *in vitro* enzymatic activity of NEDD4-1 and on G361 human melanoma cell proliferation.

Figure 1.7: Table comparing the half maximum inhibitory concentration for NEDD4-1 enzymatic activity after 1 hr treatment and half maximum inhibition of proliferation of G361 cells after 48 hrs treatment with each indolecarbinol compound

Chapter II Figures

Figure 2.1: Structures of I3C, 1-benzyl-I3C, and indolecarbinol analogues

Figure 2.2: Ribbon diagram of elastase displaying the top ranked, highest probability ligand binding sites

Figure 2.3: Predicted molecular interactions of human neutrophil elastase in complex with I3C, 1-benzyl-I3C, and indolecarbinol analogues

Figure 2.4: Binding of indolecarbinol compounds to purified human neutrophil elastase was evaluated via protein thermal shift assays

Figure 2.5: Effects of I3C, 1-benzyl-I3C, and indolecarbinol analogues on *in vitro* elastase enzymatic activity

Figure 2.6: Cell proliferation assay depicting percent of viable cells at each respective dose of indolecarbinol compound in MCF-7 cells

Figure 2.7: IC₅₀ values of indolecarbinol compounds with human neutrophil elastase and half maximum concentration for growth inhibition in MCF-7 cells

Figure 2.8: Flow cytometry profiles of MCF-7 cells with indicated doses of I3C, 1-benzyl-I3C, and indolecarbinol analogues

Chapter III Figures

Figure 3.1: Table depicting percent sequence identity of E3 ubiquitin ligase NEDD4-1 homologous proteins.

Figure 3.2: Homology model of neural cell expressed, developmentally down-regulated 4-like, isoform CRA_b

Figure 3.3: Homology model of cDNA FLJ53199, highly similar to E3 ubiquitin-protein ligase NEDD4-like

Figure 3.4: Homology model of E3 ubiquitin-protein ligase NEDD4-like protein

Figure 3.5: Homology model of E3 ubiquitin-protein ligase, HUWE1

Figure 3.6: Homology model of E3 ubiquitin-protein ligase, HERC2

Figure 3.7: Sequence alignment of NEDD4-1 HECT domain and corresponding homologous HECT domain containing proteins

Figure 3.8: Table depicting percent sequence identity of human neutrophil elastase homologous proteins

Figure 3.9: Homology model of proteinase 3

Figure 3.10: Homology model of azurocidin

Figure 3.11: Homology model of PRSSL1 protein

Figure 3.12: Sequence alignment of elastase and corresponding homologous proteins

Figure 3.13: Molecular docking model of oncogenic B-RAF V600E serine/threonine kinase highlighting indolecarbinol predicted binding site

Figure 3.14: Molecular docking model of wild-type B-RAF serine/threonine kinase highlighting indolecarbinol predicted binding site

Figure 3.15: Table depicting percent sequence identity of Oncogenic B-RAF V600E homologous proteins

Figure 3.16: Homology model V-Raf-1 murine leukemia viral oncogene

Figure 3.17: Homology model Raf proto-oncogene serine/threonine protein kinase

Figure 3.18: Homology model of highly similar to RAF proto-oncogene serine/threonine protein kinase

Figure 3.19: Sequence alignment of oncogenic B-RAF V600E and corresponding homologous proteins

Figure 3.20: Table depicting percent sequence identity of Wnt homologous proteins

Figure 3.21: Homology model of Wnt3A

Figure 3.22: Homology model of Wnt5A

Figure 3.23: Homology model of Wnt7B

Figure 3.24: Homology model of Wnt10B

Figure 3.25: Homology model of ADAM23 protein

Figure 3.26: Sequence alignment of Wnt isoforms and corresponding homologous protein, ADAM23 protein

General Introduction

Increased consumption of cruciferous vegetables such as broccoli, cabbage, cauliflower, bok choy and Brussels sprouts accompany a decline in cancers, notably colon, lung, prostate, cervix and breast [1-4]. Dietary supplements containing *Brassica* vegetables also dramatically mitigates the occurrence and frequency of carcinogen-induced and spontaneous mammary epithelial tumors as observed in various animal models [5]. Emerging pre-clinical and clinical studies have demonstrated that indole-3-carbinol (I3C), a primary bioactive constituent of cruciferous vegetables that is derived by hydrolysis from glucobrassicin in *Brassica* vegetables, has multiple anti-carcinogenic and anti-tumorigenic properties. As observed in rodent model systems and in human cancer cell xenograft tumors in athymic mice with low toxicity levels, I3C and its natural diindole condensation product have anti-tumor properties and exhibit chemo-preventative properties [6-11]. Studies with human cancer cells originating from breast, prostate, lung, colon, melanomas, leukemia, and cervical cancer cells have demonstrated that treatment with I3C and DIM impel a specific cascade of transcriptional, cell signaling, enzymatic, and metabolic cycles that result in cell cycle arrest, apoptosis, and diminished cancer cell migration and metastasis, and reduced hormone receptor signaling [6-11].

Nuclear Factors regulated by I3C

Many preclinical studies have demonstrated that I3C/DIM can exert control over various nuclear transcription factors that are implicated in maintaining cellular apoptosis and/or proliferation. Alterations in downstream events in normal and neoplastic cells can be attributed to direct or indirect interactions between I3C and/or DIM with four nuclear transcription factors which include estrogen receptor (ER), Sp1, nuclear factor κ B (NF κ B), and aryl hydrocarbon receptor (AhR).

Cancer processes influenced by I3C/DIM

Mounting evidence has proven that I3C modifies cancer risk and tumor cell behavior by influencing different cancer processes. This may be a consequence of the expression of specific target proteins that exert control over important cell signaling cascades that modulate cell growth and proliferation.

A loss of cell cycle control has been typically identified as an impediment in tumor development and proliferation. Previous studies have established that I3C induces a G1 cell cycle arrest in human breast and prostate cancer cells [12-14]. In breast cancer cells, the IC₅₀ for I3C is approximately 55 μ M [15]. It has also been observed that this is the dosage at which several prostate and cervical cancer cell lines respond through suppression in cell division [16]. CDKs and the phosphorylation of Rb have been implicated in regulating the transition from the G1 to the S phase and hence have been identified as potential I3C targets [17]. In ER-negative (BT20, MDA-MB-231 and BT539) and ER-positive (MCF-7, 734B and BT474) human breast cancer cells, as well as in prostate cancer cells (PC-3) [18, 19] it has previously been established that I3C

can elevate CDK inhibitors like p21WAF1/CIP1 (p21) and p27 (Kip1) and reduce CDK6 levels and activity. Hyperphosphorylation of the Rb protein was also hampered by I3C [14].

The induction of a G1 cell cycle arrest independent of ER status, may be attributed I3C-engendered changes in Sp1. The interaction between Sp1 and a composite DNA-binding site within the CDK6 promoter can be agitated by I3C. The binding of Sp1 to the consensus Sp1-responsive elements of p21 in its promoter region can also be instigated by DIM which in turn elevates p21 expression in breast cancer cells no matter what the ER status may be [18].

I3C is also capable of impeding the evolution of multidrug resistance (MDR) in cancer cells which may be a consequence of the suppression in the expression of MDM-1 gene transcript P-glycoprotein (P-gp) [20]. Resistance to the anticancer drugs like doxorubicin or vinblastine [21] can be assuaged by ingesting dietary I3C (300-500 mg/kg/day). In human and mouse tumor cell lines, the P-gp overexpression has been observed during cell growth arrest [22]. Knowing that P-gp increases cellular efflux of anticancer drugs and hence mitigates their effects on the target sites, the fact that I3C is able to abate the P-gp expression levels in drug resistance could have substantial ramifications in cancer therapeutics.

Apoptosis

It has been demonstrated that growth inhibition induced by I3C may be a consequence of elevated apoptosis. I3C exerts control over the Bax/Bcl-2 ratio because it can diminish the mitochondrial membrane potential [23]. Antiapoptotic Bcl-2 factors are predominately located in the mitochondrial outer membrane, nuclear envelope and endoplasmic reticulum. The mitochondrial release of apoptosis-associated factors like apoptosis-protease-activated factor 1, apoptosis-inducing factor and cytochrome c are determined by the ratio of Bax/Bcl-2 in the mitochondria. I3C mediates mitochondrial depolarization and the release of apoptotic factors via Bax oligomer-generated outer membrane channels which is a consequence of Bax migration from the cytosol to the mitochondria [23].

The activity of the cell death enzyme, caspase-3, is also sensitive to I3C's inhibitory effects. In a study conducted by Zhang and Malejka-Giganti [24], after feeding I3C (5 or 25 mg/kg of body weight) to female Sprague-Dawley rats with 7,12-dimethylbenz[alpha]anthracene induced mammary tumors for a period of 4 days, the mammary gland activity of caspase3 increased 3.6-fold. In Her-2/neu-overexpressed breast cancer cells, I3C and DIM also induced apoptotic effects [25].

Carcinogen bioactivation

The preclinical anticancer characteristics of I3C can be attributed to its ability to activate various enzymes such as cytochrome P-450 (CYP)-dependent monooxygenases, GSTs, and epoxide hydrolases. I3C can impede cervical carcinogenesis because of increased

CYP-catalyzed estrogen metabolism. The liver degrades estradiol which is a natural ligand for the estrogen receptor which then leads to higher levels of 2-hydroxyestrone (2-OHE) or 16 α -hydroxyestrone (16 α -OHE) and 4-hydroxyestrone. 16 α -OHE is associated with the proliferation of some tumor cell lines, more notably those connected to the cervix. 2-OHE outcompetes estradiol for ER binding and therefore enhances antiestrogenic and antiproliferative effects. When cervical cells are treated with I3C (100 mg/kg), higher levels of 2-OHE relative to 16 α -OHE are synthesized [26]. It has also been reported that doses of 200-400 mg of I3C per day are able to dramatically increase the magnitude of 2-OHE in relation to 16 α -OHE and to induce a regression in cervical intraepithelial neoplasia in humans [27].

When rats ingested 25 mg of I3C per kilogram body weight, the capacity of hepatic microsomes to transform estradiol to various metabolites like 2-OHE surged [28]. Correspondingly, when rats ingested DIM (5 mg/kg body weight) it proved to be more than 20-fold more potent than I3C in hampering DMBA-induced mammary tumor growth, but with minimal change in CYP1A1 activity [29]. Ultimately, the I3C-induced antiestrogenic effects may not be implicated in modified CYP-dependent estrogen metabolism.

Regulation of CYP1A1 expression

CYP1A1 has been implicated in procarcinogen activation. Within the promoter region of the CYP1A1 gene, there are consensus dioxin response elements or XREs that are capable of associating with the nuclear AhR complex. Studies that implemented a high-throughput reporter gene system that enabled the stable transformation of H4IIE cells to include the luciferase gene under the control of the CYP1A1 promoter were performed to ascertain the relative effects of I3C on CYP1A1 induction [30]. The results from this study demonstrated that I3C only moderately induced CYP1A1 with a sevenfold increase in activity at a concentration of 100 μ M. Furthermore, in T47D human breast cancer cells, treating cells with concentrations of 125 and 31 μ M for I3C and DIM respectively where indoles were bound to aryl hydrocarbon receptor, did not trigger the activity of CYP1A1-detoxifying enzymes like ethoxyresorufin O-deethylase (EROD) [31]. Hence, these studies demonstrate that tumors are not induced when cells are physiologically exposed to indoles because of their procarcinogen activation promise.

It has been established that liver and mammary mRNA levels for CYP1A1 and CYP2B1/2 escalated upon I3C treatment (250 mg/kg body weight) for 4 or 10 days [28]. Treatment of 5 mg/kg body weight of I3C led to enhanced EROD (CYP1A1) and methoxyresorufin O-demethylase (CYP1A2) activities. Treatment of 25 mg/kg or higher resulted in higher activity of benzyloxyresorufin O-dealkylase (CYP2B1) or pentoxyresorufin O-dealkylase (CYP2B1/2).

In AhR-responsive T47D human breast cancer cells, the I3C and DIM inhibitory effects on 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) promoted EROD activity which can be attributed to AhR antagonist effects. Indoles competing for binding with TCDD for AhR may trigger these effects. Therefore, I3C has the potential to serve as both an AhR

agonist and antagonist *in vivo* which would consequently influence CYP-dependent metabolism of xenobiotics and endobiotics during treatment.

Effects of I3C on 2-amino-1-methyl-6-phenylimidazo [4,5-b]pyridine metabolism

I3C is capable of reducing adducts from 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP) which is a heterocyclic amine that can trigger the formation of lymphomas, intestinal tumors, mammary adenocarcinomas and hepatocellular adenomas in rodents [32-34]. When F344 rats were fed with 0.1% I3C, PhIP-DNA adducts were mitigated by 40-100% relative to their controls [32]. These results are a testament to I3C's anticancer role in defending against PhIP-engendered colon carcinogenesis [32].

Activation of DNA repair by I3C

I3C has been implicated in preventing DNA strand breaks when compounds like benzopyrene and hydrogen peroxide are utilized. The pre-treatment of human colon adenocarcinoma LS-174 cells for 24 hours with a combination of indole[3,2-b]carbazole and sulforaphane prior to treatment with benzopyrene resulted in a 20% reduction in the amount of single-stranded DNA breaks [35].

More recent studies have demonstrated in keratinocyte cells that DIM can incite DNA damage-inducible gene expression [36]. I3C/DIM treatment also upregulates BRCA1 which is a tumor suppressor gene that is implicated in the DNA repair process. Indoles can exert control over BRCA1 protein function which subsequently can impede estrogen-dependent mammary epithelial proliferation. The GADD45 promoter region activity is also triggered by BRCA1.

Summary

I3C or DIM treatment has the capacity to inhibit tumor cell growth through a plexus of cancer processes. A multitude of studies have proven that cell growth-related genes are affected by the interplay between I3C/DIM and the promoter activities of several transcription factors such as ER α , Sp1, NF κ B and AhR. Interactions between ER α and AhR or mechanisms implicating Sp1 domains have emerged as possible manners upon which I3C can confer its inhibitory and antiproliferative effects. While the specificity of the dietary indolecarbinol targets need to be more clearly defined, it is not known if the indolecarbinol compounds exert their control over a single target protein or through a multitude of functional protein complexes. Garnering an enhanced understanding of I3C/DIM target proteins in tumor cell proliferation and/or apoptosis, will illuminate promising intervention approaches in treating cancer.

References

1. Lam TK, Gallicchio L, Lindsley K, Shiels M, Hammond E, Tao XG, Chen LW, Robinson KA, Caulfield LE, Herman JG *et al*: Cruciferous Vegetable Consumption and Lung Cancer Risk: A Systematic Review. *Cancer Epidem Biomar* 2009, 18(1):184-195.
2. Witte JS, Longnecker MP, Bird CL, Lee ER, Frankl HD, Haile RW: Relation of vegetable, fruit, and grain consumption to colorectal adenomatous polyps. *Am J Epidemiol* 1996, 144(11):1015-1025.
3. Kristal AR, Lampe JW: Brassica vegetables and prostate cancer risk: A review of the epidemiological evidence. *Nutr Cancer* 2002, 42(1):1-9.
4. Fowke JH, Shu XO, Dai Q, Shintani A, Conaway CC, Chung FL, Cai QY, Gao YT, Zheng W: Urinary isothiocyanate excretion, Brassica consumption, and gene polymorphisms among women living in Shanghai, China. *Cancer Epidem Biomar* 2003, 12(12):1536-1539.
5. Sato T, Takahashi S, Mizumoto T, Harao M, Akizuki M, Takasugi M, Fukutomi T, Yamashita J: Neutrophil elastase and cancer. *Surg Oncol* 2006, 15(4):217-222.
6. Aggarwal BB, Ichikawa H: Molecular targets and anticancer potential of indole-3-carbinol and its derivatives. *Cell Cycle* 2005, 4(9):1201-1215.
7. Ahmad A, Sakr WA, Rahman KMW: Anticancer Properties of Indole Compounds: Mechanism of Apoptosis Induction and Role in Chemotherapy. *Curr Drug Targets* 2010, 11(6):652-666.
8. Kim YS, Milner JA: Targets for indole-3-carbinol in cancer prevention. *J Nutr Biochem* 2005, 16(2):65-73.
9. Moiseeva EP, Almeida GM, Jones GDD, Manson MM: Extended treatment with physiologic concentrations of dietary phytochemicals results in altered gene expression, reduced growth, and apoptosis of cancer cells. *Mol Cancer Ther* 2007, 6(11):3071-3079.
10. Safe S, Papineni S, Chintharlapalli S: Cancer chemotherapy with indole-3-carbinol, bis(3'-indolyl)methane and synthetic analogs. *Cancer Lett* 2008, 269(2):326-338.
11. Sundar SN, Kerekatte V, Equinozio CN, Doan VB, Bjeldanes LF, Firestone GL: Indole-3-carbinol selectively uncouples expression and activity of estrogen receptor subtypes in human breast cancer cells. *Mol Endocrinol* 2006, 20(12):3070-3082.
12. Aronchik I, Chen T, Durkin KA, Horwitz MS, Preobrazhenskaya MN, Bjeldanes LF, Firestone GL: Target Protein Interactions of Indole-3-Carbinol and the Highly Potent Derivative 1-Benzyl-I3C With the C-Terminal Domain of Human Elastase Uncouples Cell Cycle Arrest From Apoptotic Signaling. *Mol Carcinogen* 2012, 51(11):881-894.
13. Cover CM, Hsieh SJ, Tran SH, Hallden G, Kim GS, Bjeldanes LF, Firestone GL: Indole-3-carbinol inhibits the expression of cyclin-dependent kinase-6 and induces a G(1) cell cycle arrest of human breast cancer cells independent of estrogen receptor signaling. *J Biol Chem* 1998, 273(7):3838-3847.

14. Chinni SR, Li YW, Upadhyay S, Koppolu PK, Sarkar FH: Indole-3-carbinol (I3C) induced cell growth inhibition, G1 cell cycle arrest and apoptosis in prostate cancer cells. *Oncogene* 2001, 20(23):2927-2936.
15. Fares FA, Ge XK, Yannai S, Rennert G: Dietary indole derivatives induce apoptosis in human breast cancer cells. *Adv Exp Med Biol* 1998, 451:153-157.
16. Nachshon-Kedmi M, Yannai S, Haj A, Fares FA: Indole-3-carbinol and 3,3'-diindolylmethane induce apoptosis in human prostate cancer cells. *Food Chem Toxicol* 2003, 41(6):745-752.
17. Sherr CJ: Cancer cell cycles. *Science* 1996, 274(5293):1672-1677.
18. Hong CB, Kim HA, Firestone GL, Bjeldanes LF: 3,3'-Diindolylmethane (DIM) induces a G(1) cell cycle arrest in human breast cancer cells that is accompanied by Sp1-mediated activation of p21(WAF1/CIP1) expression. *Carcinogenesis* 2002, 23(8):1297-1305.
19. Firestone GL, Bjeldanes LF: Indole-3-carbinol and 3,3'-diindolylmethane antiproliferative signaling pathways control cell-cycle gene transcription in human breast cancer cells by regulating promoter-Sp1 transcription factor interactions. *J Nutr* 2003, 133(7):2448s-2455s.
20. Arora A, Shukla Y: Modulation of vinca-alkaloid induced P-glycoprotein expression by indole-3-carbinol. *Cancer Lett* 2003, 189(2):167-173.
21. Christensen JG, LeBlanc GA: Reversal of multidrug resistance in vivo by dietary administration of the phytochemical indole-3-carbinol. *Cancer Res* 1996, 56(3):574-581.
22. Tanimura H, Kohno K, Sato S, Uchiumi T, Miyazaki M, Kobayashi M, Kuwano M: The Human Multidrug Resistance-1 Promoter Has an Element That Responds to Serum Starvation. *Biochem Biophys Res Commun* 1992, 183(2):917-924.
23. Sarkar FH, Rahman KMW, Li YW: Bax translocation to mitochondria is an important event in inducing apoptotic cell death by indole-3-carbinol (I3C) treatment of breast cancer cells. *J Nutr* 2003, 133(7):2434s-2439s.
24. Zhang XX, Malejka-Giganti D: Effects of treatment of rats with indole-3-carbinol on apoptosis in the mammary gland and mammary adenocarcinomas. *Anticancer Res* 2003, 23(3b):2473-2479.
25. Rahman KMW, Aranha O, Glazyrin A, Chinni SR, Sarkar FH: Translocation of Bax to mitochondria induces apoptotic cell death in Indole-3-carbinol (I3C) treated breast cancer cells. *Oncogene* 2000, 19(50):5764-5771.
26. Yuan F, Chen DZ, Liu K, Sepkovic DW, Bradlow HL, Auborn K: Anti-estrogenic activities of indole-3-carbinol in cervical cells: Implication for prevention of cervical cancer. *Anticancer Res* 1999, 19(3a):1673-1680.
27. Bell MC, Crowley-Nowick P, Bradlow HL, Sepkovic DW, Schmidt-Grimminger D, Howell P, Mayeaux EJ, Tucker A, Turbat-Herrera EA, Mathis JM: Placebo-controlled trial of indole-3-carbinol in the treatment of CIN. *Gynecol Oncol* 2000, 78(2):123-129.
28. Horn TL, Reichert MA, Bliss RL, Malejka-Giganti D: Modulations of P450 mRNA in liver and mammary gland and P450 activities and metabolism of estrogen in liver by treatment of rats with indole-3-carbinol. *Biochem Pharmacol* 2002, 64(3):393-404.

29. Chen I, McDougal A, Wang F, Safe S: Aryl hydrocarbon receptor-mediated antiestrogenic and antitumorigenic activity of diindolylmethane. *Carcinogenesis* 1998, 19(9):1631-1639.
30. Cui X, Palamanda J, Norton L, Thomas A, Lau YY, White RE, Cheng KC: A high-throughput cell-based reporter gene system for measurement of CYP1A1 induction. *J Pharmacol Toxicol* 2002, 47(3):143-151.
31. Chen I, Safe S, Bjeldanes L: Indole-3-carbinol and diindolylmethane as aryl hydrocarbon (Ah) receptor agonists and antagonists in T47D human breast cancer cells. *Biochem Pharmacol* 1996, 51(8):1069-1076.
32. He YH, Friesen MD, Ruch RJ, Schut HAJ: Indole-3-carbinol as a chemopreventive agent in 2-amino-1-methyl-6-phenylimidazo[4, 5-b]pyridine (PhIP) carcinogenesis: Inhibition of PhIP-DNA adduct formation, acceleration of PhIP metabolism, and induction of cytochrome P450 in female F344 rats. *Food Chem Toxicol* 2000, 38(1):15-23.
33. Ochiai M, Imai H, Sugimura T, Nagao M, Nakagama H: Induction of intestinal tumors and lymphomas in C57BL/6N mice by a food-borne carcinogen, 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine. *Jpn J Cancer Res* 2002, 93(5):478-483.
34. Kitamura Y, Yamagishi M, Okazaki K, Son HY, Imazawa T, Nishikawa A, Iwata T, Yamauchi Y, Kasai M, Tsutsumi K *et al*: Lack of significant inhibitory effects of a plant lignan tracheloside on 2-amino-1-methyl-6-phenylimidazo [4,5-b]pyridine (PhIP)-induced mammary carcinogenesis in female Sprague-Dawley rats. *Cancer Lett* 2003, 200(2):133-139.
35. Bonnesen C, Eggleston IM, Hayes JD: Dietary indoles and isothiocyanates that are generated from cruciferous vegetables can both stimulate apoptosis and confer protection against DNA damage in human colon cell lines. *Cancer Res* 2001, 61(16):6120-6130.
36. Carter TH, Liu K, Ralph W, Chen DZ, Qi M, Fan SJ, Yuan F, Rosen EM, Auburn KJ: Diindolyl methane alters gene expression in human keratinocytes in vitro. *J Nutr* 2002, 132(11):3314-3324.

Chapter 1

Indole-3-carbinol (I3C) analogues are potent small molecule inhibitors of NEDD4-1 ubiquitin ligase activity that disrupt proliferation of human melanoma cells

Abstract

The HECT domain-containing E3 ubiquitin ligase NEDD4-1 (Neural precursor cell Expressed Developmentally Down regulated gene 4-1) is frequently overexpressed in human cancers and displays oncogenic-like properties through the ubiquitin-dependent regulation of multiple protein substrates. However, little is known about small molecule enzymatic inhibitors of HECT domain-containing ubiquitin ligases. We now demonstrate that indole-3-carbinol (I3C), a natural anti-cancer phytochemical derived from cruciferous vegetables such as cabbage and broccoli, represents a new chemical scaffold for the generation of small molecule enzymatic inhibitors of NEDD4-1. Using *in vitro* ubiquitination assays, I3C, its stable synthetic derivative 1-benzyl-I3C and five novel synthetic analogues were shown to directly inhibit NEDD4-1 ubiquitination activity. Compared to I3C, which has an IC₅₀ of 280 μM, 1-benzyl-I3C was a significantly more potent NEDD4-1 enzymatic inhibitor with an IC₅₀ of 12 μM. Compounds 2242 and 2243, the two indolecarbinol analogues with added methyl groups that results in a more nucleophilic benzene ring π system, further enhanced potency with IC₅₀s of 2.7 μM and 7.6 μM, respectively. Protein thermal shift assays that assess small ligand binding, in combination with *in silico* binding simulations with the crystallographic structure of NEDD4-1, showed that each of the indolecarbinol compounds bind to the catalytic HECT domain of NEDD4-1. The indolecarbinol compounds inhibited human melanoma cell proliferation in a manner that generally correlated with their effectiveness as NEDD4-1 enzymatic inhibitors. Taken together, we propose that I3C analogues represent a novel set of anti-cancer compounds for treatment of human melanomas and other cancers that express indolecarbinol-sensitive target enzymes.

Introduction

E3 ubiquitin ligases play critical roles in controlling numerous cellular pathways by catalyzing the attachment of ubiquitin protein chains to specific lysine residues of protein substrates that selectively signal protein degradation, localization, endocytosis, subcellular trafficking and/or activity. The diversity and number of the ubiquitination products determine the outcomes for individual protein substrates [1]. Of the several hundred known E3 ubiquitin ligases, approximately 30 have a distinguishing HECT (Homologous to E6-AP C-Terminus) domain in their structure that includes the catalytic site for ubiquitin protein ligation [2]. The NEDD4-1 (Neural precursor cell Expressed Developmentally Down regulated gene 4-1) ubiquitin ligase is one of nine members of the NEDD4 gene family of HECT domain-containing E3 ubiquitin ligases [3]. The structure of NEDD4-1 includes an N-terminal C2 calcium-dependent lipid binding domain involved in membrane interactions, a central domain that directs protein-protein interactions through several WW (double tyrosine residue) motifs and the C-terminal catalytic HECT domain [4]. NEDD4-1 is frequently overexpressed in different types of human cancers [5] and can exert effects on the development, progression, survival and proliferation of human cancers [5] through ubiquitination of a variety of downstream protein substrates including those linked to growth factor receptor signaling and tumor suppressor activities [6]. Based on these observations, the targeting NEDD4-1 has been proposed as a potential therapeutic strategy for treatment of human cancers [5], however, to date relatively little is known about small molecule enzymatic inhibitors of NEDD4-1 or any other HECT domain containing E3 ubiquitin ligases.

We have been investigating the anti-cancer properties of indole-3-carbinol (I3C), which is derived by hydrolysis from glucobrassicin produced in cruciferous vegetables of the *Brassica* genus such as cabbage, broccoli and Brussels sprouts. I3C is a natural indolecarbinol phytochemical that can activate multiple anti-proliferative cascades in several types of human cancers, including melanoma, breast, prostate, lung, colon, leukemia, and cervical cancer [7-17]. The cellular pathways affected by I3C involve transcriptional, enzymatic, metabolic and cell signaling processes that can lead to an arrest of cell cycle progression, apoptosis, loss of cell survival signaling, down-regulation of cancer cell migration, modulation of hormone receptor signaling, and inhibition of tumor growth [18-30]. A key advance in understanding the I3C anti-proliferative mechanism is our discovery that I3C triggers distinct and overlapping sets of anti-proliferative signaling events by direct interactions with specific target proteins [31-34]. The serine protease elastase was the first identified I3C target protein, with I3C functioning as a direct noncompetitive inhibitor of elastase enzymatic activity [31]. In human breast cancer cells, cell surface-associated elastase cleaves the CD40 member of the tumor necrosis factor receptor to stimulate a pro-proliferative cascade resulting in the activation of the NFkB [35-38]. The I3C inhibition of elastase activity prevents CD40 cleavage and shifts the down stream signaling in a manner that disrupts NFkB-dependent cell cycle and proliferative gene expression [32]. These observations suggest that I3C may mediate its anti-proliferative effects in other types of cancer cells by directly interacting with other classes of target proteins with enzymatic activities.

We recently documented that I3C arrests the proliferation and induces apoptosis of human melanoma cells [34, 39]. The most sensitive melanoma cells expressed an oncogenic form of BRAF and the wild type PTEN (phosphatase and tensin homologue detected on chromosome 10), which is a lipid phosphatase that acts a tumor suppressor protein [40-42]. Concomitant with the proliferative arrest, I3C down-regulated the NEDD4-1 mediated ubiquitination of wild type PTEN, which resulted in higher levels of this tumor suppressor protein [34]. Furthermore, siRNA knock down of NEDD4-1 mimicked the I3C stabilization of PTEN protein and isothermal titration calorimetry demonstrated that I3C directly binds to the purified HECT catalytic domain of NEDD4-1 [34]. Taken together, these observations suggest the possibility that I3C may act as a direct inhibitor of NEDD4-1 ubiquitin ligase activity.

In silico binding simulations between I3C and the crystallographic structure of NEDD4-1, using shape and electrostatics as restrictive parameters, predicted that I3C binds to a small subdomain of the N-lobe of the catalytic HECT domain, but not to the C2 and WW domains [34]. These modeling results suggest that I3C could provide a chemical starting point to develop highly potent anti-cancer compounds based on enhanced target protein interactions. 1-benzy-indole-3-carbinol (1-benzyl-I3C), which has a benzyl moiety attached to the indole ring nitrogen, was synthesized and shown to have a more potent anti-proliferative response in human breast cancer cells compared to I3C and is a more potent elastase inhibitor [33]. The enhanced efficacy of 1-benzyl-I3C is due to significantly increased hydrophobic character and greater stability compared to I3C due to the steric effect of the bulky 1-benzyl substituent that hampers oligomerization of the 1-benzyl-indol-3-yl methyl cation [32]. We now demonstrate that I3C represents a chemical scaffold to develop highly potent small molecule inhibitors of NEDD4-1 ubiquitin ligase activity. Furthermore, we show that the indolecarbinol compounds bind to NEDD4-1 and robustly inhibit the proliferation of human melanoma cells with efficacies that generally correlated with their effectiveness in inhibiting NEDD4-1 activity, which implicates their potential use as a novel set of anti-cancer compounds for treatment of human melanomas and other cancers.

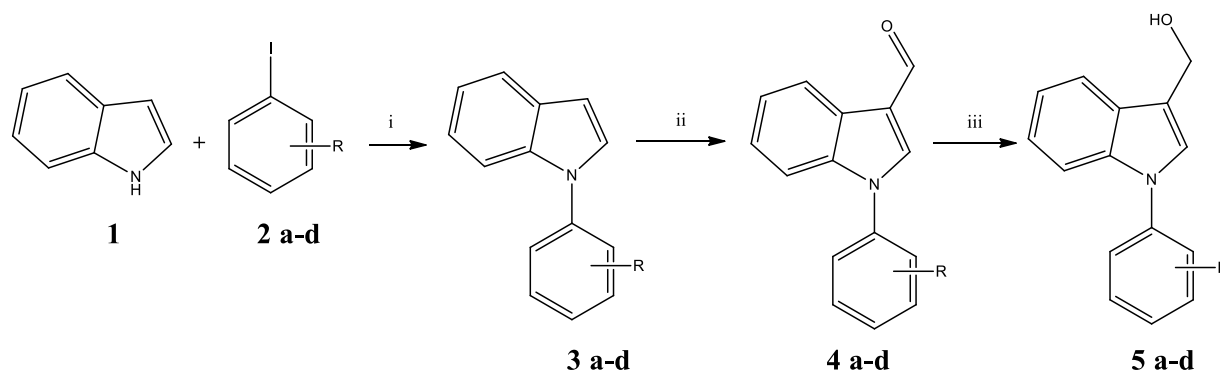
Materials and Methods

Synthesis and characterization of I3C analogues LCTA-2160, LCTA-2242, LCTA-2243, LCTA-2244 and LCTA-2163

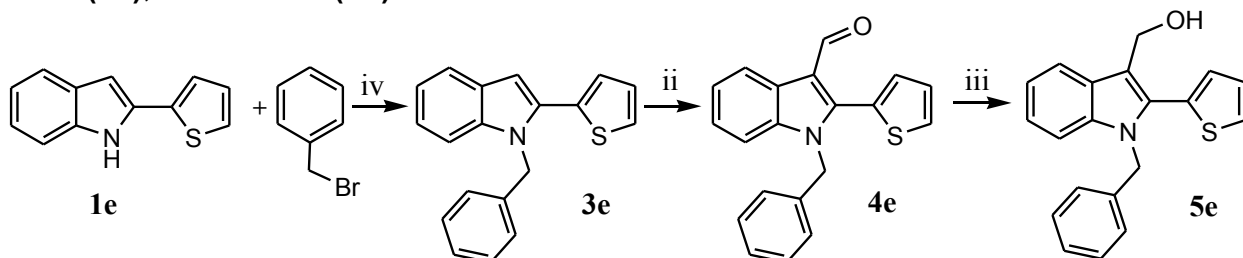
General information

NMR spectra were obtained using a Varian VXR-400 instrument operated at 400 MHz (^1H NMR). TLC was performed on Silica Gel F254 plates (Merck) using a systems A (petroleum ether), B (petroleum ether-ethyl acetate 10:1) and C (petroleum ether-ethyl acetate 1:1). Column chromatography was performed on Silica Gel Merck 60. Melting points were determined on a Buchi SMP-20 apparatus and are uncorrected. High-resolution mass spectra were recorded with electron-spray ionization on a Bruker Daltonics microOTOF-QII instrument. HPLC was performed using Shimadzu LC-20AD (Japan), column Kromasil (Sweden) 100-C18, size 5 μm , 4.6x250 mm, loop 20 μkl , elution: (I) $\text{H}_2\text{O} \rightarrow \text{CH}_3\text{CN}$ gradient from 10% to 95% CH_3CN at 20 min. (II)- H_3PO_4 0.01 M in water $\rightarrow \text{CH}_3\text{CN}$ gradient from 10% to 95% CH_3CN at 20 min. All solutions were dried over Na_2SO_4 and evaporated at reduced pressure on a Buchi-R200 rotary evaporator at the temperature below 35°C . All products were vacuum dried at room temperature. All solvents, chemicals and reagents were obtained commercially from SigmaAldrich or Merck and used without purifications.

Compounds **3** (LCTA-2160; LCTA-2242; LCTA-2243; LCTA-2244) were synthesized as shown in Scheme 1. Compound LCTA-2163 was prepared by Scheme 2 starting from 2-(thiophen-2-yl)-indole.



i) CuI ; L-Proline; K_2CO_3 ; DMSO; ii) POCl_3 ; DMF; iii) NaBH_4 ; ethanol; iv) DMSO, KOH
Scheme 1. Synthesis of compounds **5 a-d**: LCTA-2160 (**5a**); LCTA-2242 (**5b**); LCTA-2243 (**5c**); LCTA-2244 (**5d**)



Scheme 2. Synthesis of LCTA-2163 (**5e**)

1-Phenyl-1*H*-indole (**3a**). As previously described [43], for N-phenylation of indoles a mixture of indole 1 (0.585 g, 5 mmol), iodobenzene **2a** (1.26 g, 6.2 mmol), K₂CO₃ (1.38 g, 10 mmol), L-proline (0.23 g, 2 mmol), CuI (0.19 g, 1 mmol) in 20 ml of DMSO was heated at 130°C for 3h. The cooled mixture was diluted by water (100 ml) and extracted by diethyl ether (30 ml). Organic layer was separated, washed by brine, dried over Na₂SO₄ and concentrated in vacuo. The residual oil was loaded on a silica gel column and eluted with petroleum ether to afford the **3a** (0.48 g, 50% yield) as yellow oil. R_f 0.27 (A)

The compounds **3b**, **3c**, **3d** were similarly prepared from indole and the appropriated iodarenes : 1-*p*-Tolyl-1*H*-indole (**3b**), 40% yield, yellow oil, R_f 0.33 (A) ; 1-*o*-Tolyl-1*H*-indole (**3c**), 50% yield, yellow oil, R_f 0.33 (A); 1-(3,4-Dimethoxyphenyl)-1*H*-indole (**3d**), 48% yield, yellow oil, R_f 0.15 (A).

1-Benzyl-2-(thiophen-2-yl)-1*H*-indole (**3e**). To a stirred solution of 2-(thiophen-2-yl)-1*H*-indole (**1e**) (prepared by cyclization of 2-acetylthiophene phenylhydrazone as previously described [44] (1g, 5 mmol) in DMSO (30 ml) finely ground KOH (0.84 g, 15 mmol) was added. After 30 min incubation benzyl bromide (**2e**) (1g, 6 mmol) was added and reaction mixture allowed to stir at 20°C for 12 h. After that mixture was diluted by water (100 ml) and extracted by diethyl ether (30 ml). Organic layer was separated, washed by brine, dried over Na₂SO₄ and concentrated in vacuo. The residual oil was loaded on a silica gel column and eluted with petroleum ether to afford the **3e** (1.15 g, 80% yield) as yellow oil.

1-Phenyl-1*H*-indole-3-carbaldehyde (**4a**). To a stirred solution of **3a** (0.48 g, 2.5 mmol) in 20 ml DMF at 0°C was added dropwise POCl₃ (0.46 g, 3 mmol). After complete addition the temp was increased to 20°C and kept at this temp for 12 h. The solution was diluted by water (200 ml) and extracted by diethyl ether (50 ml). Aqueous layer was separated and treated by 5N NaOH to pH 10. The oil emulsion obtained was extracted by diethyl ether (50 ml), washed by brine, dried over Na₂SO₄, concentrated in vacuo and crystallized from methanol to produce **4a** as colorless crystals (0.41 g, 75% yield), Mp 70-72°C, R_f 0.61 (B).

The compounds **4b-e** were similarly prepared from the appropriated indoles **3**: 1-*p*-Tolyl-1*H*-indole-3-carbaldehyde (**4b**), 80% yield; colorless crystals, Mp 112-115°C, R_f 0.65 (B); 1-*o*-Tolyl-1*H*-indole-3-carbaldehyde (**4c**), 65% yield, colorless crystals, Mp 74-76°C, R_f 0.65 (B); 1-(3,4-Dimethoxyphenyl)-1*H*-indole-3-carbaldehyde (**4d**), 60% yield, colorless crystals, Mp 122-124°C, R_f 0.44 (B); 1-Benzyl-2-(thiophen-2-yl)-1*H*-indole-3-carbaldehyde (**4e**), 60% yield; yellow crystals, Mp 205-207°C, R_f 0.62 (B)

(1-Phenyl-1*H*-indol-3-yl)methanol (**5a**) LCTA-2160. NaBH₄ (0.1 g, 2.7 mmol) was added to solution of **4a** (0.41 g; 1.8 mmol) in ethanol (25 mL) and stirred for 2 h at room temperature. The solvent was removed in vacuo at the temperature below 25°C, the residue obtained was solved in water (100 mL) and the solution was extracted with

diethyl ether (3 × 20 mL). The combined organic phases were washed by brine, dried over Na₂SO₄ and concentrated in vacuo at the temperature below 25°C to produce **5a** as colorless oil (0.28 g, 70% yield).

δ (ppm) [DMSO-d₆]: 4.70 (2H, d, J 4Hz, 3-CH₂), 4.92 (1H, t, J 4Hz, OH), 7.14 (2H, m, H-5,6), 7.40 (7H, m, H-2, 4, 2', 3',4',5',6'), 7.69 (1H, d, J 8Hz, H-7). HRMS [M+Na]⁺, found 246.0865, [C₁₅H₁₃NO+Na]⁺ requires 246,0895. HPLC: (I) Rt 10.51 min, purity 96%. R_f 0.65 (C).

The compounds **5 b-e** were similarly prepared from the appropriated indole-3-carbaldehydes **4**: (1-*p*-Tolyl-1*H*-indol-3-yl)methanol (**5b**), yield 70%, colorless crystals, Mp 72-75°C. δ (ppm) [DMSO-d₆]: 2.37 (3H, s, Ar-CH₃), 4.71 (2H, d, J 4Hz, 3-CH₂), 4.94 (1H, t, J 4Hz, OH), 7.16 (2H, m, H-5,6), 7.41 (6H, m, H-2, 4, 2', 3',5',6'), 7.71 (1H, d, J 8Hz, H-7). HRMS [M+Na]⁺, found 260.1020, [C₁₆H₁₅NO+Na]⁺ requires 260,1051. HPLC: (II) Rt 12.78 min, purity 99%. R_f 0.58 (C).

(1-*o*-Tolyl-1*H*-indol-3-yl)methanol (**5c**): yield 65%, colorless crystals, Mp 43-45°C. δ (ppm) [DMSO-d₆]: 2.01 (3H, s, Ar-CH₃), 4.72 (2H, d, J 4Hz, 3-CH₂), 4.95 (1H, t, J 4Hz, OH), 6.91 (1H, d, J 8Hz, H-4), 7.10 (2H, m, H-5,6), 7.36 (5H, m, H-2, 3',4',5',6'), 7.72 (1H, d, J 4Hz, H-7). HRMS [M+Na]⁺, found 260.1024, [C₁₆H₁₅NO+Na]⁺ requires 260,1051. HPLC: (II) Rt 12.03 min, purity 95%. R_f 0.60 (C).

(1-(3,4-Dimethoxyphenyl)-1*H*-indol-3-yl)methanol (**5d**): yield 65%, colorless crystals, Mp 96-98°C. δ (ppm) [DMSO-d₆]: 3.72 (3H, s, OCH₃), 3.84 (3H, s, OCH₃), 4.68 (2H, d, J 4Hz, 3-CH₂), 4.90 (1H, t, J 4Hz, OH), 6.65 (1H, d, J 12Hz, 5'-H), 6.80 (1H,s, H-2'), 6.98 (1H, d, J 8Hz, H-4), 7.07 (2H, m, H-5,6), 7.24 (2H,m, H-2,6'). HRMS [M+Na]⁺, found 306.1084, [C₁₇H₁₇NO₃+Na]⁺ requires 306,1106. HPLC: (II) Rt 10.06 min, purity 99%. R_f 0.44 (C).

(1-Benzyl-2-(thiophen-2-yl)-1*H*-indol-3-yl)methanol (**5e**): yield 60%, brownish oil. δ (ppm) [DMSO-d₆]: 4.60 (2H, d, J 8Hz, 3-CH₂), 4.85 (1H, t, J 8Hz, OH), 5.42 (2H, s, CH₂-Ph), 6.90 (1H, d, J 8Hz, H-4), 7.23 (10H, m, H-5,6 of indole, 2-5 of Ph and 3-5 of thiophene), 7.74 (1H, d, J 8Hz, H-7). HRMS [M+Na]⁺, found 342.0946, [C₁₇H₁₇NO₃+Na]⁺ requires 342.0923. HPLC: (I) Rt 15.28 min, purity 95%. R_f 0.66 (C).

Biochemical Studies

In Vitro Protein Thermal Shift Assays

All reactions were set up in 20 μL reactions in 96 well plates with total human NEDD4-1^{Hect} protein concentrations of 2.5 μM and 10x SYPRO orange dye (Invitrogen) [45]. The NEDD4-1 HECT domain was pre-incubated for 5 minutes with either I3C, 2244, 2160, or 2163 at 250 μM while concentrations of 25 μM was used for compounds 1-benzyl-I3C, 2242, and 2243. Thermal melting experiments were performed on the Viiia7 instrument (Life Technologies) melt curve program with a heat ramp rate of 0.05°C/s and a temperature range beginning at 25°C and ending at 95°C. Melting temperatures

were analyzed with Protein Thermal Shift Software (Life Technologies) to identify the midpoint of the transition with an analysis of the first derivative. All experiments were performed in triplicate.

In Vitro NEDD4-1 Enzymatic Assays

Purified NEDD4-1 protein and the indicated concentrations of indolecarbinol compounds were pre-incubated for 1 hour. The reaction mixtures were then transferred to an ELISA microplate where the remaining components of the ubiquitination reaction (E1, E2, and ubiquitin) were added. This plate captures poly-ubiquitin chains formed in an E3-dependent reaction and was initiated with ATP at room temperature for 1 hour. The plates were washed 3 times and incubated with biotin-linked ubiquitin antibody and then subsequently incubated with streptavidin-HRP. The E3LITE Customizable Ubiquitin Ligase Kit from LifeSensors (UC101) is an ELISA-like highly sensitive assay that relies on enhanced chemiluminescence to quantitatively measure ubiquitin E2 conjugation and E3 ligase activity. No additional non-native tagging or labeling of ubiquitin was required. The assay is based on the ability of an ubiquitin-binding domain to preferentially bind polyubiquitin relative to monoubiquitin, thus measuring activity of E3 ligase that builds ubiquitin chains in a dose-dependent manner [53-56].

Statistical analysis

GraphPad Prism 6.0 software (GraphPad Software Inc., San Diego, CA, USA) was used to perform data analysis in all experiments. Data are presented as mean \pm SD from 3 independent experiments. Statistical analysis comparing the ubiquitin ligase activity for no added NEDD4-1 and untreated NEDD4-1 was analyzed by an unpaired t-test. Significance was assigned at a * p value < 0.05 ; ** $p < 0.01$; *** $p < 0.001$.

Biological Studies

Melanoma cell culture and proliferation assay

The human G-361 melanoma cell line was purchased from American Type Culture Collection (ATCC) (Manassas, VA), and was authenticated according to the ATCC guidelines. Cells were cultured in Modified McCoy's 5A cell media supplemented with 10% fetal bovine serum (Gemini Bio Products, West Sacramento, CA), 2 mM L-glutamine, and 2.5 ml of 10,000 U/ml penicillin/streptomycin mixture (Gibco, Life Technologies, Carlsbad, CA). Melanoma cells were seeded on a 48-well plate in triplicates and upon 80-90% confluency were treated with I3C, 1-benzyl-I3C and the 1-benzyl-I3C analogues for 48 hours. Subsequently, inhibition of proliferation was assessed using the Dojindo Cell counting Kit -8 as per the protocol in the user's manual. Briefly, 50 μ l of the CCK-8 solution was added to each well along with 450 μ l media and incubated for 2-3 hours. The absorbance was read at 450nm and % inhibition was calculated for each condition standardizing DMSO to zero.

Western blot analysis

Western Blot analyses of samples electrophoretically fractionated on 8-10% acrylamide gels were carried out as previously described [32, 33]. ECL Lightening reagents were used to visualize the primary antibody bound protein bands in nitrocellulose membranes and the results captured on ECL Autoradiography Film (GE Healthcare, Piscataway, NJ). The western blots employed the following primary antibodies: MITF-M (Thermo-Scientific, Waltham, MA), and HSP 90 (BD Bio-sciences, San Jose, CA).

In Silico Computational Methods

Protein Structure Preparation

The coordinates for the NEDD4-1 HECT domain protein structure were obtained from the RCSB Protein Data Bank (PDB). 5C91 [46] was used to examine the interactions between the NEDD4-1 HECT domain and indolecarbinol compounds. The protein structure was prepared using the Schrödinger's Protein Preparation Wizard module (Schrödinger Suite 2015-4 Protein Preparation Wizard; Epik version 3.4, Schrödinger, LLC, New York, NY, 2015; Impact version 6.9, Schrödinger, LLC, New York, NY, 2015; Prime version 4.2, Schrödinger, LLC, New York, NY, 2015.) [47]. Hydrogen atoms were added and the side chain structures of Gln and Asn were flipped in order to yield the maximum degree of hydrogen bond interactions. The protein was subsequently minimized using the OPLS force field in the MacroModel module in Schrödinger (MacroModel, version 11.0, Schrödinger, LLC, New York, NY, 2015.) with backbone atoms fixed.

Induced fit docking methods

The NEDD4-1 HECT domain protein structure (PDB accession number: 5C91) was applied with the induced-fit docking (IFD) method in the Schrödinger software suite (Small-Molecule Drug Discovery Suite 2015-4: Schrödinger Suite 2015-4 Induced Fit Docking protocol; Glide version 6.9, Schrödinger, LLC, New York, NY, 2015; Prime version 4.2, Schrödinger, LLC, New York, NY, 2015) [48-50]. I3C, 1-benzyl-I3C, and their corresponding analogs were prepared using LigPrep (version 3.6, Schrödinger, LLC, New York, NY, 2015) and were optimized with the OPLS force field in the MacroModel module in Schrödinger. Ligands were docked to the rigid protein using the soften-potential docking in the Glide program with the van der Waals radii scaling of 0.5 for the protein. The resulting top 20 poses of ligands were used to sample the protein plasticity using the Prime program (version 4.2, Schrödinger, LLC, New York, NY, 2015) in the Schrödinger suite [51, 52]. Flexibility of the protein was taken under consideration by having residues that were at least one atom within 5 Å of any of the 20 ligand poses subjected to a conformational search and minimization. Residues outside the zone were fixed. The new 20 protein conformations were subsequently utilized for redocking. In the final stage, the ligand poses were redocked using Glide SP into structures within 30.0 kcal/mol of the top 20 structures. The binding affinity of each complex was reported in the GlideScore. A more negative GlideScore is indicative of a more favorable binding interaction. Figures

were made using the Pymol program (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.)

Results

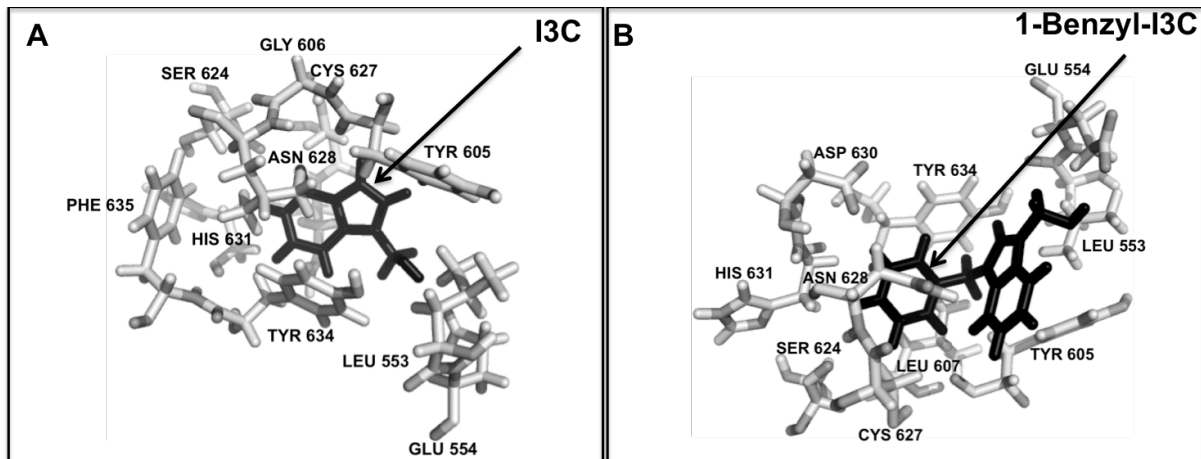
In silico computational modeling of molecular interactions and direct binding of I3C and 1-Benzyl-I3C to the HECT domain of the ubiquitin ligase NEDD4-1

The natural compound I3C and its highly potent synthetic derivative 1-benzyl-I3C [33, 57], which has a benzyl moiety attached to the indole ring nitrogen, were initially used to assess the potential interactions of indolecarbinol compounds and their effects on the enzymatic activity of NEDD4-1, a HECT domain-containing ubiquitin ligase [58]. *In silico* molecular binding simulations using the Schrödinger Induced Fit Docking protocol were performed to evaluate and compare potential predicted binding sites within NEDD4-1^{HECT} for I3C and 1-benzyl-I3C. Proteins recognize ligands based on 3D structure and electrostatic complementarities, and the Induced Fit Docking protocol was implemented for binding simulations because this program incorporates protein flexibility within a tested ligand-protein complex to account for the dependence of active site geometry upon conformational changes generated by a bound ligand. The previously reported structure of NEDD4-1^{HECT} co-crystallized with an indole-based small molecule (PDB ID: 5C91) [46] was used as the foundation structure for docking I3C or 1-benzyl-I3C to NEDD4-1^{HECT}. Predicted interactions with specific amino acid residues of the NEDD4-1^{HECT} domain are within 4 Å of either I3C or 1-benzyl-I3C.

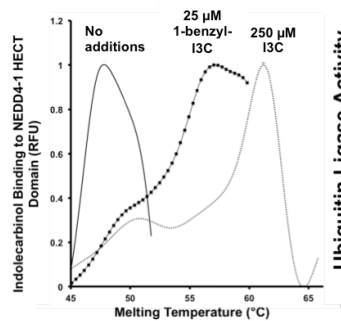
Structural analysis shows that the NEDD4-1^{HECT} domain is a bi-lobed domain consisting of an N-terminal N-lobe that interacts with the E2 ubiquitin ligase and a C-terminal C lobe containing the catalytic cysteine and which is free to rotate around the flexible hinge that tethers it to the N lobe [58]. As shown in Figure 1.1A, computer modeling predicts that the hydrophobic indole core of I3C is oriented toward a pocket of the N-lobe and is enveloped by residues Tyr 634, Try 605, Leu 607, Asn 628, His 631, Cys 627 and Leu 553. In particular, the indole moiety of I3C is predicted to make aromatic edge-to-face interactions with residues Tyr 605 and Tyr 634 (Figure 1A). Notably, Tyr 605 is an essential residue for noncovalent ubiquitin binding and poly-ubiquitination chain elongation by NEDD4-1 [46, 58]. Similarly to I3C, the indole moiety in 1-benzyl-I3C makes edge-to-face interactions with residues Tyr 605 and Tyr 634 and is securely positioned against residues Leu 553 and Glu 554. The additional phenyl moiety points toward a complementary cavity formed by Leu 607, His 631, Ser 624, Cys 627, Asn 628, and Asp 630 which may act to further stabilize and enhance 1-benzyl-I3C interactions with NEDD4-1 (Figure 1.1B).

Protein Thermal Shift assays were used to directly test whether I3C and 1-benzyl-I3C bind to the purified bacterially-synthesized HECT Domain of NEDD4-1 (NEDD4-1^{HECT}). Indolecarbinol-mediated alterations in thermostability of NEDD4-1^{HECT} are indicative of direct interactions between each potential ligand and the protein. Purified NEDD4-1^{HECT} at a concentration of 2.5 µM was incubated with 250 µM I3C, 25 µM 1-benzyl-I3C or with the vehicle control and each mixture was then treated with Sypro orange dye, which binds to hydrophobic regions of proteins and emits a fluorescent signal as the protein unfolds [45]. A 10-fold lower concentrations of 1-benzyl-I3C compared to I3C was used because this I3C derivative was previously shown to be significantly more

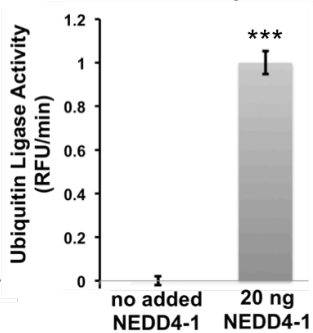
potent than I3C in its ability to interact with human neutrophil elastase [33, 57]. The binding reactions were slowly heated from 25°C to 95°C, and the unfolding of the HECT domains continuously recorded from the fluorescent signals. Conditions that thermally stabilize the protein will typically shift the melting temperature upwards by 4° C to greater than 20° C. Generally as an empirical estimate, a greater than 4° C shift in melting temperature corresponds to binding of a ligand with a K_d of approximately <1 μ M [59], although the magnitude of the temperature shift does not necessarily correlate with ligand affinity. As shown in Figure 1C, in the absence of added indolecarbinol compounds, the first derivative NEDD4-1^{HECT} melting profile yielded an apparent T_m of 48°C. Exposure to 250 μ M I3C or 25 μ M 1-benzyl-I3C stabilized the NEDD4-1^{HECT} secondary structure based on concomitant increases in melting temperatures or ΔT_m values of 13.3 °C for I3C and 10.2 °C for 1-benzyl-I3C. These results demonstrate that both indolecarbinol compounds directly bind to NEDD4-1^{HECT}.



C Protein Thermal Shift Assay



D Ubiquitin Ligase Activity



E

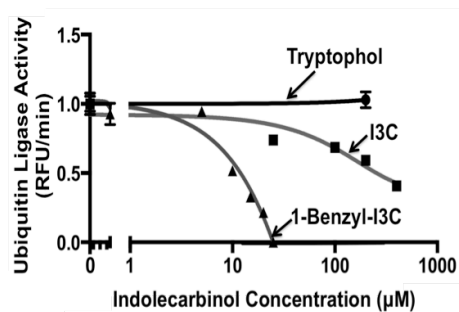


Figure 1.1. *In silico* computational modeling of predicted molecular interactions of I3C and 1-benzyl-I3C with human NEDD4-1 HECT domain. (A & B) The 3D crystal structure of NEDD4-1 HECT domain (PDB accession number: 5C91) and the corresponding indolecarbinol compounds (shown in 1A: I3C; 1B: 1-benzyl-I3C) were processed with the Protein Preparation Wizard in the Schrödinger suite and were evaluated using the Schrödinger induced fit docking (IFD) protocol to represent conformational changes in the ligand binding site stimulated by ligand binding. Close contact residues are labeled and are within 4 Å of each indolecarbinol compound. Models were visualized using PyMol. (C) Binding of I3C or 1-benzyl-I3C to purified NEDD4-1 HECT domain was evaluated via protein thermal shift assays. The first derivative melt profiles show melting temperatures of NEDD4-1 HECT domain in the absence or presence of either 25 µM 1-benzyl-I3C or 250 µM I3C. Data is expressed as average of triplicate experiments. (D) Quantification of purified NEDD4-1 E3 ligase activity was determined via an ELISA based polyubiquitination assay as described in the Methods. Assay functionality was validated by assessing polyubiquitination activity in the absence or presence of NEDD4-1. (E) Purified NEDD4-1 was pre-incubated with the indicated concentrations of I3C, 1-benzyl-I3C, or tryptophol for 1 hr and subsequently transferred to an ELISA microplate where the remaining contents of the reaction (E1, E2, and ubiquitin) were added. The *in vitro* formation of poly-ubiquitin chains was analyzed in an ELISA microplate and the enzymatic reactions were initiated with ATP. The washed plates were first incubated with biotin-linked ubiquitin antibody and then incubated with streptavidin-HRP to monitor NEDD4-1 auto-ubiquitination as the readout for NEDD4-1 activity. Enhanced chemiluminescent (ECL) reagent was then added to each well and incubated for 5 min. The relative light units (RLUs) were recorded using an LMax II/II384 microplate luminometer. NEDD4-1 enzymatic activity was standardized to the untreated NEDD4-1 protein and the data are expressed as the average of triplicate ± SD.

I3C and enhanced potency 1-benzyl-I3C represent a new class of enzymatic inhibitors of NEDD4-1 ubiquitination activity

Because the HECT domain of NEDD4-1 contains the catalytic site that mediates the ubiquitin ligase activity of this enzyme, the predicted locations of the *in silico* defined interaction sites of I3C and 1-benzyl-I3C and the fact that both indolecarbinol compounds directly bind to the HECT domain suggested the possibility that these indolecarbinol compounds could alter NEDD4-1 enzymatic activity. Relatively little is known about small molecular inhibitors of this important ubiquitin ligase. Therefore, an *in vitro* polyubiquitylation assay was optimized to test the potential effects and efficacy of I3C and 1-benzyl-I3C on NEDD4-1 activity. In order to validate and standardize the assay, ubiquitin ligase activity was initially monitored in the absence or presence of 20 ng of NEDD4-1, which is an E3 ubiquitin ligase. The reaction mixtures were then transferred to an ELISA microplate where the remaining contents of the reaction (E1, E2, and ubiquitin) were added. The plate captures poly-ubiquitin chains formed in an E3-dependent reaction that was initiated with ATP at room temperature for 1 hour. The washed plates were incubated with biotin-linked ubiquitin antibody and then incubated with streptavidin-HRP to monitor NEDD4-1 auto-ubiquitylation as the readout for NEDD4-1 activity. As shown in Figure 1D, no auto-ubiquitination activity was observed in the absence of NEDD4-1 whereas in the presence of 20 ng NEDD4-1, ubiquitin ligase activity was high at a relative rate of approximately 1.02 RFU/min confirming the functionality of the assay.

To directly determine the potential effects of indolecarbinol compounds on NEDD4-1 ubiquitin ligase activity, a range of concentrations of either I3C or 1-benzyl-I3C were tested in the *in vitro* polyubiquitylation assay in comparison to the biologically inactive I3C analog, tryptophol, which contains an ethanol group attached to the C3 ring carbon instead of a hydroxy methyl group and acts as a negative control. The reaction mixtures were preincubated with purified human NEDD4-1 protein for 1 hour and the formation of poly-ubiquitin chains was analyzed in an ELISA microplate in the presence of an E1 ligase, an E2 ligase and ubiquitin and the enzymatic reactions were initiated with ATP. The washed plates were first incubated with biotin-linked ubiquitin antibody and then incubated with streptavidin-HRP to monitor NEDD4-1 auto-ubiquitination as the readout for NEDD4-1 activity. As shown in Figure 1.1E, a high concentration of tryptophol had no effect on NEDD4-1 auto-ubiquitination activity, whereas I3C inhibited NEDD4-1 ubiquitin ligase activity with a half maximal inhibition IC_{50} of 280 μ M. In comparison to the natural indolecarbinol I3C, 1-benzyl-I3C displayed significantly more potent *in vitro* inhibitory effects on NEDD4-1 ubiquitin ligase activity with an IC_{50} of 12 μ M, which represents an approximate 23-fold higher efficacy than I3C (Figure 1.1E). These results establish I3C and the highly potent 1-benzyl-I3C as small molecule inhibitors of NEDD4-1 enzymatic activity and further implicate indolecarbinol compounds as a novel class of inhibitors of HECT-domain ubiquitin ligases.

Anti-proliferative responsiveness of I3C and 1-Benzyl-I3C in human G361 melanoma cells

Before analyzing the effects of an expanded set of indolecarbinol derivatives on NEDD4-1, it was important to determine whether the differences in effectiveness of I3C and 1-benzyl-I3C to inhibit NEDD4-1 enzymatic activity generally reflected the relative efficacies of these indolecarbinol compounds on anti-proliferative responses in melanoma cells. The Microphthalmia Associated Transcription Factor isoform M (MITF-M), a basic helix-loop-helix leucine zipper transcription factor, is highly involved with melanoma cell proliferation, cell cycle control and cell survival through its transcriptionally regulated gene products [60-62]. We recently observed that I3C down regulates MITF-M expression by disruption of oncogenic BRAF signaling through the Brn2 transcriptional regulator of MITF-M [39]. Therefore, human G361 melanoma cells, which display an oncogenic BRAF/wild type PTEN genotype, were treated with a range of concentrations of either I3C or 1-benzyl-I3C for 48 hours and the MITF-M protein levels determined by western blot analysis. As shown in Figure 1.2A (left panel), I3C and 1-benzyl-I3C strongly down regulated MITF-M protein levels, although importantly the quantification of the relative levels of MITF-M protein by densitometry compared to the HSP90 demonstrated that 1-benzyl-I3C displayed significantly enhanced efficacy compared to I3C (Figure 1.2A, right panel).

MITF-M induces expression of several G₁ acting cell cycle genes [63, 64], such as CDK4, CDK2 and cyclin D1, suggesting that the I3C and 1-benzyl-I3C-mediated loss of MITF-M protein would be predicted to cause a G₁ cell cycle arrest of melanoma cells at concentrations indicative of their efficacy to down regulate MITF-M. To test this possibility, G361 melanoma cells were treated with 200 μ M I3C, 10 μ M 1-Benzyl-I3C or with the DMSO vehicle control for 48 hours and the potential cell cycle effects were analyzed by flow cytometry of propidium iodide-stained nuclear DNA. As shown in Figure 2B, indicative of a proliferative state, G361 cells treated with the DMSO vehicle control displayed cells in all phases of the cell cycle including approximately 50.1% cells in the G₁ phase. In contrast, treatment with 200 μ M I3C increased the percent of the percentage of cells with a G₁ DNA content in the overall cell population to 72.4%, which is consistent with a G₁ cell cycle arrest (Figure 1.2B, middle panel). Notably, treatment with 10 μ M 1-benzyl-I3C induced a G₁ cell arrest with 88.2 % of the total cells arrested in the G₁ phase of the cell cycle with relatively small numbers of cells remaining in the S and G₂/M cell cycle phases (Figure 1.2B, right panel). Taken together, our results show that 1-benzyl-I3C exhibits a significantly more effective anti-proliferative response compared to its parental compound I3C that generally reflects the increased efficacy for the inhibition of NEDD4-1 enzymatic activity. Therefore, we analyzed the interactions with NEDD4-1 and anti-proliferative response of additional synthetic indolecarbinol analogues with varying chemical properties.

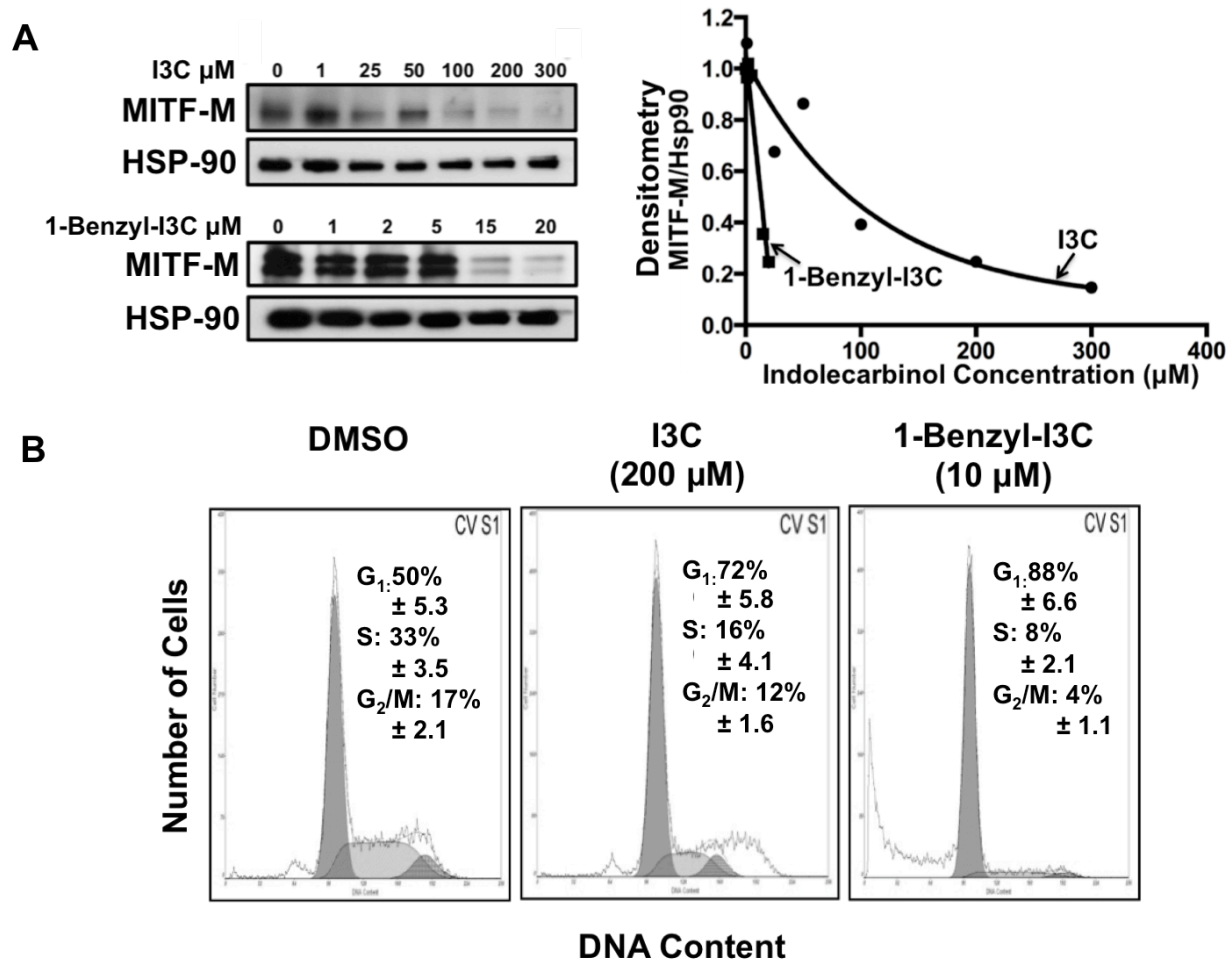


Figure 1.2. Effects of I3C and 1-benzyl-I3C on MITF-M protein levels and cell cycle in G361 human melanoma cells. (A) Western blot analysis of the effects of I3C and 1-benzyl-I3C on protein expression of MITF and in G361 cells. G361 cells were treated for 48 hours with the indicated concentrations of I3C or 1-benzyl-I3C. Cells were lysed, electrophoretically fractionated and western blots probed for MITF-M protein. HSP-90 was used as a loading control (left panel). The relative levels of MITF-M protein under each condition were quantified by densitometry of the western blots and normalized to the HSP-90 loading control protein (right panel). (B) G361 melanoma cells treated for 48 hr with the DMSO vehicle control, 200 μM I3C, or 10 μM 1-benzyl-I3C and the cell population DNA contents were quantified by flow cytometry of propidium stained nuclei. Representative flow histogram of each set of cells is shown in the panel.

Synthesis of I3C and 1-benzyl-I3C derivatives and in silico predicted interactions with the HECT domain of the NEDD4-1 ubiquitin ligase

To understand the molecular interactions between indolecarbinol compounds and the HECT domain of NEDD4-1, indolecarbinol analogues were designed to selectively alter chemical properties but maintain the benzene moiety that enables the 6 local $C^{\delta-}-H^{\delta+}$ bond dipoles around the benzene ring to create an overall charge distribution that is composed of a negative charge in the center of the benzene ring that can bind cations. Also, each of the synthetic derivatives maintains the C-3 hydroxy methyl substituent that is important for the biological activity of the natural I3C parental compound [65]. In two analogs, methyl substituents were added at the para and ortho positions on the benzene ring forming (1-p-tolyl-1H-indol-3-yl)-carbinol (Figure 1.3, compound 2242) and (1-o-tolyl-1H-indol-3-yl)-carbinol (Figure 1.3, compound 2243), respectively. Each methyl group is electron donating, thus making the π system more nucleophilic and showing a higher propensity to participate in electrophilic substitution reactions. Two electron donating methoxy substituents were attached to the para and meta positions in the benzene ring forming (1-(3,4-dimethoxyphenyl)-1H-indol-3-yl) carbinol (Figure 1.3, compound 2244). The lone pairs of oxygen electrons on the dimethoxy groups are adjacent to the π system thereby increasing the electron density on the ring through a resonance donating effect that causes the π system to be more nucleophilic. In another indolecarbinol analogue, the linker carbon was eliminated from the benzyl group forming (1-phenyl-1H-indol-3-yl) carbinol (Figure 1.3, compound 2160), which makes the overall benzene moiety more rigid and less likely to participate in nucleophilic interactions. In the fifth derivative, a thiophene moiety was added to carbon 2 in the indole ring forming benzyl-2(thiophen-2-yl)-1H-indol-3-yl-carbinol (Figure 1.3, compound 2163). The thiophene group should add hydrophobic character but also significantly increases the bulkiness and decreases the flexibility of this analogue.

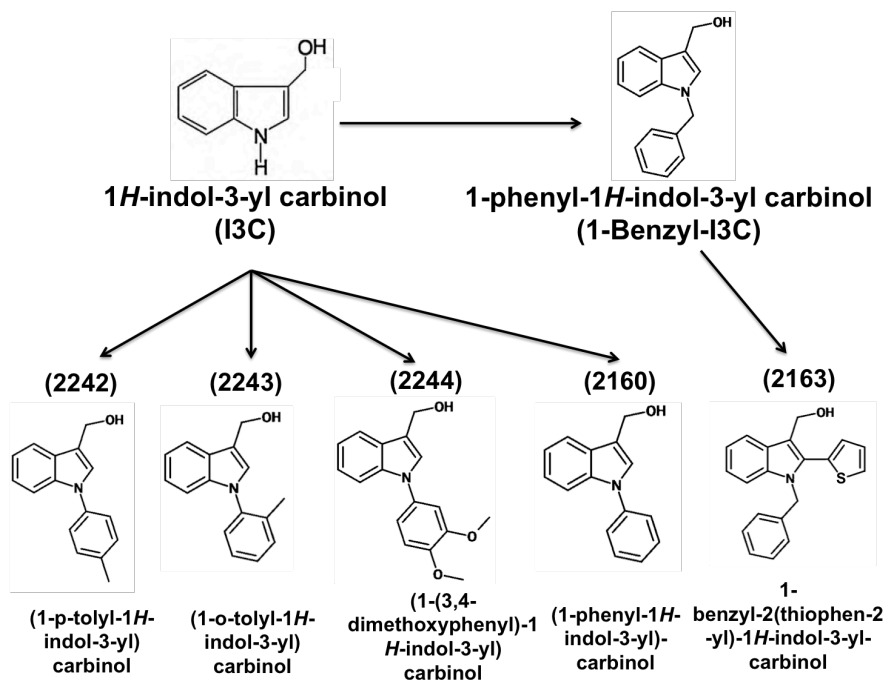


Figure 1.3. Structures of I3C, 1-benzyl-I3C, and indolecarbinol analogues. Each of the synthetic analogs maintains the C-3 hydroxy methyl substituent. Methyl substituents were added at the para and ortho positions on the benzene ring forming (1-*p*-tolyl-1H-indol-3-yl)-carbinol (compound 2242) and (1-*o*-tolyl-1H-indol-3-yl)-carbinol (compound 2243), respectively. Two methoxy substituents were attached to the para and meta positions in the benzene ring forming (1-(3,4-dimethoxyphenyl)-1H-indol-3-yl) carbinol (compound 2244). The linker carbon was eliminated from the benzyl group forming (1-phenyl-1H-indol-3-yl) carbinol (compound 2160). A thiophene moiety was added to carbon 2 in the indole ring forming benzyl-2(thiophen-2-yl)-1H-indol-3-yl-carbinol (compound 2163).

In silico computational modeling was performed to evaluate potential binding sites within NEDD4-1^{HECT} for each of the synthetic indolecarbinol analogues using the Schrödinger Induced Fit Docking protocol as described earlier. Simulations that dock each compound with NEDD4-1^{HECT} domain implicate several differences in selectivity between the synthetic indolecarbinol derivatives (Figure 1.4). For comparison, the predicted interactions of 1-benzyl-I3C is shown in Figure 4A. Each of indolecarbinol analogs is calculated to bind within the same binding cavity as I3C and 1-benzyl-I3C, which is adjacent to the N-lobe of the NEDD4-1 HECT domain. The π -system of the bound phenyl group in compound 2242 is predicted to engage in an aromatic CH- π association with the aryl ring hydrogen of Phe 637. Potential hydrophobic associations are also observed between the compound's phenyl moiety and Met 600 (Figure 1.4B). Contrastingly, the methyl substituent placed in the ortho position on the phenyl ring of compound 2243 orients the compound such that it is predicted to make aryl-aryl interactions with Tyr 634 and van der Waals associations with Asn 628 and Glu 554 (Figure 1.4C). Compound 2244 is predicted to be anchored by hydrophobic residues Ile 596 and Leu 553 and is able to make aryl-aryl interactions with Tyr 557 and Tyr 634 via its 3,4, di-methoxy phenyl moiety. Compound 2244 is also predicted to associate with Arg 558 through cation $\cdots\pi$ bonds (Figure 1.4D). 2160 is predicted to engage in aryl-aryl interactions with Tyr 634 and Tyr 605 via the compound's indole moiety while its phenyl moiety is calculated to be anchored by hydrophobic residues Leu 607, Ile 638, and Ile 596 (Figure 1.4E). The thiophene moiety in compound 2163 is embedded between negatively charged residue Glu 554 and also engages in T-shaped edge-to-face interactions with Tyr 605. 2263 is also predicted to interact with hydrophobic residues Phe 637 and Met 600 via its phenyl moiety and is anchored between Tyr 634, Asn 628, and His 631 via its indole moiety. (Figure 1.4F).

All predicted interactions of the indolecarbinol analogs with specific amino acid residues of the NEDD4-1^{HECT} domain are within 4 Å of each compound. There are several amino acid residues of the HECT domain that are predicted to interact with all or most of the indolecarbinol analogs (Leu 553, Tyr 605, Leu 607, Cys 627, Asn 628, Tyr 634). The *in silico* analysis also uncovered predicted amino acid residue interactions that are relatively specific for selective sets of the indolecarbinol analogues (Tyr 557 and Phe 608 for Compounds 2242, 2243 and 2244; Ser 624 and His 631 for I3C, 1-benzyl-I3C and compound 2163; Glu 599 for compounds 2242, 2160 and 2163; Ile for compounds 2244 and 2160; Phe 601 for compounds 2243 and 2244; Phe 635 for I3C and compound 2242; Asn 602 for compounds 2243 and 2244; Asp 630 for 1-benzyl-I3C). Calculated docking scores for the *in silico* defined interactions of each indolecarbinol derivative with the NEDD4-1^{HECT} domain yielded negative values between approximately -11 kJ/mol and -6 kJ/mol, implicating that the ligand-protein structures are predicted to be generally stable. Simulations that used the C2 or WW domain structure of NEDD4-1 displayed no binding interactions with the indolecarbinol compounds (data not shown).

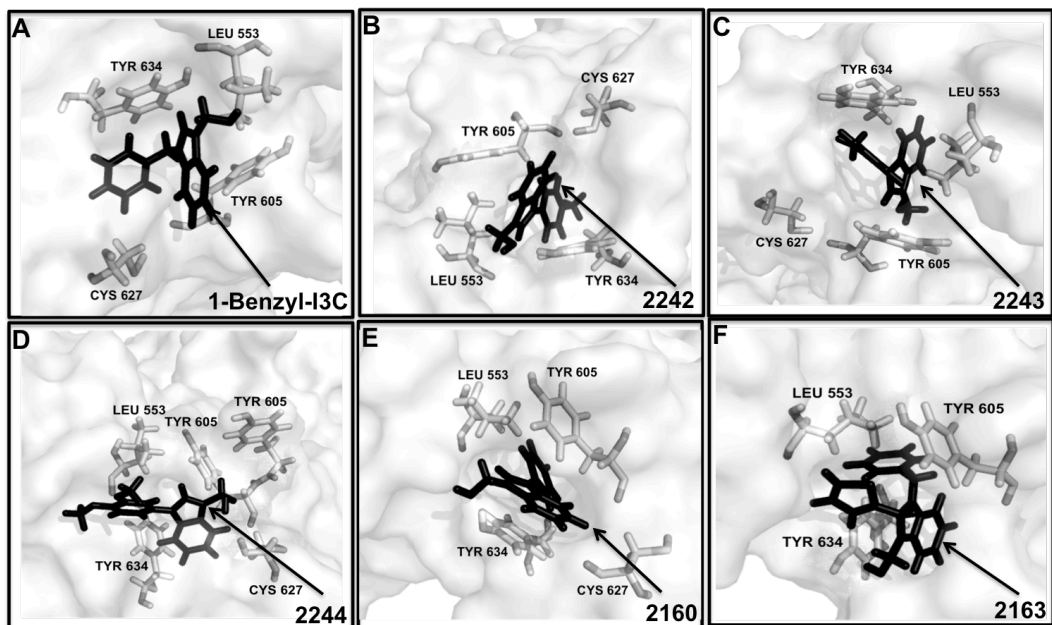


Figure 1.4. *In silico* computational models of predicted molecular interactions of (A) 1-benzyl-I3C, (B) 2242, (C) 2243, (D) 2244, (E) 2160, and (E) 2163 with the human NEDD4-1 HECT domain x-ray crystal structure (PDB accession number: 5C91). The structures of NEDD4-1 HECT domain and the corresponding indolecarbinol compounds (shown in 1A: I3C; 1B: 1-benzyl-I3C) were processed with the Protein Preparation Wizard in the Schrödinger suite and were evaluated using the Schrödinger induced fit docking (IFD) protocol to represent conformational changes in the ligand binding site stimulated by ligand binding. Close contact residues are labeled and are within 4 Å of each indolecarbinol compound. Models were visualized using PyMol.

Binding of indolecarbinol analogs to the purified NEDD4-1^{HECT} domain of NEDD4-1

The direct binding of each of the indolecarbinol analogs to the purified bacterially synthesized NEDD4-1^{HECT} domain was analyzed by Protein Thermal Shift assays described earlier for I3C and 1-benzyl-I3C. For comparative purposes, the Protein Thermal Shift melting profile for the NEDD4-1^{HECT} in the absence of added indolecarbinols (T_m of 48°C) and after addition of 1-benzyl-I3C (T_m of 58°C) from Figure 1C is incorporated into Figure 5. Exposure to all but one of the indolecarbinol compounds stabilized to different extents the NEDD4-1^{HECT} secondary structure based on concomitant increases in melting temperature (Figure 1.5). From the protein melting temperatures profiles of NEDD4-1^{HECT} (Figure 1.6B), the compounds listed in descending melting temperatures or ΔT_m values are 1-benzyl-I3C (10°C), compound 2244 (9°C), compound 2242 (7°C), compound 2243 (5°C) and compound 2160 (1.7°C). Treatment of NEDD4-1^{HECT} with compound 2163 displayed a reduced (-2°C) peak melting temperature compared to the untreated NEDD4-1^{HECT}. These results show that each of the indolecarbinol compounds directly bind to the purified NEDD4-1^{HECT} domain, although the nature of the protein binding sites may differ, especially for the 2163 compound.

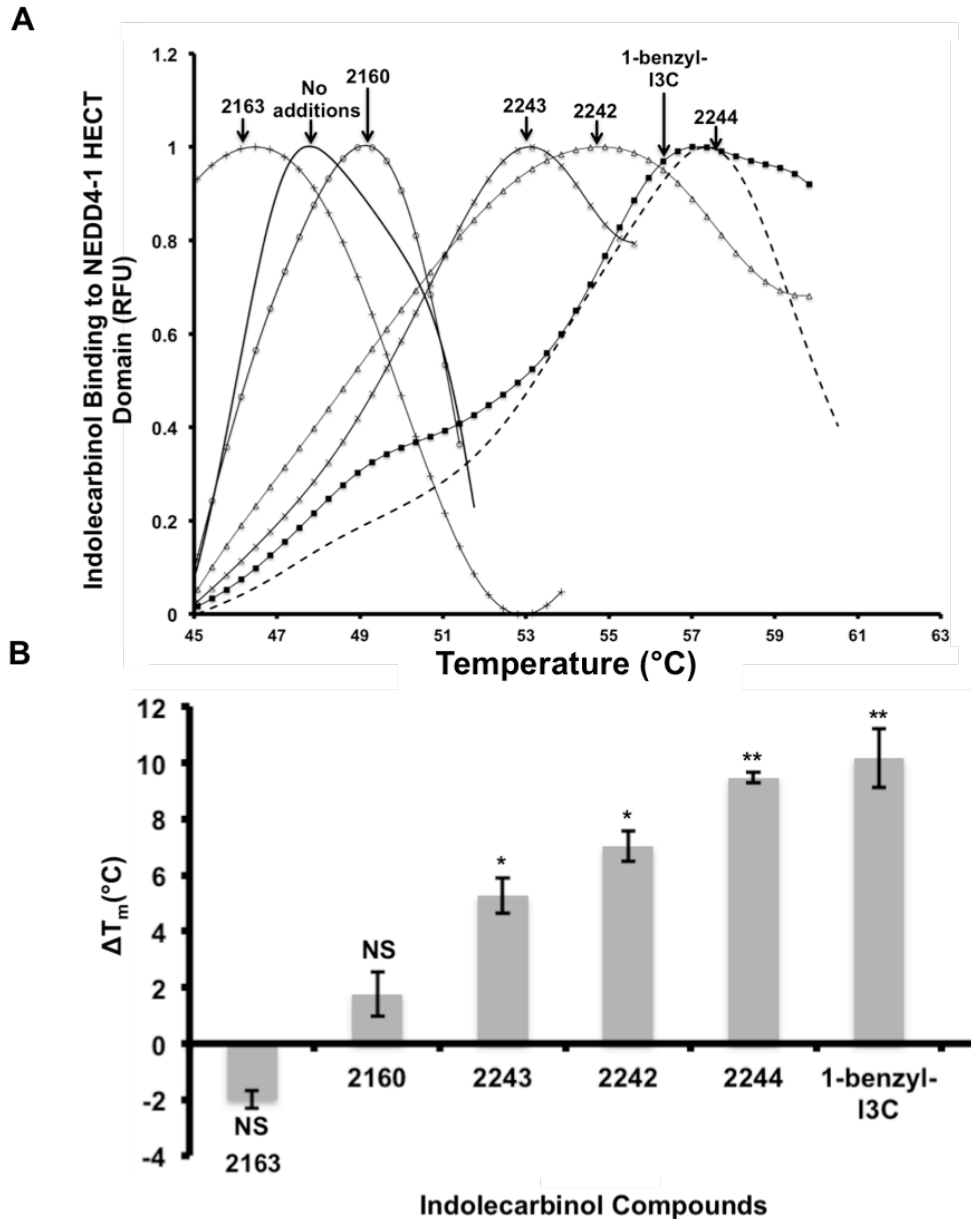


Figure 1.5. Binding of indolecarbinol compounds to purified NEDD4-1 HECT domain was evaluated via protein thermal shift assays. (A) The first derivative melt profiles show melting temperatures of NEDD4-1 HECT domain in the absence or presence of either 25 μ M 1-benzyl-I3C, 2242, or 2243, or 250 μ M I3C, 2244, 2160, or 2163. An increase in melting temperature is indicative of enhanced protein thermostability due to indolecarbinol binding. (B) Bar graph depicts changes in melting temperature of NEDD4-1 HECT domain in the presence of each respective indolecarbinol compound. Data is expressed as average of triplicate experiments and significance was assigned at a * p value < 0.05; ** p < 0.01; *** p < 0.001 where each group was compared to control protein with no additions; NS, no significant difference compared to control.

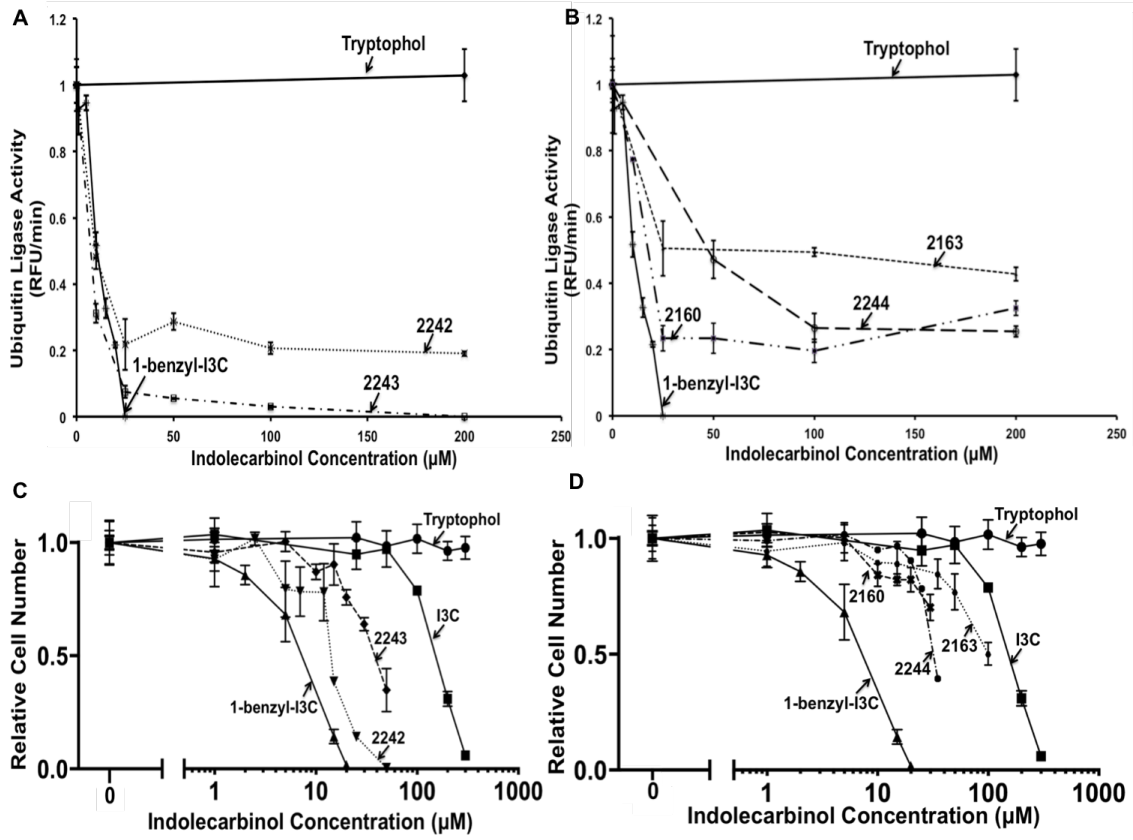


Fig. 1.6. Effects of indolecarbinol compounds on the *in vitro* enzymatic activity of NEDD4-1 and on G361 human melanoma cell proliferation. (A & B) Purified NEDD4-1 was pre-incubated with the indicated concentrations of each indolecarbinol compound for 1 hr and each mixture subsequently transferred to an ELISA microplate where the remaining contents of the ubiquitination reaction (E1, E2, and ubiquitin) were added. The formation of poly-ubiquitin chains was analyzed in an ELISA microplate and the enzymatic reactions were initiated with ATP. The washed plates were first incubated with biotin-linked ubiquitin antibody and then incubated with streptavidin-HRP to monitor NEDD4-1 auto-ubiquitination as the readout for NEDD4-1 activity. Enhanced chemiluminescent (ECL) reagent was then added to each well and incubated for 5 min. The relative light units (RLUs) were recorded using an LMax II/II384 microplate luminometer. NEDD4-1 enzymatic activity was standardized to the untreated NEDD4-1 protein and the data are expressed as the average of triplicate \pm SD. Comparison of Tryptophol, 1-benzyl-I3C, 2242, and 2243 is shown in panel A and comparison of Tryptophol, 1-benzyl-I3C, 2244, 2160, and 2163 is shown in panel B. (C & D) The concentration dependent effects of 48 hr treatment of indolecarbinol compounds on the proliferation of G361 melanoma cells. Absorbance at 450 nm monitored the relative cell number and the values were determined by standardizing the average of each treatment triplicate to the average value observed in vehicle control treated melanoma cells. The data show the average of 3 independent experiments. Comparison of Tryptophol, I3C, 1-benzyl-I3C, 2242, and 2243 is shown in panel C and the comparison of Tryptophol, I3C, 1-benzyl-I3C, 2244, 2160, and 2163 is shown in panel

Inhibitory effects of the indolecarbinol analogs on the *in vitro* enzymatic activity of the NEDD4-1 ubiquitin ligase and anti-proliferative efficacy of human melanoma cells

Each of the indolecarbinol analogues was preincubated with purified human NEDD4-1 protein for 1 hour and the *in vitro* polyubiquitylation assay described earlier was used to assess the effects on NEDD4-1 ubiquitin ligase activity. Compared to the effects of 1-benzyl-I3C from Figure 1E, the concentration-dependent effects of compounds 2242 and 2243 on NEDD4-1 enzymatic activity are shown in the left panel of Figure 6A, and the concentration-dependent effects of compounds 2244, 2160 and 2163 are shown in the right panel of Figure 6A. As a negative control, the level of NEDD4-1 enzymatic activity in the presence of 200 μM tryptophol, a biologically inactive indolecarbinol [65], is also shown. Of the tested indolecarbinol analogues, compounds 2242 and 2243 were the most effective inhibitors of NEDD4-1 ubiquitin ligase activity. Compounds 2242 and 2243 each have electron donating methyl substituents added to different positions on benzene ring (see Figure 1.3). Their efficacies were similar to or slightly better than 1-benzyl-I3C and significantly greater than that of I3C (see Figure 1.1) with observed IC_{50} of 13 μM for compound 2242 μM , IC_{50} of 8 μM for compound 2243 and IC_{50} of 12 μM for 1-benzyl-I3C (summarized in Figure 1.7). The high potency of 2243, 2242, and 1-benzyl-I3C may be attributed to the addition of the phenyl moiety on the indole nitrogen of I3C that is predicted to form stabilizing cation- π interactions with basic residues within the HECT domain of NEDD4-1.

In contrast, compounds 2160, 2244 and 2163 have weaker inhibitory effects on NEDD4-1 catalytic activity compared to 1-benzyl-I3C with experimentally defined IC_{50} s of 30 μM , 55 μM and 150 μM respectively. Based on the computational models, the substituents of the compounds 2163 and 2244 are solvent-exposed, and hence, unfavorable solvation of the bulky and hydrophobic groups may account for the reduced inhibitory activities compared to the 1-benzyl-I3C. Furthermore, the additional thiophene group attached to the indole ring in compound 2163 likely increases the compound's steric hindrance and prevents it from binding tightly to the protein, which likely accounts for the inhibitory effects on NEDD4-1 activity of 2163. These observations identify indolecarbinol analogs as a novel set of small molecule inhibitors of NEDD4-1 ubiquitin ligase activity.

Because of the connection between NEDD4-1 activity and the proliferation of G361 melanoma cells, the anti-proliferative effectiveness of 48 hour treatment with a concentration range of each of the indolecarbinol analogs was evaluated in comparison to 1-benzyl-I3C, I3C and the inactive indolecarbinol tryptophol. Melanoma cell proliferation was determined by CCK-8 proliferation assay that measures the number of cells. As shown in Figure 6C, the two most potent indolecarbinols were 1-benzyl-I3C and compound 2242 which displayed half maximal inhibition of melanoma cell proliferation of 15 μM and 18 μM , respectively (see Figure 1.7). The anti-proliferative efficacies of compounds 2243, 2244 and 2160 (half maximal responses of 25 μM , 40 μM and 47 μM , respectively) were intermediate between I3C and the two the highly potent molecules 1-benzyl-I3C and compound 2242. The anti-proliferative potency of

compound 2163 displayed a generally similar half-maximal response as I3C (see Figure 1.7). Tryptophol had no effect on melanoma cell proliferation, which is consistent with its lack of effects in other cancer cell types [32, 65]. Comparison of half maximal anti-proliferative responses and half maximal inhibition of NEDD4-1 enzymatic activities showed an overall correlation of the efficacies of both effects for the indolecarbinol analogues (Figure 1.7). One notable deviation of this correlation is with compound 2243 in which the relative inhibition of NEDD4-1 activity is more effective compared to its relative anti-proliferative response, which is likely caused by factors such as the efficiency of cellular import and stability of the indolecarbinol analogues.








Indolecarbinol Compound	Structure	Half Maximum Inhibitory Concentration for NEDD4-1 Enzymatic Activity (μM)	Half Maximum Concentration for Growth Inhibition in G361 Cells (μM)
I3C		280	107
1-benzyl-I3C		12	15
2242		13	18
2243		8	25
2244		55	40
2160		32	47
2163		150	106

Fig. 1.7. Table comparing the half maximum inhibitory concentration for NEDD4-1 enzymatic activity after 1 hr treatment and half maximum inhibition of proliferation of G361 cells after 48 hrs treatment with each indolecarbinol compound. The data was processed using GraphPad Prism 6 non-linear regression analysis.

Discussion

Regulation of cellular ubiquitination reactions through the activation or inhibition of individual ubiquitin ligases is an emerging potential strategy to target dysfunctional or hyperactive cellular processes in human cancer cells. One such intriguing therapeutic target is NEDD4-1, a HECT domain-containing ubiquitin ligase, which can display oncogenic-like properties through the targeted degradation of its protein substrates [1, 6]. However, relatively little is known about small inhibitors of HECT domain-containing ubiquitin ligases, although conceivably the use of such inhibitors could potentially be used in the development of new anti-cancer strategies. Our results now demonstrate that the I3C and its synthetic derivatives are potent inhibitors of NEDD4-1 ubiquitin ligase activity and that the I3C indolecarbinol structure represents a novel chemical platform of small molecule inhibitors of HECT-domain-containing ubiquitin ligases. We examined the *in vitro* binding and inhibition of enzymatic activity of purified NEDD4-1 by I3C, 1-benzyl-I3C and of five additional synthetic indolecarbinol analogues that display selective alterations in the physical and chemical properties of the resulting indolecarbinol structure [31-33]. The enhanced ability of the indolecarbinol compounds to inhibit NEDD4-1 activity required an optimal nucleophilic π system within the benzene ring, whereas, functionality of the indolecarbinol analogues decreased with the loss of flexibility of the chemical scaffold. Also, each of the synthetic indolecarbinol compounds maintains the C-3 hydroxy methyl substituent that is important for the biological activity of the I3C parental compound [65]. The overall efficacy of the anti-proliferative effects in melanoma cells of the tested indolecarbinol compounds generally correlated with their *in vitro* inhibition of NEDD4-1 ubiquitin ligase activity, implicating the potential importance of the indolecarbinol structural scaffold in developing highly potent and stable anti-cancer molecules.

In silico computational simulations predict stable interactions of each of the indolecarbinol compounds with the catalytic HECT domain of NEDD4-1, whereas, the same analysis shows that the C2 and WW domains of NEDD4-1 do not contain any predicted indolecarbinol binding sites. NEDD4-1 enzymatic activity was strongly inhibited by treatment with 1-benzyl-I3C, compound 2242 and compound 2243, each of which have significant electron donating moieties attached to the nitrogen position of the indole ring of I3C. Adding electron donating alkyl substituents to either the para or ortho positions on the benzene ring, forming compounds 2242 and 2243 respectively, produced the most potent inhibitors of NEDD4-1 enzymatic activity. This result suggests the importance of an optimal benzene ring nucleophilic π system for inhibiting enzymatic activity of NEDD4-1. In order to provide structural information for the indolecarbinol compounds in complex with NEDD4-1, we performed induced fit docking in the Schrödinger software suite to better understand the binding affinity of the protein-ligand complexes. While other docking programs have in the past been employed in structure-based drug design to optimize small molecule drug candidates, these programs assume that a flexible ligand is docked to a rigid receptor binding pocket, however, protein structural flexibility ought to be addressed due to the high dimensionality of the protein-ligand complex and to the dual nature of ligand binding sites which are defined by regions of high stability and high flexibility [66]. Indeed, the induced fit docking simulations predict highly stable interactions with the polar amino acid Asn 628 and

hydrophobic amino acids Leu 607, Tyr 634, Tyr 605, Leu 553, Ile 638, Meth 600, and Phe 637 by compound 2242 and compound 2243 within the HECT domain of NEDD4-1. The addition of two electron donating methoxy substituents to the para and meta positions in the benzene ring forming compound 2244 resulted in an approximate 4.5-fold less potency compared to 1-benzyl-I3C but remained approximately 5-fold more potent than I3C. *In silico* modeling suggests that one reason for this reduced potency is the loss of predicted π stacking edge-to-face interactions with Tyr 634 and Tyr 605. Also, relative to 1-benzyl-I3C, compound 2244 is predicted to flip its orientation such that the indole ring is positioned where 1-benzyl-I3C's phenyl moiety was predicted to be located.

Compound 2160, in which the single carbon linker between the indolecarbinol nitrogen and the benzene ring was eliminated, reduced the 1-benzyl-I3C potency of NEDD4-1 activity inhibition by approximately 2.6-fold. The loss of the linker carbon causes the overall benzene moiety to be more rigid and less likely to participate in nucleophilic interactions, which further implicates the importance of maintaining the nucleophilic π system within the benzene ring for the inhibition of NEDD4-1 activity. The rigidity of compound 2160's benzene moiety causes it to adjust its position such that the benzene moiety is now predicted to lose its network of associations with critical residues Tyr 634, Asn 628, His 631 and Asp 630 that is predicted with 1-benzyl-I3C and instead forms weaker hydrophobic interactions with Ile 638 and Phe 637. The addition of the thiophene moiety to carbon 2 in the indole ring of 1-benzyl-I3C, forming compound 2163, generated an inhibitor of NEDD4-1 activity that is approximately equivalent to I3C and thereby reversed the enhanced efficacy of 1-benzyl-I3C. *In silico* modeling predicts that compound 2163's benzene moiety can maintain hydrophobic associations with residue Leu 607 but its thiophene moiety displaces interactions with Tyr 605 and Tyr 634 which is necessary for securely fastening compound 2163 at this pocket, and thus likely accounts for the loss of inhibitory activity of the 1-benzyl-I3C scaffold. Thus, although the thiophene moiety adds hydrophobic character, it also significantly increases the bulkiness and decreases the flexibility of the compound.

The predicted interactions of the indolecarbinol compounds with the NEDD4-1 HECT domain appears to be relatively specific for NEDD4-1 because the HECT domains in the HERC1, E6-AP, and ITCH families of ubiquitin ligases, which are only partially homologous to the NEDD4-1 HECT domain [67], show no *in silico* predicted interactions with I3C (data not shown). Consistent with the *in silico* docking analysis, protein thermal shift assays demonstrated that each of indolecarbinol compounds bind *in vitro* to the purified NEDD4-1 HECT domain. The C2 and HECT domains of NEDD4-1, which are responsible for membrane localization and ubiquitin ligase activity respectively, bind each other resulting in an auto-inhibition of NEDD4-1 activity [68-70]. Conceivably, the indolecarbinol interactions with NEDD4-1 HECT domain could potentially inhibit NEDD4-1 activity by stabilizing the C2-HECT autoinhibitory intramolecular interactions or by preventing ubiquitin binding and poly ubiquitin chain elongation by forming tight aromatic associations with Tyr 605. An alternative hypothesis is that the indolecarbinol compounds may alter functional interactions of one or more NEDD4-1 upstream regulators that are thought to have context-specific effects

on NEDD4-1 protein stabilization and activity depending on the transformed or stressed state of the cells [71, 72].

The indolecarbinol inhibition of NEDD4-1 enzymatic activity and disruption of downstream signaling probably accounts, in part, for the anti-proliferative effects of indolecarbinol molecules. Based on the known actions of I3C in human cancer cells [31-33], it is highly likely that synthetic indolecarbinol analogues of I3C regulate multiple cell signaling pathways in melanoma cells. In this regard, the most sensitive melanoma cells to the anti-proliferative effects of I3C express an oncogenic BRAF and a wild type form of PTEN [34, 39]. Oncogenic BRAF signaling maintains high levels of cellular Erk/MAPK activity whereas, PTEN down-regulates AKT signaling, which ultimately abolishes cell survival networks, anti-apoptotic and cell invasion properties of human melanoma cells [42, 73, 74]. The mechanistic relationship between NEDD4-1 ubiquitin ligase activity and PTEN protein has not been definitively established because of cell type-dependent differences in the resulting effects on PTEN protein ubiquitination, such as on protein stability and localization, as well as one report suggesting that even though PTEN is a substrate of NEDD4-1, this ubiquitin ligase is regulated by PTEN signaling [75]. Several studies conclude that the oncogenic functions of NEDD4-1 are inversely linked to the tumor suppression functions of PTEN [5]. We have observed in wild type PTEN expressing melanoma cells, which also express an oncogenic BRAF, that the down regulation of NEDD4-1 activity either by I3C treatment or by siRNA knockdown of NEDD4-1 resulted in higher levels of wild type PTEN protein [34]. NEDD4-1 is frequently up regulated in a variety of human cancers [76], for example, in a majority of non-small cell lung cancer tumor tissues NEDD4-1 over expression correlated with the loss of PTEN protein [77]. In NHA-E6/E7/hTERT cells, expression of high levels of FoxM1B up regulated NEDD4-1, which led to the down regulation of PTEN and the hyperactivation of AKT [78].

The enhanced efficacy of the I3C analogues for their anti-proliferative properties is likely due to a combination of higher affinity and qualitatively more effective interactions with specific target proteins such as NEDD4-1 as well as potentially from enhanced cellular stability and more efficient uptake. In this regard, the addition of a benzene moiety to the position one nitrogen of the indole ring, forming 1-benzyl-I3C, prevents the self-condensation that occurs to different extents with the parental I3C compound [57]. Similarly, the highly potent compounds 2242 and 2243 are likely to be more stable in a cellular context and both are potentially highly promising anti-cancer molecules. Treatment with I3C activates distinct and overlapping sets of anti-proliferative signaling events in different cancer cell types [79], and the cell type selective responses of indolecarbinol compounds can be attributed to the presence and activity of individual indolecarbinol target proteins that regulate specific sets of cellular cascades. For example, we previously demonstrated that the I3C and 1-benzyl-I3C inhibition of elastase enzymatic activity triggers anti-proliferative signaling in human breast cancer cells [33, 57] and I3C can inhibit oncogenic BRAF enzymatic activity and disrupt downstream signaling events in human melanoma cells [39]. We are currently attempting to dissect common structural features between the indolecarbinol target proteins that account for their sensitivity to I3C and its highly potent derivatives. Furthermore,

because the vast majority of human melanoma cells express an oncogenic form of BRAF, the simultaneous disruption of NEDD4-1 and BRAF signaling by I3C and its derivatives implicates indolecarbinol analogs as intriguing new chemical foundations to develop highly potent, stable and target-specific anti-cancer compounds.

References

1. Ravid T, Hochstrasser M: Diversity of degradation signals in the ubiquitin-proteasome system. *Nat Rev Mol Cell Biol* 2008, 9(9):679-690.
2. Rotin D, Kumar S: Physiological functions of the HECT family of ubiquitin ligases. *Nat Rev Mol Cell Biol* 2009, 10(6):398-409.
3. Bernassola F, Karin M, Ciechanover A, Melino G: The HECT family of E3 ubiquitin ligases: multiple players in cancer development. *Cancer Cell* 2008, 14(1):10-21.
4. Yang B, Kumar S: Nedd4 and Nedd4-2: closely related ubiquitin-protein ligases with distinct physiological functions. *Cell Death Differ* 2010, 17(1):68-77.
5. Ye X, Wang L, Shang B, Wang Z, Wei W: NEDD4: a promising target for cancer therapy. *Curr Cancer Drug Targets* 2014, 14(6):549-556.
6. Katz M, Shtiegman K, Tal-Or P, Yakir L, Mosesson Y, Harari D, Machluf Y, Asao H, Jovin T, Sugamura K *et al*: Ligand-independent degradation of epidermal growth factor receptor involves receptor ubiquitylation and Hgs, an adaptor whose ubiquitin-interacting motif targets ubiquitylation by Nedd4. *Traffic* 2002, 3(10):740-751.
7. Aggarwal BB, Ichikawa H: Molecular targets and anticancer potential of indole-3-carbinol and its derivatives. *Cell Cycle* 2005, 4(9):1201-1215.
8. Ahmad A, Sakr WA, Rahman KMW: Anticancer Properties of Indole Compounds: Mechanism of Apoptosis Induction and Role in Chemotherapy. *Curr Drug Targets* 2010, 11(6):652-666.
9. Moiseeva EP, Almeida GM, Jones GD, Manson MM: Extended treatment with physiologic concentrations of dietary phytochemicals results in altered gene expression, reduced growth, and apoptosis of cancer cells. *Mol Cancer Ther* 2007, 6(11):3071-3079.
10. Hsu JC, Dev A, Wing A, Brew CT, Bjeldanes LF, Firestone GL: Indole-3-carbinol mediated cell cycle arrest of LNCaP human prostate cancer cells requires the induced production of activated p53 tumor suppressor protein. *Biochem Pharmacol* 2006, 72(12):1714-1723.
11. Melkamu T, Zhang XX, Tan JK, Zeng Y, Kassie F: Alteration of microRNA expression in vinyl carbamate-induced mouse lung tumors and modulation by the chemopreventive agent indole-3-carbinol. *Carcinogenesis* 2010, 31(2):252-258.
12. Kim DS, Jeong YM, Moon SI, Kim SY, Kwon SB, Park ES, Youn SW, Park KC: Indole-3-carbinol enhances ultraviolet B-induced apoptosis by sensitizing human melanoma cells. *Cell Mol Life Sci* 2006, 63(22):2661-2668.
13. Machijima Y, Ishikawa C, Sawada S, Okudaira T, Uchihara JN, Tanaka Y, Taira N, Mori N: Anti-adult T-cell leukemia/lymphoma effects of indole-3-carbinol. *Retrovirology* 2009, 6.
14. Qi M, Anderson AE, Chen DZ, Sun S, Auburn KJ: Indole-3-carbinol prevents PTEN loss in cervical cancer in vivo. *Mol Med* 2005, 11(1-12):59-63.
15. Jin L, Qi M, Chen DZ, Anderson A, Yang GY, Arbeit JM, Auburn KJ: Indole-3-carbinol prevents cervical cancer in human papilloma virus type 16 (HPV16) transgenic mice. *Cancer Res* 1999, 59(16):3991-3997.

16. Cover CM, Hsieh SJ, Tran SH, Hallden G, Kim GS, Bjeldanes LF, Firestone GL: Indole-3-carbinol inhibits the expression of cyclin-dependent kinase-6 and induces a G1 cell cycle arrest of human breast cancer cells independent of estrogen receptor signaling. *J Biol Chem* 1998, 273(7):3838-3847.
17. Sarkar FH, Li Y: Indole-3-carbinol and prostate cancer. *J Nutr* 2004, 134(12 Suppl):3493S-3498S.
18. Firestone GL, Bjeldanes LF: Indole-3-carbinol and 3,3'-diindolylmethane antiproliferative signaling pathways control cell-cycle gene transcription in human breast cancer cells by regulating promoter-Sp1 transcription factor interactions. *J Nutr* 2003, 133(7 Suppl):2448S-2455S.
19. Firestone GL, Sundar SN: Minireview: Modulation of Hormone Receptor Signaling by Dietary Anticancer Indoles. *Mol Endocrinol* 2009, 23(12):1940-1947.
20. Marconett CN, Sundar SN, Poindexter KM, Stueve TR, Bjeldanes LF, Firestone GL: Indole-3-carbinol triggers aryl hydrocarbon receptor-dependent estrogen receptor (ER)alpha protein degradation in breast cancer cells disrupting an ERalpha-GATA3 transcriptional cross-regulatory loop. *Mol Biol Cell* 2010, 21(7):1166-1177.
21. Marconett CN, Sundar SN, Tseng M, Tin AS, Tran KQ, Mahuron KM, Bjeldanes LF, Firestone GL: Indole-3-carbinol downregulation of telomerase gene expression requires the inhibition of estrogen receptor-alpha and Sp1 transcription factor interactions within the hTERT promoter and mediates the G(1) cell cycle arrest of human breast cancer cells. *Carcinogenesis* 2011, 32(9):1315-1323.
22. Maruthanila VL, Poornima J, Mirunalini S: Attenuation of Carcinogenesis and the Mechanism Underlying by the Influence of Indole-3-carbinol and Its Metabolite 3,3'-Diindolylmethane: A Therapeutic Marvel. *Adv Pharmacol Sci* 2014, 2014:832161.
23. Sarkar FH, Li Y, Wang Z, Kong D: Cellular signaling perturbation by natural products. *Cell Signal* 2009, 21(11):1541-1547.
24. Xu Y, Zhang J, Dong WG: Indole-3-carbinol (I3C)-induced apoptosis in nasopharyngeal cancer cells through Fas/FasL and MAPK pathway. *Med Oncol* 2011, 28(4):1343-1348.
25. Aggarwal BB, Shishodia S: Molecular targets of dietary agents for prevention and therapy of cancer. *Biochemical Pharmacology* 2006, 71(10):1397-1421.
26. Kim YS, Milner JA: Targets for indole-3-carbinol in cancer prevention. *J Nutr Biochem* 2005, 16(2):65-73.
27. Rogan EG: The natural chemopreventive compound indole-3-carbinol: State of the science. *In Vivo* 2006, 20(2):221-228.
28. Auburn KJ, Fan S, Rosen EM, Goodwin L, Chandraskaren A, Williams DE, Chen D, Carter TH: Indole-3-carbinol is a negative regulator of estrogen. *J Nutr* 2003, 133(7 Suppl):2470S-2475S.
29. Wang TT, Milner MJ, Milner JA, Kim YS: Estrogen receptor alpha as a target for indole-3-carbinol. *J Nutr Biochem* 2006, 17(10):659-664.
30. Weng JR, Tsai CH, Kulp SK, Chen CS: Indole-3-carbinol as a chemopreventive and anti-cancer agent. *Cancer Lett* 2008, 262(2):153-163.

31. Nguyen HH, Aronchik I, Brar GA, Nguyen DH, Bjeldanes LF, Firestone GL: The dietary phytochemical indole-3-carbinol is a natural elastase enzymatic inhibitor that disrupts cyclin E protein processing. *Proc Natl Acad Sci U S A* 2008, 105(50):19750-19755.
32. Aronchik I, Bjeldanes LF, Firestone GL: Direct inhibition of elastase activity by indole-3-carbinol triggers a CD40-TRAF regulatory cascade that disrupts NF-kappaB transcriptional activity in human breast cancer cells. *Cancer Res* 2010, 70(12):4961-4971.
33. Aronchik I, Chen T, Durkin KA, Horwitz MS, Preobrazhenskaya MN, Bjeldanes LF, Firestone GL: Target protein interactions of indole-3-carbinol and the highly potent derivative 1-benzyl-I3C with the C-terminal domain of human elastase uncouples cell cycle arrest from apoptotic signaling. *Mol Carcinog* 2012, 51(11):881-894.
34. Aronchik I, Kundu A, Quirit JG, Firestone GL: The antiproliferative response of indole-3-carbinol in human melanoma cells is triggered by an interaction with NEDD4-1 and disruption of wild-type PTEN degradation. *Mol Cancer Res* 2014, 12(11):1621-1634.
35. Nemoto E, Tada H, Shimauchi H: Disruption of CD40/CD40 ligand interaction with cleavage of CD40 on human gingival fibroblasts by human leukocyte elastase resulting in down-regulation of chemokine production. *J Leukoc Biol* 2002, 72(3):538-545.
36. Tong AW, Papayoti MH, Netto G, Armstrong DT, Ordonez G, Lawson JM, Stone MJ: Growth-inhibitory effects of CD40 ligand (CD154) and its endogenous expression in human breast cancer. *Clin Cancer Res* 2001, 7(3):691-703.
37. Zhou Y, Eppenberger-Castori S, Eppenberger U, Benz CC: The NFkappaB pathway and endocrine-resistant breast cancer. *Endocr Relat Cancer* 2005, 12 Suppl 1:S37-46.
38. Madhusoodhanan R, Natarajan M, Veeraraghavan J, Herman TS, Aravindan N: NFkappaB activity and transcriptional responses in human breast adenocarcinoma cells after single and fractionated irradiation. *Cancer Biol Ther* 2009, 8(9):765-773.
39. Kundu A, Quirit JG, Khouri MG, Firestone GL: Inhibition of oncogenic BRAF activity by indole-3-carbinol disrupts microphthalmia-associated transcription factor expression and arrests melanoma cell proliferation. *Mol Carcinog* 2016.
40. Aguisa-Toure AH, Li G: Genetic alterations of PTEN in human melanoma. *Cell Mol Life Sci* 2012, 69(9):1475-1491.
41. Conde-Perez A, Larue L: PTEN and melanomagenesis. *Future Oncol* 2012, 8(9):1109-1120.
42. Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, Theurillat JP, Nickerson E, Auclair D, Li L, Place C *et al*: A landscape of driver mutations in melanoma. *Cell* 2012, 150(2):251-263.
43. Zhang H, Cai Q, Ma D: Amino acid promoted CuI-catalyzed C-N bond formation between aryl halides and amines or N-containing heterocycles. *J Org Chem* 2005, 70(13):5164-5173.
44. Holla BS, Ambekar SY: Studies in Biheterocycles .1. Formylation of 2-(2'-Thienyl) Indole. *J Indian Chem Soc* 1974, 51(11):965-966.

45. Niesen FH, Berglund H, Vedadi M: The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* 2007, 2(9):2212-2221.
46. Kathman SG, Span I, Smith AT, Xu Z, Zhan J, Rosenzweig AC, Statsyuk AV: A Small Molecule That Switches a Ubiquitin Ligase From a Processive to a Distributive Enzymatic Mechanism. *J Am Chem Soc* 2015, 137(39):12442-12445.
47. Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W: Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des* 2013, 27(3):221-234.
48. Farid R, Day T, Friesner RA, Pearlstein RA: New insights about HERG blockade obtained from protein modeling, potential energy mapping, and docking studies. *Bioorg Med Chem* 2006, 14(9):3160-3173.
49. Sherman W, Day T, Jacobson MP, Friesner RA, Farid R: Novel procedure for modeling ligand/receptor induced fit effects. *J Med Chem* 2006, 49(2):534-553.
50. Sherman W, Beard HS, Farid R: Use of an induced fit receptor structure in virtual screening. *Chem Biol Drug Des* 2006, 67(1):83-84.
51. Jacobson MP, Pincus DL, Rapp CS, Day TJ, Honig B, Shaw DE, Friesner RA: A hierarchical approach to all-atom protein loop prediction. *Proteins* 2004, 55(2):351-367.
52. Jacobson MP, Friesner RA, Xiang ZX, Honig B: On the role of the crystal environment in determining protein side-chain conformations. *J Mol Biol* 2002, 320(3):597-608.
53. Hershko A, Ciechanover A: The ubiquitin system. *Annu Rev Biochem* 1998, 67:425-479.
54. Welchman RL, Gordon C, Mayer RJ: Ubiquitin and ubiquitin-like proteins as multifunctional signals. *Nat Rev Mol Cell Biol* 2005, 6(8):599-609.
55. Ciechanover A: The ubiquitin-proteasome pathway: on protein death and cell life. *EMBO J* 1998, 17(24):7151-7160.
56. Raasi S, Varadan R, Fushman D, Pickart CM: Diverse polyubiquitin interaction properties of ubiquitin-associated domains. *Nat Struct Mol Biol* 2005, 12(8):708-714.
57. Nguyen HH, Lavrenov SN, Sundar SN, Nguyen DH, Tseng M, Marconett CN, Kung J, Staub RE, Preobrazhenskaya MN, Bjeldanes LF *et al*: 1-Benzyl-indole-3-carbinol is a novel indole-3-carbinol derivative with significantly enhanced potency of anti-proliferative and anti-estrogenic properties in human breast cancer cells. *Chem Biol Interact* 2010, 186(3):255-266.
58. Maspero E, Mari S, Valentini E, Musacchio A, Fish A, Pasqualato S, Polo S: Structure of the HECT:ubiquitin complex and its role in ubiquitin chain elongation. *EMBO Rep* 2011, 12(4):342-349.
59. Vedadi M, Niesen FH, Allali-Hassani A, Fedorov OY, Finerty PJ, Jr., Wasney GA, Yeung R, Arrowsmith C, Ball LJ, Berglund H *et al*: Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc Natl Acad Sci U S A* 2006, 103(43):15835-15840.
60. Regad T: Molecular and cellular pathogenesis of melanoma initiation and progression. *Cell Mol Life Sci* 2013, 70(21):4055-4065.

61. Levy C, Khaled M, Fisher DE: MITF: master regulator of melanocyte development and melanoma oncogene. *Trends Mol Med* 2006, 12(9):406-414.
62. Pierrat MJ, Marsaud V, Mauviel A, Javelaud D: Expression of microphthalmia-associated transcription factor (MITF), which is critical for melanoma progression, is inhibited by both transcription factor GLI2 and transforming growth factor-beta. *J Biol Chem* 2012, 287(22):17996-18004.
63. Du J, Widlund HR, Horstmann MA, Ramaswamy S, Ross K, Huber WE, Nishimura EK, Golub TR, Fisher DE: Critical role of CDK2 for melanoma growth linked to its melanocyte-specific transcriptional regulation by MITF. *Cancer Cell* 2004, 6(6):565-576.
64. Wellbrock C, Rana S, Paterson H, Pickersgill H, Brummelkamp T, Marais R: Oncogenic BRAF regulates melanoma proliferation through the lineage specific factor MITF. *PLoS One* 2008, 3(7):e2734.
65. Jump SM, Kung J, Staub R, Kinseth MA, Cram EJ, Yudina LN, Preobrazhenskaya MN, Bjeldanes LF, Firestone GL: N-Alkoxy derivatization of indole-3-carbinol increases the efficacy of the G1 cell cycle arrest and of I3C-specific regulation of cell cycle gene transcription and activity in human breast cancer cells. *Biochem Pharmacol* 2008, 75(3):713-724.
66. Luque I, Freire E: Structural stability of binding sites: Consequences for binding affinity and allosteric effects. *Proteins-Structure Function and Genetics* 2000:63-71.
67. Scheffner M, Kumar S: Mammalian HECT ubiquitin-protein ligases: biological and pathophysiological aspects. *Biochim Biophys Acta* 2014, 1843(1):61-74.
68. Wang J, Peng Q, Lin Q, Childress C, Carey D, Yang W: Calcium activates Nedd4 E3 ubiquitin ligases by releasing the C2 domain-mediated auto-inhibition. *J Biol Chem* 2010, 285(16):12279-12288.
69. Wiesner S, Ogunjimi AA, Wang HR, Rotin D, Sicheri F, Wrana JL, Forman-Kay JD: Autoinhibition of the HECT-type ubiquitin ligase Smurf2 through its C2 domain. *Cell* 2007, 130(4):651-662.
70. Snyder PM, Olson DR, McDonald FJ, Bucher DB: Multiple WW domains, but not the C2 domain, are required for inhibition of the epithelial Na⁺ channel by human Nedd4. *J Biol Chem* 2001, 276(30):28321-28326.
71. Oberst A, Malatesta M, Aqeilan RI, Rossi M, Salomoni P, Murillas R, Sharma P, Kuehn MR, Oren M, Croce CM *et al*: The Nedd4-binding partner 1 (N4BP1) protein is an inhibitor of the E3 ligase Itch. *Proc Natl Acad Sci U S A* 2007, 104(27):11280-11285.
72. Schieber C, Howitt J, Putz U, White JM, Parish CL, Donnelly PS, Tan SS: Cellular up-regulation of Nedd4 family interacting protein 1 (Ndfip1) using low levels of bioactive cobalt complexes. *J Biol Chem* 2011, 286(10):8555-8564.
73. Chakraborty R, Wieland CN, Comfere NI: Molecular targeted therapies in metastatic melanoma. *Pharmgenomics Pers Med* 2013, 6:49-56.
74. Bandarchi B, Jabbari CA, Vedadi A, Navab R: Molecular biology of normal melanocytes and melanoma cells. *J Clin Pathol* 2013, 66(8):644-648.
75. Ahn Y, Hwang CY, Lee SR, Kwon KS, Lee C: The tumour suppressor PTEN mediates a negative regulation of the E3 ubiquitin-protein ligase Nedd4. *Biochem J* 2008, 412(2):331-338.

76. Chen C, Matesic LE: The Nedd4-like family of E3 ubiquitin ligases and cancer. *Cancer Metastasis Rev* 2007, 26(3-4):587-604.
77. Amodio N, Scrima M, Palaia L, Salman AN, Quintiero A, Franco R, Botti G, Pirozzi P, Rocco G, De Rosa N *et al*: Oncogenic role of the E3 ubiquitin ligase NEDD4-1, a PTEN negative regulator, in non-small-cell lung carcinomas. *Am J Pathol* 2010, 177(5):2622-2634.
78. Dai B, Pieper RO, Li D, Wei P, Liu M, Woo SY, Aldape KD, Sawaya R, Xie K, Huang S: FoxM1B regulates NEDD4-1 expression, leading to cellular transformation and full malignant phenotype in immortalized human astrocytes. *Cancer Res* 2010, 70(7):2951-2961.
79. Gupta SC, Kim JH, Prasad S, Aggarwal BB: Regulation of survival, proliferation, invasion, angiogenesis, and metastasis of tumor cells through modulation of inflammatory pathways by nutraceuticals. *Cancer Metastasis Rev* 2010, 29(3):405-434.

Chapter 2

Indole-3-carbinol (I3C) analogues are potent small molecule inhibitors of human neutrophil elastase activity that disrupt proliferation of human breast cancer cells

Abstract

Human neutrophil elastase is a serine protease secreted by neutrophils and acts as an independent prognostic marker for late stage breast cancer. Our lab has demonstrated that indole-3-carbinol (I3C) can serve as a direct non-competitive inhibitor for elastase enzymatic activity and hence trigger the cell cycle arrest of human breast cancer cells. By employing *in vitro* protein thermostability assays and kinetic assays, we were able to prove that I3C can serve as a platform for developing novel small molecule enzymatic inhibitors of elastase. Relative to I3C which exhibits an IC₅₀ of 208 μM, compounds 2242, 2160, and 2243 inhibited elastase activity much more significantly as well with an IC₅₀ of 30 μM, 25 μM, and 17 μM respectively. The potency of compound 2242, 2160, and 2243 appeared to be sensitive to structural changes on the phenyl ring, more notably when methyl substituents were added to the para and ortho positions. The activity was abolished with the addition of bulky chemical groups like the thiophene substituent attached to the indole of 2163 and the electron donating methoxy group attached to the phenyl moiety of 2244 in the meta and ortho positions which exhibit an IC₅₀ of 250 μM and 343 μM respectively. The general potency of the anti-proliferative effects in breast cancer cells of the tested indolecarbinol compounds roughly corresponded to their *in vitro* inhibition of elastase activity which substantiates the significance of the indolecarbinol structure in generating stable and effective inhibitors. Molecular docking simulations also predicted stable interactions of each of the indolecarbinol compounds with the crystal structure of human neutrophil elastase. Therefore, based on the enhanced potency and higher specificity of inhibition of elastase activity and the ability to still sustain meager toxicity, indolecarbinol-derived therapeutics can be developed as a safer, less harmful approach for treating cancer.

Introduction

Many instances of cancer are treated through long-term use of chemotherapeutic agents that are virtually guaranteed to cause severe side effects. Therefore, there is a need to identify chemical compounds with low levels of toxicity that can be used as effective, alternative treatments. Epidemiological and physiological studies show that increased cruciferous vegetable intake (i.e. broccoli, cabbage, cauliflower) is associated with a reduced incidence in cancers particularly colon, lung, prostate, cervix and breast. Rising preclinical and clinical evidence indicate that indole-3-carbinol (I3C), a primary bioactive component in cruciferous vegetables that is derived by hydrolysis from glyco Brassica in *Brassica* vegetables, has multiple anticarcinogenic and antitumorigenic properties [1]. I3C and its natural diindole condensation product 3,3'-diindolylmethane (DIM) both possess chemo-preventative as well as strong anti-tumor properties *in vivo* in rodent model systems and in human cancer cell xenograft tumors in athymic mice with low levels of toxicity [2-4]. In indolecarbinol sensitive human cancer cells originating from breast, prostate, lung, colon, melanomas, leukemia, and cervical cancer, I3C and DIM elicit defined functionally regulated sets of transcriptional, cell signaling, enzymatic, and metabolic cascades that result in a cell cycle arrest, apoptosis, down-regulation of cancer cell migration and metastasis, and inflection of hormone receptor signaling [6-11]. The indole's potency and selectivity for a particular cellular pathway relies on the hormonal or environmental nature, the cancer cell's tissue origin and its phenotype [2]. While the mechanisms that underline tumor-specific indolecarbinol anti-proliferative responses is not clearly defined, the indole's selectivity can be attributed to the presence and activity of distinct indolecarbinol target proteins that regulate specific sets of cellular cascades.

We have established that elastase, a serine protease that acts as an independent prognostic marker for late stage breast cancer, is a target protein for I3C. Previous studies have indicated that I3C, can serve as a direct non-competitive inhibitor of human neutrophil elastase enzymatic activity that regulates the cell cycle arrest of human breast cancer cells [5]. Elastase can be observed in the cytoplasmic azurophilic granules and intercellular membranes, can be localized to the cell surface and released in the extracellular space [5]. Inhibition of the intracellular and extracellular forms of this protein mediate I3C's response in breast cancer cells. Additionally, it has been observed that I3C treatment in estrogen responsive MCF-7 breast cancer cells and in estrogen independent and highly invasive MDA-MB-231 breast cancer cells abolishes the intracellular elastase-dependent protein processing of Cyclin E [12], which triggers the accumulation of the higher molecular form of cyclin E protein which is affiliated with decreased CDK2 activity which mimics normal breast tissues [12]. The tumor necrosis factor (TNF) receptor family member CD40 has further been detected as a cell surface elastase substrate in human breast cancer cells and has been shown that upon I3C inhibition of extracellular elastase activity, processing of the 44 kDa full-length CD40 is disturbed, leading to the loss of NFkB nuclear localization and transcriptional activity and down-regulation of cyclin D1 expression [13]. More notably, the ablation of elastase via siRNA knockdown imitates the cellular responses of I3C as well as the G1

cell cycle arrest of I3C [12, 13] which is a testament to the enzyme's essential functional role as an I3C target protein.

While elastase has been identified as an I3C protein target, the high concentrations of I3C used to treat cells, its ability to dimerize and affect specificity and I3C-protein interactions in breast cancer calls for the need to develop more potent, stable and target-specific I3C derivatives. Through previous *in vitro* enzymatic studies, it was demonstrated that a derivative of I3C, 1-benzyl-I3C, which has a benzyl moiety attached to the indole ring nitrogen, recognizes elastase as a specific target protein and exhibits 1000-fold enhanced potency thus giving rise to the possibility that synthetic I3C derivatives with more robust anti-proliferative effects in human breast cancer could be produced.

The aforementioned studies which have cemented elastase as an I3C protein target serve as a platform for understanding the molecular interactions between these proteins and other I3C derived inhibitors. While the mechanistic and structural I3C-elastase have been investigated, it is not yet known how I3C and 1-benzyl-I3C analogs associate with elastase.

In this study, we show through *in silico* simulations, enzyme kinetic analysis, and cellular proliferation assays, that analogs of I3C and 1-benzyl-I3C which retain the parent compound's indole ring serve as elastase inhibitors.

Materials and Methods

In Silico Computational Methods

Protein Structure Preparation

The coordinates for the human neutrophil elastase protein structure were obtained from the RCSB Protein Data Bank (PDB). 5A8X [14] was used to examine the interactions between elastase and indolecarbinol compounds. The protein structure was prepared using the Schrödinger's Protein Preparation Wizard module (Schrödinger Suite 2015-4 Protein Preparation Wizard; Epik version 3.4, Schrödinger, LLC, New York, NY, 2015; Impact version 6.9, Schrödinger, LLC, New York, NY, 2015; Prime version 4.2, Schrödinger, LLC, New York, NY, 2015.) [15]. Hydrogen atoms were added and the side chain structures of Gln and Asn were flipped in order to yield the maximum degree of hydrogen bond interactions. The protein was subsequently minimized using the OPLS force field in the MacroModel module in Schrödinger (MacroModel, version 11.0, Schrödinger, LLC, New York, NY, 2015.) with backbone atoms fixed.

Identification of Surface Pockets using SiteMap

The site recognition software SiteMap (version 4.0, Schrödinger, LLC, New York, NY, 2016) [16] was run on the human neutrophil elastase (PDB accession number:5A8X) to isolate the top ranked 5 potential ligand-binding sites. In order to locate energetically favorable sites, SiteMap employs an algorithm similar to Goodford's GRID algorithm which uses interaction energies between grid probes and the protein. Sites were retained if they were composed of at least 15 site points. The default settings in SiteMap were applied and included a restrictive hydrophobicity definition, a standard grid (1.0 Å) and the OPLS2005 force field.

The SiteMap program calculates size, volume, degree of enclosure/exposure, degree of contact, hydrophobic/-philic character, hydrophobic/-philic balance and hydrogen-bonding possibilities (hydrogen/donors). Additionally, SiteMap determines two scores for each site and calculates it as:

$$\text{SiteScore} = 0.0733\sqrt{n} + 0.6688e - 0.201p$$

Where

n = the number of site points (capped at 100),

e = enclosure,

p = hydrophilicity of the site (capped at 1.0).

The druggability score, Dscore, is defined as:

$$\text{Dscore} = 0.094\sqrt{n} + 0.61e - 0.324p$$

Where n , e , and p are defined as above, but here p is uncapped. A SiteScore of greater than 1.0 suggests a relatively high probable ligand binding whereas a SiteScore of 0.8 can be used to distinguish between drug-binding and non-drug-binding sites. The Dscore, however, can be employed to differentiate between druggable and undruggable sites by inciting penalties on highly hydrophilic sites since ligands that bind to these particular sites would be very polar and would hence be removed from the organism

[17]. However, this does not necessarily preclude any possibility of any ligand binding, but it would be less likely to locate a high affinity drug-like ligand for such a particular site.

Glide Ligand Docking

All docking calculations were performed using the GLIDE (Grid-based Ligand Docking with Energetics) v7.2 (Glide, version 7.2, Schrödinger, LLC, New York, NY, 2016) [18] module in Schrödinger software, whereby OPLS2005 force field parameters were applied. The software relies on a grid-based ligand docking method with energetics, and searches for favorable interactions between one or more small ligand molecules and receptor molecules, often a protein. As soon as the protein and ligands are prepared in their proper forms for docking, a receptor grid file is created with a grid box of 80 x 80 x 80 Å with coordinates X= 125.0, Y= 30.0, and Z= 30. Additionally, docked ligands needed to be less than 20 Å. Default setting for ligand docking were employed, whereby SP (standard precision) was applied and ligand sampling was flexible with nitrogen inversions and ring conformations sampled. Bias sampling of torsions were used for amides only and nonplanar conformations were penalized. Epik state penalties were also included in the docking score.

Glide Algorithm

The Glide algorithm employs a 'Funnel' docking process [19-20], where a series of hierarchical filters are used as in order to achieve the best scoring ligand poses by evading false positives. Glide's 'funnel' docking process are carried out as follows:

- (1) Site point search
- (2) Ligand screening
 - (a) Diameter test
 - (b) Subset test
 - (c) Greedy score
 - (d) Refinement
- (3) Grid minimization
- (4) Final score (glide score)

Plasmids and Proteins

Human neutrophil elastase (HNE) were cloned into pMAL vectors for expression in *E. Coli* cells. To purify human neutrophil elastase (HNE) and the HECT domain of the human E3 ubiquitin ligase NEDD4-1, we expressed pMAL-TEV-HNE-his in BL21(DE3) cells, purified the proteins on amylose resin, cleaved them by TEV protease, and re-purified them on Ni-NTA agarose.

In Vitro Protein Thermal Shift Assays

All reactions were set up in 20 μL reactions in 96 well plates with total human NEDD4-1^{Hect} protein concentrations of 2.5 μM and 10x SYPRO orange dye (Invitrogen) [45]. The NEDD4-1 HECT domain was pre-incubated for 5 minutes with either I3C, 2244, 2160, or 2163 at 250 μM while concentrations of 25 μM was used for compounds 1-benzyl-I3C, 2242, and 2243. Thermal melting experiments were performed on the Via7 instrument (Life Technologies) melt curve program with a heat ramp rate of 0.05°C/s and a temperature range beginning at 25°C and ending at 95°C. Melting temperatures were analyzed with Protein Thermal Shift Software (Life Technologies) to identify the midpoint of the transition with an analysis of the first derivative. All experiments were performed in triplicates.

In Vitro Elastase Enzymatic Assays

Elastase activity was measured using Elastase EnzCheck assay kit (Molecular Probes; Invitrogen, Eugene, OR). Briefly, pure elastase at pH 7.0 was diluted in 1x reaction buffer (1 M Tris-HCl pH 8.0 containing 2 mM sodium azide) and pre-incubated with the specified concentrations of the specified drugs (tryptophol, I3C, 1-benzyl-I3C, 2160, 2163, 2242, 2243, 2244 or vehicle control DMSO) in protected 96-well plates at 25° for 1 hour. DQ-Elastin substrate labeled with fluorescent quenched BODIPY FL probe was added at the specified concentrations to the wells containing either pure human or cellular elastase to a total volume of 200 μM . The samples were gently agitated for 5 min on a shaker, and the fluorescence emitted by the elastase cleaved probe was measured at 485 nm excitation, 530 nm emission wavelength at 5 min intervals for a total of 90 min using SpectraMax (Molecular Devices, Sunnyvale, CA) M2 plate reader and SoftMax Pro v5 program (Molecular Devices, Sunnyvale, CA).

MCF-7 breast cancer cell culture and proliferation assay

The human MCF-7 breast cancer cell line was purchased from American Type Culture Collection (ATCC) (Manassas,VA), and was authenticated according to the ATCC guidelines. The cells were propagated according to ATCC protocol specifications and kept at low passage throughout the study. Cells were cultured in Modified McCoy's 5A cell media supplemented with 10% fetal bovine serum (Gemini Bio Products, West Sacramento, CA), 2 mM L-glutamine, and 2.5 ml of 10,000 U/ml penicillin/streptomycin mixture (Gibco, Life Technologies, Carlsbad, CA). Breast cancer cells were seeded on a 48-well plate in triplicates and upon 80-90% confluency were treated with I3C, 1-benzyl-I3C and the 1-benzyl-I3C analogues for 48 hours. Subsequently, inhibition of proliferation was assessed using the Dojindo Cell counting Kit -8 as per the protocol in the user's manual. Briefly, 50 μl of the CCK-8 solution was added to each well along with 450ul media and incubated for 2-3 hours. The absorbance was read at 450nm and % inhibition was calculated for each condition standardizing DMSO to zero.

Results

Molecular Modeling of Elastase Indolecarbinol Compounds Surface Pocket Identification in human neutrophil elastase crystal structure

Previous studies have demonstrated that I3C acts as a non-competitive inhibitor for human neutrophil elastase, implying that I3C binds outside the active site pocket, however, the precise binding location for I3C, 1-benzyl-I3C, and their corresponding analogs are not clearly defined (Figure 2.1). In order to characterize the structural requirements that can account for the indolecarbinol-specific inhibition of elastase enzymatic activity and to determine where each compound binds to the protein, Schrödinger's SiteMap was employed to identify potential binding regions for each ligand. SiteMap uses a scoring function to evaluate a site's proclivity for ligand binding and subsequently ranks pharmaceutically relevant binding sites. Suitable regions for ligand binding are determined by accessing ligand hydrogen-bond donors, acceptors, metal-binding functionality, and hydrophobic groups [16-17]. A SiteMap generates a SiteScore which is based on a weighted sum of properties which include the number of site points, the enclosure score, and the hydrophilic score. A score of greater than 1 suggests a site of potential ligand binding and a score of 0.8 can be differentiated between drug-binding and non-drug binding sites [16-17]. The human neutrophil elastase crystal structure (PDB accession code: 5A8X) was obtained from the Protein Data Bank and was used to model all indolecarbinol-elastase complex structures [14]. After employing the SiteMap algorithm, three sites (Figure 2.2) were identified. Because site 3 had a score below 0.8, this site was discarded as a viable ligand binding site, leaving behind site 1 and site 2 with scores of 0.99 and 0.85 respectively. As seen in Figure 2.2, site 2 is adjacent to the active site whereas site 1 is located near the C-terminus, outside of the active site cavity. Because site 1 yielded a higher SiteScore than site 2 and is positioned outside the active site, site 1 was selected as a potential candidate for docking the indolecarbinol compounds.

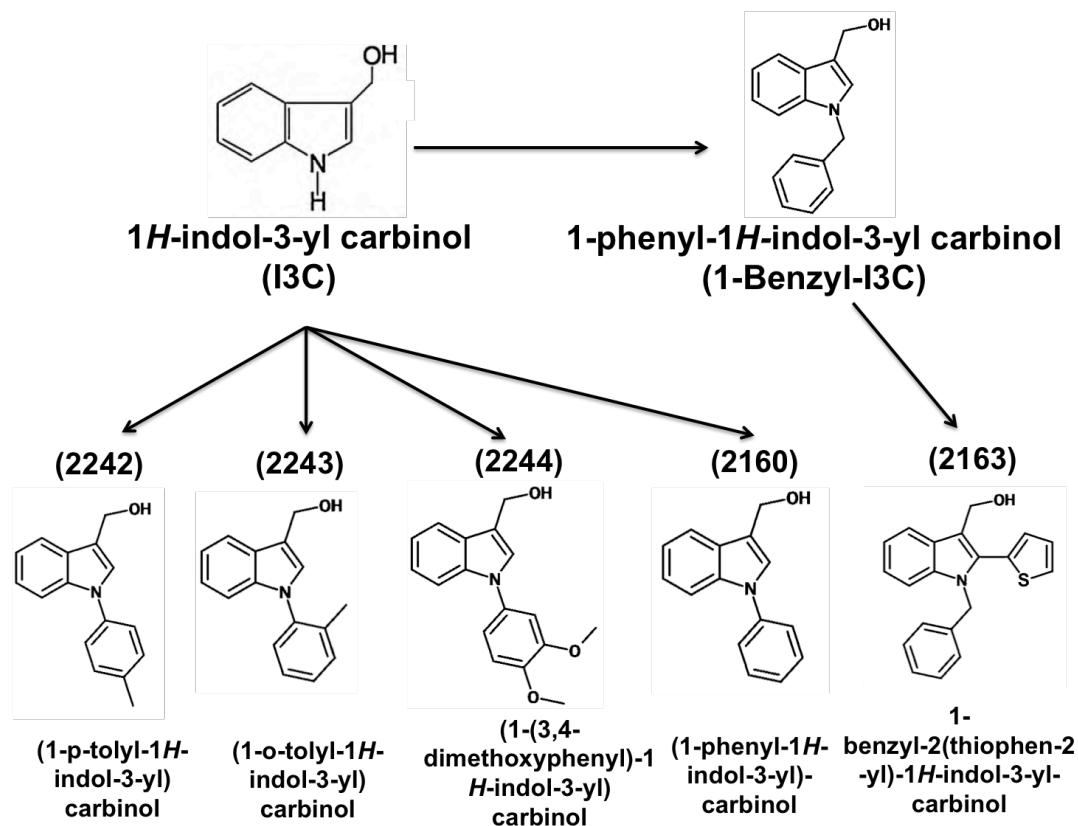


Figure 2.1. Structures of I3C, 1-benzyl-I3C, and indolecarbinol analogues. Each of the synthetic analogs maintains the C-3 hydroxy methyl substituent. Methyl substituents were added at the para and ortho positions on the benzene ring forming (1-*p*-tolyl-1H-indol-3-yl)-carbinol (compound 2242) and (1-*o*-tolyl-1H-indol-3-yl)-carbinol (compound 2243), respectively. Two methoxy substituents were attached to the para and meta positions in the benzene ring forming (1-(3,4-dimethoxyphenyl)-1H-indol-3-yl) carbinol (compound 2244). The linker carbon was eliminated from the benzyl group forming (1-phenyl-1H-indol-3-yl) carbinol (compound 2160). A thiophene moiety was added to carbon 2 in the indole ring forming benzyl-2(thiophen-2-yl)-1H-indol-3-yl-carbinol (compound 2163).



Figure 2.2. SiteMap's best predicted ligand binding sites on human neutrophil elastase. Ribbon diagram of elastase displaying the top ranked, highest probability ligand binding sites. Biochemical characterizations of each binding site is depicted by various colors. Yellow patches denote hydrophobic regions, blue denotes a hydrogen-bond donor, and red denotes a hydrogen bond acceptor.

To further characterize the indolecarbinol binding site pockets and as an initial attempt at determining precisely where each compound binds, rigid receptor docking was carried out using Schrödinger's Glide program. Here the binding affinity of each complex was recorded in GlideScore and the more negative GlideScore corresponded to more favorable binding. Calculated docking scores for the *in silico* defined interactions of each indolecarbinol compound with human neutrophil elastase yielded negative values of approximately -6 kJ/mol, indicating that the ligand-protein structures are predicted to be generally stable.

I3C derivative-elastase binding interactions were compared against I3C itself and to the more potent derivative, 1-benzyl-I3C to serve as a baseline for comparison. Structural analysis of HNE shows that it exhibits a highly basic surface due to 18 arginine residues which are balanced by only 6 acidic residues. Approximately 40% of the total amino acid side chains are hydrophobic and accessible to bulk solvent water molecules. Four Arg residues, Arg 36, Arg 147, Arg 177, and Arg 217, are located adjacent to the active site, spanning a total of 40 Å, and are separated by hydrophobic amino acids [21]. As a consequence, the protein's inherently basic nature allows for I3C, 1-benzyl-I3C, and their corresponding analogs to bind at its surface. All predicted interactions are within 4 Å of each inhibitor. Molecular modeling reveals that all indolecarbinol compounds are predicted to bind to a hydrophobic cavity, enveloped by residues Cys 168, Val 162, Cys 182, Val 163, Val 181, Leu 130, Ala 232, and Pro 230, away from the enzyme's catalytic site. The derivatives are also anchored between polar amino acid residues Gln 233, His 210, Asn 180, Thr 164 and positively charged amino acid residues Arg 178, Arg 129, Arg 128, and Arg 177. Differences in hydrogen bonding networks and pi-pi stacking interactions among the indolecarbinol analogs can be attributed to various changes in the substituents attached at the indole ring nitrogen. As depicted in Figure 2.3, Glide docking analysis predicts that the hydrophobic core of I3C is oriented in a hydrophobic pocket while the C-3 hydroxy methyl substituent that is attached to I3C's indole ring is also able to engage in hydrogen bonding interactions with the backbone of Val 163. Additionally, I3C's indole ring nitrogen is able to partake in hydrogen bonding interactions with the backbone of Arg 178 which allows I3C to be securely anchored. Contrastingly, the additional phenyl moiety in 1-benzyl-I3C allows it to associate with Arg 129 through cation- π bonds while the C-3 hydroxy methyl substituent that is attached to 1-benzyl-I3C's indole ring fastens the indole ring in a hydrophobic pocket where it is able to engage in hydrogen bonding with the backbones of Val 163 and Cys 182. While the indole ring in compounds 2242 and 2243 are positioned in the same orientation as I3C, the additional phenyl moiety on the indole ring nitrogen is securely embedded via van der Waals interactions with polar residues Asn 180 and Thr 164 and hydrophobic residues Val 181 and Cys 168. The 6 local $C^{\delta-}-H^{\delta+}$ bond dipoles around the benzene ring enables the compound to create an overall charge distribution that is composed of a negative charge in the center of the ring upon which cations bind, here more notably Arg177, Arg 129, and Arg 178. The two electron donating methoxy substituents that are attached to the para and meta positions in the benzene ring of compound 2244 are adjacent to the π system thereby increasing the electron density on the ring through a resonance donating effect that allows the π system to be more nucleophilic. The bulkiness of the additional 3,4-dimethoxy phenyl substituent displaces

the indole ring of compound 2244 and orients the indole ring in such a way that enables it to engage in stabilizing pi-pi stacking interactions with Arg 178 through cation- π bonds. In compound 2160, the linker carbon was eliminated from the benzyl moiety which makes the overall benzene ring more rigid and less likely to participate in nucleophilic interactions. The phenyl moiety associates with polar residues Val 181, Leu 130, Ala 232, and Pro 230 through van der Waals interactions and makes contact with positively charged amino acid residues His 210 and Gln 233. The orientation of the indole ring in compound 2160 positions it in a way that allows it to engage in pi-pi stacking interactions with Arg 178. A thiophene moiety was attached to carbon 2 in the indole ring of compound 2163. While the thiophene group adds hydrophobic character to this analog, it also significantly increases the bulkiness and decreases the flexibility of this compound. Here the indole ring is confined between hydrophobic residues Cys 168, Val 163, Cys 182, Val 162, and Val 181 and positively charged residues Arg 178, Arg 177, and Arg 128.

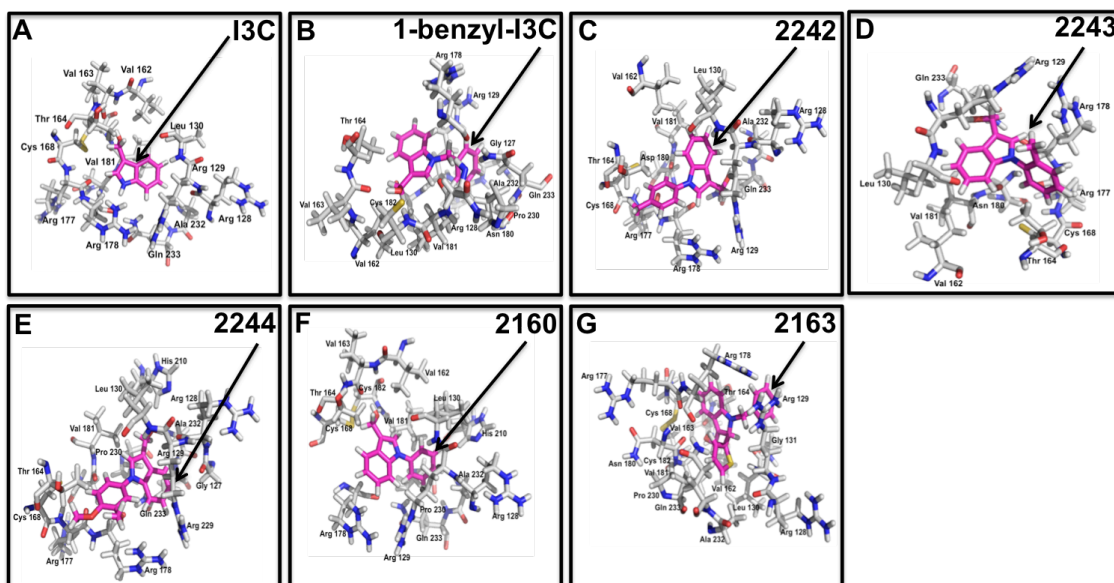


Figure 2.3. Predicted molecular interactions of human neutrophil elastase in complex with (A)I3C (B)1-benzyl-I3C (C) 2242 (D) 2243 (E) 2244 (F) 2160 and (G) 2163.

Stabilization of Human Neutrophil Elastase Upon Ligand Binding

Protein thermal shift assays were used to directly test whether the indolecarbinol compound can bind to purified human neutrophil elastase. Indolecarbinol-mediated alterations in thermostability of elastase are indicative of direct interactions between each potential ligand and the protein. Purified elastase at a concentration of 2.5 μM was pre-incubated with either 250 μM I3C, 2244, 2160, or 2163 while concentrations of 25 μM was used for compounds 1-benzyl-I3C, 2242, and 2243 or with the vehicle control and each mixture was then treated with Sypro orange dye, which binds to hydrophobic regions of proteins and emits a fluorescent signal as the protein unfolds. A 10-fold lower concentration of 1-benzyl-I3C compared to I3C was used because this I3C analog was previously shown to be significantly more potent than I3C in its ability to interact with elastase [23-24]. The binding reactions were slowly heated from 25°C to 95°C, and the unfolding of the elastase protein was continuously recorded from the fluorescent signals. Conditions that thermally stabilize the protein will typically shift the melting temperature upwards by 4°C to greater than 20°C. Generally as an empirical estimate, a greater than 4°C shift in melting temperature corresponds to binding of a ligand with a K_d of approximately $<1\mu\text{M}$, although the magnitude of the temperature shift does not necessarily correlate with ligand affinity [22]. As shown in Figure 2.4, in the absence of any added indolecarbinol compound, the first derivative elastase melting profile yielded an apparent T_m of 47°C. However, upon addition of certain indolecarbinols, the stabilization of the protein's secondary structure and concomitant increase in melting temperature were observed. Compound 2163 produced a -2°C change in HNE's melting temperature, whereas compound 2242 and compound 2244 increased the protein's temperature by 6°C and 8°C respectively. In the presence of 1-benzyl-I3C and 2243, HNE exhibited unfolding transitions at significantly higher temperatures (ΔT_m values of 12°C for both compounds) and reached maximum stability in the presence of I3C and 2160 (ΔT_m values of 13°C and 14°C for I3C and 2160 respectively).

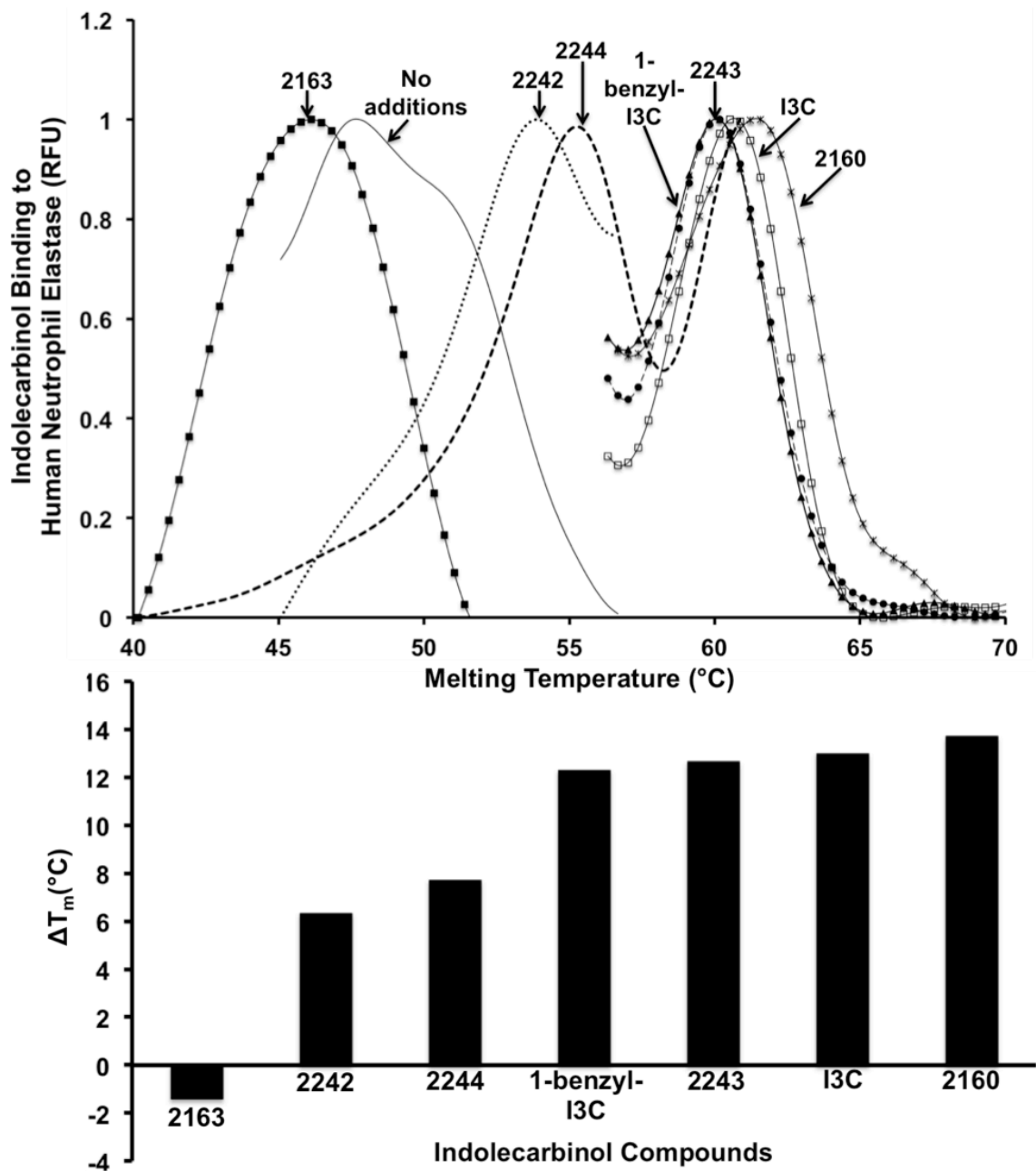


Figure 2.4. Protein Thermal Shift Assay. Incubation with 1-Benzyl-I3C and corresponding analogs and purified human neutrophil elastase. (A) First derivative melt profile of human neutrophil elastase in the absence and presence of indolecarbinols. (B) Graph depicting change in melting temperature of elastase in the presence of indolecarbinols.

Indole Specificity of Elastase Inhibition

We previously demonstrated that not only is elastase an I3C target protein and that this indolecarbinol acts as a non-competitive allosteric inhibitor of elastase enzymatic activity, but 1-benzyl-I3C serves as a more potent inhibitor of elastase, significantly inhibiting elastase activity more efficiently than I3C at lower concentrations. *In vitro* enzymatic assays were employed to determine whether derivatives of I3C and 1-benzyl-I3C similarly identifies elastase as a specific target protein and inhibits enzymatic activity. The biologically inactive derivative tryptophol, which contains an ethanol group attached to the C3 ring carbon, was used as a negative control whereas 1-benzyl-I3C was used to establish a baseline level of inhibition. For the *in vitro* reactions, pure human neutrophil elastase was used and the DQ elastin-conjugated to BODIPY FL dye was used as the substrate whereby a fluorescent cleavage product is released upon elastase-specific processing. The effects of each indolecarbinol on elastase activity was assessed by measuring the fluorescence signal in each *in vitro* processing reaction. As depicted in Figure.2.5, I3C inhibited elastase with a half maximal inhibition at IC_{50} 208 μ M whereas 1-benzyl-I3C inhibited elastase activity more significantly with an IC_{50} of 13 μ M as previously reported from our lab. 2163 and 2244 were the weakest compounds to inhibit elastase activity with a half maximal inhibition of 250 μ M and 343 μ M respectively. Interestingly, compounds 2242, 2160, and 2243 inhibited elastase activity much more significantly with and an IC_{50} of 30 μ M, 25 μ M, and 17 μ M respectively. The potency of compound 2242, 2160, and 2243 appeared to be sensitive to structural changes on the phenyl ring, more notably when methyl substituents were added to the para and ortho positions. The activity was abolished with the addition of bulky chemical groups like the thiophene substituent attached to the indole of 2163 and the electron donating methoxy group attached to the phenyl moiety of 2244 in the meta and ortho positions. As suggested by the computational models, the substituents of the compounds 2163 and 2244 are solvent-exposed, and hence, unfavorable solvation of the bulky and hydrophobic groups may justify the activity loss.

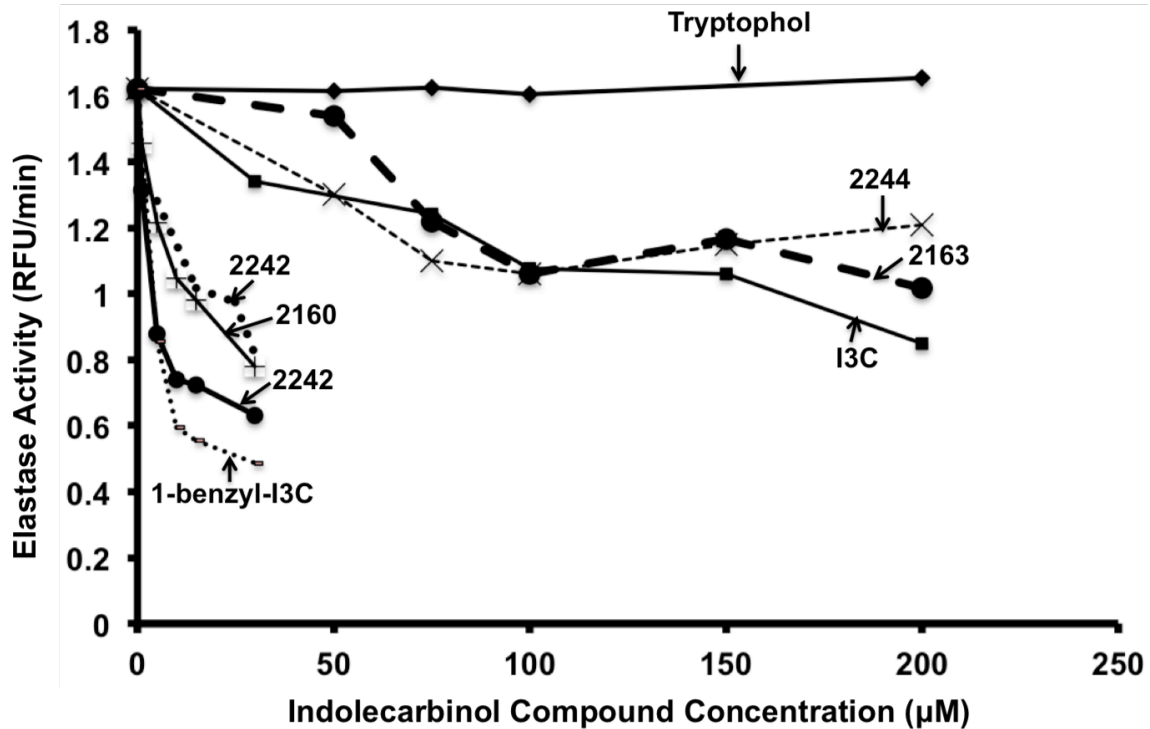


Figure 2.5. *In vitro* elastase enzymatic activity. (A) Kinetic enzymatic assay comparing the effects of elastase on the substrate elastin when incubated with I3C and various derivatives. Purified human neutrophil elastase was pre-incubated with the indicated concentrations of tryptophol, I3C, 1-benzyl-I3C, 2160, 2163, 2242, 2243, and 2244, and then incubated with the DQ-Elastin substrate labeled with the fluorescent quenched probe BODIPY FL. The fluorescence intensity was measured at 485 nm excitation and 530 nm emission wavelengths.

Because of the connection between elastase enzymatic activity and the proliferation of MCF-7 cells, the anti-proliferative effectiveness of 48 hour treatment with a concentration range of each indolecarbinol analog was evaluated in comparison to 1-benzyl-I3C, I3C and the inactive indolecarbinol tryptophol (Figure 2.6). Breast cancer cell proliferation was determined by a CCK-8 proliferation assay that measures the number of cells. As observed in Figure 2.7, the most effective indolecarbinol compounds were 1-benzyl-I3C and compound 2244 which displayed a half maximal inhibition of breast cancer cell proliferation of 0.90 μM and 32 μM respectively. The anti-proliferative efficacies of compounds 2160, 2243, and 2242 (half maximal responses of 41 μM , 58 μM , 72 μM , respectively) were intermediate between I3C and the two highly potent compounds 1-benzyl-I3C and compound 2244. I3C had the least profound anti-proliferative effects with a half-maximal response of 157 μM . Tryptophol had no effect on breast cancer cell proliferation, which is consistent with its lack of effects in other cancer types. (Cells treated with compound 2163 could not be assessed because of immediate cell death upon treatment.) A distinct correlation in the efficacies of the indolecarbinol compounds can be observed between the half maximal anti-proliferative responses and half maximal inhibition of elastase enzymatic activities. One notable deviation of this correlation is with compound 2244 in which the relative inhibition of elastase activity is less effective compared to its relative anti-proliferative response which may likely be a consequence of compound 2244 binding to a different target protein that controls cellular proliferation in breast cancer.

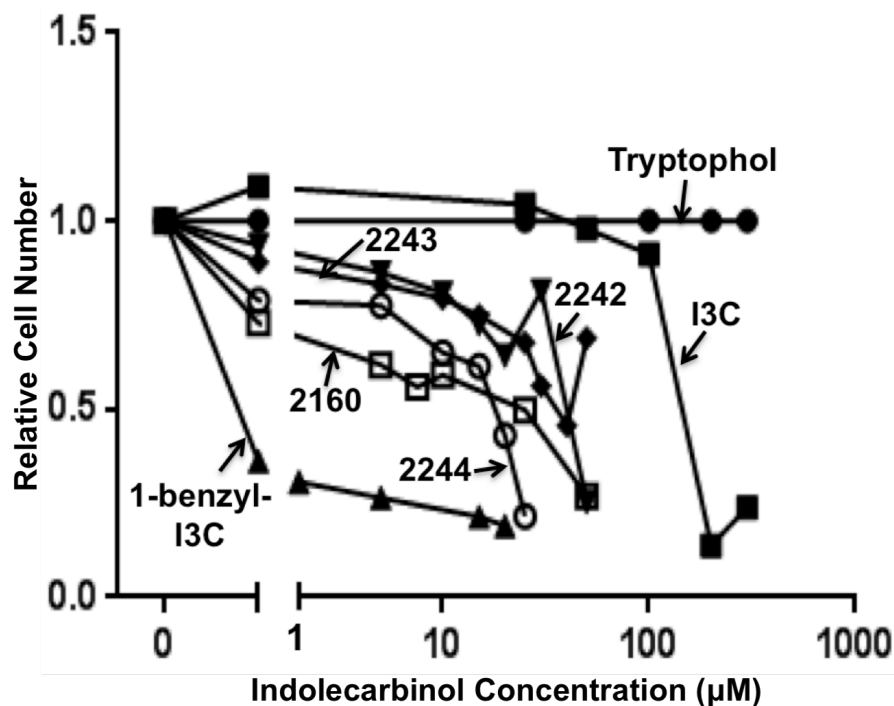


Figure 2.6. Cell proliferation assay depicting percent of viable cells at each respective dose of indolecarbinol compound in MCF-7 cells. (Assay performed by Kevin Poindexter)








Indolecarbinol Compound	Structure	Half Maximum Inhibitory Concentration for Elastase Enzymatic Activity (μM)	Half Maximum Concentration for Growth Inhibition in MCF-7 Cells (μM)
I3C		208	157
1-benzyl-I3C		13	0.9
2242		30	72
2243		17	58
2244		343	32
2160		25	41
2163		250	-

Figure 2.7. IC_{50} values of indolecarbinol compounds with human neutrophil elastase and half maximum concentration for growth inhibition in MCF-7 cells.

Efficacy of the anti-proliferative responses of the indolecarbinol compounds

To determine the indolecarbinol compound specificity of elastase activity inhibition, MCF-7 breast cancer cells were treated with either 200 μ M tryptophol, which is a biologically inactive I3C analog, 200 μ M I3C, 15 μ M 1-benzyl-I3C, 50 μ M 2242, 40 μ M 2243, 25 μ M 2244, 50 μ M 2160, or the DMSO vehicle control for 48 hours and the potential cell cycle effects were evaluated by flow analyzed by flow cytometry of propidium iodide-stained nuclear DNA. As shown in Figure XX, MCF-7 cells that were treated with the DMSO vehicle control displayed cells in all phases of the cell cycle including approximately 56% in the G1 phase, which is indicative of a proliferative state. Correspondingly, the cell cycle profile after tryptophol treatment was similar to the vehicle control-treated cells (DMSO). Contrastingly, treatment with 200 μ M I3C increased the percentage of cells with a G1 DNA content in the overall cell population to 75% which is consistent with a G₁ cell cycle arrest. More notably, treatment with 15 μ M 1-benzyl-I3C, 40 μ M 2243 or 25 μ M 2244 induced a G₁ cell cycle arrest with 64%, 65%, or 68% of the total cells arrested in the G1 phase of the cell cycle respectively, while a relatively small amount of cells remained in the S and G₂/M cell cycle phases (Fig. 2.8). Treatment with 50 μ M 2242 or 50 μ M 2160, however, were less effective in inducing a G1 cell cycle arrest in the G1 phase of the cell cycle, with only 59% or 63% of cells in the G1 phase of the cell cycle respectively. Taken together, our results show that 1-benzyl-I3C, 2243, and 2244 display a significantly more potent anti-proliferative response compared to its parental compound I3C that generally demonstrates the increased efficacy for the inhibition of elastase activity.

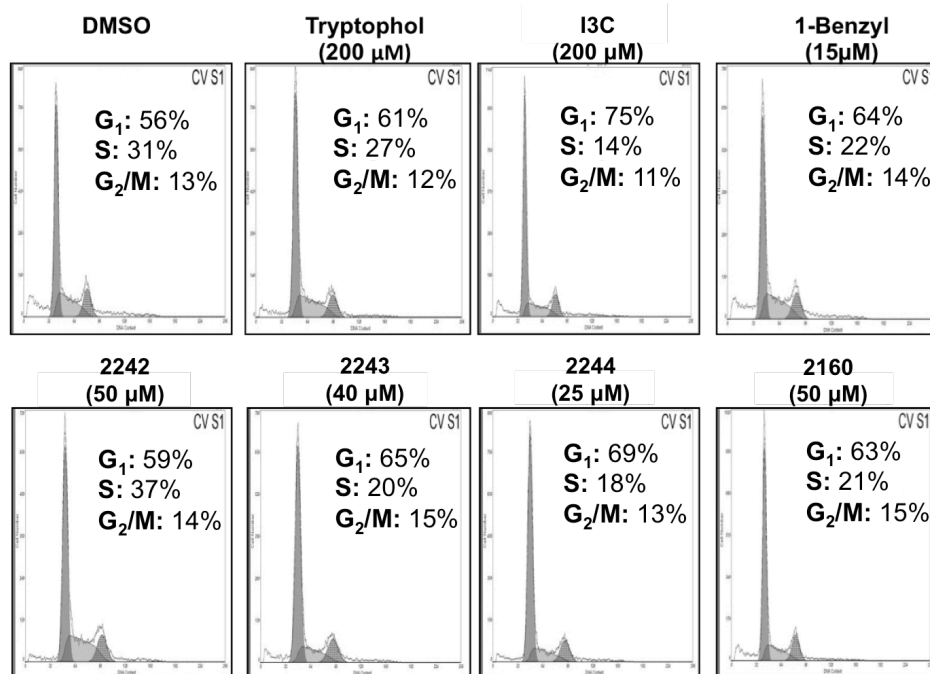


Figure 2.8. Flow cytometry profiles of MCF-7 cells with indicated doses of I3C, 1-benzyl-I3C, and indolecarbinol analogues. (Assay performed by Kevin Poindexter)

Discussion

The quest to identify target proteins for natural indolecarbinol compounds that mediate anti-carcinogenic responses in human cancer has been an encumbrance for the understanding of the anti-cancer mechanism of phytochemicals. It has been previously established that increased consumption of Brassica vegetables is directly correlated to a decline in human cancer risk and in experimental animals a reduction in tumor frequency [25, 26]. Studies conducted in our lab have proven that in human breast cancer cells, I3C, which is a primary component in Brassica vegetables, promotes a G1 cell cycle arrest through a cellular pathway that diminishes CDK6 transcription and inhibits CDK2 kinase activity [27-30]. Further investigation from our lab has recognized human neutrophil elastase as the first identified I3C target protein in human breast cancer cells and delineated its mechanism of inhibition which is directly associated to regulation of breast cancer cellular proliferation [31, 32]. After performing an siRNA knockdown of elastase, the I3C cellular responses as well as the I3C-induced G1 cell cycle arrest were mimicked, which testifies to elastase's role as an I3C target protein. Aronchik et al had performed *in silico* studies using shape and electrostatics as restrictive parameters and employed the known crystal structure of human neutrophil elastase to make predictions about potential indolecarbinol binding sites. Glide analysis indicated that both I3C and 1-benzyl-I3C engaged in selective hydrogen bonding to residues in the C-terminal domain, outside the active site of elastase which is consistent with the data that prove I3C serves as an allosteric non-competitive inhibitor. The computational docking model was utilized as an aide for synthesizing a novel I3C-resistant form of elastase by abridging the C-terminal domain which resulted in a $\Delta 205$ elastase construct which disrupted the interaction between I3C and benzyl-I3C and their corresponding contact residues. While the truncated elastase remained enzymatically active, it was resistant to I3C and 1-benzyl-I3C's inhibitory effects. Additionally, expression of the $\Delta 205$ elastase construct in breast cancer cells rendered a cellular phenotype that was resistant to I3C and 1-benzyl-I3C's dependent loss of CD40 protein processing, ablated NF κ B nuclear localization and promotion of a G1 cell cycle arrest [33]. The aforementioned studies therefore validate the molecular docking predictions. Disentangling the I3C and 1-benzyl-I3C engendered cell cycle and apoptotic responses in human cancer cells through the physical modification of elastase proves that individual target proteins can discriminately mediate the indolecarbinol sensitivity of specific anti-proliferative signaling.

One major limitation in employing I3C is the requirement for relatively high doses. Additionally in cancer cells, I3C condenses into several major and minor dimeric and trimeric products that collectively generate an additive physiological response [26, 34]. When breast cancer cells are treated with I3C, the compound remains stable, however, an extensive portion is transformed into DIM, its self-condensation product [35] and together, I3C and DIM can induce a robust anti-proliferative effect. I3C and DIM independently suppress the growth of reproductive cancer cells through distinct mechanisms. Therefore, a fraction of the I3C responses in reproductive cancer cells could be a consequence DIM's effects. While other labs have made attempts in synthesizing more powerful I3C-derived compounds, they have only achieved limited

success because many compounds exhibit a generally low efficacy of inhibition with compounds demonstrating only a 5-10 fold increase in efficiency [36] and only one compound exhibiting as much as 100-fold increased strength for apoptosis [37]. Implementing a systematic structure-activity approach in developing more potent I3C-derived compounds, our lab has previously established that the addition of substituents to the nitrogen position 1 in the indole ring with more hydrophobic alkyl groups exhibit the highest efficacy [38]. The results of these studies have inaugurated a novel I3C analog, 1-benzyl-I3C, which manifested a significantly enhanced hydrophobic character. Furthermore, 1-benzyl-I3C is more stable than I3C which may be a consequence of the compound's additional phenyl moiety which impedes the 1-benzyl-indol-3-yl methyl cation oligomerization. While the metabolism of 1-benzyl-I3C has not sufficiently been elucidated, it can be gleaned from the structure that the generation of cydroxylated or epoxydated compound in the phenyl ring *in vitro* is not feasible, though the potential for metabolic degradadation *in vivo* to some degree could lead to an N-debenzylated product or to a hydroxylated or oxidated compound. In spite of this, our lab has proved that 1-benzyl-I3C remains stable and does not deteriorate even after a period of 10 days in an ethanol-citrate phosphate buffer solution at a pH of 5.5 [32]. Therefore, these studies suggest that the condensation of 1-benzyl-I3C into dibenzyl-DIM *in vitro* (at neutral or greater pH) transpires very slowly and thus, the biological effects of 1-benzyl-I3C are not a result of its transformation into a DIM-type analog.

Studies conducted in our lab have stabled that 1-benzyl-I3C has a higher potency than I3C, with an approximately 1000-fold elevated efficacy of the cell cycle arrest of both human estrogen responsive and estrogen-independent human breast cancer cells. Additionally, 1-benzyl-I3D has an IC_{50} value in the hundreds of nano-molar range unlike I3C which has an IC_{50} value in the micromolar range, and hence can overcome the issue of employing relatively high doses of I3C that is necessary for inhibiting the growth of human breast cancer cells.

1-benzyl-I3C is also able to sustain the I3C anti-proliferative responses that are observed in estrogen-sensitive and independent human breast cancer cells, notably with a G1 cell cycle arrest regulation of G1-acting cell cycle components thus implying that the additional phenyl moiety substituent does not disturb I3's biological characteristics. More notably, upon 1-benzyl-I3C treatment in human breast cancer cells, CDK2 enzymatic activity was strongly suppressed and CDK6 transcript and protein levels were mitigated which accounts for a loss of cellular CDK6 activity. Moreover, in human breast cancer cells, 1-benzyl-I3 dramatically diminished CDK4 kinase activity in comparison to the moderate effects of I3. Correspondingly, in contrast to I3C, 1-benzyl-I3C can stimulate levels of the p21 and p27 CDK inhibitors, and hence, 1-benzyl-I3C has a stronger and more comprehensive efficacy of its cell cycle effects. The approximately 1000-fold enhanced potency of 1-benzyl-I3C towards the restriction of human breast cancer cellular proliferation may be a consequence of the collective inhibition of cellular kinase activities of all three G1-activating cylin/CDK protein complexes.

Additionally, because 1-benzyl-I3C can efficiently restrict the growth of hormone responsive and hormone unresponsive breast cancer cells, this implies that this

compound harbors powerful potential for treating a comprehensive range of various breast cancer types which has been an intrinsic obstacle with the available therapeutic treatments.

1-benzyl-I3C also proved to be a powerful compound in an *in vivo* environment and was able to induce an anti-proliferative response in human breast cancer cell-derived tumors in athymic mice.

Based on the inhibition of breast cancer cell growth in culture and *in vivo* in tumors [32], 1-benzyl-I3C is a more effective I3C-based analog. This may be attributed to a combination of its cellular stability and its ability to more forcefully inhibit elastase enzymatic activity, thus making 1-benzyl-I3C a powerful compound in selectively inhibiting elastase. In light of these studies that established 1-benzyl-I3C as a more potent indole-based inhibitor, 1-benzyl-I3C thus exhibits immense potential in serving as a scaffold for generating novel, more powerful, chemotherapeutic agents.

Our results demonstrate that I3C, 1-benzyl-I3C, and its corresponding synthetic analogs are effective inhibitors of human neutrophil elastase activity and the parent indolecarbinol ring structure represent a novel class of small molecule inhibitors of elastase. We assessed the *in vitro* binding and inhibition of enzymatic activity of purified elastase by I3C, 1-benzyl-I3C and five additional synthetic indolecarbinol-derived compounds that exhibit selective modifications in both the physical and chemical properties of the final indolecarbinol structure [31-33]. The nucleophilic π system within the phenyl ring is necessary for the indolecarbinol compounds to effectively inhibit elastase activity. It was observed that as the rigidity of the chemical scaffold increased, the functionality of the indolecarbinol compounds decreased. It was also important that each new, synthetic indolecarbinol compound retain the C-3 hydroxy methyl substituent in order for the I3C parental structure to sustain its biological activity [38]. Furthermore, the general potency of the anti-proliferative effects in breast cancer cells of the tested indolecarbinol compounds roughly corresponded to their *in vitro* inhibition of elastase activity which substantiates the significance of the indolecarbinol structure in generating stable and effective inhibitors.

Molecular docking simulations predicted stable interactions of each of the indolecarbinol compounds with the crystal structure of human neutrophil elastase. Treatment with 1-benzyl-I3C and compound 2243 significantly inhibited elastase enzymatic activity. The electron donating substituents attached to the nitrogen of the indole ring of I3C may contribute to the compounds' inhibitory effects. The results are a testament to the importance of an optimal benzene ring nucleophilic π system for inhibiting elastase enzymatic activity. To enhance our understanding of the binding of the protein-ligand complexes and to gain structural information for the indolecarbinol compounds in complex with elastase, we first employed Maestro's SiteMap in the Schrödinger software suite to identify possible indolecarbinol binding sites based on size, functionality, and degree of solvent exposure. The scoring function implemented in SiteScore can assess a site's penchant for ligand binding and can rank sites of promise while also disregarding sites that are not likely to be pharmaceutically relevant. Glide

grids in the Schrödinger software suite were subsequently implemented for virtual screening of the indolecarbinol compounds. The Glide docking simulations predicted highly cohesive interactions with hydrophobic residues Leu 130, Val 162 Cys 168, Val 181, Pro 230, positively charged residues Arg 128, Arg 129, Arg 177, Arg 178, and polar residues Thr 164, Asn 180, and Gln 233 and compound 2242, 2243, and 2160 within a region outside of the elastase active site. Compound 2244 harbors two electron donating methoxy substituents oriented in the para and meta positions in the benzene ring and results in an approximate 27-fold less efficacy compared to 1-benzyl-I3C. The molecular docking analysis suggests that one explanation for this reduced effectiveness is the inability of the compound to be stably anchored to the hydrophobic pocket and hence loses interactions with its benzene substituent and hydrophobic residues Leu 130, Val 181, Pro 230, and Ala 232. Instead the position the compound 2244 has been flipped in such a way that the indole ring is now oriented towards the hydrophobic pocket and the phenyl moiety is pointed away. *In silico* modeling predicts that compound 2163 can still maintain hydrophobic contacts with Leu 130, Pro 230, and Ala 232 through its thiophene moiety, however, the bulkiness of this additional substituent displaces the indole ring from its pocket which thus may account for the loss of inhibitory activity of the 1-benzyl-I3C scaffold.

In line with the computational docking studies, protein thermostability assays demonstrated that each indolecarbinol compound binds *in vitro* to purified elastase. Since it has been previously established that I3C serves as a non-competitive inhibitor and molecular docking studies predict that all indolecarbinol compounds bind outside the enzyme active site, it is possible that the newly synthesized indolecarbinol compounds inhibit elastase activity through a similar mechanism, however additional studies need to be performed in order to ascertain each compound's exact mechanism of inhibition.

The anti-proliferative effects of the indolecarbinol compounds can be attributed to the inhibition of elastase activity and the resulting disruption of down stream signaling. In breast cancer tissues, elastase levels are typically higher and are classified as an independent prognostic factor for breast cancer associated with a poor clinical disease outcome. Elastase is an essential target in the treatment of breast cancer and thus, the control of elastase activity brings forth beneficial therapeutic promise. The circulating fractions of elastase as well as the binding of its physiological inhibitor α 1-anti-trypsin are what exerts control over the active and inactive elastase equilibrium under physiological conditions [39-41]. In various diseased states which include pulmonary ailments such as chronic obstructive pulmonary disease, cystic fibrosis, and breast cancer, low levels of α 1-anti-trypsin or elevated levels of expression or activity of elastase can cause a loss of inhibitory balance and promote inordinate elastase substrate processing [32, 42, 43]. While there have been many classes of synthetic and natural compound that have been examined for their promise to inhibit elastase proteolytic activity both *in vitro* and *in vivo* and to mitigate the pathogenesis of elastase-dependent disorders, many of these treatments require daily intravenous injections and evoke many deleterious side effects which include liver toxicity [44].

Therefore, based on the enhanced potency and higher specificity of inhibition of elastase activity and the ability to still sustain meager toxicity, indolecarbinol-derived therapeutics can be developed as a safer, less harmful approach for treating cancer. The identification of more indolecarbinol target proteins associated with anti-proliferative pathways that do not implicate elastase for regulation will be beneficial in the evolution of novel anti-cancer clinical approaches employing indolecarbinol-based analogs either alone or in combination with other known inhibitors for treatment of indolecarbinol-sensitive human cancers.

References

1. Kim YS, Milner JA. Targets for indole-3-carbinol in cancer prevention. *J Nutr Biochem* 2005; 16: 65-73.
2. Aggarwal BB, Ichikawa H. Molecular targets and anticancer potential of indole-3-carbinol and its derivatives. *Cell Cycle* 2005; 4: 1201-15.
3. Moiseeva EP, Almeida GM, Jones GD, Manson MM. Extended treatment with physiologic concentrations of dietary phytochemicals results in altered gene expression, reduced growth, and apoptosis of cancer cells. *Mol Cancer Ther* 2007; 6: 3071-9.
4. Safe S, Papineni S, Chintharlapalli S. Cancer chemotherapy with indole-3-carbinol, bis(3'-indolyl)methane and synthetic analogs. *Cancer Lett* 2008; 269: 326-38.
5. Aronchik I, Chen T, Durkin KA, et al. Target protein interactions of indole-3-carbinol and the highly potent derivative 1-benzyl-I3C with the C-terminal domain of human elastase uncouples cell cycle arrest from apoptotic signaling. *Mol Carcinog*; 51: 881-94.
6. Brew CT, Aronchik I, Hsu JC, et al. Indole-3-carbinol activates the ATM signaling pathway independent of DNA damage to stabilize p53 and induce G1 arrest of human mammary epithelial cells. *Int J Cancer* 2006; 118: 857-68.
7. Melkamu T, Zhang X, Tan J, Zeng Y, Kassie F. Alteration of microRNA expression in vinyl carbamate-induced mouse lung tumors and modulation by the chemopreventive agent indole-3-carbinol. *Carcinogenesis*; 31: 252-8.
8. Machijima Y, Ishikawa C, Sawada S, et al. Anti-adult T-cell leukemia/lymphoma effects of indole-3-carbinol. *Retrovirology* 2009; 6: 7.
9. Kim DS, Jeong YM, Moon SI, et al. Indole-3-carbinol enhances ultraviolet B-induced apoptosis by sensitizing human melanoma cells. *Cell Mol Life Sci* 2006; 63: 2661-8.
10. Qi M, Anderson AE, Chen DZ, Sun S, Auburn KJ. Indole-3-carbinol prevents PTEN loss in cervical cancer in vivo. *Mol Med* 2005; 11: 59-63.
11. Brew CT, Aronchik I, Kosco K, McCammon J, Bjeldanes LF, Firestone GL. Indole-3-carbinol inhibits MDA-MB-231 breast cancer cell motility and induces stress fibers and focal adhesion formation by activation of Rho kinase activity. *Int J Cancer* 2009; 124: 2294-302.
12. Nguyen HH, Aronchik I, Brar GA, Nguyen DH, Bjeldanes LF, Firestone GL. The dietary phytochemical indole-3-carbinol is a natural elastase enzymatic inhibitor that disrupts cyclin E protein processing. *Proc Natl Acad Sci U S A* 2008; 105: 19750-5.
13. Aronchik I, Bjeldanes LF, Firestone GL. Direct inhibition of elastase activity by indole-3-carbinol triggers a CD40-TRAF regulatory cascade that disrupts NF-kappaB transcriptional activity in human breast cancer cells. *Cancer Res*; 70: 4961-71.
14. von Nussbaum F, Li VM, Meibom D, Anlauf S, Bechem M, Schamberger J, et al. Potent and Selective Human Neutrophil Elastase Inhibitors with Novel Equatorial Ring Topology: in vivo Efficacy of the Polar Pyrimidopyridazine BAY-8040 in a

- Pulmonary Arterial Hypertension Rat Model. *ChemMedChem* 2016; 11(2): 199-206.
15. Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W: Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des* 2013, 27(3):221-234.
 16. Halgren, T. Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inf. Model* 2009; 49: 377-389.
 17. Patschull AO, Gooptu B, Ashford P, Daviter T, Nobeli I. In silico assessment of potential druggable pockets on the surface of α 1-antitrypsin conformers. *PLoS One* 2012; 7(5): e36612.
 18. Friesner, RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med Chem* 2006; 49:6177-6196.
 19. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem* 2004; 47: 1750-1759.
 20. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shaw DE, Shelley M, Perry JK, Francis P, Shenkin PS. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem* 2004; 47:1739-1749.
 21. Korkmaz B, Horwitz MS, Jenne DE, Gauthier F. Neutrophil Elastase, Proteinase 3, and Cathepsin G as Therapeutic Targets in Human Diseases. *Pharmacol Rev* 2010; 62(4): 726-759.
 22. Vedadi M, Niesen FH, Allali-Hassani A, Fedorov OY, Finerty PJ, Jr., Wasney GA, Yeung R, Arrowsmith C, Ball LJ, Berglund H *et al*: Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc Natl Acad Sci U S A* 2006; 103(43):15835-15840.
 23. Aronchik I, Chen T, Durkin KA, Horwitz MS, Preobrazhenskaya MN, Bjeldanes LF, Firestone GL. Target protein interactions of indole-3-carbinol and the highly potent derivative 1-benzyl-I3C with the C-terminal domain of human elastase uncouples cell cycle arrest from apoptotic signaling. *Mol Carcinog* 2012; 51(11):881-894.
 24. Nguyen HH,, Lavrenov SN, Sundar SN, Nguyen DH, Tseng M, Marcoett CN, Kung J, Staub RE, Preobrazhenskaya MN, Bjeldanes LF *et al*. 1-benzyl-indole-3-carbinol is a novel indole-3-carbinol derivative with significantly enhanced potency of anti-proliferative and anti-estrogenic properties in human breast cancer cells. *Chem Biol Interact* 2010; 186(3):255-266.
 25. Lopez-Otin C, Diamandis EP: Breast and prostate cancer: An analysis of common epidemiological, genetic, and biochemical features. *Endocr Rev* 1998, 19(4):365-396.
 26. Aggarwal BB, Ichikawa H: Molecular targets and anticancer potential of indole-3-carbinol and its derivatives. *Cell Cycle* 2005, 4(9):1201-1215.
 27. Firestone GL, Bjeldanes LF: Indole-3-carbinol and 3-3'-diindolylmethane

- antiproliferative signaling pathways control cell-cycle gene transcription in human breast cancer cells by regulating promoter-Sp1 transcription factor interactions. *J Nutr* 2003, 133(7):2448s-2455s.
28. Cover CM, Hsieh SJ, Tran SH, Hallden G, Kim GS, Bjeldanes LF, Firestone GL: Indole-3-carbinol inhibits the expression of cyclin-dependent kinase-6 and induces a G(1) cell cycle arrest of human breast cancer cells independent of estrogen receptor signaling. *J Biol Chem* 1998, 273(7):3838-3847.
 29. Cram EJ, Liu BD, Bjeldanes LF, Firestone GL: Indole-3-carbinol inhibits CDK6 expression in human MCF-7 breast cancer cells by disrupting Sp1 transcription factor interactions with a composite element in the CDK6 gene promoter. *J Biol Chem* 2001, 276(25):22332-22340.
 30. Garcia HH, Brar GA, Nguyen DHH, Bjeldanes LF, Firestone GL: Indole-3-carbinol (I3C) inhibits cyclin-dependent kinase-2 function in human breast cancer cells by regulating the size distribution, associated cyclin E forms, and subcellular localization of the CDK2 protein complex. *J Biol Chem* 2005, 280(10):8756-8764.
 31. Aronchik I, Bjeldanes LF, Firestone GL: Direct Inhibition of Elastase Activity by Indole-3-Carbinol Triggers a CD40-TRAF Regulatory Cascade That Disrupts NF-kappa B Transcriptional Activity in Human Breast Cancer Cells. *Cancer Res* 2010, 70(12):4961-4971.
 32. Nguyen HH, Aronchik I, Brar GA, Nguyen DHH, Bjeldanes LF, Firestone GL: The dietary phytochemical indole-3-carbinol is a natural elastase enzymatic inhibitor that disrupts cyclin E protein processing. *P Natl Acad Sci USA* 2008, 105(50):19750-19755.
 33. Aronchik I, Chen T, Durkin KA, Horwitz MS, Preobrazhenskaya MN, Bjeldanes LF, Firestone GL: Target Protein Interactions of Indole-3-Carbinol and the Highly Potent Derivative 1-Benzyl-I3C With the C-Terminal Domain of Human Elastase Uncouples Cell Cycle Arrest From Apoptotic Signaling. *Mol Carcinogen* 2012, 51(11):881-894.
 34. Riby JE, Feng CL, Chang YC, Schaldach CM, Firestone GL, Bjeldanes LF: The major cyclic trimeric product of indole-3-carbinol is a strong agonist of the estrogen receptor signaling pathway. *Biochemistry-Us* 2000, 39(5):910-918.
 35. Staub RE, Feng CL, Onisko B, Bailey GS, Firestone GL, Bjeldanes LF: Fate of indole-3-carbinol in cultured human breast tumor cells. *Chem Res Toxicol* 2002, 15(2):101-109.
 36. Brandi G, Paiardini M, Cervasi B, Fiorucci C, Filippone P, De Marco C, Zaffaroni N, Magnani M: A new indole-3-carbinol tetrameric derivative inhibits cyclin-dependent kinase 6 expression, and induces G(1) cell cycle arrest in both estrogen-dependent and estrogen-independent breast cancer cell lines. *Cancer Res* 2003, 63(14):4028-4036.
 37. Weng JR, Tsai CH, Kulp SK, Wang D, Lin CH, Yang HC, Ma Y, Sargeant A, Chiu CF, Tsai MH *et al*: A potent indole-3-carbinol-derived antitumor agent with pleiotropic effects on multiple signaling pathways in prostate cancer cells. *Cancer Res* 2007, 67(16):7815-7824.
 38. Jump SM, Kung J, Staub R, Kinseth MA, Cram EJ, Yudina LN, Preobrazhenskaya MN, Bjeldanes LF, Firestone GL: N-Alkoxy derivatization of indole-3-carbinol increases the efficacy of the G1 cell cycle arrest and of 13C-

- specific regulation of cell cycle gene transcription and activity in human breast cancer cells. *Biochem Pharmacol* 2008, 75(3):713-724.
39. Korkmaz B, Attucci S, Jourdan ML, Juliano L, Gauthier F: Inhibition of neutrophil elastase by alpha 1-protease inhibitor at the surface of human polymorphonuclear neutrophils. *J Immunol* 2005, 175(5):3329-3338.
 40. Brower MS, Harpel PC: Alpha-1-Antitrypsin Human-Leukocyte Elastase Complexes in Blood - Quantification by an Enzyme-Linked Differential Antibody Immunosorbent-Assay and Comparison with Alpha-2-Plasmin Inhibitor Plasmin Complexes. *Blood* 1983, 61(5):842-849.
 41. Vogelmeier C, Hubbard RC, Fells GA, Schnebli HP, Thompson RC, Fritz H, Crystal RG: Antineutrophil Elastase Defense of the Normal Human Respiratory Epithelial Surface Provided by the Secretory Leukoprotease Inhibitor. *J Clin Invest* 1991, 87(2):482-488.
 42. Akizuki M, Fukutomi T, Takasugi M, Takahashi S, Sato T, Harao M, Mizumoto T, Yamashita J: Prognostic significance of immunoreactive neutrophil elastase in human breast cancer: Long-term follow-up results in 313 patients. *Neoplasia* 2007, 9(3):260-264.
 43. Yamashita J, Ogawa M, Shirakusa T: Free-Form Neutrophil Elastase Is an Independent Marker Predicting Recurrence in Primary Breast-Cancer. *J Leukocyte Biol* 1995, 57(3):375-378.
 44. Tamakuma S, Ogawa M, Aikawa N, Kubota T, Hirasawa H, Ishizaka A, Taenaka N, Hamada C, Matsuoka S, Abiru T: Relationship between neutrophil elastase and acute lung injury in humans. *Pulm Pharmacol Ther* 2004, 17(5):271-279.

Chapter 3

**Protein homology modeling unveils novel targets for natural and synthetic
indolecarbinol compounds**

Abstract

New and accurate predictions can be made regarding interactions between proteins and small molecule ligands by delineating the properties of proteins across various homologous proteins. In ligand-protein docking, homology models have progressively been employed and have been expanding the amount of protein targets that are feasible for synthesizing novel small molecule inhibitors. Here we focused on protein indolecarbinol binding sites and their corresponding biological relevance for protein function. In order to make predictions about additional indolecarbinol target proteins, we obtained the crystal structures of the four I3C target proteins to date (human neutrophil elastase, ubiquitin E3 ligase NEDD4-1, oncogenic BRAF V600E serine/threonine kinase, WNT) and utilized them as the starting template structure for synthesizing other homologous protein models. After examining the indolecarbinol binding sites between the homology models, it is evident that the homologous proteins that harbor greater than 50% sequence identity share a distinct structural architecture that confers binding to the indolecarbinol compounds, but generally, homologous proteins that share 30% sequence identity or below tend to lose essential I3C contact residues. The results presented here have provided invaluable insight regarding various I3C target proteins and have unlocked the potential for structure-based drug design. The quest to locate additional I3C target proteins has been facilitated through scrupulous structural comparisons of other homologous proteins and their corresponding indolecarbinol binding sites.

Introduction

Comparative Homology Modeling and Small Molecule docking background

The dearth of structural information surrounding protein-ligand complexes has hampered the endeavors in elucidating the binding selectivity of ligands with proteins. When no high-resolution structure is available, small-molecule docking into homology models can be employed for structure-based drug design and understanding of protein-ligand complexes. Homology modeling has increasingly become utilized as the primary means for gathering 3D coordinates of proteins as the amount of modeling software and known protein structures are gradually escalating. In order to enhance our understanding of protein function, general knowledge of a protein's three-dimensional structure is required, however, the foundation for this knowledge is based on a protein's sequence [1]. In the Swissprot and TrEMBL (<http://us.expasy.org/sprot/>), there are 2 million protein sequences, 30,000 proteins of which have had their structures solved (<http://www.rcsb.org/pdb/>). While the rate of experimentally procured structures is progressively expanding, the amount of new sequences is surging at a faster rate than the number of structures solved. Hence, in order to resolve this disparity, a number of computational methods have been employed.

Homology modeling is a method for identifying the similarity of residues for a given query protein sequence at its analogous topological position on the corresponding template protein. When experimental data is not present for a particular protein, a known 3D structure of a homologous protein can be used as the foundation structure to build a model and provide invaluable structural information and insight into the protein's mechanistic function.

Determining the 3D structure of a protein from its amino acid sequence can be accomplished by comparative or homology modeling [2]. Homology modeling methods rests on the fact that evolutionary related proteins exhibit similar structures. In the Protein Data Bank (PDB) (<http://www.rcsb.org/pdb>) [3], there are currently 123,456 protein structures that are available and so homology modeling is an appealing, alternative method for producing a sound 3D model of a protein from its amino acid sequence as discussed in many meetings of the bi-annual critical assessment of techniques for protein structure prediction (CASP) [4]. The conformational space is searched in homology modeling by minimally disrupting existing experimentally solved structures. This basis for this technique relies on the fact that the structural conformation of a protein is more highly maintained than its amino acid sequence and that minor to moderate alterations in sequence typically results in slight variation in the 3D structure [5].

If no appropriate template is available, then *de novo* or *ab-initio* methods can be employed to create three-dimensional protein models without the use of a homologous template structure. While *ab-initio* methods have been employed for making structure predictions, its application has not proven feasible [6] due to the force field inaccuracies and the need for extensive conformation sampling. Folding and unfolding exhibit a free energy difference of approximately several kcal/mol which corresponds to several

atomic van der Waals interactions [7]. This becomes a difficult obstacle to overcome in any energy function because of the inaccuracy of the force field which hence makes conformation sampling even more challenging since it could bias the energy minimization to the conformation of global energy minimum. Consequently, this would be distinctly different from the protein's actual conformation. Homology modeling, on the other hand, has proven successful as structural genomics initiatives are becoming more prevalent. Proteins that exhibit high sequence similarity and hence are evolutionary related can be allocated into different families. Proteins that reside in the same family share the same three-dimensional architecture which thus confers a structural representation of all proteins in a family even when only one structure of a member is available.

In the Pfam database (<http://pfam.xfam.org>) there are 16,306 families. For a query protein sequence, the steps for homology modeling include 1) determine the evolutionary related proteins of a known structure from the Protein Data bank to model the target protein 2) align the query sequence to the template structure and identify corresponding residues of the target sequence and template structure 3) on the basis of the alignment, generate the three-dimensional model 4) examine and refine the model and assess the model's quality [8, 9]. The accuracy and quality of a model relies on how closely related the target protein is to the template protein. In fact there is a direct association between the sequence identity of two proteins and the variations of their C α atoms of their shared structure [10]. If the sequence identity is greater than 40%, 90% of main-chain atoms could be constructed with an RMSD (root-mean-square distance) error of approximately 1 Å [9]. While areas of the target protein can be aligned onto the template structure, the three-dimensional structural coordinates of the homology model can be ascertained purely on the basis of the information yielded by the template structure itself [11-13]. Specialized methods are necessary for areas of the target protein that do not correspond to regions with a template. Loops and side-chains account for the greatest structural difference when proteins share greater than 40% sequence identity. However, when the sequence identity is between 30-40%, obtaining a reliable model of a query sequence to a template sequence becomes more difficult because insertions and deletions are more prevalent and the alignment becomes even more elaborate because of the issue of selecting an appropriate homologue structure. In homology models where the sequence identity is less than 40%, 80% of main-chain backbone atoms are predicted to exhibit an RMSD of 3.5 Å [14-17].

Of all the known sequences, about 57% of them have at least one structural motif that is associated with at least one known protein structure [18]. The likelihood of identifying a known structure for an arbitrarily chosen sequence from a genome is between the range of 30% to 65% [19, 20]. With the increasing number of unique folds identified, that probability is rising. Presently, among the globular proteins, there are approximately 1000 to 5000 folds, with an emergence of 200 folds garnered from the structure deposition [21] and with the advances in homology modeling, there are now greater than 1.1 million proteins that have at least one of their domains determined. While homology modeling is less precise than experimental means, it has still proven beneficial in hypothesis generation with regards to ligand binding site prediction [22, 23] structure-based drug design [24].

Sequence Alignment

For successful homology modeling, homologues with known structures must first be selected from the PDB. Homology detection is achieved by comparing the query sequence with all the sequences in the PDB and is carried out by a dynamic programming method [25] and its corresponding derivatives [26, 27]. BLAST (<http://www.ncbi.nlm.nih.gov/blast/>) is a software that examines sequence databases and identifies the most favorable local alignments to the query sequence. A sequence identity that is below 30% exhibit less reliable hits from BLAST. To circumvent these obstacles, different approaches such as template consensus sequences [28, 29] and profile analysis [30-32] have been employed.

Profile methods have surfaced as the chief method in distant homology searches. The accuracy of sequence alignments has been enhanced and has broadened the limits of searchable sequence similarity through the use of position-specific profile search methods like PSI-BLAST [33] and hidden Markov models [34]. A pair-wise scan of the database is done for sequence profile methods such as PSI-BLAST. A position specific score matrix is then generated once the sequences are aligned. In the subsequent round of database searching, the matrix then replaces the query sequence. This step is repeated until there are no new alignment that emerges. Evolutionary information for both the query and target proteins can be included to improve the sequence profile method whereby the query protein profile is compared with the target protein profiles. A profile represents a series of vectors where each vector exhibits how often an amino acid appears in a specific position in the multiple sequence alignment. The score is calculated by obtaining the log of the probability for an amino acid present in this vector. For profile-profile alignments, two frequency vectors are implemented and this is accomplished by employing a, a probabilistic model [35], a dot-product [36], or an information theoretical measure [37]. As investigated by Ohlson et al [38] profile-profile methods have been shown to be 30% more successful than standard sequence-profile methods because they are capable of identifying superfamily-related proteins and generally have better quality alignments. Applying a probabilistic scoring function also confers a greater advantage because of its ability to generate good alignments and demonstrate good fold recognition while still applying the same gap-penalties [7].

Information garnered from multiple structure alignment, secondary structure prediction, and environment has improved the profile method acuteness. The fact that structural information across proteins in the same family is maintained alludes to the importance of residues at specific locations in regards to a protein's function and folding requirement. Applying multiple sequence alignments provides a means for laying the structural foundation onto a sequence and hence, these resulting sequence profiles have been used to recognize homologous core motifs [39]. Once a homology model of a target protein has been generated, *in silico* ligand docking studies can be executed to predict essential binding interactions and small-molecule binding free energy [40]. Information provided by these studies can enhance our understanding of binding modes of different target proteins for I3C, 1-benzyl-I3C, and their corresponding analogs.

Rational structure-based drug design relies on the knowledge of 3D protein-ligand complexes. Comparing the structural landscape between proteins that share a high sequence homology can be beneficial and provide insight into the functional relationships among these proteins. While the majority of methods compare protein structures when their amino acid sequence is maintained in alignments, differences in overall protein topologies, can yield new information regarding structural links between proteins.

E3 Ubiquitin Ligase background

Ubiquitylation is important for controlling biological signaling. Misfolded or damaged proteins are ubiquitylated which ultimately leads to its degradation by the 26S proteasome prior to its arrival at its subcellular site. A ligand-activated receptor may become ubiquitylated which triggers the endocytosis of the receptor and subsequently removes it from the membrane, thus terminating the signal which was created by the activated receptor. The ubiquitylation of a cell-cycle inhibitor leads to its degradation which enables cells to continue through the cell cycle. Hence, protein ubiquitylation can modulate signaling by maintaining protein cellular levels, regulating its subcellular localization, and mitigating the accumulation of damaged proteins and in some cases by inhibiting protein degradation. The involvement of ubiquitin ligases in a variety of cellular processes is a testament to how critical their role is in regulating cellular homeostasis which causes them to be stringently regulated by multiple enzymatic steps and various accessory proteins.

The consecutive activities of three enzymes modulate ubiquitin conjugation [41]. The ATP-driven reaction, whereby a thiol ester is generated with Ub's C-terminal Gly residue (Gly 76), is catalyzed by the activating enzyme, E1. Ubiquitin is subsequently transferred to the Cys residue of E2, a ubiquitin conjugating enzyme. Lastly, E3, a Ub-protein ligase then passes the ubiquitin from the E2 to a lysine residue on the E3's protein substrate. The E3 is essential for identifying and selecting a distinct structural domain on its corresponding protein substrate [41, 42].

E3 protein-ubiquitin ligases are essential for choosing specific proteins for conjugation to ubiquitin. In general, E3 ligases are structurally characterized by having at least two functional domains—one of which that is responsible for modulating the interaction with the E2 ubiquitin conjugating enzyme and the other which is involved in the identification of specific substrate proteins. E3 ligases are comprised of two primary classes which include proteins with a HECT (homologous to E6-AP carboxyl terminus) catalytic domain [43] and proteins with a zinc-binding RING finger (really interesting novel gene) [44] adaptor domain. Members from the first family harbor a conserved region of approximately 350 residues and within this domain, there is an essential Cys residue that is important for catalytic activity [43, 45, 46]. Members that belong to the second family of E3 ligases are characterized by having multiple Cys and His residues that mediate the binding of two zinc ions to create a globular E2-binding domain [42]. There are three additional E3 ligase domains which include the U Box [47, 48], PHD [49-51] and HUL-1. While the two types of E3s associate with similar structural motifs on their corresponding E2s [52, 53], the two varieties of E3s don't share any structural

homologies nor do they operate in the same manner during catalysis. For HECT E3 mediated catalysis, the E2-bound Ub is forwarded to the E3 Cys before it interacts with its substrate Lys residue [46]. RING E3s, on the other hand, serve as scaffolds that bridge the anchored E2 and the substrate in such a way as to promote the direct interaction of the E2-linked ubiquitin to the substrate Lys [53].

NEDD4 Family

Proteins from the Nedd4 family are classified as HECT E3 ligases [54, 55]. In 1992, a screen for genes developmentally down-regulated in the early embryonic mouse central nervous system identified Nedd4. The Nedd4 family members are characterized by a catalytic HECT domain at the C-terminus, and an N-terminal region which is involved in substrate recognition which includes a C2 domain and various WW domains interspersed throughout. HECT domains exhibit enzymatic activity and are directly involved in catalyzing the covalent attachment of ubiquitin to their substrate proteins via a ubiquitin-HECT thioester intermediate [56-58]. HECT E3 ligases are prevalent in all eukaryotic organisms, with the *Saccharomyces cerevisiae* genome encoding 5 HECT E3s and the human genome encoding 28. The genomes of some pathogenic bacteria, while they do not encode ubiquitin, they do encode HECT domain-like E3s that are injected into host cells where the ubiquitin-conjugation system is utilized [59-61].

The substrate specificity of HECT E3s are determined by their N-terminal extensions. Human HECT E3s have been divided into three subfamilies based on the presence of specific amino acid sequences within these N-terminal extensions. The subfamilies include Nedd4/Nedd4-like E3s, which harbor WW domains, HERC (HECT and RCC1-like domain) E3s containing RLDs (RCC1-Like Domains), and “other” HECT E3s that exhibit neither RLDs nor WW domains [57, 58].

Structural and Functional Roles of HECT domain

The HECT domain modulates the interaction with E2, primarily Ubch7 and members of the Ubch5 subfamily of E2s and establishes a thioester complex with ubiquitin through a conserved cysteine residue [57, 58]. HECT E3s are responsible for catalyzing the final step in ubiquitin chain formation and elongation.

HECT domain structure

The HECT domain exhibits a bi-lobal structure with the C-terminal lobe harboring the catalytic cysteine residue and the N-terminal lobe serving as the E2 binding domain. A flexible hinge region connects the lobes and promotes the appropriate arrangement of the catalytic cysteine towards the ubiquitin-E2 thioester bond to enable transthioesterification of ubiquitin to the HECT domain. The catalytic cysteine residue of an unloaded E2 and the catalytic cysteine of the HECT domain is separated by a distance of approximately 41 Å in the case of E6AP and approximately 16 Å in the case of WWP1 [52, 62, 63] in the absence of ubiquitin and is too large for transthioesterification to occur. Once a ubiquitin-loaded E2 is bound however, the distance is drastically reduced (approximately 8 Å for a complex between NEDD4-

1/NEDD4L and ubiquitin loaded Ubch5B [64]. Hence, the HECT-E2 complex surface is dependent on the ubiquitin-loading status of the E2 and employs non-covalent interactions between the N lobe and the E2 and ubiquitin and the C lobe [64].

HECT domain activity regulation

The E3 HECT domain activity is regulated in two manners. First, the HECT E3 can associate with its substrate protein via specific protein-protein interaction domains/motifs that reside in the N-terminal to the HECT domain. Some interaction motifs in the HECT E3s are also capable of associating with regulatory proteins that can either promote (“adaptor and/or auxiliary proteins”) or hamper (“inhibitory proteins”) the interaction of substrates with their corresponding E3s. Ndfip1 and Ndfip2 and several members of the α -arrestin protein family (Arrdc) for instance, interact with the WW domains of Ned4 family members through their PY motifs [64-67] thus facilitating the ubiquitination of their substrate proteins [66, 67].

The second level of regulation is based on the catalytic activity of the HECT domain which includes the association with its corresponding E2. SMURF2 and Itch which are both classified under the NEDD4-like family of E3s harbor WW domains an N-terminal C2 domain. In comparison to other HECT domains, the SMURF2 HECT domain needs an additional protein called SMAD7 to facilitate the binding of both the HECT domain of SMURF2 and Ubch7. Previous NMR studies have revealed that the catalytic cysteine of the HECT domain is not accessible to Ubch7 because it is restrained by an intramolecular association of the C2 domain with the HECT domain. SMAD7 abrogates the inhibitory effect of the C2 domain and is regulated by extracellular stimuli such as TGF β [68]. Co-precipitation studies have demonstrated that for Itch, the interaction between the HECT domain and the WW domain causes Itch to become inactive. Additionally, the phosphorylation of serine and threonine residues in the N terminus of Itch triggers a conformational change which leads to the disjunction of the WW-HECT domain association and ultimately to the activation of Itch [69]. Overall, intra- and/or intermolecular interactions are responsible for modulating the activity of some HECT domains to enable their corresponding substrate proteins to be ubiquitinated when necessary.

Ubiquitin chain formation

There are seven lysine residues residing in ubiquitin that can be utilized for ubiquitin-ubiquitin conjugation. Homo-polymeric (where a specific lysine residue of ubiquitin is employed for conjugation throughout a chain such as K48 or K63) and hetero-polymeric chains (where various lysine residues of ubiquitin are altered with one chain and/or one ubiquitin moiety and can be adjusted at several lysines) can be amassed. Linear chains can also be assembled though the alpha-amino group of the N-terminal residue of ubiquitin. The different manners in which a substrate protein is ubiquitinated can render different fates. Modification with a K48-linked or K11-linked ubiquitin chain, for instance, targets proteins for degradation by the proteasome, while mono-ubiquitination or modification with K63-linked ubiquitin chains are not related to any proteolytic roles [70-

73]. Ubiquitin chain assembly progresses in a sequential manner and it is the C lobe of the HECT domain that confers the E3 ligase's ability to generate distinct ubiquitin chains.

Nedd4-like E3s

There are nine human members that are classified as HECT E3s which include NEDD4, NEDD4-1/NEDD4L, ITCH, SMURF1, SMURF2, WWP1, WWP2, NEDL1 and NEDL2. NEDD4 (Neural precursor cell-expressed developmentally downregulated gene 4) was first cloned as a developmentally regulated gene in the mouse central nervous system [74, 75].

NEDD4

Some potential substrates for NEDD4 have been revealed which includes proteins like Hgs, a sorting adaptor, implying that HECT E3s may be implicated in vesicular sorting and trafficking [76, 77].

Mice that lack or exhibit impaired NEDD4 are phenotypically growth retarded and display developmental abnormalities in the nervous and cardiovascular systems and also harbor deficiencies in the neuromuscular junctions and T cell function. Because of the intricate phenotypes that are correlated with NEDD4 knockout, this implies that Nedd4 has multiple targets *in vivo*. Animals that do not have NEDD4 die during fetal development between E14 and E18 and are less than half the size of their wild-type littermates [78].

Reduced cell surface expression and signaling through insulin and IGF-1 receptors can cause growth retardation in Nedd4 deficient animals. Mitogenic activity as well as levels of Grb10, which is an inhibitor of both receptors, are increased as observed in fibroblasts that were retrieved from Nedd4^{-/-} embryos. While it has not been clearly established, Grb10 is not a direct Nedd4 target, however, Grb10 does directly associate with NEDD4 via the NEDD4 C2 domain and the lethality in NEDD4^{-/-} mice can be moderately restored by the maternal inheritance of an impaired Grb 10 allele [78].

T cell function is also regulated by Nedd4. While Nedd4^{-/-} animals exhibit normal T cell development in the thymus, their T cells in the periphery are hypo-responsive, do not proliferate well in response to antigen and are less inclined to generate IL-2 [79]. B cells that are deficient in NEDD4 go through class switching at a reduced frequency [79]. The lower activation of NEDD4^{-/-} T cells may be attributed to increased levels of Cbl-b which is a RING type ubiquitin ligase and a known NEDD4 target [79].

Studies involving NEDD4 knockout mice have demonstrated that NEDD4 modulates neurite growth and arborization in neurons [80, 81]. NEDD4 interacts with TNIK, a serine/threonine kinase, and Rap2A to create a complex that regulates the Nedd4-mediated ubiquitination of Rap2A [80] thus leading to diminished Rap2A function and facilitation of dendrite growth. Studies conducted by Drinjakovic *et al* [81] revealed that PTEN downregulation induced by NEDD4-mediated ubiquitination regulates branching of retina ganglion axons. Additionally, NEDD4 is essential for the proper formation and functioning of neuromuscular junctions [82]. It has been demonstrated in NEDD4^{-/-}

embryos that the size of skeletal muscle fibers and the number of motor neurons are significantly curtailed. In NEDD4 skeletal muscle-specific knockout mice, it has been established that a deficiency in NEDD4 leads to escalated mass of type II fast twitch fibers of denervated gastrocnemius muscle after tibial nerve transection which implies that in vivo NEDD4 is responsible for denervation-induced skeletal muscle atrophy [83]. Nedd4 is also crucial for vascular development [84]. These vascular defects in NEDD4^{-/-} embryos may also arise because of elevated levels of thrombospondin-1 (Tsp-1) which is an inhibitor of angiogenesis.

In neuroblastoma and pancreatic cancer cells, NEDD4 is responsible for ubiquitinating and degrading N- and c-Myc oncoproteins [85]. By directly binding to the NEDD4 promoter, the class III histone deacetylase SIRT2 augments N-Myc and c-Myc protein stability, thus facilitating cancer cell proliferation and causing histone H4 lysine 16 to be deacetylated which subsequently represses NEDD4 gene expression. NEDD4 gene expression is actually reactivated by SIRT2 inhibitors which lowers N-Myc and c-Myc protein expression, and impairs neuroblastoma and pancreatic cancer cell proliferation [85]. Based on these studies, this suggests that NEDD4-mediated Myc control may be an effective therapeutic target for neuroblastoma and pancreatic cancer.

It has also been demonstrated that NEDD4 (and NEDD4-1 as well) can associate with viral proteins to regulate the budding of viruses such as the Ebola virus [86, 87] and retroviruses [88-93]. Ubiquitination of viral matrix proteins such as the Epstein-Barr virus LMP2A and LMP2A-associated proteins like Lyn allows them to be identified by the ESCRT machinery where they are trafficked through the host cell vesicular transport machinery [94].

Serine Protease background

Approximately 35 to 75% of the population of human circulating leukocytes are comprised of polymorphonuclear neutrophils [95]. In mammals, these are the most prevalent type of white blood cells. Neutrophils originate from pluripotent stem cells in the bone marrow and travel through the bloodstream and constitute 1.5 to 5×10^9 cells/liter. Because they contain granules and have a multi-lobular nucleus, human polymorphonuclear neutrophils have been classified as granulocytes. Neutrophils are the chief regulators of the inflammatory response and they are significant players in the innate immune defense system and defend against invading pathogens [96]. They mediate short-term phagocytosis during the advent of infection [95]. Neutrophils are the first inflammatory cells to exit the vasculature during the acute phase of inflammation where they subsequently gravitate towards sites of inflammation which is followed by various inflammatory stimuli [96]. In order to protect the host from invading pathogens, neutrophils use complementary oxidative and non-oxidative pathways [96].

Neutrophil elastase and proteinase 3 are primary components of neutrophil azurophilic granules and are implicated in the non-oxidative pathway of intracellular and extracellular pathogen destruction. These neutrophil serine proteases internalize phagocytized microorganisms coupled with microbicidal peptides and the membrane-

associated NADPH oxidase system, which generates reactive oxygen metabolites, within phagolysosomes [97]. Activated neutrophils also secrete neutrophil extracellular traps (NET) that are comprised of a web-like DNA structure and serve as an additional extracellular antimicrobial mechanism. NETs kill pathogens extracellularly and impede additional spread by acting as physical barriers with their outer structure composed of chromatin bound to positively charged molecules such as histones and NSPs. These NET-associated NSPs are involved in pathogen killing by impairing bacterial virulence factors extra-cellularly [98, 99].

Neutrophil serine proteases are also implicated in various inflammatory human ailments which include chronic lung diseases such as cystic fibrosis, chronic obstructive pulmonary disease, acute lung injury, and acute respiratory distress syndrome [100, 101]. These diseases are a consequence of the aggregation and activation of neutrophils in the airway which lead to extensive secretion of active NSPs which hence spur matrix loss and inflammation. Gene mutations and modified cellular trafficking in NSPs also trigger different human disorders. For instance, mutations in the ELA2/ELANE gene encoding HNE lead to human cyclic neutropenia and severe congenital neutropenia [102, 103]. Additionally, neutrophil membrane-bound proteinase 3 (mPR3) is the primary target antigen of anti-neutrophil cytoplasmic autoantibodies (ANCA) which correlates to Wegener granulomatosis [104]. Mutation of the gene (CTSC) encoding dipeptidyl peptidase I (DPPI) affects all corresponding proteases which in turn triggers various granular hematopoietic serine proteases [105, 106]. CTSC mutations lead to Papillon-Lefevre syndrome and palmoplantar keratosis [107, 108].

Biological Functions of Elastase and Proteinase 3

Antimicrobial Roles in Infections

Human neutrophil elastase and proteinase 3 are directly implicated in the intracellular killing of phagocytosed bacteria in phagolysosomes coupled with myeloperoxidase and reactive oxygen species created by the NAPH oxidase complex [109]. Extracellular killing can also take place via trapping of bacteria in NET that is comprised of filamentous DNA structures and peppered with cationic proteases which include NSP secreted by activated neutrophils [98]. HNE cleaves the outer membrane Protein A of *Escherichia coli* [110] and other virulence factors of *Salmonella enterica*, *Yersinia enterocolitica*, and *Shigella flexneri* [111] which is how HNE executes its antimicrobial activity on gram-negative bacteria.

Proteolytic activity of NSPs is not compulsory for their antimicrobial activity. Because of their inherently positive surface charge, NSPs are able to bind tightly to bacterial membranes. Binding to the bacterial membrane alone may actually impede bacterial protein synthesis and triggers membrane disjunction and depolarization [112].

Processing and Activation of Cellular Receptors

NSPs modulate the immune response through the cleavage and activation of specific cellular receptors. HNE and PR3 can cleave the N-terminal extracellular domains of protease-activated receptors (PARs), which belong to a subfamily of related G-protein-coupled receptors [113, 114]. These receptors are pervasively expressed in different tissues and cells, but more notably, in platelets and endothelial cells. PAR extracellular domains are processed through the exposure of a bound ligand that enables the receptor autoactivation and the ensuing induction of an intracellular signaling cascade through phospholipase C [113, 114]. To date, four PARs have been identified whereby three, PAR-2, PAR-3, and PAR-4, can be activated by thrombin. PR3 and HNE can cleave PAR-1 which hampers their activation by thrombin [115].

Elastase and Proteinase 3 Structural Characteristics

Human neutrophil elastase and proteinase 3 are comprised of two homologous β -barrels and a C-terminal α -helix [116-118]. Segment linkers affix six antiparallel β -sheets within each barrel. The catalytic triad (residues Ser 195, Asp 102, and His 57 (chymotrypsin numbering) are positioned at the intersection of the two β -barrels and the active site clefts crosses perpendicular to this terminal. Nucleophil attack on the carbonyl carbon (C = O) of the substrate scissile bond is performed by Ser195 thus instigating the onset of the catalytic reaction. After the dipeptide at the amino terminal is cleaved by dipeptidyl peptidase I (DPPI), the free ammonium group of the first N-terminal residue (Ile16) generates an internal salt bridge with the side-chain carboxylate of Asp194 that enables substrate access to the active site S1 pocket [119]. The activation domain of the neutrophil serine proteases are composed of Ile16 and the three flexible surface loops (217-225 loop, 180 loop, and autolysis loop) of the C-terminal β -barrel. Human neutrophil elastase and proteinase 3 are comprised of predominantly positively charged residues and are consequently extremely cationic proteases. The activation domain in HNE and PR3 harbor an assemblage of positively charged residues [96]. In proteinase 3, negatively charged residues are dispersed throughout the positive cluster in the activation domain. Loop 217-225 on PR3 is composed of a hydrophobic patch of residues which potentially accounts for the specific binding of mature PR3 binding to the neutrophil membrane as opposed to its proform which lacks this binding ability [96].

While human neutrophil elastase and proteinase 3 share a common fold, their differences in substrate cleavage stem from substitutions in the S1 pocket and from differences in the surface loops adjacent to the active site. As observed from the solvent-accessible surfaces, the charge distributions in the substrate binding region vary greatly.

Substrate specificity for each chymotrypsin-like serine protease is determined by the S1 pocket characteristics. The S1 pocket of HNE is comprised of amino acid residues Val 190, Phe 192, Ala 213, Val 26, Phe 228, and the disulfide bridge Cys 191-Cys 220 which renders the pocket hydrophobic and hemispherical. PR3 also exhibits a hemispheric, but smaller S1 pocket which can be attributed to the Val/Ile substitution at position 190 [117]. Human neutrophil elastase and proteinase3 have a stronger

preference for small hydrophobic residues like Val, Cys, Ala, Met, and Ile in their S1 pocket. The S2 subsite of HNE is enveloped by residues Phe 215, Leu 99, and the imidazole ring of the catalytic triad His 57 which is predominately hydrophobic. The S2 subsite of PR3, however, is a buried polar pocket with higher polarity which is a consequence of the Leu-to-Lys substitution at position 99 [117]. At the P2 position of HNE, medium-sized hydrophobic side chains like Pro are more strongly preferred and the presence of Lys 99 in PR3 causes it to preferentially hold negatively charged residues at the P2 site. The Leu-to-Lys substitution at position 99 in PR3 renders PR3 to be smaller and more polar than HNE. The S1' and S2' sub-sites in HNE are hydrophobic and are enshrouded by residues Cys 42-Cys 58 and Phe41-Leu143, respectively. In PR3, the S2' sub-site resides near charged residues Asp 61 and Arg 143, neither of which is maintained in HNE [117]. The S1' and S3' sites in PR3 are smaller and more polar as a result of the displacement of Asp 61 in loops 60 to the negatively charged side chain in S1' and S3'.

Human Azurocidin Structure

Azurocidin is a single polypeptide glycoprotein that is generated as a 251 amino acid precursor from which 26 amino acid residues are cleaved from the N-terminus and 3 residues are removed from the C-terminus [120, 121]. The mature polypeptide is comprised of 222 amino-acid residues with a calculated molecular mass of 24 kDa. As observed through sequence analysis, azurocidin has 45% homology to human neutrophil elastase and 42% homology to proteinase 3.

Because His 57 and Ser 195 from the catalytic triad have been replaced by serine and glycine residues respectively, this has hampered azurocidin's ability to bind to heparin-binding proteins such as hHBP and pHBP with chemotactic properties for monocytes and fibroblasts [122]. Azurocidin exhibits two domains typical for trypsin-like proteinases, each containing a β -barrel of six antiparallel β -strands [123]. It shares structural homology to neutrophil elastase with minor differences in the loop region and a varied surface charge distribution. On opposite sides of the protein, basic and acidic amino-acid residues generate separate patches, however, the four disulfide bridges are maintained.

Azurocidin Biosynthesis and Processing

During the promyelocyte stage of neutrophil differentiation and after processing, azurocidin genes along with neutrophil elastase and proteinase 3 genes are constitutively expressed in a cell-specific manner and the proproteins are then targeted to azurophilic granules [124]. The bulk of azurocidin resides in azurophilic granules which is a consequence of expression in early stages of neutrophil differentiation, however, more recent data has revealed that there exists a secretable storage of molecules that are located in secretory vesicles which are separate from azurophilic granules [125, 126]. These structures are synthesized during the late stages of neutrophil differentiation which alludes to the notion of late expression of azurocidin.

Both azurocidin and proteinase 3 expression can be engendered in non-myeloid cell types as well. It has been shown that human endothelial cells have the ability to express proteinase 3 [127] and azurocidin [128]. Tumor necrosis factor α (TNF- α) and interleukin-1 β (IL-1 β) also induce azurocidin expression in the corneal epithelium [129]. Azurocidin becomes active after the co-translational removal of the signal peptide which subsequently releases a short N-terminal propeptide. Azurocidin propeptide is composed of seven amino-acid residues which are processed in two steps. Five residues are initially cleaved by an unknown enzyme which is followed by the removal of the Leu-Asp dipeptide. Both cleavages take place in the post-Golgi compartment. C-terminal peptide cleavage subsequently follows the N-terminal processing. The extraction of the N-terminal propeptide transforms the inactive zymogens of neutrophil elastase to proteolytically active forms. However, for azurocidin, the association between propeptide removal and the gain of biological activity is ambiguous and additional studies need to be performed [124].

Azurocidin is an essential mediator of the inflammatory response. Being a part of neutrophilic azurophilic granules, azurocidin is implicated in oxygen-independent killing mechanisms that function in phagocytosing neutrophils. It triggers the contraction of endothelial cells and is released extracellularly. Neutrophil extravasation can develop because of gaps generated between the cells of the endothelial lining. Once azurocidin is secreted it attracts monocytes which instigates their influx in to inflammation sites. Hence, these characteristics make azurocidin a potential new target in therapies against the detrimental effects of the inflammatory response.

BRAF V600E

4% of incident cancers can be attributed to melanoma and the mortality rate is increasing. To date, surgery remains the most viable treatment option for early disease while adjuvant therapy and interferon alpha has proven to be beneficial in some stage II and stage III cases Kirkwood [130]. Advanced melanoma, however, harbors a poor prognosis and has few therapies.

BRAF is a serine/threonine protein kinase that activates the MAP kinase/ERK-signaling pathway. It is most easily activated by Ras and generally, the basal kinase activity of BRAF is greater than that of other family members which may account for the periodic mutational activation of BRAF in human tumors. In melanoma, BRAF mutations is typically observed in patients whose tumors appear on skin without chronic, sun-induced damage. Incidentally, approximately 50% of melanomas exhibit activating BRAF mutations. Over 90% of BRAF mutations that are present in melanomas are at codon 600 and among these, greater than 90% are at a single nucleotide mutation which results in a substitution of glutamic acid to valine [130] (BRAFV600E which corresponds to nucleotide 1799 T>A and codon GTG>AGG). The second most common mutation is BRAFV600K whereby valine is substituted by lysine represents 5-6% of the population of mutations and is followed by BRAFV600R and then BRAFV600D.

A chief breakthrough of the past few years was the finding that RAF kinases could homo- and heterodimerize [131, 132] and interestingly, the active RAF kinase structure is a side-to-side dimer in which only one monomer must have catalytic activity [133]. Dimerization is enhanced by Ras and is prone to negative feedback regulation by Erk.

Raf mutations exhibit the highest frequency in melanoma and contributes to genomic instability which advances the proliferation of cancer cells. Mutated BRAF is separate from upstream growth stimuli and signals as a monomer. The most persistent BRAF mutation, BRAFV600E, initiates constitutive activation of the kinase and leads to insensitivity to negative feedback mechanisms [134, 135].

BRAFV600E is associated with various mechanisms of melanoma progression but more notably in the activation of the downstream MEK/ERK pathway, evasion of senescence and apoptosis, unchecked replicative potential, angiogenesis (through MEK-dependent activation of HIF-1 α and VEGF), tissue invasion and metastasis (through upregulation of several proteins implicated in integrin signaling, migration, cell contractility, tumor- and microenvironment-derived interleukin-8), and the evasion of immune response [135]. With more advances in the understanding of melanoma, this has spurred the advent of potential targets for therapy with BRAF mutations being the primary target.

BRAF Structure

Mutated V600E BRAF enhances activation by 500-fold and triggers the constitutive activation of MEK/ERK cascade in tumor cells. When no extracellular stimuli is present, this oncogenic BRAF enables the activation of this signaling which renders the cell self-sufficient in growth cues in this pathway.

The kinase domain in the BRAF exhibits a bilobal structure with a catalytic cleft disjointing a small and large lobe. Within the inactive BRAF state, a conserved motif which consists of an aspartate, phenylalanine, and glycine (DFG), which resides in the activation domain, shows a flipped conformation that positions the region of activation towards the P-loop of the N-lobe [136]. Hydrophobic interactions are facilitated by the association of glycine-rich segments G595-V600 near the G463-V470. Consequently, this leads to a conformational change that restricts access to the catalytic cleft. In this inactive state, the residues implicated in phosphotransfer reactions are aligned, but the ATP and peptide substrate recognition regions are disordered [136]. The active conformation requires a shift in the orientation of the DFG motif/activation domain [137]. Once the activation domain is phosphorylated, this leads to the disruption of the hydrophobic associations between the activation domain and the P-loops, which consequently flips the DFG segment to its active conformation, hence aligning the ATP and peptide recognition domains and enabling entry to the catalytic cleft [136].

Typically, the valine at position 600 in BRAF interacts with F467 of the P-loop. V600 resides in the activation segment adjacent to the DFG motif [137]. The substitution of valine to glutamate at the 600 position, disturbs the hydrophobic associations, thwarting the conformation that maintains the inactive position of the DFG motif and re-orienting

the DFG motif into its active state which ultimately renders the BRAF V600E in a constitutively active state [137, 138].

BRAF inhibition history

Initial attempts at inhibiting mutated BRAF in melanoma included a non-selective BRAF broad-spectrum kinase inhibitor called Sorafenib which exhibited a bi-aryl urea structure. Phase I studies displayed activity against melanoma and so treatment was further developed in combination with carboplatin and paclitaxel which had previously been used in treating lung cancer. Since response rates exceeded 30% in phase II studies, this developed was eventually continued into phase III trials, but it ultimately failed in both second and first line trials [139]. More studies revealed that sorafenib lacks selectivity and potency for RAF but it is a highly effective inhibitor of VEGFR2, VEGFR3, and other kinases [140].

Vemurafenib: phase I and II results

Vemurafenib (PLX4032/RG7204/RO5185426), a selective and potent inhibitor of oncogenic mutant BRAF, recently yielded positive results in phase I [141] and phase II trials [142]

The phase I study involved a 55-solid cancer patients dose-escalation trial with a subsequent 32-melanoma patients extension phase involving the maximum dose that could be given without deleterious side effects. The starting dose of Vemurafenib was 160 mg daily and was largely well tolerated. The response rate was 56% in patients who had melanoma with the BRAFV600E mutation. The study further demonstrated a substantial impact on (progression-free survival) PFS of greater than 7 months and showed that the maximum tolerated dose to be at 960 mg twice-daily. Remarkably, responses were detected in all sites of disease which included bone, liver and small bowel [141].

The phase II trial had tested patients who had previously received treatment for melanoma with the BRAFV600E mutation and had determined how effective vemurafenib was with respect to the overall response rate, duration of response, and overall survival. 132 patients were enrolled and were given a dose of 960 mg of vemurafenib orally twice-daily until toxic effects were unacceptable and revealed a confirmed response rate of 53% with a median period of response of 6.7 months and a median overall survival of 15.9 months, which was an unrivaled result in melanoma patients [143].

At 6 months, the overall survival rate was 77%, 58% at 12 months, and 43% at 18 months [143].

In a 2-arm randomized phase III trial in treatment-naïve patients, the results had revealed a relative reduction of 63% in the risk of death and of 74% in the risk of tumor progression in patients. While patients did experience some adverse side effects with

treatment, vemurafenib was the first personalized compound that established an improvement in progression-free survival and overall survival in metastatic melanoma exhibiting the BRAF V600 mutation [144]. It is also the first compound in melanoma that demonstrated an anti-proliferative effect through the inhibition of a specific molecular protein target.

B-raf inhibitors: drug resistance

While overall, vemurafenib treatment displayed a relatively high success rate in patients, soon after the first indication of objective response that was seen with vemurafenib, signs of disease progression were quickly discernible in some patients and the range of response duration employing this therapy became expansive [145].

Tumors can potentially survive and circumvent the effects of BRAF inhibitor therapy by activating both MAPK-dependent and MAPK-independent pathways. In many instances, biochemical studies have demonstrated the manifestation of NRAS mutations, the appearance of MEK mutations, and higher expression levels of COT kinase [146].

PI3K pathway signaling activation and upregulation of platelet-derived growth factor (PDGF) receptor and insulin-like growth (IGF) receptor are additional mechanisms by which tumors can adapt in the presence of BRAF inhibitor therapy.

With the continued disease progression after vemurafenib treatment, clinical trials have been underway to test combination treatment or succession of therapies in melanoma patients who have higher propensity for recurrence or who have progressed [147]. Experimental and initial clinical results have demonstrated prolonged progression-free survival with the use of combination therapy such as vemurafenib coupled with a MEK inhibitor. This treatment alternative has proven successful and appears to reduce the emergence of skin toxicities. These new developments have paved the way for evolution of targeted drugs that allow us to customize treatment to each individual patient.

B-raf inhibitors in combination trials

In order to determine the safety, tolerability, and effectiveness of vemurafenib and other drugs in combination such as the Cytotoxic T Lymphocyte Antigen 4 (CTLA-4) inhibitor mAb ipilimumab, a clinical phase I/II trial has been implemented [147]. The two drugs couple target therapy and immunotherapy methods and employs different mechanisms of action, where ipilimumab maintains an active immune response. B-Raf inhibitors and ipilimumab also act at different paces. Vemurafenib operates quickly with a swift metabolic cessation, however, disease progression reappears after a median of 6-8 months. The mAb, on the other hand, works slowly which enables the disease to become chronic.

In the 1b phase of the BRIM-7 trial, a dose-escalation study was being performed to ascertain the safety, tolerability and pharmacokinetics of vemurafenib in combination with GDC-0973, which is a selective inhibitor for MEK1/2, in patients harboring the

BRAFV600E positive metastatic melanoma who have progressed after vemurafenib treatment alone [143]. Employing combined therapy may be beneficial because of the additive or synergistic effects on progression-free survival. Combinational therapy also circumvents the toxicities that are associated with the activation of the MEK pathway when BRAF inhibitors are used as primary agents. Treating melanoma patients who exhibit the BRAFV600E mutation with a MEK inhibitor is an alternative approach for mitigating the skin toxicities of RAF inhibitor therapy by synergistically dampening the ERK pathway activity [148]. Phase I and II trials have already demonstrated that this type of treatment is safe and has the capacity to reduce the toxicity of either inhibitor alone [148].

Wnt Protein Background

The recognition of wnt signaling in cancer began with the discovery that mammary tumors emerging from mice infected with the murine mammary tumor virus were frequently instigated by the activation of the murine int-1 gene which was later called Wnt1. It later came to light that the Wnt gene resembled the fly wingless gene which is a secreted factor that modulates a signaling cascade that included GSK3 and armadillo, which is the fly version of mammalian β -catenin [149]. This became significant when it was discovered that the human tumor suppressor adenomatous polyposis coli (APC) proteins was observed in conjunction with β -catenin. Wnt became implicated in cancer signaling when it was discovered that APC could down-regulate β -catenin and Wnt-1 could upregulate it.

In the Wnt signaling pathway, APC is the most frequently mutated gene in human cancers [149]. Genetic abnormalities in APC are the source of familial adenomatous polyposis which is a heritable syndrome whereby the affected individuals develop hundreds of polyps in the large intestine early in their life and become afflicted with colorectal cancer [150]. Loss of function in both APC alleles leads to tumorigenesis and that loss is structurally connected to the protein's ability to coordinate β -catenin protein stability [149].

Mutations in APC eliminate all binding sites for Axin, which serve as a scaffold that also interacts with β -catenin, and activates the protein kinases GSK3 and CKI, both crucial for tagging β -catenin for destruction via the E3 ubiquitin ligase β -TRCP [149]. Axins I and II are tumor suppressors that are mutated more commonly in hepatocellular and some colorectal, and some familial cancer syndromes as well [149, 151]. It has also been observed that mutations within β -catenin itself prevent it from being tagged for destruction by the kinases, [152] which is prevalent in high frequency in hepatocellular cancers and medulloblastoma.

WNT family of proteins

The WNT family of proteins are essential in the regulation of a wide array of processes which include: cell proliferation; survival; migration and polarity; specification of cell fate; and self-renewal in stem cells [153]. Regulation of the WNT ligand levels are essential

for WNT signal transduction and hence a disturbance in ligand levels can lead to defects in embryonic development.

Cell signaling begins when a Wnt family protein binds to an appropriate receptor or a receptor group on a cell surface. Frizzled (Fz) family receptors were among the first variety of transmembrane molecules identified that serve as receptors for Wnt-ligands. Fz receptors are seven-pass transmembrane proteins that are essential for nearly all Wnt signaling, in which the N-terminal Fz cysteine rich domain (CRD) serves as the Wnt binding domain.

Wnt ligands are also able to bind to additional cell surface receptors which are members of a family of LRP receptor molecules (LRP-receptor-related protein) that contain a single transmembrane domain. In vertebrates, The LRP family members consist of LRP5 and LRP6. Phenotypes from LRP5/6 are knocked out in mice mimic the effects of some Wnt gene silencing. Silencing of LRP6 expression, for instance, disrupts the development of middle and posterior brain (similar to Wnt1 blocking), and displaces limbs in the ventral direction (similar to Wnt7a knockout) and hampers nerve tissue volume expansion (similar to Wnt3a silencing) [154]. Therefore, LRP5/6 has been implicated in the canonical Wnt signaling pathway.

Ror1, Ror2, and Ryk, which belong to the family of receptor tyrosine kinases (RTK), are also involved in the noncanonical Wnt signaling pathway. A physical interaction between Ryk, Wnt1 and Wnt3a which leads to the activation of the Wnt signaling pathway has been demonstrated in the HEK293T cells [154].

Canonical Wnt signaling pathway

The canonical Wnt signaling cascade, activated by Wnt family proteins, has been implicated in a broad range of biological processes which include embryogenesis and tissue regeneration. On the cellular level, Wnt signaling also regulates proliferation and differentiation and maintenance of stem cell populations [155].

Activation of the canonical Wnt signaling cascade begins with β -catenin stabilization. β -catenin accumulates in the cytoplasm and is subsequently translocated into the nucleus where it acts as a coactivator of TCF/LEF-dependent genes of transcription [156].

The Wnt signaling pathway is also activated through the formation of a triple complex which is comprised of the Wnt-ligand, Fz family receptor, and the LRP5/6 coreceptor. This triple complex formation leads to the translocation of a number of proteins on the membrane which include Dishevelled (Dvl), Axin and GSK-3 β .

Wnt/Ca²⁺ signaling pathway

Once Wnt binds to the Frizzled receptor, this the heterotrimeric G-protein dissociates into G α - and G β/γ -subunits. The released G β/γ complex can then activate phospholipase C (PLC) where it is translocated on the membrane and hydrolyzes to phosphatidylinositol(4,5)-bisphosphates (PIP2) up to inositol(1,4,5)-triphosphates (IP3)

and diacylglycerol (DAG). DAG triggers PKC kinase, while IP₃ stimulates the release of Ca²⁺ ions from intracellular compartments. The increase in Ca²⁺ cytoplasmic levels subsequently activate a cascade of Ca²⁺-dependent effector molecules which include Ca²⁺/calmodulin-dependent kinase II (CaMKII), nuclear factor associated with t-cells (NFAT), and calcineurin [157]. The Wnt/Ca²⁺ pathway is important in the regulation and organization of the cytoskeleton and cell motility [158].

Wnt Signaling in Cell Polarization

Wnt signaling provides and maintains planar cell polarity (PCP) and is stimulated by the ligands Wnt7a and Wnt11. Wnt/PCP-signaling regulates Rac and Rho which are small GTPases.

Wnt proteins are hydrophobic which is primarily due to the post-translational addition of palmitate and/or palmitoleic acid to one or two residues (namely Cys77 and/or Ser209 in Wnt3a). For Wnt intracellular trafficking, acylation is essential for Wnt activity, however, Wnt's precise role is not clear. Studies have suggested that the Wnt lipid group(s) may be responsible for directly associating the Fz-CRD [159] and may also regulate binding to WIF. Genetic evidence implies that acyltransferases (*porcupine* in *Drosophila*) are necessary for Wnt palmitoylation in Wnt-secreting cells [160]. Acylation is speculated to localize Wnts to cell membranes while Wnts that are in circulation may be bound to carrier proteins that protect the lipid from solvent [161].

Structure of XWnt-Fz8-CRD

The WNT family of proteins are comprised of 19 secreted, highly modified glycoproteins that are about 40 kDa in size and are essential in the regulation of a wide array of processes which include cell proliferation, survival, migration and polarity, specification of cell fate, and self-renewal in stem cells [153]. Wnt proteins are approximately 350 residues, cysteine-rich secreted glycoproteins that activate cell surface receptors on responder cells to activate at least three different signaling pathways which include the canonical β -catenin pathway, the non-canonical planar cell polarity (PCP) and the Ca²⁺ pathways [162]. Posttranslational lipid modifications are necessary for their biological activity [162]. Removal of the palmitic acid increases the protein's hydrophilicity and signal activity of Wnt ligands thereby demonstrating the importance of the lipid moiety.

Cell signaling begins when a Wnt family protein binds to an appropriate receptor or a receptor group on a cell surface. Frizzled (Fz) family receptors have been identified among the first variety of transmembrane molecules that serve as receptors for Wnt-ligands. Fz receptors are seven-pass transmembrane proteins that are essential for nearly all Wnt signaling in which the N-terminal Fz cysteine rich domain (CRD) serves as the Wnt binding domain.

Wnt proteins are hydrophobic which is primarily due to the post-translational addition of palmitate and/or palmitoleic acid to one or two residues (namely Cys77 and/or Ser209 in Wnt3a). For Wnt intracellular trafficking, acylation is essential for wnt activity,

however, wnt's precise role is not clear. Studies have suggested that the Wnt lipid group(s) may be responsible for directly associating the Fz-CRD [159] and may also regulate binding to WIF. Genetic evidence implies that acyltransferases (*porcupine* in *Drosophila*) are necessary for Wnt palmitoylation in Wnt-secreting cells [160]. Acylation is speculated to localize Wnts to cell membranes while Wnts that are in circulation may be bound to carrier proteins that protect the lipid from solvent [161].

More recently, Janda *et al.*, crystallized the glycosylated *Xenopus* Wnt8 (Xwnt8) in complex with the cysteine rich domain (CRD) of Fz [162]. The crystal structure reveals that XWnt8 cinches the Fr8-CRD at two opposing regions using extended thumb and index fingers jutting from a main "palm" region to associate with "site 1" and "site 2," respectively, spanning about 2000 Å² of surface area. Xwnt8 is composed of an N-terminal α-helical domain (NTD) that extends from residues 1-250 (helices A through F) where the lipid-modified thumb resides, and a C-terminal cysteine-rich region (CTD) which extends from residues 261-338 [162]. Each region interacts with the Fz8-CRD domain whose conformation remains unchanged when compared to the unbound structure. The XWnt8 NTD is composed of a seven α-helical bundle palm, which host two inter-helical loop insertions that are sustained by 4 disulfide bonds. The CTD's primary feature is a long 40 amino acid β-strand hairpin that is secured by an intricate web of disulfide bonds.

While the functional role for Wnt lipid modification is not clear, it has been demonstrated that it is important for full biological activity. The crystal structure of XWnt8-Fz8-CRD (PDB Accession Number: 4FOA) reveals that XWnt8 lipidation is directly implicated in Fz8-CRD binding at binding site 1 [162]. Located on the tip of the thumb extending from the XWnt8 NTD, a 15 Å tube of electron density was observed at the hydroxyl group of Ser 187. The size of the electron density suggests that it corresponds to a 14-carbon lipid chain. This lipid moiety enshrouds the contact interface and buries about 580 Å² of the total surface area, associating with 11 Fz8 residues and treading across the cleft on the Fz8-CRD surface.

Within the Wnt family of proteins, there is a high degree of conservation of apolar amino acids in the CRD contacting the acyl group, which suggests that the lipid-binding site is conserved across the Fz-CRD. This conserved region among the other Fz-CRD could potentially regulate lipid-binding affinity and could render some Wnt specificity. Lipid binding may be attributed to the high degree of surface hydrophobicity coupled with the complementarity of the lipid-CRD interface [162] whereby the lipid and apolar residues residing in the Fz-CRD core are able to come together as a consequence of solvent exclusion [162].

The thumb loop region which is comprised of residues 181-188 associates with the Fz8-CRD and at the extreme apex of the thumb loop (residues 186-188), several main chain amino acid residues for van der Waals contacts with the Fz8-CRD. At the foot of the thumb loops, a salt bridge spans across Lys 182 of Wnt and Glu 64 of Fz8 and a hydrogen bond is formed with Asn 58 of Fz. In all Wnts, the Lys and Arg are conserved at this particular location and the Glu or Asp are conserved at the Glu 64 region in 8 out

of 10 mammalian Fz-CRD [162]. It has been speculated that the lipid site 1 binding can be attributed to the lipid-in-groove contact with the residues at the base subsequently contributing.

Because Wnt is a morphogen, several studies have suggested that acylation may be implicated in compartmentalizing wnt into the cell membrane to increase local concentrations curb Wnt availability to specific target tissues [163, 164]. The XWnt8 crystal structure corroborates the notion that the lipid is accessible for fastening Wnt to the cell membrane.

Site 2 is on the opposing side of Fz8-CRD from site 1 and is comprised of residues from Cys 315 to Cys 325 disulfide at the tip of the top of the XWnt8 CTD index finger. Site 2 on the XWnt8 index finger is a twisted β strand with disulfide bonds interspersed from Gly 299 to the C-terminal Cys 338. Site 2 is comprised of hydrophobic residues Cys 315, Phe 317, Trp 319, a tandem Cys 329 – Cys 321 disulfide bond, and Val 323 forming a vast network of van der Waals contacts with the mainchain and apolar residues on the Fz8-CRD. The side chain of Trp 319 from XWnt8 is embedded in a pocket on the Fz8-CRD surface and associates with Fz8-CRD main chain residues which include 150152, and the side chain of conserved Phe 86. The residues in site 2 on XWnt8 are all varied throughout the Wnts. However, Tyr 48 and Cys 148 in the Fz8-CRD are conserved and generate van der Waals interactions with XWnt8. However, for site 1, the contact residues for Wnt and Fz are conserved apolar amino acids. Residues in Fz8-CRD that are varied may correspond to various Wnt sub-type preferences. Interestingly, Met 149 at the center of site 2 is conserved in 5 of 10 mammalian Fz-CRD, but is disparate in Fz1, 2, 3, 6, and 7.

Methods

Homology modeling

The sequence of various homologues for human neutrophil elastase, NEDD4-1, WNT, and oncogenic V600E B-RAF were obtained from the universal protein resource (Uniprot) [165]. Using the default parameters, the BLASTp program which is embedded within Uniprot was employed to search for templates for sequence alignment [166]. The 3D structures of human neutrophil elastase, NEDD4-1 HECT domain, WNT, and oncogenic V600E B-RAF were downloaded from the PDB (PDB accession numbers: 5A8X, 2XBF, 4F0A, and 3OG7 respectively) as the template structures. Prime 2016-3 in the Schrödinger Suite (Schrödinger, LLC, New York, NY). Subsequently, the SSpro program bundled with prime was utilized to predict the secondary structure of each corresponding homologue for each target protein. The alignment between the target and template sequence was performed using the ClustalW method in Prime and subsequent loop refinement was executed.

Ligand Preparation

LigPrep 2016-3 in the Schrödinger Suite was employed to build the 3D structures for each ligand. The force field used was OPLS_2005 and their ionization states were created at a pH of 7.0 ± 2.0 using Epik2.2 in the Schrödinger Suite. For each structure, all tautomers are considered. Double bonds that can transition to single bonds in any tautomer can transition between E and Z forms as long as there are no additional topological constraints that could impede this shift. The tautomer utility in LigPrep creates suitable combinations of all tautomers for each structure. Therefore, to account for the different tautomer conformations, double bonds in the structure have been removed and all combinations of 180 degree rotations about double bonds within each tautomer are created.

Docking Grid Generation

The 3D structures for the target protein homology models were optimized and prepared using the Protein Preparation Wizard in the Schrödinger Suite. Hydrogen atoms were added and bond orders were assigned to each protein. The structures were subsequently minimized to arrive at the converged root mean square deviation (RMSD) of 0.3 Å with the OPLS_2005 force field. SiteMap 2016-3 in the Schrödinger Suite was used to search for potential ligand binding sites located within the protein. Contour maps of hydrophobic or hydrophilic fields were created. Regions of donor, acceptor, and metal-binding were also identified within the contour maps. The Receptor Grid Generation in Schrödinger Suite was applied to generate the docking grid. The grid enclosing box was centered around the workspace ligand. The grid size was adjusted so that it was large enough to encapsulate all of the possible ligand binding regions in the protein.

Molecular Docking

The optimized 3D structures of the indolecarbinol compounds were docked into the docking grid in the 3D homology model of each target protein using Glide 2016-3 in Schrödinger Suite with standard precision (SP).

Results

NEDD4-1 homology modeling

To understand the specificity of I3C and 1-benzyl-I3C recognition for each target protein (human E3 ubiquitin ligase NEDD4-1, human neutrophil elastase, oncogenic B-RAF V600E serine/threonine kinase, and human Wnt), a comprehensive comparison of each target protein's surface topology and their corresponding homologues was performed. Homology models were constructed from their template structures which were obtained from the Protein Data Bank (<http://www.rcsb.org/pdb/home/home.do>). Here, the 3D structures of the indolecarbinol-binding sites and the contacting amino acid residues have been extensively examined. Distinctive molecular patterns between each ligand and interaction site on each target protein can be observed.

Homologues of human NEDD4-1 were obtained by entering the template FASTA sequence (Accession code: NP_006145.2) into the BLASTp program provided by the universal protein resource (Uniprot) [165, 166]. Previous studies have reported that 3D structures are classified as similar if the sequence identity between two proteins is higher than 25% [16, 167]. To build the 3D structural homology models, the human NEDD4-1 Hect domain crystal structure was used as a template (PDB accession code: 2XBF). Five homologues (neural cell expressed, developmentally down-regulated 4-like isoform CRA_b, highly similar to E3 ubiquitin-protein ligase NEDD4-like protein, , E3 ubiquitin protein ligase HUWE1, and E3 ubiquitin-protein ligase HERC2) from the BLASTp search were selected and ranged in sequence similarity to NEDD4-1 from 26.5%-64.1% (Figure 3.1).

Uniprot ID	Protein Name	Sequence Identity to NEDD4-1 (%)
A0A024R281	Neural cell expressed, developmentally down-regulated 4like, isoform CRA_b	64.1
B7Z6K0	cDNA FLJ53199, highly similar to E3 ubiquitin-protein ligase NEDD-4like protein	62.6
K7ERN1	E3 ubiquitin-protein ligase NEDD4-like	60.0
Q7Z6Z7	E3 ubiquitin-protein ligase HUWE1	54.7
O95714	E3 ubiquitin-protein ligase HERC2	26.5

Figure 3.1. E3 ubiquitin ligase NEDD4-1 homologous proteins. Table depicts % sequence identity of each corresponding homologous protein to NEDD4-1. Proteins are identified by their UniProt ID.

The alignment of the NEDD4-1 homologues and the HECT domain template sequence was executed using ClustalW in Prime. The model subsequently underwent non-templated loop refinement using the default sampling method.

Molecular Docking of NEDD4-1 homologous proteins and indolecarbinol compounds

The different binding poses of the I3C, 1-benzyl-I3C, and their corresponding analogs were assessed and ranked based on their docking scores. The pose that exhibited the lowest score implies the most reasonable binding mode for each indolecarbinol compound.

Previous studies have established that NEDD4-1 is an I3C target protein [168]. Within the human genome, there are 28 members of the HECT E3 family in which the catalytic HECT domain is positioned at the C terminus of the proteins and various domains located at the N-terminus confer substrate binding. The human HECT E3 ligases can be generally divided into three families that include the NEDD4, HERC, and other HECT members. Given the relatively high sequence similarity that is shared between members of the various HECT families, several homologues of the NEDD4-1 protein were selected and their corresponding sequences were then aligned to the crystal structure of the NEDD4-1 HECT domain (PDB accession code: 2XBF) to generate homology models. The neural cell expressed, developmentally down-regulated 4-like, isoform CRA_b (Uniprot ID: A0A024R281) had a sequence identity of 64.1% and exhibited the highest ranked homology in the BLAST search for human NEDD4-1 protein. The subsequent homologues selected for comparison progressively shared less sequence similarity so as to note the changes in the indolecarbinol binding pockets across various HECT families. cDNA FLJ53199, highly similar to E3 ubiquitin-protein ligase NEDD4-like protein (Uniprot ID: B7Z6K0), E3 ubiquitin-protein ligase NEDD4-like (Uniprot ID: K7ERN1), E3 ubiquitin-protein ligase HUWE1 (Uniprot ID: Q7Z6Z7), and E3 ubiquitin-protein ligase HERC2 (Uniprot ID: O95714) sequences were used in the homology, molecular docking studies and exhibited sequence similarities of 62.6%, 60.0%, 54.7%, and 26.5% respectively (Figure 3.1). Because the full-length human NEDD4-1 structure has not yet been crystalized, the HECT domain which includes residues 519-900 (as notated on the human full-length NEDD4-1 isoform 1 sequence) was used as a template to build the homology models. Therefore, the N-terminal region of the selected homologues could not be aligned with the template structure because homologue sequences are larger than the template sequence. Prime was used to synthesize the initial homology models of the HECT family members, whereby the model that exhibited the lowest energy was utilized for further optimization.

Previous crystallographic studies conducted by Kathman *et al.* [169] revealed an indole-based inhibitor that was covalently bound to the catalytic HECT domain of NEDD4-1 (PDB accession code: 5C91). This co-crystallized structure was hence used as a structural guide for docking I3C, 1-benzyl-I3C, and their corresponding analogs onto the NEDD4-1 HECT homology models. The homology models were superimposed onto the crystal structure of the NEDD4-1 HECT domain in adducts with the indole-based

inhibitor, 4YU, to obtain a starting docking position for I3C, 1-benzyl-I3C, and their derivatives (Figure 3.2A). The 7 indolecarbinol compounds were docked onto the HECT domain of the homology models. In the 5 homology models selected to perform molecular docking studies, all harbored a catalytic HECT domain located on the C-terminus. (For the sake of clarity, I3C contact points will be described in detail and will serve as a means of comparison between each homology model examined. For the neural cell expressed, developmentally down-regulated 4like, isoform CRA_b (Uniprot ID: A0A024R281) homology model, I3C was predicted to bind to a hydrophobic pocket comprised of residues Tyr 588, Phe 507, Leu 561, Tyr 559, and Cys 581 (Figure 3.2B). Here the indole ring of I3C is able to engage in stable pi-pi stacking interactions with Tyr 559 and the hydroxyl moiety that is attached to the indole ring forms a hydrogen bond interaction with the side chain of Glu 508. The remaining indolecarbinol compounds are embedded in the same hydrophobic pocket with slight deviations in orientations based on the various moieties attached at the indole ring and their interacting residues can be summarized in Figure 3.2C.

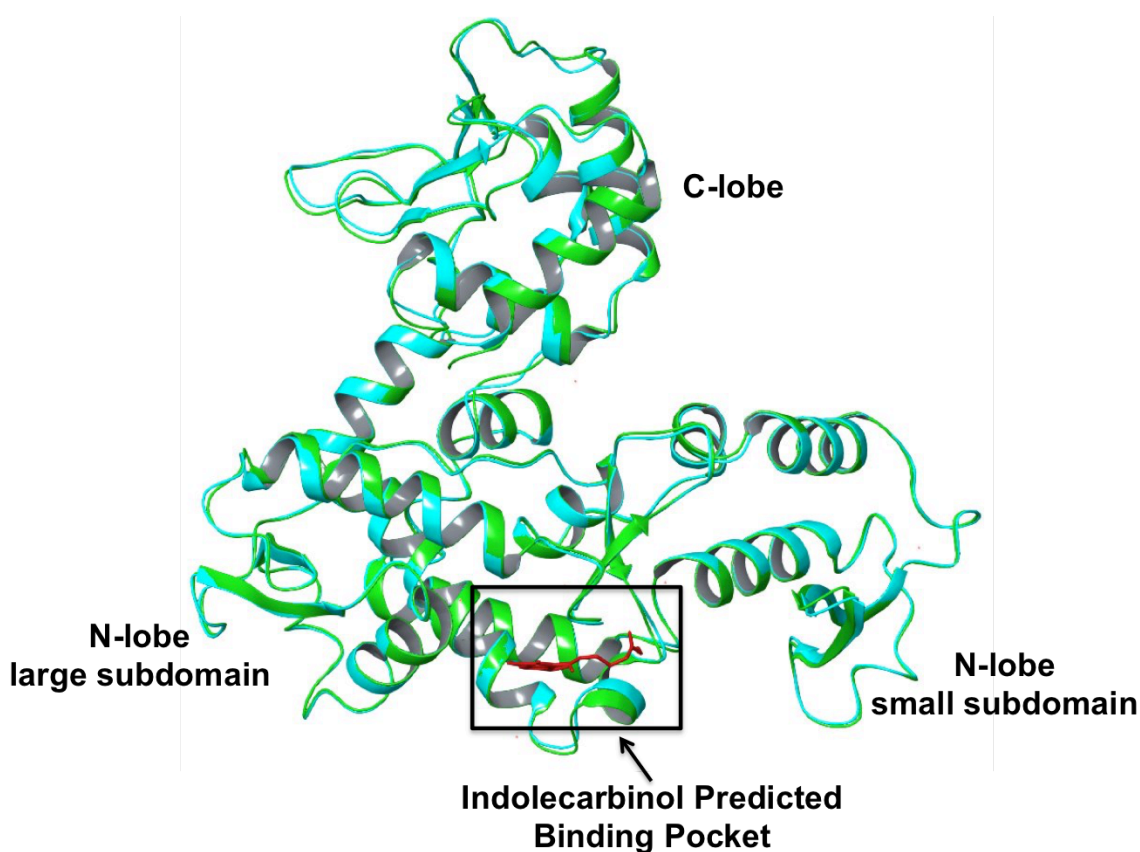


Figure 3.2A. Homology model of Neural cell expressed, developmentally down-regulated 4-like, isoform CRA_b (Uniprot ID: A0A024R281). Ribbon diagram of homology model shown in teal, superimposed onto human NEDD4-1 Hect domain crystal structure (PDB accession code: 5C91) shown in green. Previously crystallized indole-based small molecule is depicted in red and shows region of potential indolecarbinol binding.

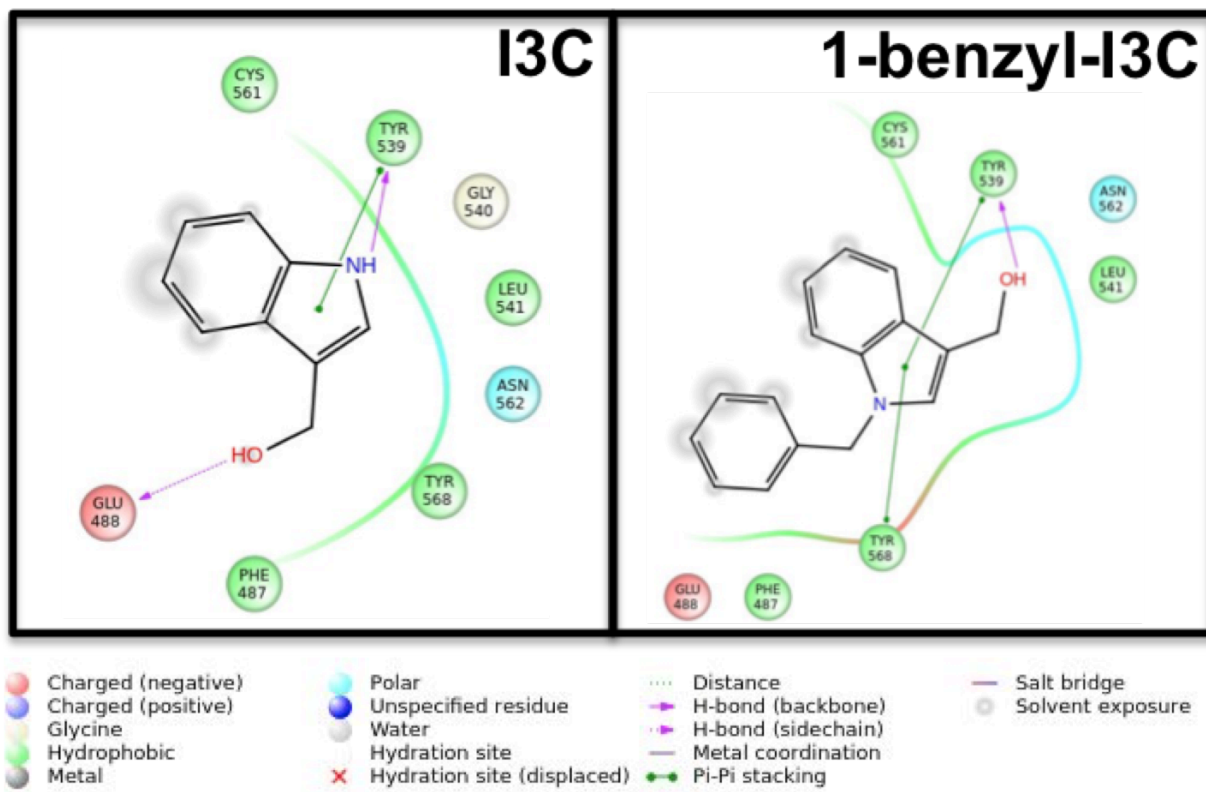
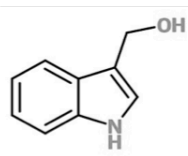
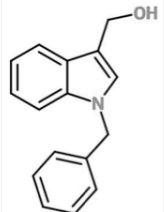
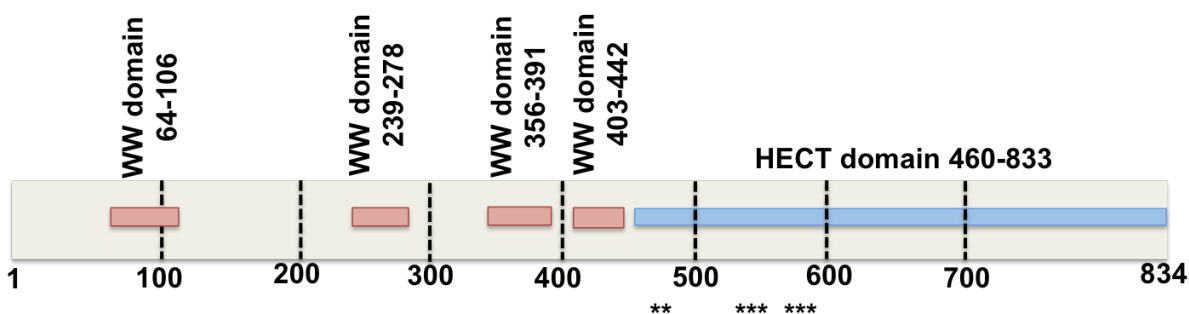


Figure 3.2B. Ligand interaction diagram of I3C and 1-benzyl-I3C in adducts with Neural cell expressed, developmentally down-regulated 4-like, isoform CRA_b. Predicted molecular docking interactions diagram depicting close contact residues within 4 Å of either I3C or 1-benzyl-I3C.

Indolecarbinol Compound	Structure	Predicted Interacting Amino Acid Residue	Amino Acid Biological Property
I3C		Phe 487 Tyr 539 Leu 541 Cys 561 Tyr 568	Hydrophobic
		Asn 562	Polar
		Glu 488	Charged (negative)
		Gly 540	Unique
1-benzyl-I3C		Phe 487 Tyr 539 Leu 541 Cys 561 Tyr 568	Hydrophobic
		Asn 562	Polar
		Glu 488	Charged (negative)



I3C Predicted Contact Points: Phe 487, Glu 488, Tyr 539, Gly 540, Leu 541, Cys 561, Asn 562, Tyr 568 (denoted by *)

Figure 3.2C. Table and diagram of predicted molecular docking interactions of I3C and 1-benzyl-I3C with Neural cell expressed, developmentally down-regulated 4-like, isoform CRA_b. Table summarizes I3C and 1-benzyl-I3C interactions with their close contact residues and each residue's corresponding biological property. All interactions described are within 4 Å of each indolecarbinol analog. Figure below shows I3C predicted contact points within each subdomain of the homology model.

In the cDNA FLJ53199, highly similar to E3 ubiquitin-protein ligase NEDD-4like homology model (Uniprot ID: B7Z6K0) (Figure 3.3A), the indole ring engages in pi-pi stacking interactions with Tyr 559 and is securely fastened by hydrophobic residues Phe 507, Leu 561, Cys 581, and Tyr 588 and polar residue Asn 582. The hydroxyl moiety makes hydrogen bond interactions with the side-chain of Glu 508 and the nitrogen on the indole ring also engages in a hydrogen bond interaction with the backbone of Tyr 559 (Figure 3.3B & C).

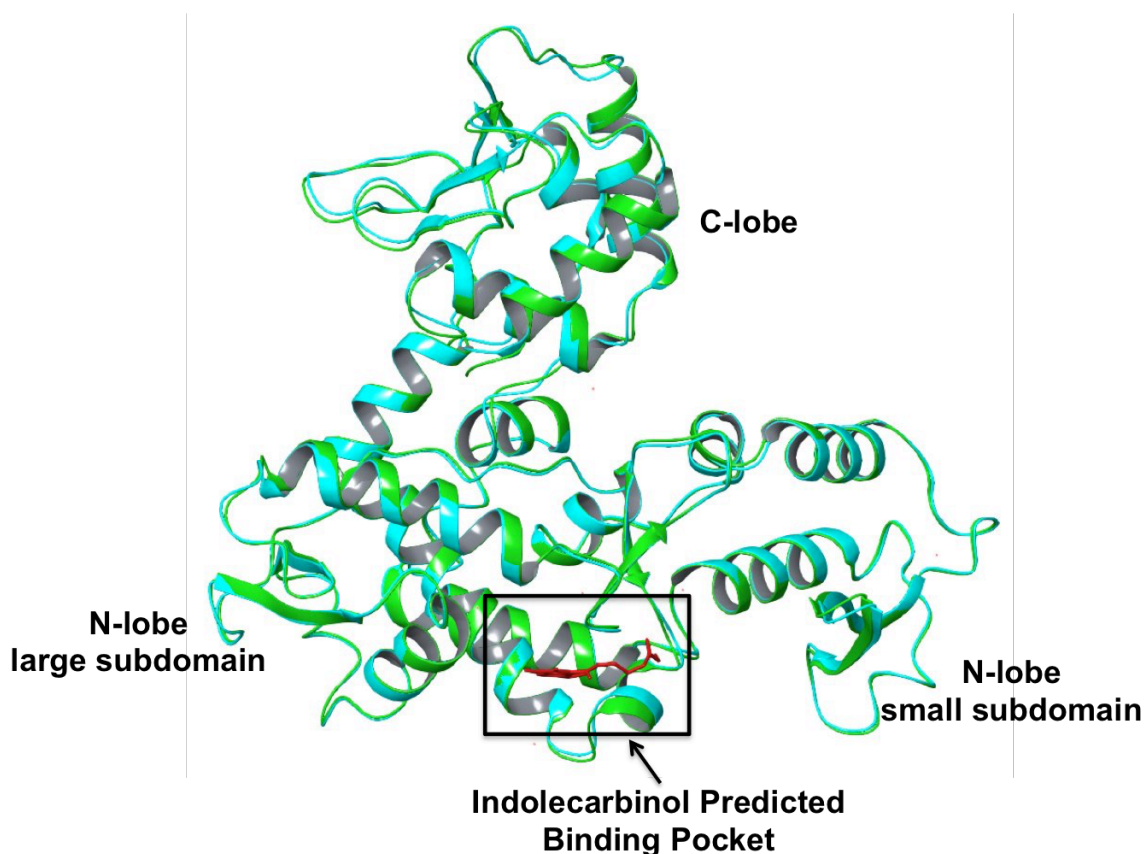


Figure 3.3A. Homology model of cDNA FLJ53199, highly similar to E3 ubiquitin-protein ligase NEDD4-like (Uniprot ID: B7Z6K0). Ribbon diagram of homology model shown in teal, superimposed onto human NEDD4-1 Hect domain crystal structure (PDB accession code: 5C91) shown in green. Previously crystallized indole-based small molecule is depicted in red and shows region of potential indolecarbinol binding.

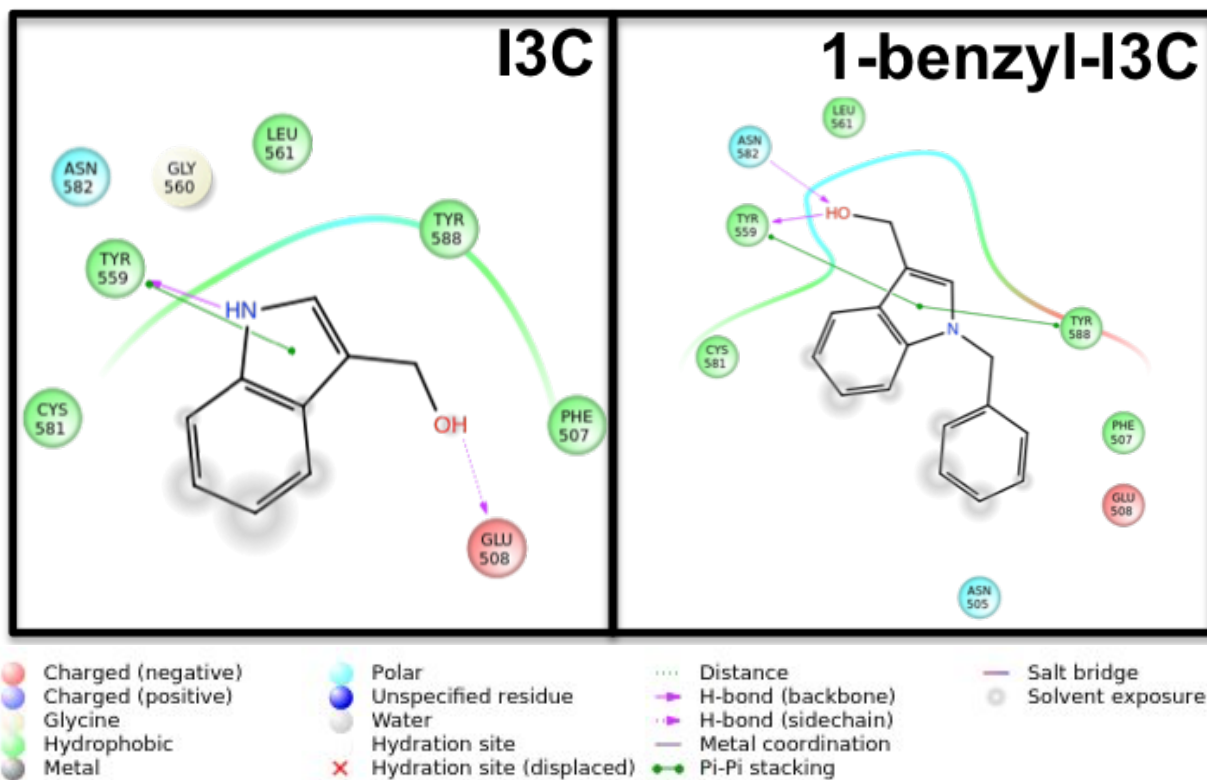
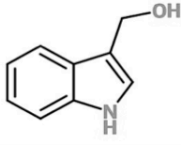
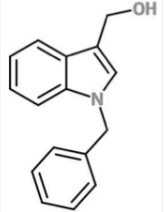
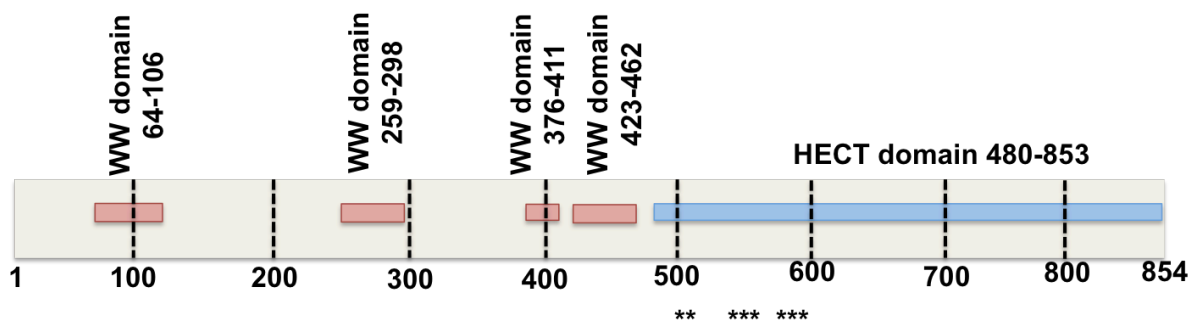


Figure 3.3B. Ligand interaction diagram of I3C and 1-benzyl-I3C in adducts with cDNA FLJ53199, highly similar to E3 ubiquitin-protein ligase NEDD4-like. Predicted molecular docking interactions diagram depicting close contact residues within 4 Å of either I3C or 1-benzyl-I3C.

Indolecarbinol Compound	Structure	Predicted Interacting Amino Acid Residue	Amino Acid Biological Property
I3C		Cys 581 Tyr 559 Leu 561 Tyr 588 Phe 507	Hydrophobic
		Asn 582	Polar
		Glu 508	Charged (negative)
1-benzyl-I3C		Cys 581 Tyr 559 Leu 561 Tyr 588 Phe 507	Hydrophobic
		Asn 582 Asn 505	Polar
		Glu 508	Charged (negative)



I3C Predicted Contact Points: Phe 507, Glu 508, Tyr 559, Gly 560, Leu 561, Cys 581, Asn 582, Tyr 588 (denoted by *)

Figure 3.3C. Table and diagram of predicted molecular docking interactions of I3C and 1-benzyl-I3C with cDNA FLJ53199, highly similar to E3 ubiquitin-protein ligase NEDD4-like. Table summarizes I3C and 1-benzyl-I3C interactions with their close contact residues and each residue's corresponding biological property. All interactions described are within 4 Å of each indolecarbinol analog. Figure below shows I3C predicted contact points within each subdomain of the homology model.

For the E3 ubiquitin-protein ligase NEDD4-like (Uniprot ID: K7ERN1) (Figure 3.4A), I3C was predicted to be enveloped by hydrophobic residues Phe 292, Leu 346, and Cys 366. Additionally, the indole ring on I3C stably associates with hydrophobic residues Tyr 344 and Tyr 373 through pi-pi stacking interactions. The hydroxyl moiety engages in a hydrogen bonding interaction with the backbone of Tyr 344. I3C is also anchored in between negatively charged residue Glu 293 and polar residue Asn 367 (Figure 3.4B & C).

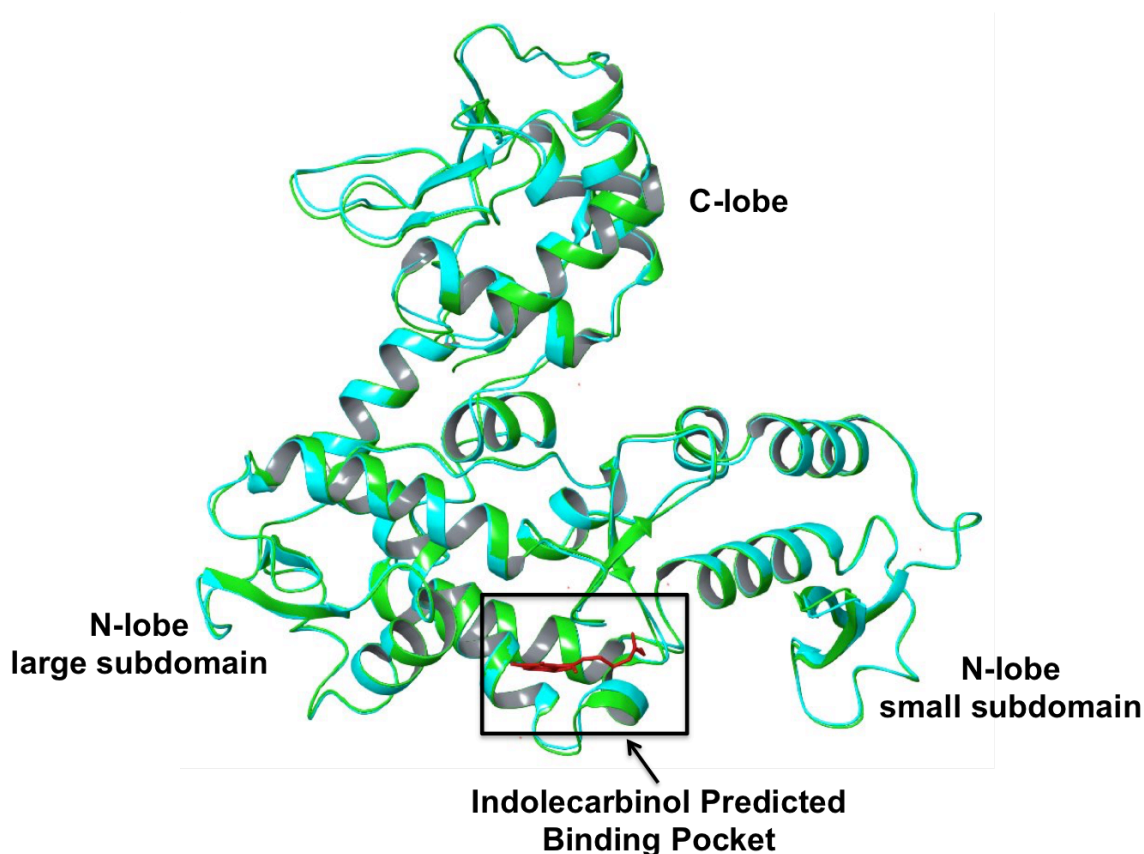


Figure 3.4A. Homology model of E3 ubiquitin-protein ligase NEDD4-like (Uniprot ID: K7ERN1). Ribbon diagram of homology model shown in teal, superimposed onto human NEDD4-1 Hect domain crystal structure (PDB accession code: 5C91) shown in green. Previously crystallized indole-based small molecule is depicted in red and shows region of potential indolecarbinol binding.

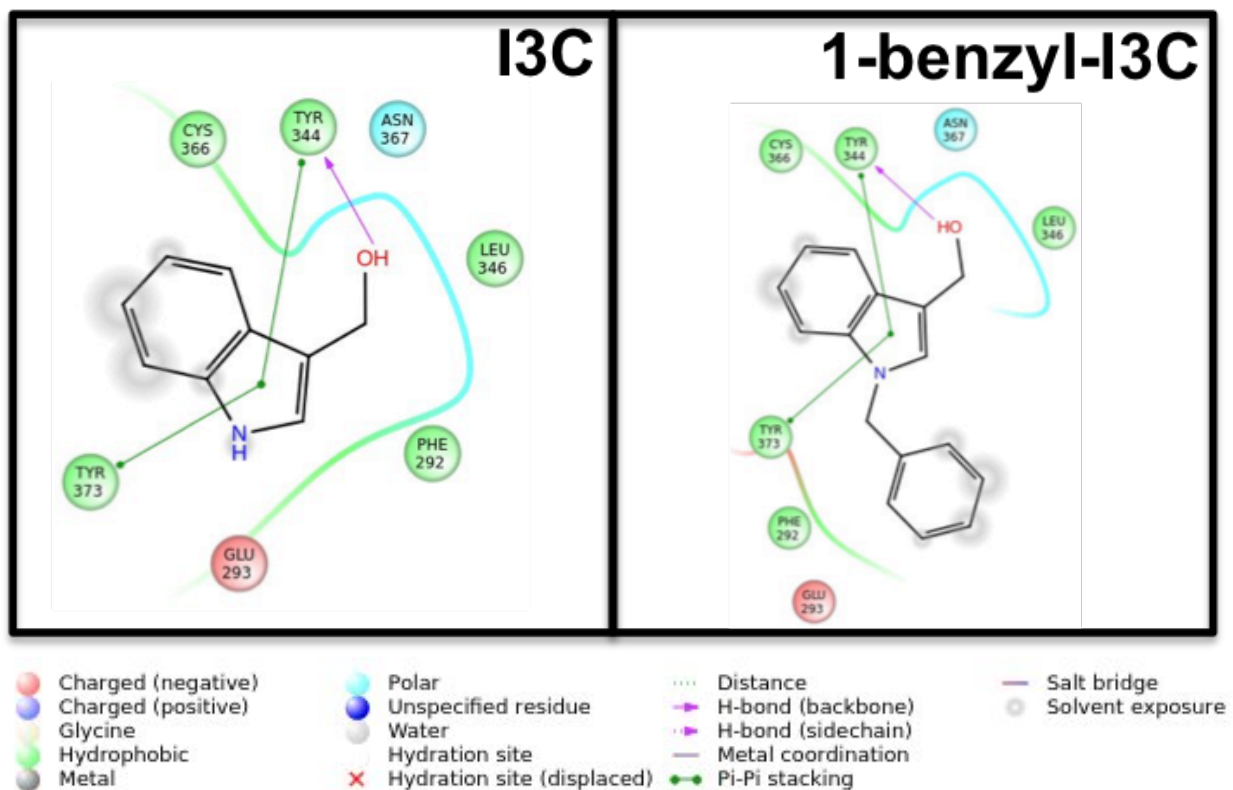
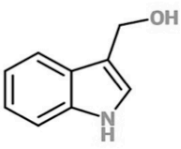
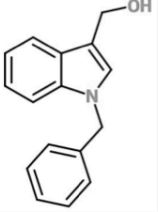
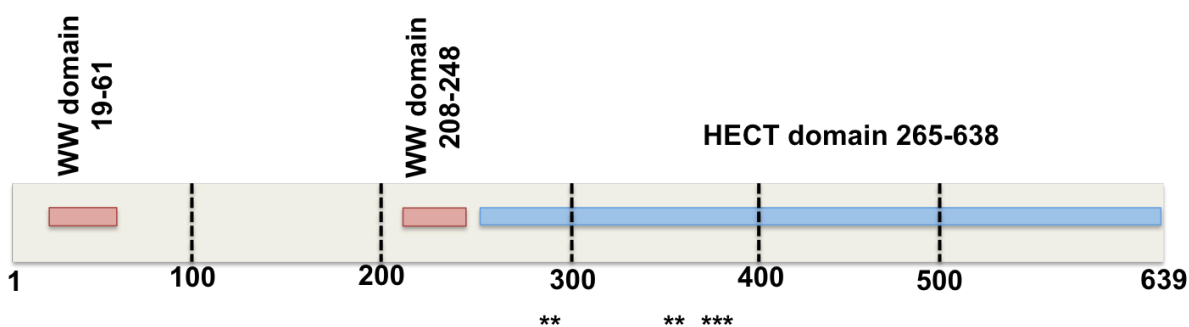


Figure 3.4B. Ligand interaction diagram of I3C and 1-benzyl-I3C in adducts with E3 ubiquitin-protein ligase NEDD4-like homology model. Predicted molecular docking interactions diagram depicting close contact residues within 4 Å of either I3C or 1-benzyl-I3C.

Indolecarbinol Compound	Structure	Predicted Interacting Amino Acid Residue	Amino Acid Biological Property
I3C		Phe 292 Tyr 344 Leu 346 Cys 366 Tyr 373	Hydrophobic
		Asn 367	Polar
		Glu 293	Charged (negative)
1-benzyl-I3C		Phe 292 Tyr 344 Leu 346 Cys 366 Tyr 373	Hydrophobic
		Asn 367	Polar
		Glu 293	Charged (negative)



I3C Predicted Contact Points: Phe 292, Glu 293, Tyr 344, Leu 346, Cys 366, Asn 367, Tyr 373 (denoted by *)

Figure 3.4C. Table and diagram of predicted molecular docking interactions of I3C and 1-benzyl-I3C with E3 ubiquitin-protein ligase NEDD4-like. Table summarizes I3C and 1-benzyl-I3C interactions with their close contact residues and each residue's corresponding biological property. All interactions described are within 4 Å of each indolecarbinol analog. Figure below shows I3C predicted contact points within each subdomain of the homology model.

Similarly, the E3 ubiquitin-protein ligase, HUWE1 (Uniprot ID: Q7Z6Z7) homology model exhibited a hydrophobic pocket akin to the NEDD4-1 HECT domain (PDB accession code: 5C91) (Figure 3.5A). I3C was predicted to bind to a pocket comprised of hydrophobic residues Phe 4027, Tyr 4031, Met 4073, Tyr 4078, Ala 4079, Leu 4080, Cys 4099, Tyr 4106, and Phe 4109 negatively charged residues Glu 4028 and Glu 4072, and polar residue Asn 4100. The hydroxyl moiety on the indole ring is able to associate with both Tyr 4078 and Asn 4100 through hydrogen bond interactions and the nitrogen on the indole ring can also engage in a hydrogen bond interaction with the sidechain of Glu 4028 (Figure 3.5B & C).

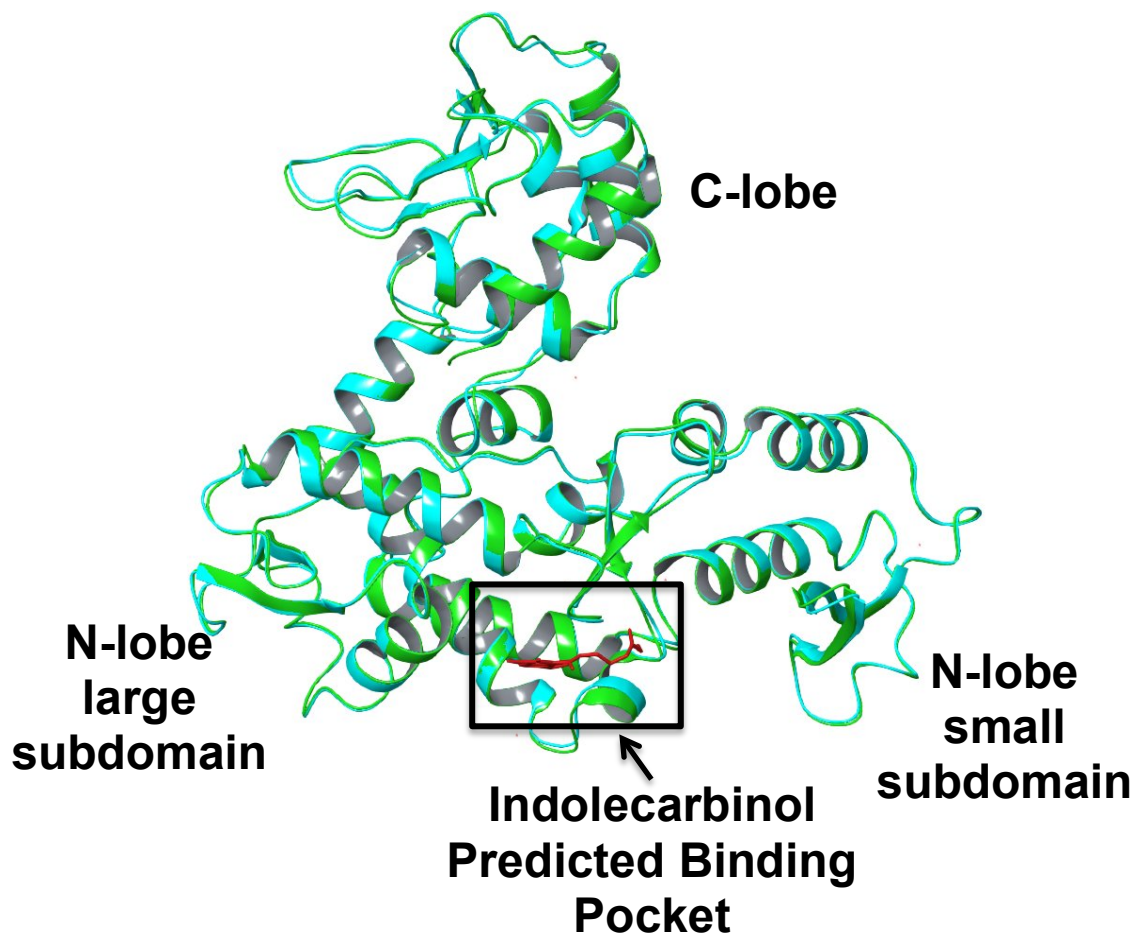


Figure 3.5A. Homology model of E3 ubiquitin-protein ligase, HUWE1 (Uniprot ID: Q7Z6Z7) homology model. Ribbon diagram of homology model shown in teal, superimposed onto human NEDD4-1 Hect domain crystal structure (PDB accession code: 5C91) shown in green. Previously crystallized indole-based small molecule is depicted in red and shows region of potential indolecarbinol binding.

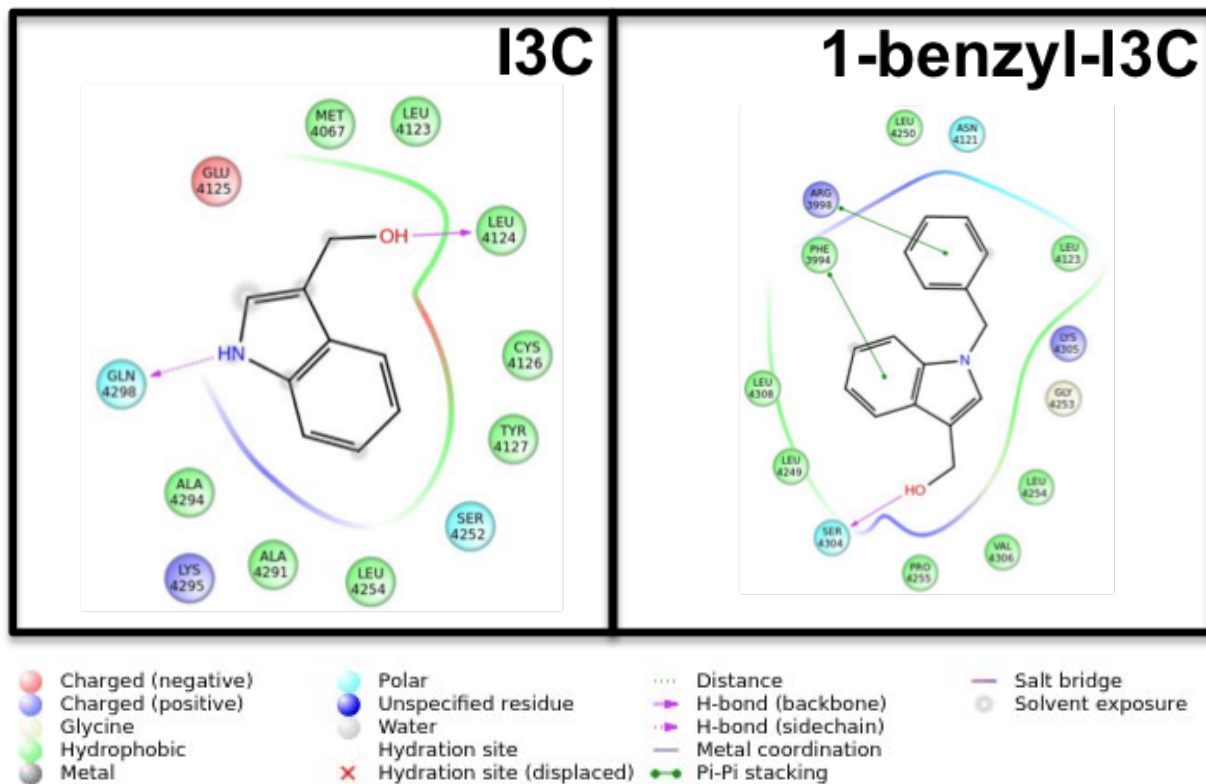
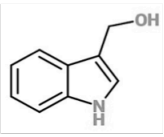
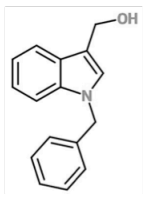
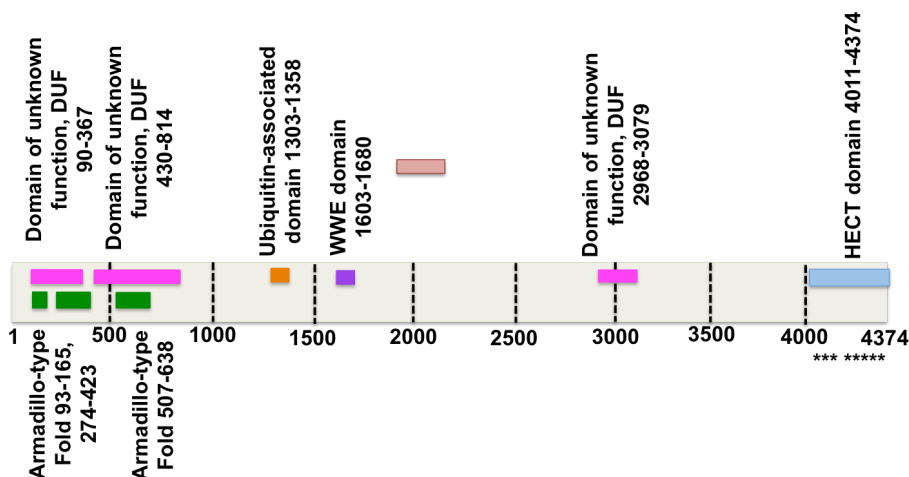


Figure 3.5B. Ligand interaction diagram of I3C and 1-benzyl-I3C in adducts with E3 ubiquitin-protein ligase, HUWE1 homology model. Predicted molecular docking interactions diagram depicting close contact residues within 4 Å of either I3C or 1-benzyl-I3C.

Indolecarbinol Compound	Structure	Predicted Interacting Amino Acid Residue	Amino Acid Biological Property
I3C		Met 4067 Leu 4123 Leu 4124 Cys 4126 Tyr 4127 Leu 4254 Ala 4291 Ala 4294	Hydrophobic
		Ser 4252 Gln 4298	Polar
		Glu 4125	Charged (negative)
		Lys 4295	Charged (positive)
1-benzyl-I3C		Phe 3994 Leu 4123 Leu 4249 Leu 4250 Leu 4254 Pro 4255 Val 4306 Leu 4308	Hydrophobic
		Asn 4121 Ser 4304	Polar
		Arg 3998 Lys 4305	Charged positive)



I3C Predicted Contact Points: Phe 4027, Glu 4028, Tyr 4031, Glu 4072, Met 4073, Tyr 4078, Ala 4079, Leu 4080, Cys 4099, Asn 4100, Tyr 4106, Phe 4109 (denoted by *)

Figure 3.5C. Table and diagram of predicted molecular docking interactions of I3C and 1-benzyl-I3C with E3 ubiquitin-protein ligase, HUWE1. Table summarizes I3C and 1-benzyl-I3C interactions with their close contact residues and each residue's corresponding biological property. All interactions described are within 4 Å of each indolecarbinol analog. Figure below shows I3C predicted contact points within each subdomain of the homology model.

SiteMap [170] was employed to search and examine the potential ligand binding sites in the target proteins. The program synthesizes information on the binding site characteristics and begins its calculations with a search that identifies one or more regions on or near the protein surface that may be reasonable for ligand binding to a protein. Site points are generated and is comprised of a grid of points which is used to find the sites. Contour maps or site maps are then created to yield hydrophobic and hydrophilic maps. These hydrophilic maps are then separated into donor, acceptor, and metal-binding regions. Each site is then examined by determining various properties such as hydrophobic regions that have the capacity to fit a larger hydrophobic group. These site maps can be implemented to generate a target for ligand docking with Glide and to assess potential ligand binding sites by unveiling how well the ligand poses exhibit shape complementarity to the protein site. Here, the 7 indolecarbinol compounds tested followed the most favorable pose with the lowest free energy conformation.

SiteMap was employed to identify the top-ranking surface clefts in the E3 ubiquitin-protein ligase HERC2 homology model (UniProt ID: O95714) because the predicted indolecarbinol binding pocket that was observed in the opposing NEDD4-1 homology models appeared to be lacking essential residues that were necessary for anchoring the ligands in place (Figure 3.6). Superimposing the HERC2 homology model on to the NEDD4-1 HECT domain co-crystallized small-molecule indole inhibitor structure (PDB accession code: 5C91) and then subsequently docking I3C, 1-benzyl-I3C, and their corresponding analogs onto the 4YU binding site did not yield feasible, low free energy scores that would suggest a region of high binding affinity for the compounds. Therefore, SiteMap was implemented to locate a more stable binding region for the indolecarbinol compounds. Ten potential binding sites were determined. The best-ranked site yielded a SiteScore of 1.236 where a score of greater than 1 suggests a conceivable binding site. This site is adjacent to the C-lobe and docking analysis shows I3C embedded in a pocket comprised of polar residues Thr 4672, Ser 4677, and Thr 4727, hydrophobic residues Pro 46798, Leu 4729, and positively charged residue Arg 4728 (Figure 3.6B & C).

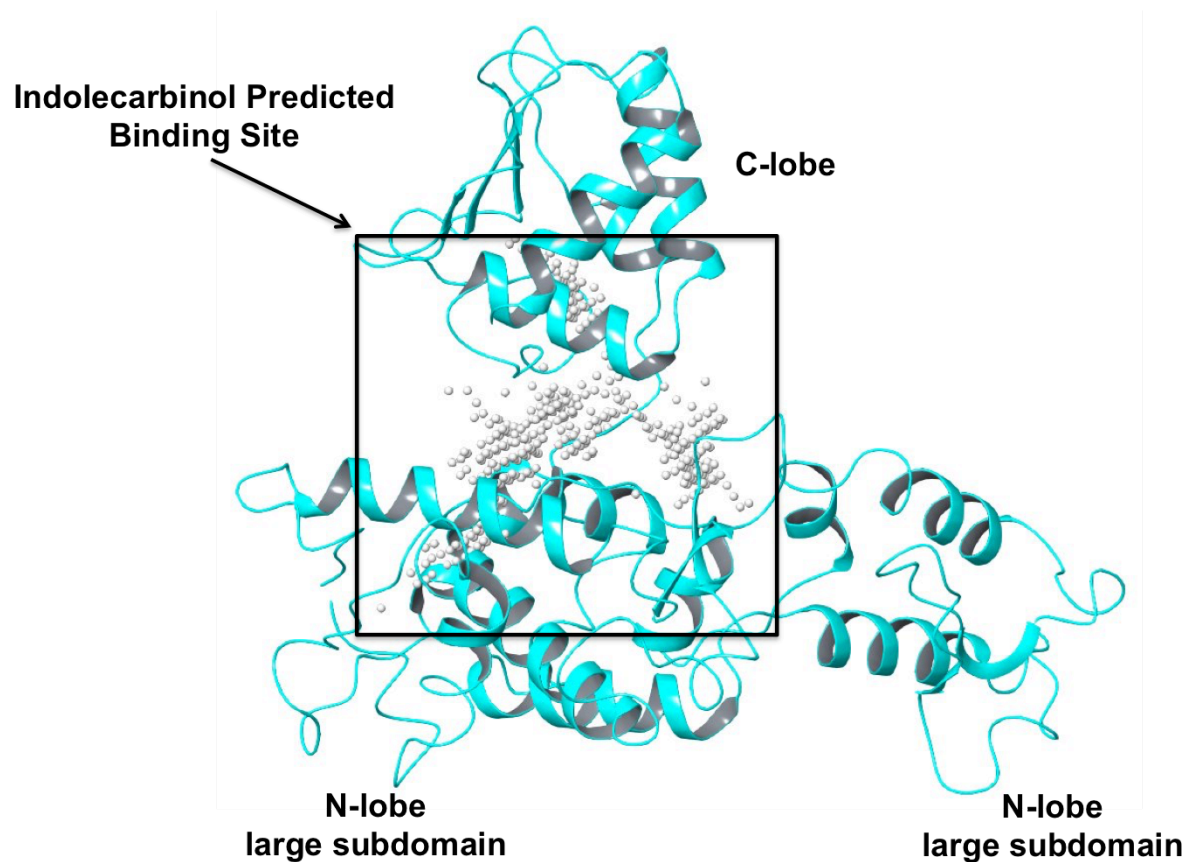


Figure 3.6A. Homology model of E3 ubiquitin-protein ligase HERC2 (UniProt ID: O95714) . Ribbon diagram of homology model. Box encircles SiteMap's best predicted indolecarbinol binding site.

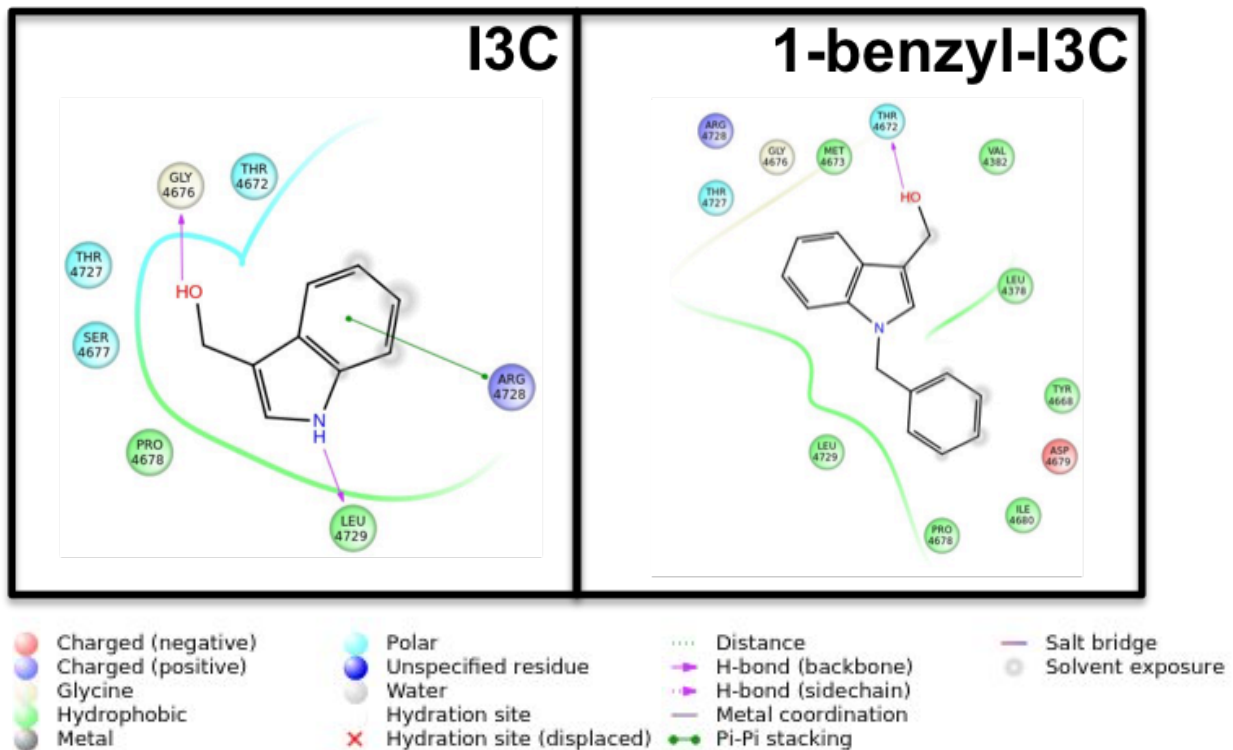
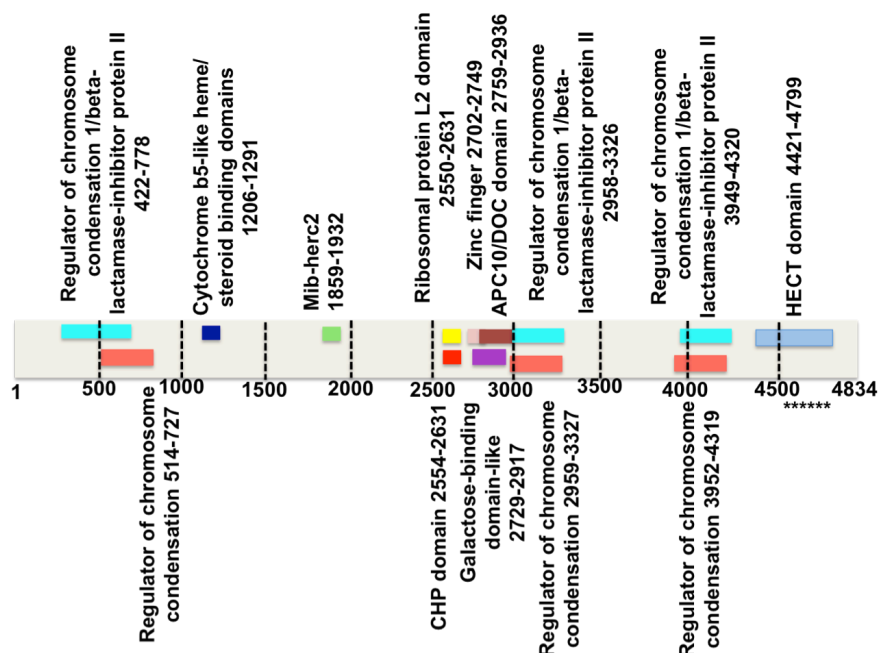


Figure 3.6B. Ligand interaction diagram of I3C and 1-benzyl-I3C in adducts with E3 ubiquitin-protein ligase HERC2 homology model. Predicted molecular docking interactions diagram depicting close contact residues within 4 Å of either I3C or 1-benzyl-I3C.

Indolecarbinol Compound	Structure	Predicted Interacting Amino Acid Residue	Amino Acid Biological Property
I3C		Pro 4678 Leu 4729	Hydrophobic
		Thr 4672 Ser 4677	Polar
		Arg 4728	Charged (positive)
1-benzyl-I3C		Leu 4378 Val 4382 Tyr 4668 Met 4673 Pro 4678 Ile 4680 Leu 4729	Hydrophobic
		Thr 4672 Thr 4727	Polar
		Arg 4728	Charged (positive)
		Asp 4679	Charged (negative)
		Gly 4676	Unique



I3C Predicted Contact Points: Thr 4672, Gly 4676, Ser 4677, Pro 4678, Thr 4727, Leu 4729 (denoted by *)

Figure 3.6C. Table and diagram of predicted molecular docking interactions of I3C and 1-benzyl-I3C with E3 ubiquitin-protein ligase, HERC2. Table summarizes I3C and 1-benzyl-I3C interactions with their close contact residues and each residue's corresponding biological property. All interactions described are within 4 Å of each indolecarbinol analog. Figure below shows I3C predicted contact points within each subdomain of the homology model.

As previously reported in the NEDD4-1 HECT domain in complex with indole small molecule (4YU) (PDB accession code: 5C91), it was observed that the indole small molecule inhibitor is able to association with the non-catalytic Cys 627, which is responsible for elongating poly-ubiquitin chains. The indole core is also anchored in a pocket of the N-lobe which is composed of residues Leu 553, Glu 554, Asn 602, Tyr 605, Leu 607, and Tyr 634. In the NEDD4-1 homology models that exhibited a sequence identity of greater than 50%, it is evident that hydrophobic residues such as Tyr, Cys, and Leu, and Phe are critical for stably fastening the indolecarbinol compounds within the small sub-domain of the N-lobe. Interestingly, among the selected homology models that did share a relatively high sequence identity of greater than 50%, the predicted binding pockets within these regions were all comprised of residues that are also present in the 4YU binding site of the NEDD4-1 HECT domain crystal structure (PDB accession code: 5C91) (Figure 3.7), which again is a testament of the significance of these residues.

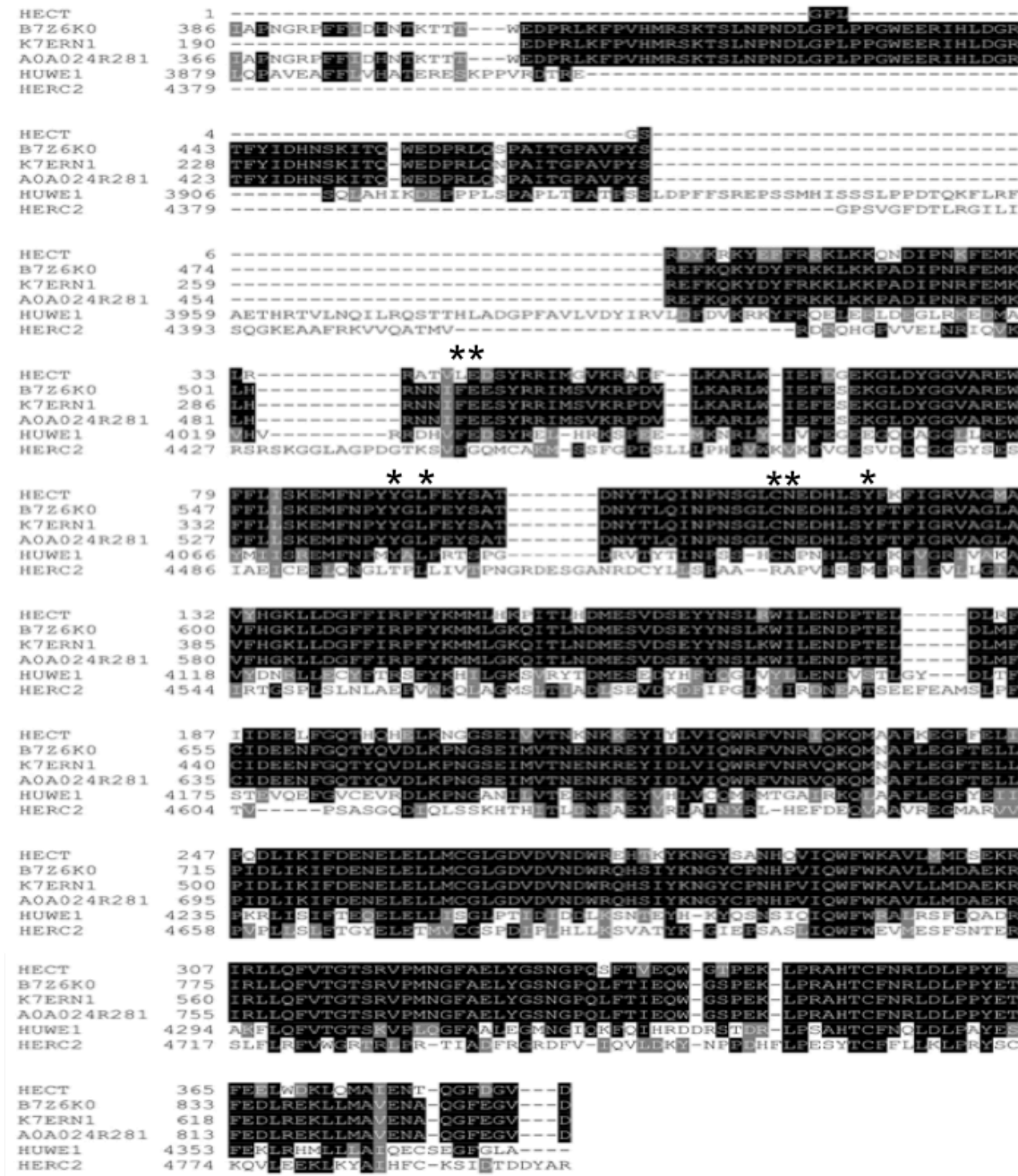


Figure 3.7. Sequence alignment of NEDD4-1 HECT domain and corresponding homologous HECT domain containing proteins. E3 ubiquitin Nedd4-like proteins are identified by their UniProt ID. * denotes predicted I3C amino acid contact points. Residues highlighted in black correspond to conserved sequences, regions in grey correspond to amino acid residues that are not conserved, but share the same biological property, while residues not highlighted show areas that are not conserved.

Elastase Homology Modeling

While previous studies have indicated that I3C acts as a non-competitive inhibitor for human neutrophil elastase [171], thus establishing that I3C binds outside the active site, predictions can be made about indolecarbinol potential binding sites on other highly homologous proteins. Three homologues proteinase 3 (Uniprot ID: D6CHE9), azurocidin (Uniprot ID: P20160), and PRSSL1 protein (Uniprot ID: B7ZMF6) from the BLASTp search were selected and ranged in sequence similarity to NEDD4-1 from 35.7%-54.6% (Figure 3.8).

Uniprot ID	Protein Name	Sequence Identity to HNE (%)
D6CHE9	Proteinase 3	54.6
P20160	Azurocidin	45.7
B7ZMF6	PRSSL1 protein	35.7

Figure 3.8. Elastase homologous proteins. Table depicts % sequence identity of each corresponding homologous protein to elastase. Proteins are identified by their UniProt ID.

Molecular Docking of elastase homologous proteins and indolecarbinol compounds

Proteinase 3, which is a polymorphonuclear leukocyte serine protease that is responsible for degrading elastin, fibronectin, laminin, vitronectin, and collagen types I, III, and IV and has been implicated in causing of emphysema. Proteinase 3 (Uniprot ID: D6CHE9) shares a 54.6% sequence identity to human neutrophil elastase. Because a crystal structure between proteinase 3 and an indole compound is not currently available in the Protein Data Bank, SiteMap was used to identify potential surface clefts (Figure 3.9A) and from here on was also employed to determine potential binding sites on the remaining homology models examined. The indole ring on I3C is oriented towards a pocket located in the C-terminal and is comprised of hydrophobic residues Pro 104, Val 105, Pro 106, Val 136, Val 137, Cys 141, Pro 143, Ile 146, Cys 147, positively charged residues Arg 142 and Arg 194, and polar residues Thr 138 and Asn 145. The indole parent ring is securely fastened into this pocket by engaging in hydrogen bond interactions with the backbone of Val 137 and Cys 147 through its hydroxyl moiety. The nitrogen on the indole ring is also able to associate with Val 105 through hydrogen bond interactions (Figure 3.9B & C).

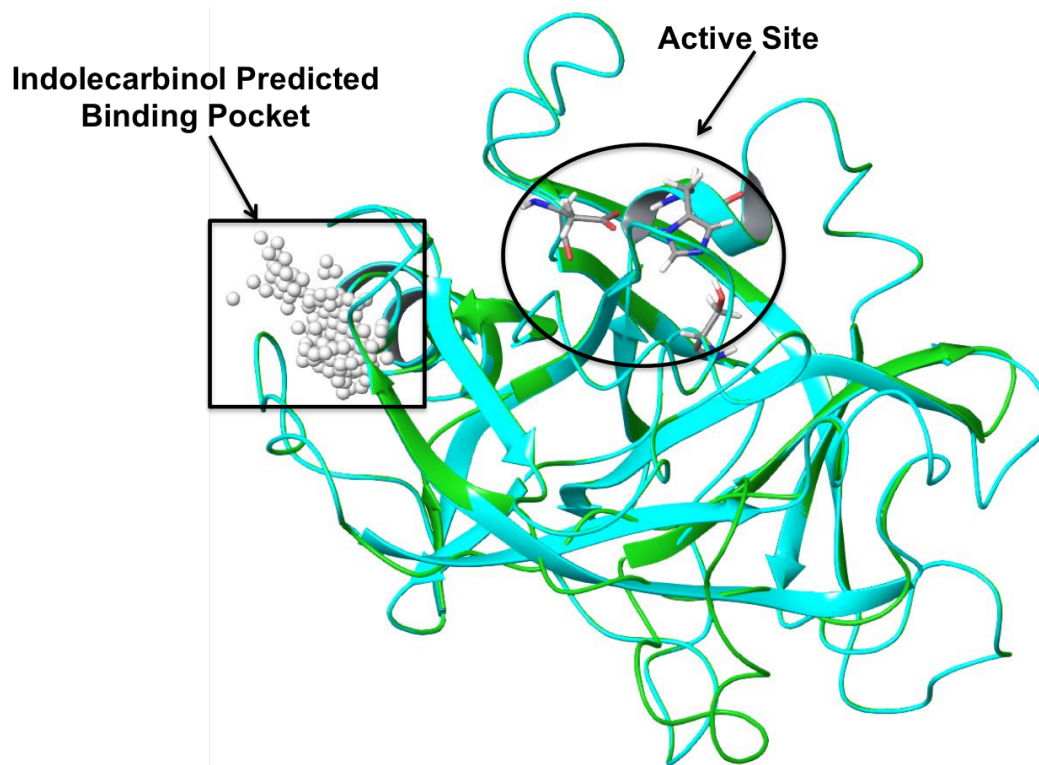


Figure 3.9A. Homology model of proteinase 3 (Uniprot ID: D6CHE9). Ribbon diagram of homology model shown in teal, superimposed onto human neutrophil elastase crystal structure (PDB accession code: 5A8X) shown in green. Box encircles SiteMap's best predicted indolecarbinol binding site.

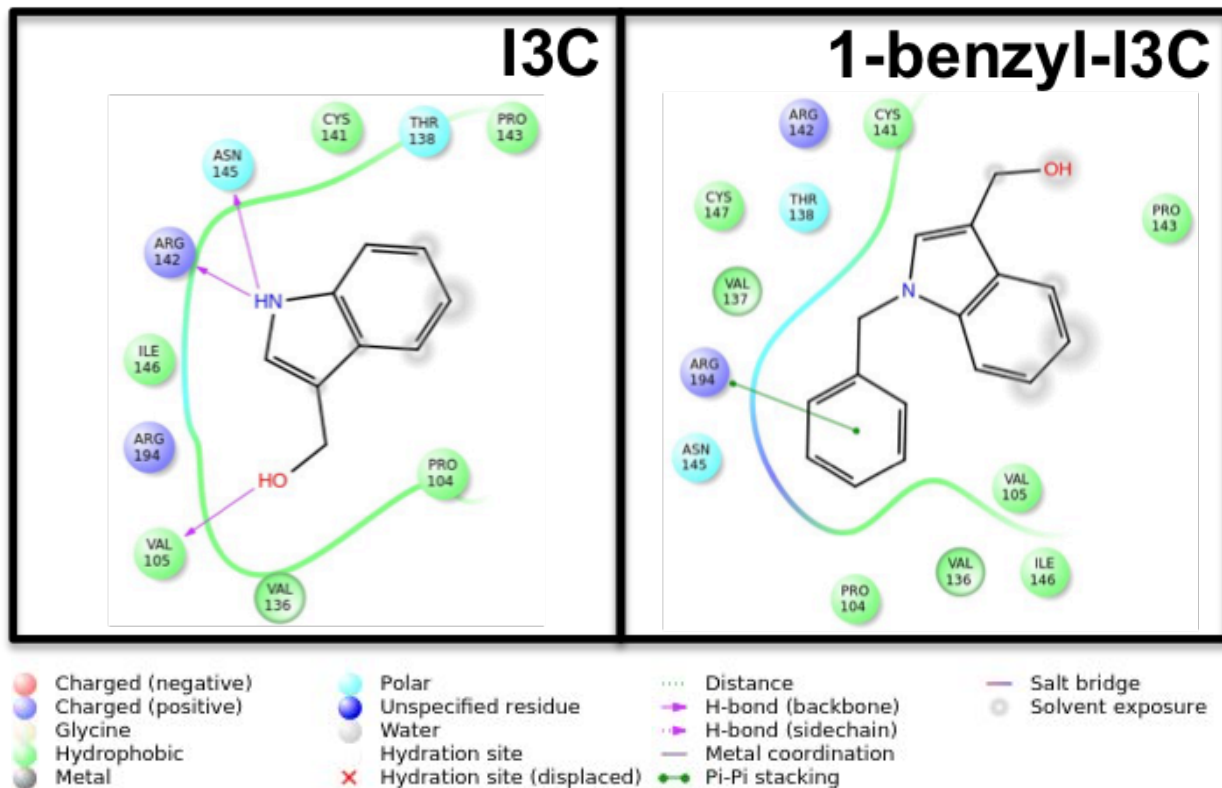
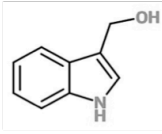
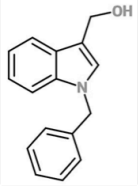
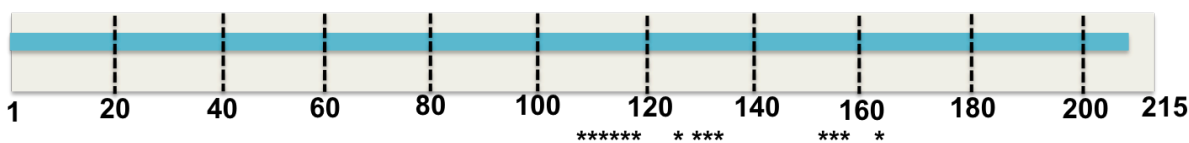


Figure 3.9B. Ligand interaction diagram of I3C and 1-benzyl-I3C in adducts with proteinase 3 homology model. Predicted molecular docking interactions diagram depicting close contact residues within 4 Å of either I3C or 1-benzyl-I3C.

Indolecarbinol Compound	Structure	Predicted Interacting Amino Acid Residue	Amino Acid Biological Property
I3C		Pro 104 Val 105 Pro 106 Val 137 Cys 141 Pro 143 Ile 146 Cys 147	Hydrophobic
		Thr 138 Asn 145	Polar
		Arg 142 Arg 194	Charged (positive)
1-benzyl-I3C		Pro 104 Val 105 Val 136 Val 137 Cys 141 Pro 143 Ile 146 Cys 147	Hydrophobic
		Thr 138 Asn 145	Polar
		Arg 142 Arg 194	Charged (positive)

Peptidase S1A, chymotrypsin family 1-207



I3C Predicted Contact Points: Ala 113, Met 114, Gly 115, Gly 117, Arg 118, Val 119, Ala 126, Gln 130, Glu 131, Leu 132, Gly 156, Ile 157, Cys 158, Asp 161 (denoted by *)

Figure 3.9C. Table and diagram of predicted molecular docking interactions of I3C and 1-benzyl-I3C with proteinase 3. Table summarizes I3C and 1-benzyl-I3C interactions with their close contact residues and each residue's corresponding biological property. All interactions described are within 4 Å of each indolecarbinol analog. Figure below shows I3C predicted contact points within each subdomain of the homology model.

Azurocidin (Uniprot ID: P20160) is a neutrophil granule-derived antibacterial and monocyte and fibroblast-specific chemotactic glycoprotein and functions in killing microorganisms and regulating multifunctional inflammation. It exhibits a 45.7% sequence similarity to human neutrophil elastase (Figure 3.10A). The highest ranked binding site showed I3C embedded in a region in the C-terminal, composed of predominantly hydrophobic residues such as Val 143, Val 174, Pro 176, Cys 180, Pro 182, Val 185, and Cys 186, polar residues Thr 142, Thr 175, and Asn 184, negatively charged residue Glu 144, and positively charged residue 181. They hydroxyl moiety is able to associate with Thr 175 and Cys 186 through hydrogen bond interactions (Figure 3.10B & C).

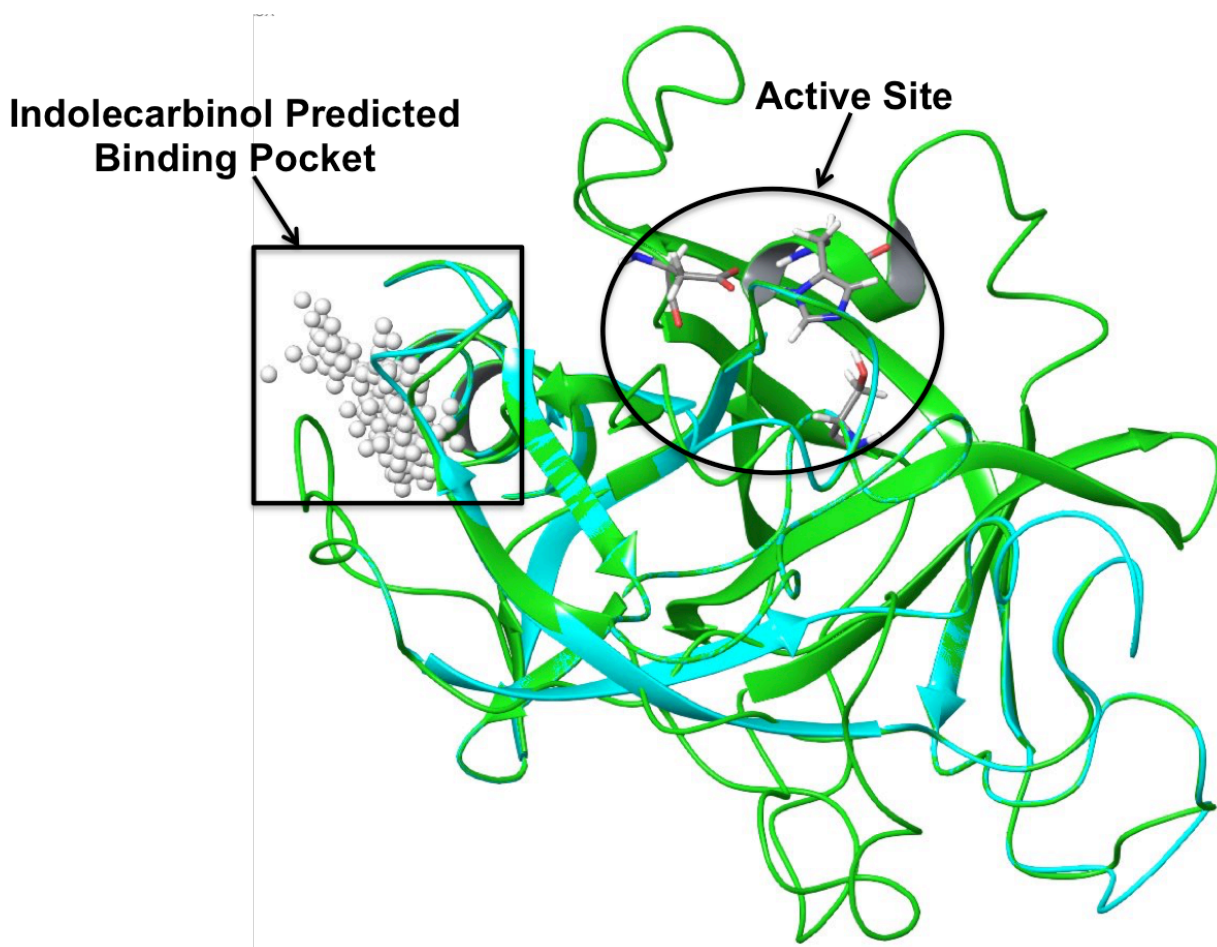


Figure 3.10A. Homology model of azurocidin (Uniprot ID: P20160). Ribbon diagram of homology model shown in teal, superimposed onto human neutrophil elastase crystal structure (PDB accession code: 5A8X) shown in green. Box encircles SiteMap's best predicted indolecarbinol binding site.

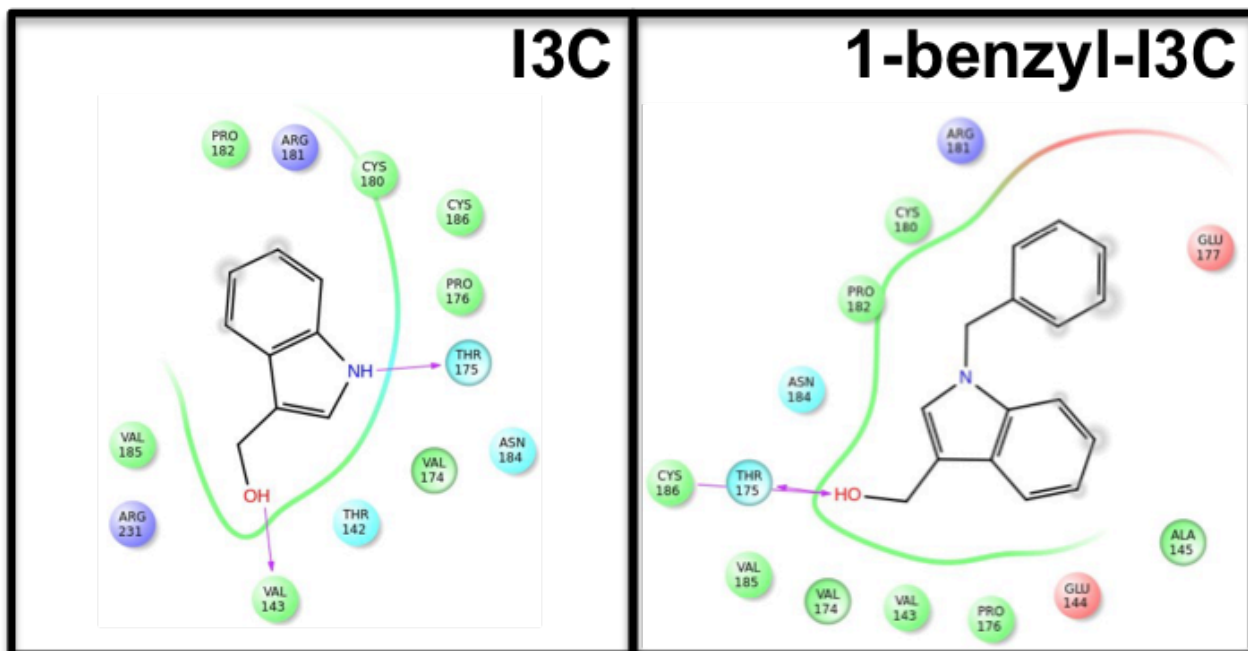
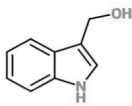
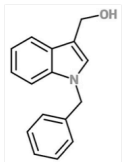
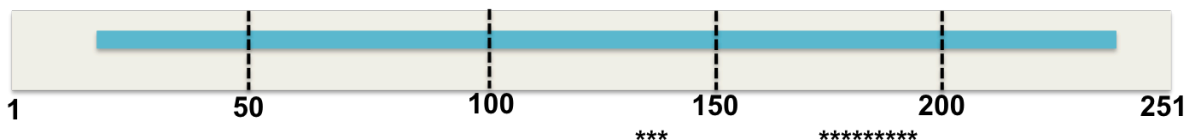


Figure 3.10B. Ligand interaction diagram of I3C and 1-benzyl-I3C in adducts with azurocidin homology model. Predicted molecular docking interactions diagram depicting close contact residues within 4 Å of either I3C or 1-benzyl-I3C.

Indolecarbinol Compound	Structure	Predicted Interacting Amino Acid Residue	Amino Acid Biological Property
I3C		Val 143 Val 174 Pro 176 Cys 180 Pro 182 Val 185 Cys 186	Hydrophobic
		Thr 142 Thr 175 Asn 184	Polar
		Arg 181	Charged (positive)
		Glu 144	Charged (negative)
1-benzyl-I3C		Val 143 Val 174 Pro 176 Cys 180 Pro 182 Val 185 Cys 186	Hydrophobic
		Thr 142 Thr 175 Asn 184	Polar
		Arg 181 Arg 231	Charged (positive)
		Glu 177	Charged (negative)

Peptidase S1A, chymotrypsin family 17-245



I3C Predicted Contact Points: Thr 142, Val 143, Glu 144, Val 174, Thr 175, Pro 176, Cys 180, Arg 181, Pro 182, Asn 184, Val 185, Cys 186 (denoted by *)

Figure 3.10C. Table and diagram of predicted molecular docking interactions of I3C and 1-benzyl-I3C with azurocidin. Table summarizes I3C and 1-benzyl-I3C interactions with their close contact residues and each residue's corresponding biological property. All interactions described are within 4 Å of each indolecarbinol analog. Figure below shows I3C predicted contact points within each subdomain of the homology model.

The PRSSL1 protein homology model (Uniprot ID: B7ZMF6) which exhibits serine-type endopeptidase activity and shares 35.7% sequence identity to human neutrophil elastase shows a predicted binding site that is located in a different region that was observed in human neutrophil elastase, proteinase 3, and azurocidin (Figure 3.11A). The indole ring is fastened in between hydrophobic residues Leu 92, Leu 170, Pro 171, and Pro 172 and polar residues Gln 55, His 56, and Ser 93 and is also able to engage in hydrogen bond interactions with Gly 54 through its hydroxyl moiety (Figure 3.11B & C).

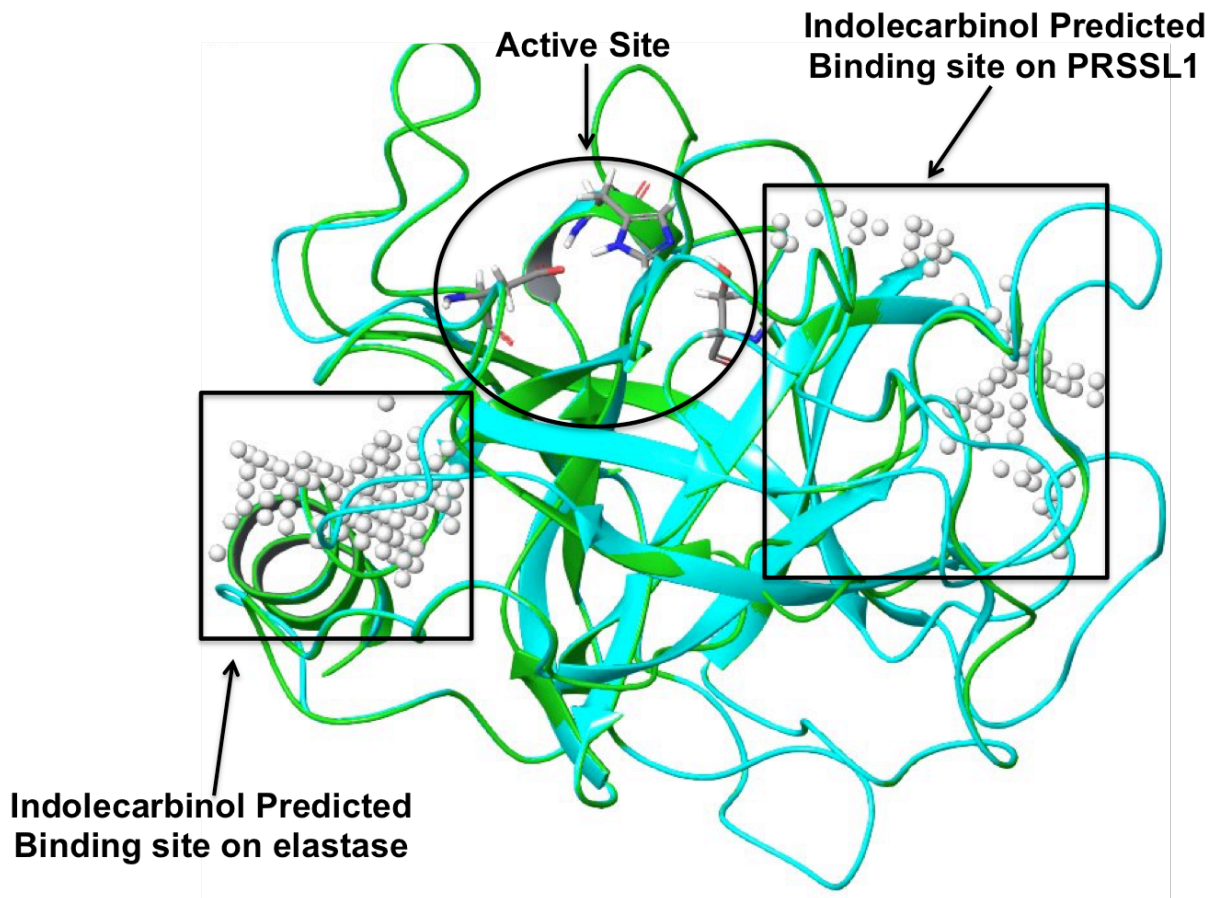


Figure 3.11A. Homology model of PRSSL1 protein (Uniprot ID: B7ZMF6). Ribbon diagram of homology model shown in teal, superimposed onto human neutrophil elastase crystal structure (PDB accession code: 5A8X) shown in green. Boxes encircle SiteMap's best predicted indolecarbinol binding site and predicted indolecarbinol binding site on elastase.

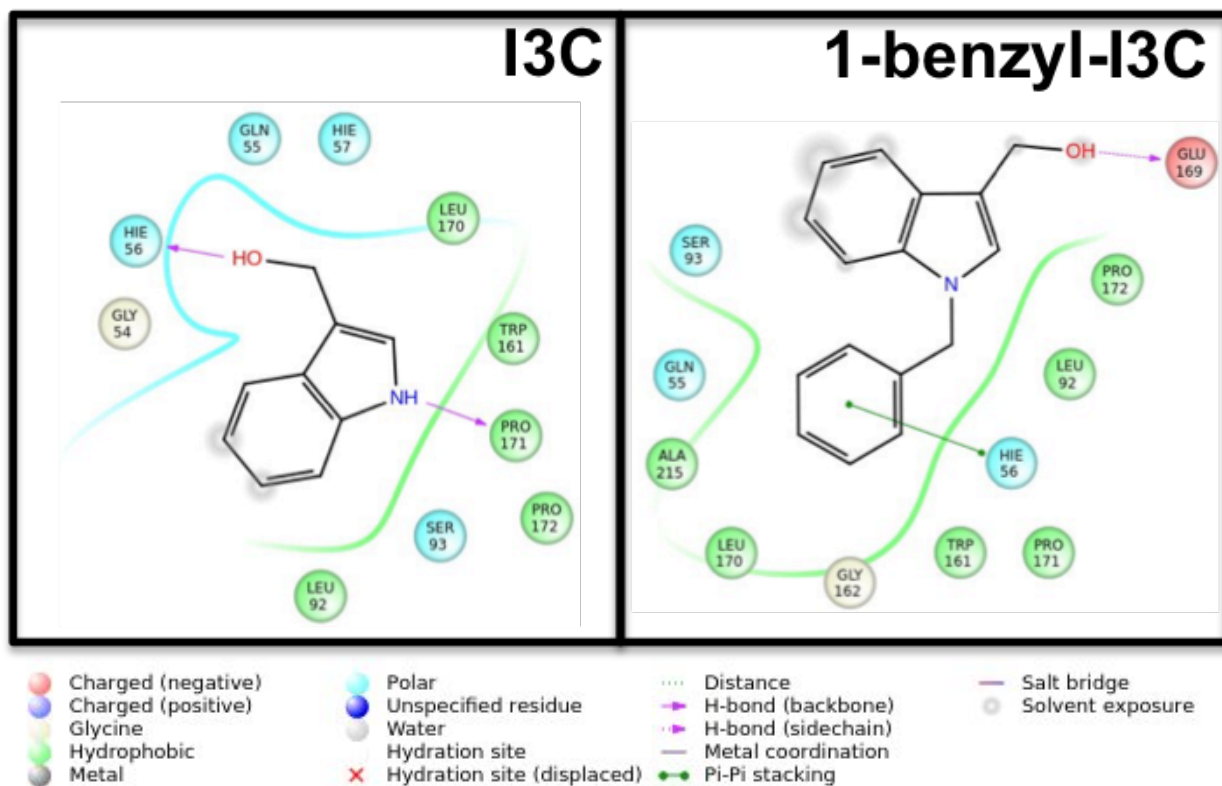
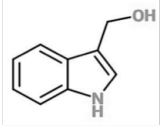
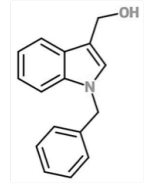
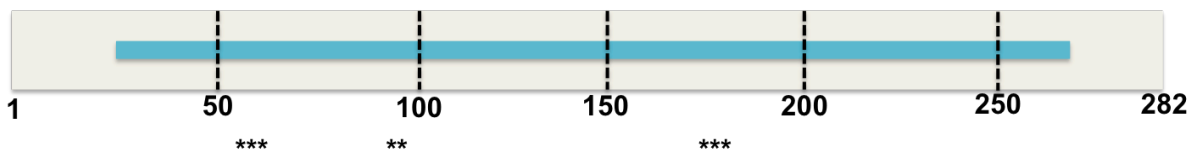


Figure 3.11B. Ligand interaction diagram of I3C and 1-benzyl-I3C in adducts with PRSSL1 homology model. Predicted molecular docking interactions diagram depicting close contact residues within 4 Å of either I3C or 1-benzyl-I3C.

Indolecarbinol Compound	Structure	Predicted Interacting Amino Acid Residue	Amino Acid Biological Property
I3C		Leu 92 Leu 170 Pro 171 Pro 172	Hydrophobic
		Gln 55 His 56 Ser 93	Polar
		Gly 54	Unique
1-benzyl-I3C		Leu 92 Trp 161 Leu 170 Pro 171 Pro 172	Hydrophobic
		Gln 55 His 56 Ser 93	Polar
		Gln 169	Charged (negative)
		Gly 54	Unique

Peptidase S1A, chymotrypsin family 27-262



I3C Predicted Contact Points: Gly 54, Gln 55, His 56, Leu 92, Ser 93, Leu 170, Pro 171, Pro 172 (denoted by *)

Figure 3.11C. Table and diagram of predicted molecular docking interactions of I3C and 1-benzyl-I3C with PRSSL1 protein. Table summarizes I3C and 1-benzyl-I3C interactions with their close contact residues and each residue's corresponding biological property. All interactions described are within 4 Å of each indolecarbinol analog. Figure below shows I3C predicted contact points within each subdomain of the homology model.

When the crystal structure of human neutrophil elastase (PDB accession code: 5A8X), and the homology models for proteinase 3 and azurocidin are superimposed onto each other, interestingly, it can be observed that the predicted indolecarbinol binding site that is incidentally present outside the active site of elastase aligns with best-ranked predicted indolecarbinol binding sites on proteinase 3 and azurocidin. In both these sites present that are present on their corresponding homology model, the pocket is predominantly composed of hydrophobic residues like Val, Pro, and Cys. The PRSSL1 protein which exhibits a lower sequence similarity to elastase (35.7%) and consequently lacks the necessary residues that is integral for securely anchoring the indolecarbinol compounds. While the best-ranked predicted binding site lies outside the PRSSL1 protein binding site, it is positioned on the opposite realm of the predicted binding sites for elastase, proteinase 3, and azurocidin, and is composed of hydrophobic residues like Leu and Pro and polar residues like His, Gln, and Ser (Figure 3.12).

```

elastase      1  -----IVGRRARPHAWPFMVSLQL---RGGH
proteinase   1  -----MASLQMRGNPGSH
azurocidin   1  MTRLTV-----LALLAGLASSRAGSSPLLDIVGGRKARPROFPFLASTQN---QGRH
PRSSL1       1  MG-IGLRGWRPLLTVATALMLPVKPPGSWGAIIGGHEVTPHSRPYMASVRF---GGQH

elastase     25  FCGATLIAPNFVMSAAHCVANVNVRAVRVVLGAHNLSRRE-PTROVFVQRIFE-NGYDP
proteinase   14  FCGGTLIHPSFVLTAAHCLRDIQRILVNVVLGAHNVRTQE-PTQCHFSVAQVFL-NNYDA
azurocidin   51  FCGGALIHARFVMTAASCFOSONPGVSTVVLGAYDLRRRERQSRQTFSSSMSE-NGYDP
PRSSL1       57  HCGGFLLRARVWVSAAHCFSHRDLRTGLVVLGAHVLSAE-PTQQVFGIDAITTHPDYHP

elastase     83  VNLNDIVILQLNGSATINANVOVAQLPAQGR-RIGNGVQCLAMGWGLLGRNRGIASVLO
proteinase   72  ENKLNIDILLIQLSSPANLSASVATVOLPODQ-PVPHGTQCLAMGWGRVGAHPPAQVLO
azurocidin   110 QONLNDLMLLQLDREANLTSSVTILPLPLQNA-TVEAGTRCQVAGWGSQRSGGRLSRFPR
PRSSL1       116 MTHANDICLLRLNGSAVLGPAVGLLRLPGRRARREPTAGTRCRVAGWGFVSDFEELPPGLM

elastase     142 ELNVTVVT----SICRR----SNVCTLV-RGRQAGVCFGDSGSPLVCNGLIHGIASFVR
proteinase   131 ELNVTVVT----FFCRP----HNICTEV-PRRKAGICFGDSGGPLICDGLIQGIDSFVI
azurocidin   169 FVNVTVTPE----DQCRP----NNVCTGV-LTRRGICNGDGGTPLVCEGLAHGVASFSL
PRSSL1       176 EAKVRVLDPDVFNSWKGHLTLLMLCTRSGDSHRRGFCSDSGGPLVCRNRAHCLVVSFSG

elastase     192 GGCASGLYPDAFAPVAQFVNWIDSIIQ-----
proteinase   181 WGCATRLFPDFTRVALYVDWIRSTLRRVEAKGR-P-----
azurocidin   220 GPCCGR--GPDFFTRVALFRDWIDGVLNNP--GPGPA-----
PRSSL1       236 LWCGDPKTPDVYTVQVSAFVAWIWDVVRSSPQPGPLPGTTRPPGEAA

```

Figure 3.12. Sequence alignment of elastase and corresponding homologous proteins. * denotes predicted I3C amino acid contact points. Residues highlighted in black correspond to conserved sequences, regions in grey correspond to amino acid residues that are not conserved, but share the same biological property, while residues not highlighted show areas that are not conserved.

Molecular Docking of Oncogenic B-RAF V600E serine/threonine kinase and Wild-type B-RAF

Studies conducted by Kundu *et al.* [172] demonstrated that I3C directly inhibits oncogenic BRAF-V600E kinase activity with no effect on the wild-type BRAF protein. This ultimately leads to loss of down-stream BRAF-V600E signaling, reduced MITF-M gene expression, and enhanced sensitivity of oncogenic BRAF expressing melanoma cells to the anti-proliferative effects of I3C. Understanding I3C's selectivity towards oncogenic BRAF V600E may pave the way for individual treatment in personalized medicine. Structural comparisons of oncogenic B-RAF and wild-type BRAF and various homologues of oncogenic B-RAF were executed to detect visible patterns in indolecarbinol binding sites. Vemurafenib, an FDA approved B-Raf enzyme inhibitor, has been shown to reduce disease progression and death and is selective for the oncogenic B-RAF V600E kinase. Combination treatment of I3C and Vemurafenib on melanoma cells have proven to cooperatively reduce MITF-M expression and mitigate melanoma cell proliferation, thus suggesting the possibility of employing I3C-derived compounds coupled with Vemurafenib for the evolution of combinational therapeutic approaches. Molecular docking analysis was initially performed on the oncogenic BRAF V600E kinase and the wild-type BRAF as a baseline for comparison.

Since I3C treatment in melanoma cells will be aimed more for combinational treatment with Vemurafenib, docking of I3C, 1-benzyl-I3C, and their corresponding analogs was performed on the crystal structure of BRAF V600E kinase in adducts with the kinase inhibitor, Vemurafenib (PDB accession code: 3OG7) (Figure 3.13A). Docking studies revealed that I3C is predicted to bind to the opposing monomer not bound to Vemurafenib and is embedded in a pocket away from the V600E mutation. I3C is anchored in a pocket comprised of hydrophobic residues Ile 463, Val 471, Ala 481, Leu 514, Trp 531, Cys 532, and Phe 583, polar residues Thr 529, Gln 530, positively charged residue 483, and negatively charged residue Asp 594. The hydroxyl moiety on the indole ring associates with the backbone of Cys 532 through hydrogen bonding interactions (Figure 3.13B & C).

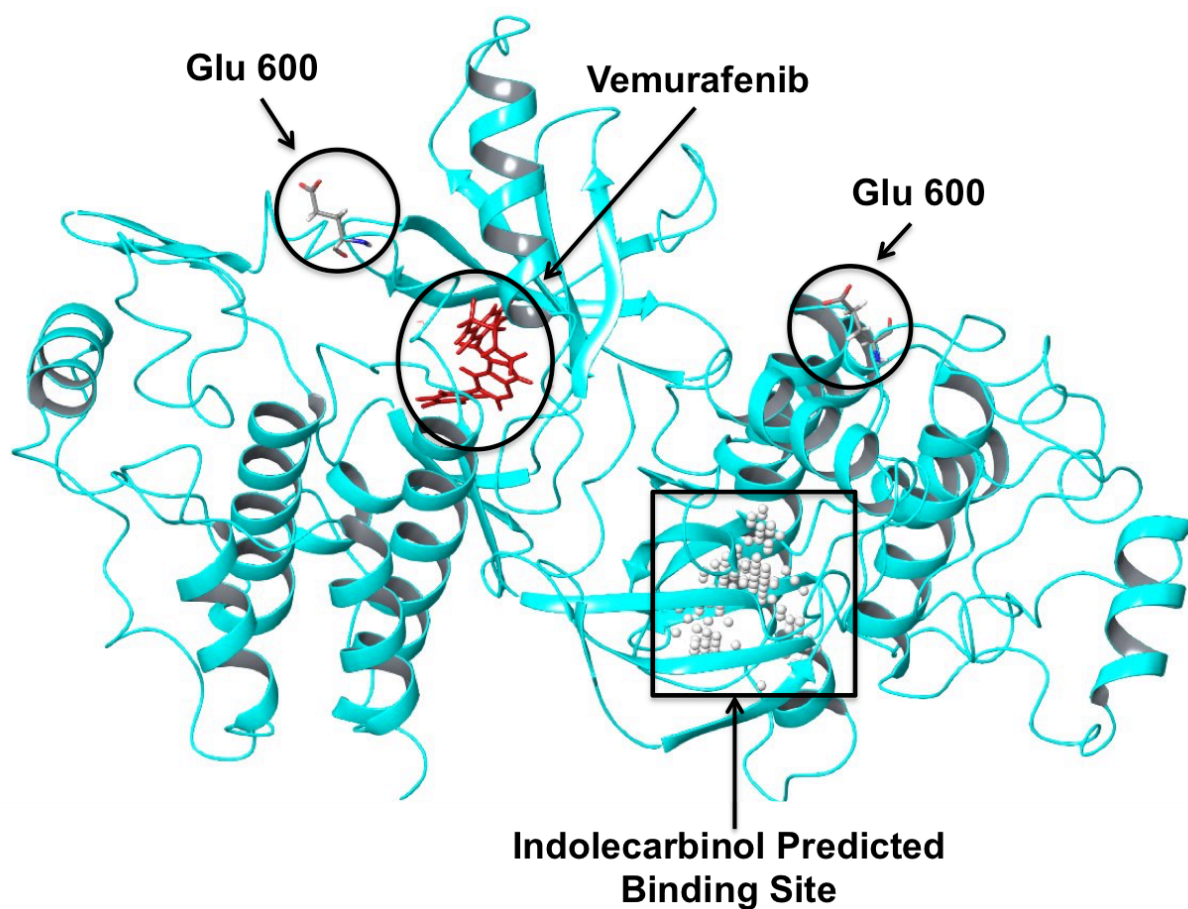


Figure 3.13A. Molecular docking model of oncogenic B-RAF V600E serine/threonine kinase (PDB accession code: 3OG7) highlighting indolecarbinol predicted binding site. Ribbon diagram of oncogenic B-RAF *in silico* model. Oncogenic B-Raf inhibitor, Vemurafenib, is depicted in red. Mutated glutamate 600 residue is shown as sticks on each protomer. Box encircles SiteMap's best predicted indolecarbinol binding site.

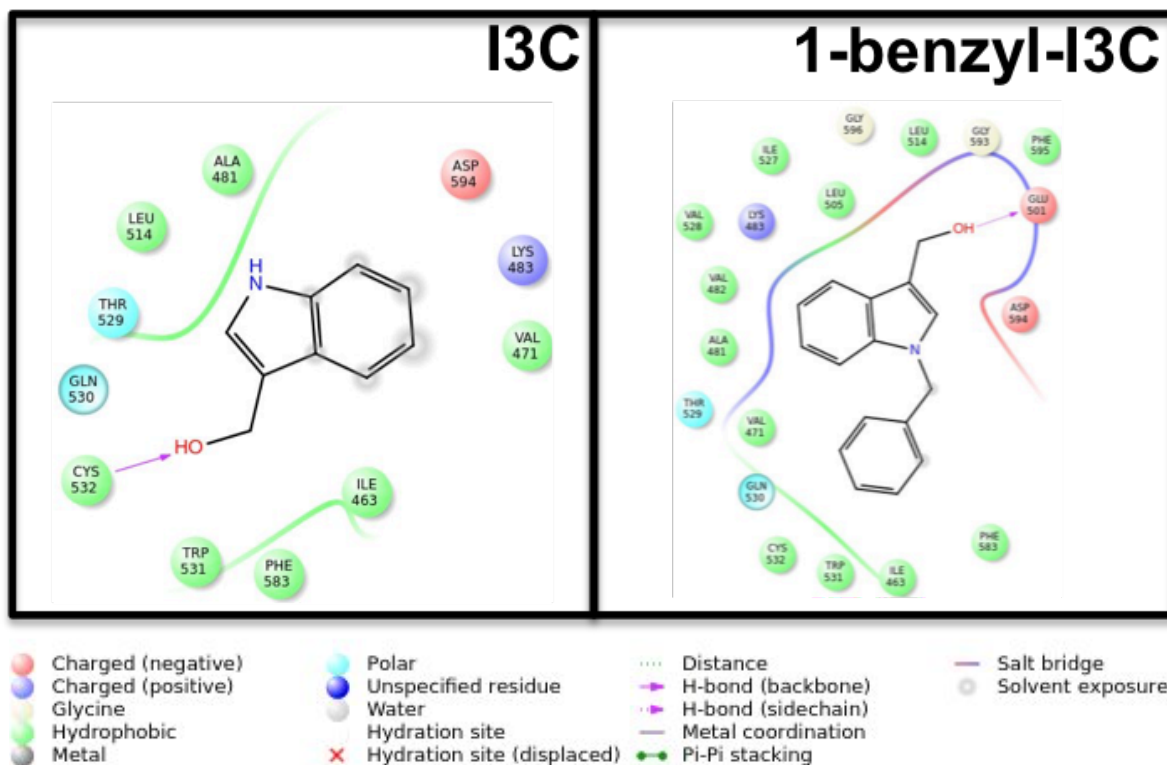
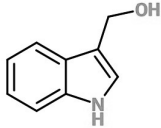
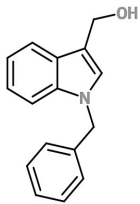
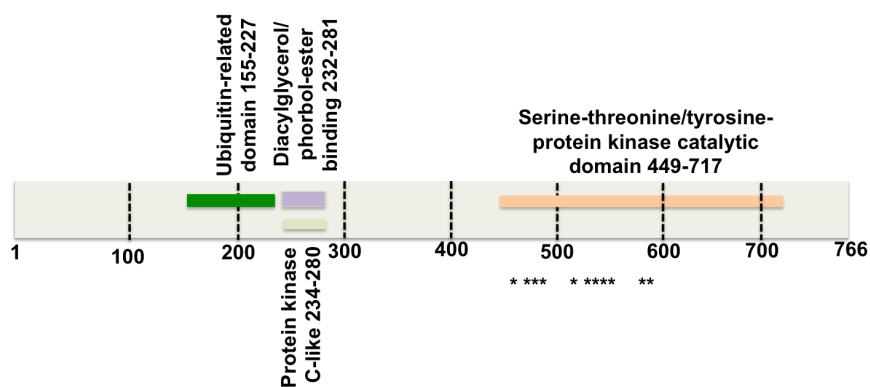


Figure 3.13B. Ligand interaction diagram of I3C and 1-benzyl-I3C in adducts with oncogenic B-RAF V600E *in silico* model. Predicted molecular docking interactions diagram depicting close contact residues within 4 Å of either I3C or 1-benzyl-I3C.

Indolecarbinol Compound	Structure	Predicted Interacting Amino Acid Residue	Amino Acid Biological Property
I3C		Ile 463 Val 471 Ala 481 Leu 514 Trp 531 Cys 532 Phe 583	Hydrophobic
		Thr 529 Gln 530	Polar
		Asp 594	Charged (negative)
		Lys 483	Charged (positive)
1-benzyl-I3C		Phe595 Leu 514 Leu 505 Ile 527 Val 528 Val 482 Ala 481 Val 471 Cys 532 Trp 31 Ile 463 Phe 583	Hydrophobic
		Thr 529 Gln 530	Polar
		Glu 501 Asp 594	Charged (negative)
		Lys 483	Charged (positive)



I3C Predicted Contact Points: Ile 463, Val 471, Ala 481, Lys 483, Leu 514, Thr 529, Gln 530, Trp 531, Cys 532, Phe 583, Asp 594 (denoted by *)

Figure 3.13C. Table and diagram of predicted molecular docking interactions of I3C and 1-benzyl-I3C with oncogenic B-RAF. Table summarizes I3C and 1-benzyl-I3C interactions with their close contact residues and each residue's corresponding biological property. All interactions described are within 4 Å of each indolecarbinol analog. Figure below shows I3C predicted contact points within each subdomain of the *in silico* model.

After using SiteMap to identify potential I3C binding regions on wild-type BRAF kinase (Figure 3.14A), it was observed that the indole ring on I3C was oriented towards a pocket composed of hydrophobic residues Ile 462, Val 470, Ala 480, Leu 513, Trp 530, Cys 531, Phe 594, polar residues Thr 528, Gln 529, and positively charged residue Lys 482. The hydroxyl moiety engages in hydrogen bonding interactions with the sidechain of Thr 528 (Figure 3.14B & C). While the I3C predicted binding pocket shares similar contact residues, the location the I3C binding site is different from the site observed on the BRAF V600E model. When the wild-type BRAF model is superimposed onto the BRAF V600E model, it is evident that I3C is predicted to bind to the same binding region as Vemurafenib.

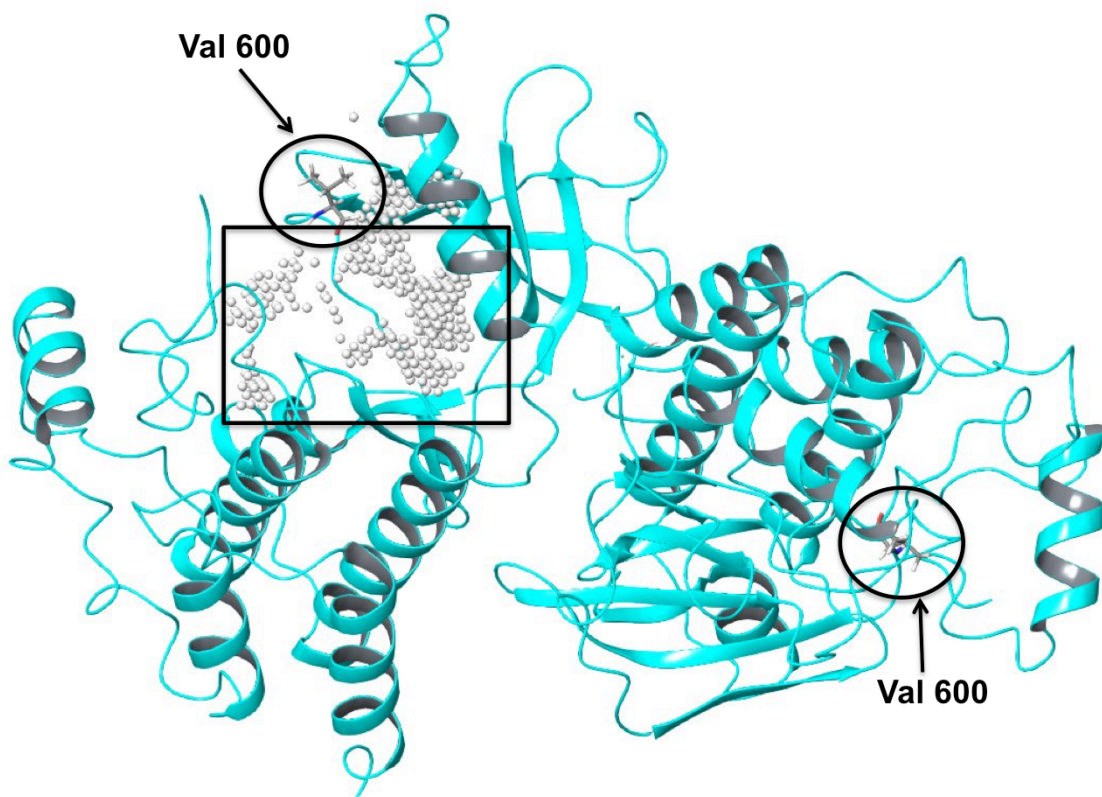


Figure 3.14A. Molecular docking model of wild-type B-RAF serine/threonine kinase (PDB accession code: 1UWH) highlighting indolecarbinol predicted binding site. Ribbon diagram of wild-type B-RAF *in silico* model. Valine 600 residue is shown as sticks on each protomer. Box encircles SiteMap's best predicted indolecarbinol binding site.

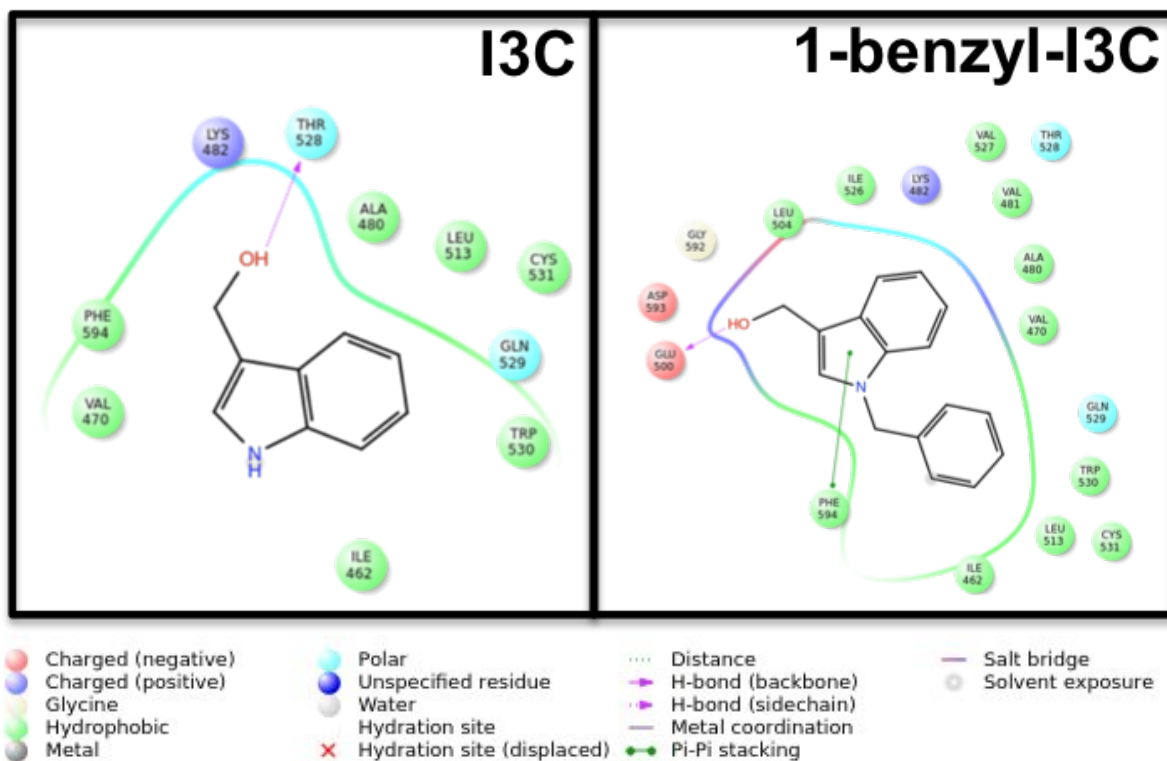
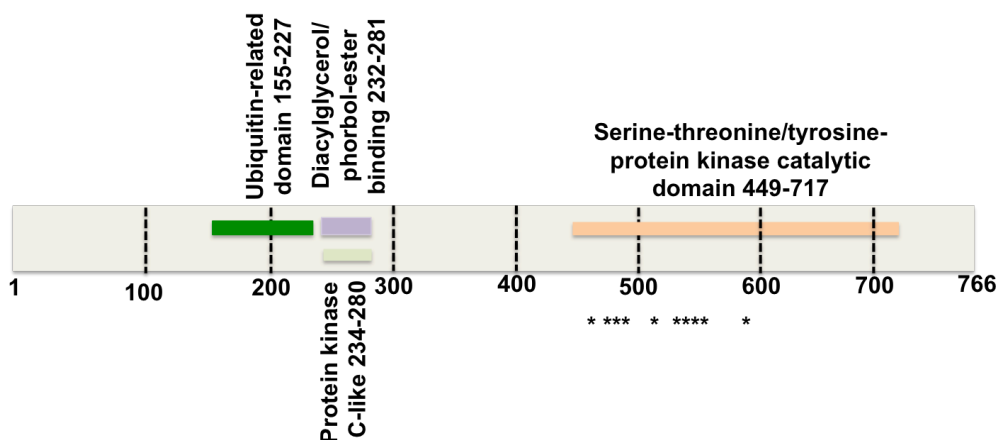


Figure 3.14B. Ligand interaction diagram of I3C and 1-benzyl-I3C in adducts with wild-type B-RAF *in silico* model. Predicted molecular docking interactions diagram depicting close contact residues within 4 Å of either I3C or 1-benzyl-I3C.

Indolecarbinol Compound	Structure	Predicted Interacting Amino Acid Residue	Amino Acid Biological Property
I3C		Ile 462 Val 470 Phe 594 Ala 480 Leu 513 Cys 531 Trp 530	Hydrophobic
		Thr 528 Gln 529	Polar
		Lys 482	Charged (positive)
1-benzyl-I3C		Phe 594 Leu 504 Ile 526 Val 527 Val 481 Ala 480 Val 470 Trp 530 Cys 531 Leu 513 Ile 462	Hydrophobic
		Thr 528 Gln 529	Polar
		Glu 500 Asp 593	Charged (negative)
		Lys 482	Charged (positive)



I3C Predicted Contact Points: Ile 462, Val 470, Ala 480, Lys 482, Leu 513, Thr 528, Gln 529, Trp 530, Cys 531, Phe 594 (denoted by *)

Figure 3.14C. Table and diagram of predicted molecular docking interactions of I3C and 1-benzyl-I3C with wild-type B-RAF. Table summarizes I3C and 1-benzyl-I3C interactions with their close contact residues and each residue's corresponding biological property. All interactions described are within 4 Å of each indolecarbinol analog. Figure below shows I3C predicted contact points within each subdomain of the *in silico* model.

Oncogenic B-RAF V600E homology modeling

Homology models for various homologues of oncogenic BRAF V600E were generated to ascertain whether or not patterns for indolecarbinol binding sites could be visibly discerned. Four homologues V-raf1 murine leukemia viral oncogene-like protein isoform 3 (UniProt ID: L7RRS6), RAF proto-oncogene serine/threonine-protein kinase (UniProt: H7C155), highly similar to RAF proto-oncogene serine/threonine-protein kinase (cDNA FLJ57286) (UniProt ID: B4E0X2), and kinase suppressor of Ras 2 (UniProt ID: E9PB13) from the BLASTp search were selected and ranged in sequence similarity to oncogenic B-RAF from 35.9%-75.7% (Figure 3.15).

Uniprot ID	Protein Name	Sequence Identity to V600E Oncogenic B-Raf Kinase, Chain A (PDB Accession code: 3OG7) (%)
L7RRS6	V-raf-1 murine leukemia viral oncogene-like protein 1 isoform 3	75.7
H7C155	RAF proto-oncogene serine/threonine-protein kinase	75.7
B4E0X2	cDNA FLJ57286, highly similar to RAF proto-oncogene serine/threonine-protein kinase	75.5
E9PB13	Kinase suppressor of Ras 2	35.9

Figure 3.15. Oncogenic B-RAF V600E homologous proteins. Table depicts % sequence identity of each corresponding homologous protein to oncogenic B-RAF. Proteins are identified by their UniProt ID.

V-Raf-1 murine leukemia viral oncogene (Uniprot ID: L7RRS6) exhibits a 75.7% sequence similarity to oncogenic BRAF kinase (Figure 3.16A). Here, I3C is securely fastened by hydrophobic residues Ala 373, Ile 355, Leu 406, Trp 423, Cys 424, and Phe 475, polar residues Thr 421 and Gln 422, negatively charged residue Asp 486, and positively charged residue Lys 375 (Figure 3.16B & C).

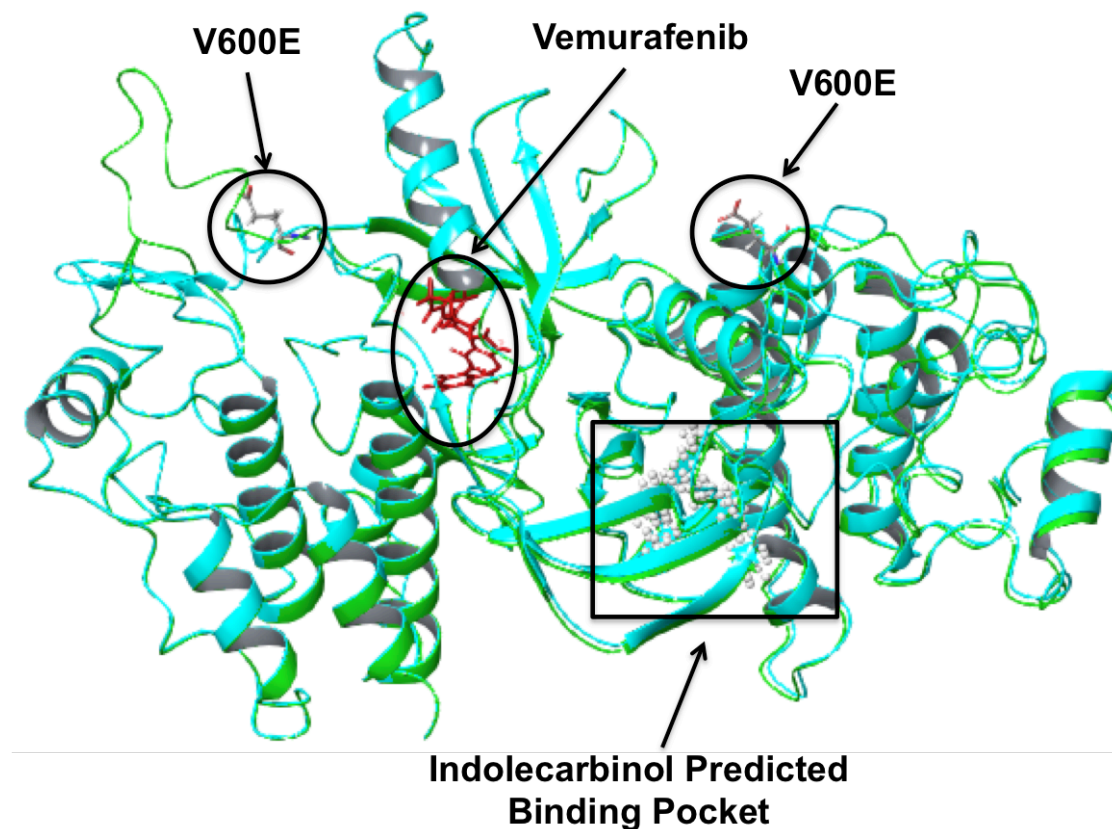


Figure 3.16A. Homology model V-Raf-1 murine leukemia viral oncogene (Uniprot ID: L7RRS6). Ribbon diagram of homology model shown in teal, superimposed onto oncogenic B-RAF crystal structure (PDB accession code: 3OG7) shown in green. Oncogenic B-Raf inhibitor, Vemurafenib, is depicted in red. Mutated glutamate 600 residue is shown as sticks on each protomer. Box encircle SiteMap's best predicted indolecarbinol binding site on the homology model.

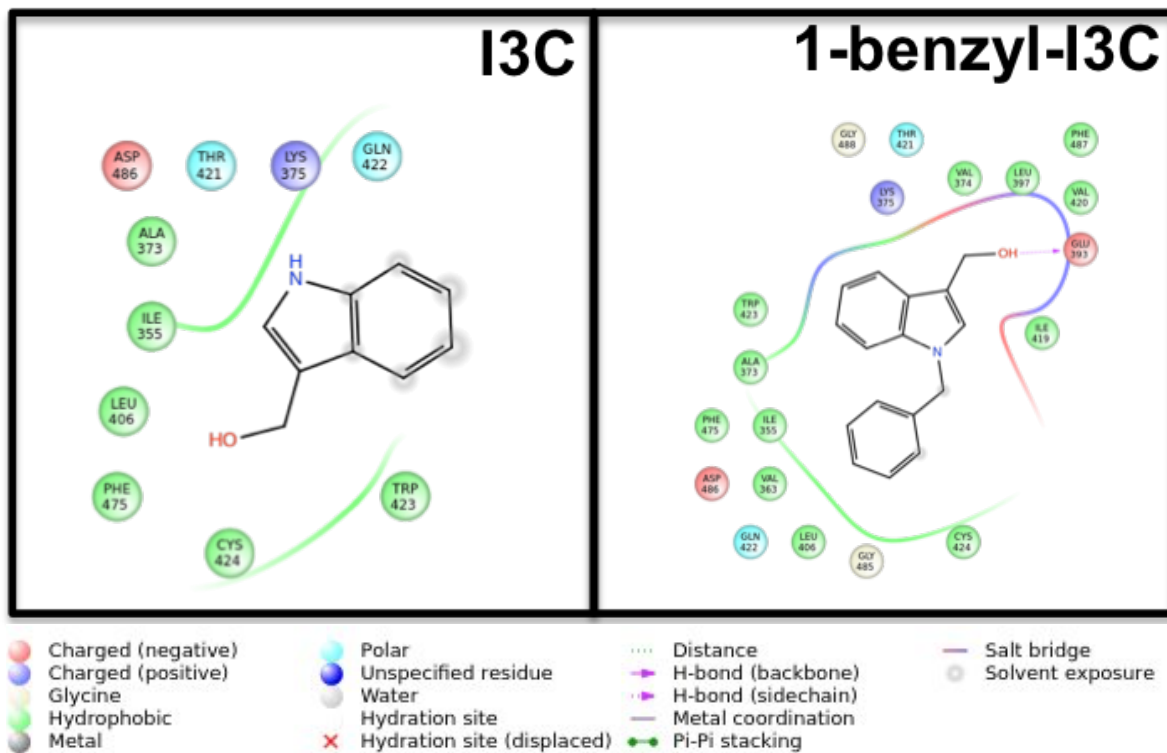
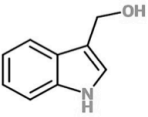
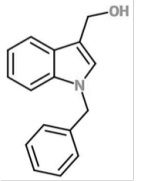
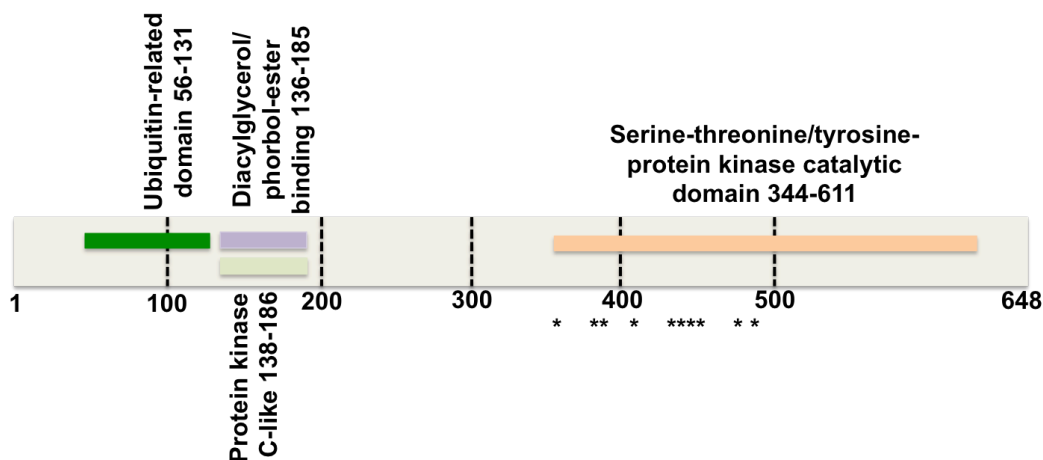


Figure 3.16B. Ligand interaction diagram of I3C and 1-benzyl-I3C in adducts with V-raf-1 murine leukemia viral oncogene-like protein 1 homology model. Predicted molecular docking interactions diagram depicting close contact residues within 4 Å of either I3C or 1-benzyl-I3C.

Indolecarbinol Compound	Structure	Predicted Interacting Amino Acid Residue	Amino Acid Biological Property
I3C		Ala 373 Ile 355 Leu 406 Phe 475 Cys 424 Trp 423	Hydrophobic
		Thr 421 Gln 422	Polar
		Asp 486	Charged (negative)
		Lys 375	Charged (positive)
1-benzyl-I3C		Cys 424 Leu 406 Val 363 Ile 355 Phe 475 Ala 373 Trp 423 Val 374 Leu 397 Phe 487 Val 420 Ile 419	Hydrophobic
		Thr 421 Gln 422	Polar
		Glu 393 Asp 486	Charged (negative)
		Lys 375	Charged (positive)



I3C Predicted Contact Points: Ile 355, Ala 373, Lys 375, Leu 406, Thr 421, Gln 422, Trp 423, Cys 424, Phe 475, Asp 486 (denoted by *)

Figure 3.16C. Table and diagram of predicted molecular docking interactions of I3C and 1-benzyl-I3C V-raf-1 murine leukemia viral oncogene-like protein 1 homology model. Table summarizes I3C and 1-benzyl-I3C interactions with their close contact residues and each residue's corresponding biological property. All interactions described are within 4 Å of each indolecarbinol analog. Figure below shows I3C predicted contact points within each subdomain of the homology model.

In the Raf proto-oncogene serine/threonine protein kinase (Uniprot ID: H7C155) which exhibits 75.7% sequence identity to oncogenic BRAF, I3C was predicted to bind a region similar to the V-Raf-1 murine leukemia viral oncogene homology model (Figure 3.17A). Here, I3C is adjacent to hydrophobic residues Ala 373, Ile 355, Leu 406, Trp 423, Cys 424, Phe 475, polar residues Thr 421 and Gln 422, negatively charged residue Asp 486, and positively charged residue 375 (Figure 3.17B & C).

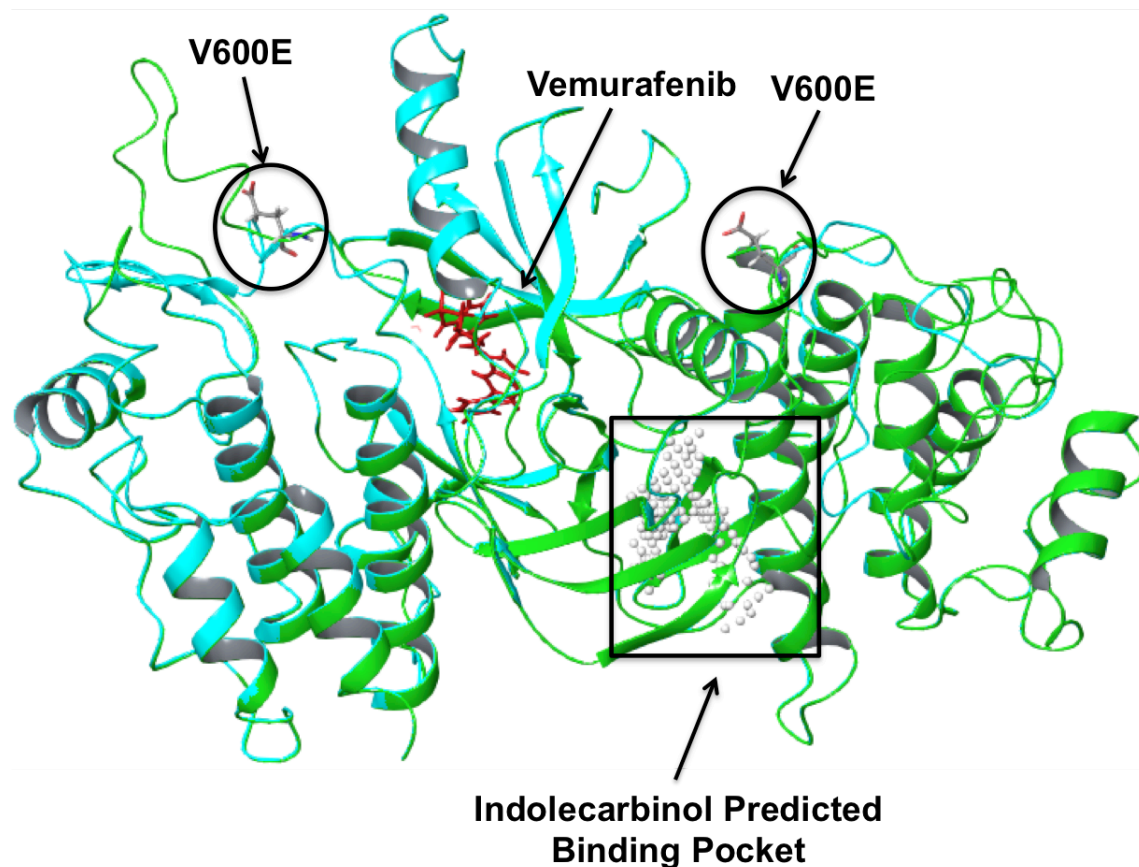


Figure 3.17A. Homology model Raf proto-oncogene serine/threonine protein kinase (Uniprot ID: H7C155). Ribbon diagram of homology model shown in teal, superimposed onto oncogenic B-RAF crystal structure (PDB accession code: 3OG7) shown in green. Oncogenic B-Raf inhibitor, Vemurafenib, is depicted in red. Mutated glutamate 600 residue is shown as sticks on each protomer. Box encircle SiteMap's best predicted indolecarbinol binding site on the homology model.

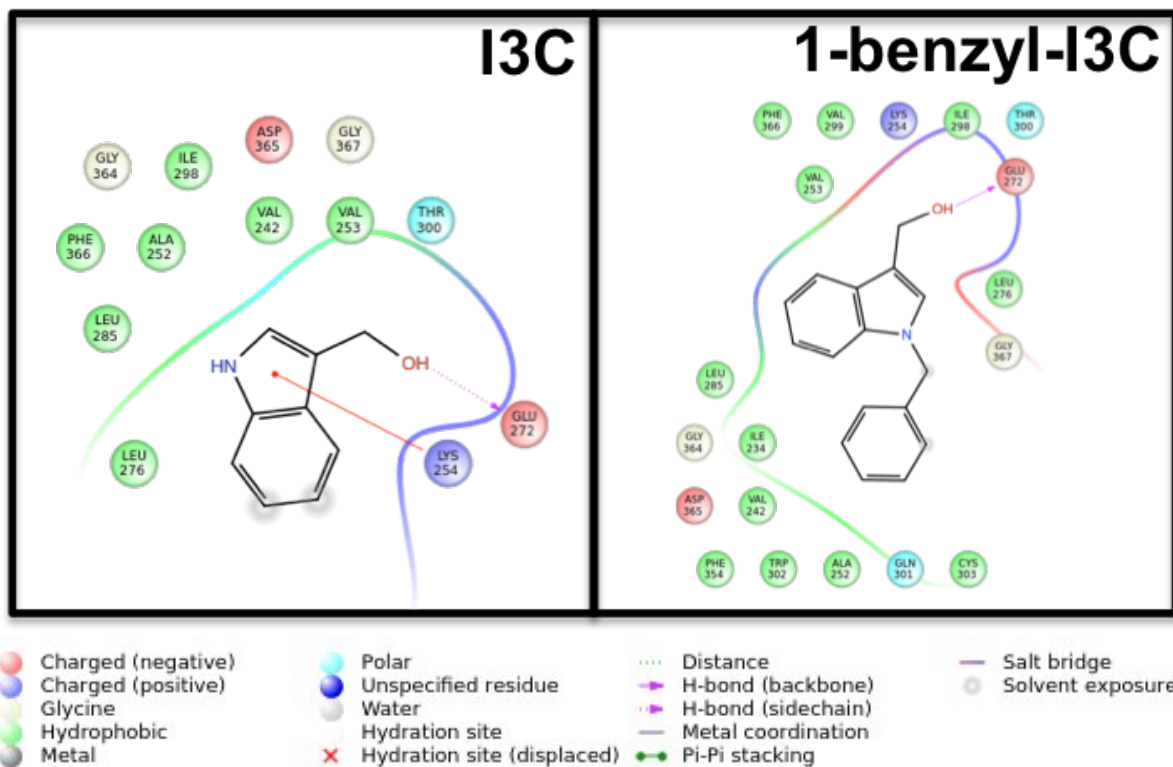
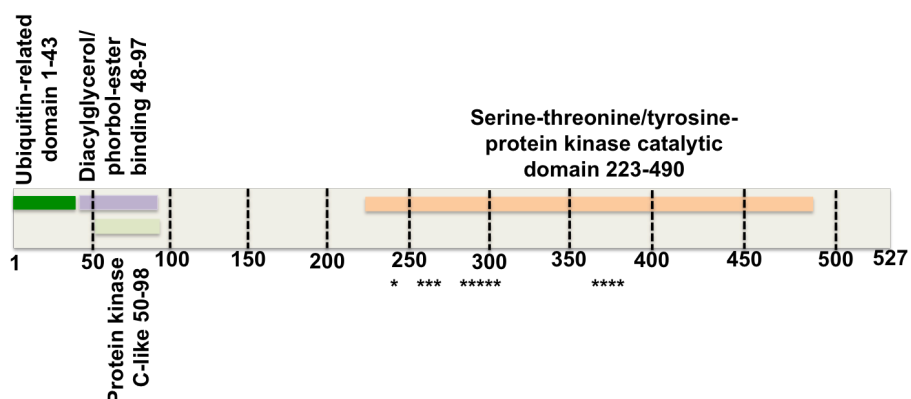


Figure 3.17B. Ligand interaction diagram of I3C and 1-benzyl-I3C in adducts with Raf proto-oncogene serine/threonine protein kinase (Uniprot ID: H7C155) homology model. Predicted molecular docking interactions diagram depicting close contact residues within 4 Å of either I3C or 1-benzyl-I3C.

Indolecarbinol Compound	Structure	Predicted Interacting Amino Acid Residue	Amino Acid Biological Property
I3C		Leu 276 Leu 285 Phe 366 Ala 252 Ile 298 Val 242 Val 253	Hydrophobic
		Thr 300	Polar
		Glu 272 Asp 365	Charged (negative)
		Lys 254	Charged (positive)
		Gly 364 Gly 367	Unique
1-benzyl-I3C		Leu 276 Cys 303 Ala 252 Trp 302 Phe 354 Val 242 Ile 234 Leu 285 Val 253 Phe 366 Val 299 Ile 298	Hydrophobic
		Thr 300 Gln 301	Polar
		Glu 272 Asp 365	Charged (negative)
		Lys 254	Charged (positive)
		Gly 367	Unique



I3C Predicted Contact Points: Val 242, Ala 252, Val 253, Lys 254, Glu 272, Leu 276, Leu 285, Ile 298, Thr 300, Gly 364, Asp 365, Phe 366, Gly 367 (denoted by *)

Figure 3.17C. Table and diagram of predicted molecular docking interactions of I3C and 1-benzyl-I3C with Raf proto-oncogene serine/threonine protein kinase (Uniprot ID: H7C155) homology model. Table summarizes I3C and 1-benzyl-I3C interactions with their close contact residues and each residue's corresponding biological property. All interactions described are within 4 Å of each indolecarbinol analog. Figure below shows I3C predicted contact points within each subdomain of the homology model.

For the highly similar to RAF proto-oncogene serine/threonine protein kinase (Uniprot ID: B4E0X2) which shares 75.5% sequence identity to BRAF V600E (Figure 3.18A), I3C was had its indole ring oriented towards a hydrophobic pocket comprised of Val 282, Ala 292, Val 293, Leu 316, Leu 325, Ile 338, and Phe 406, positively charged residue Lys 294, negatively charged residue Glu 312, Asp 405, and polar residue Thr 340. Here, the hydroxyl substituent on the indole ring is able to associate with the sidechain of Glu 312 through hydrogen bond interactions (Figure 3.18B & C).

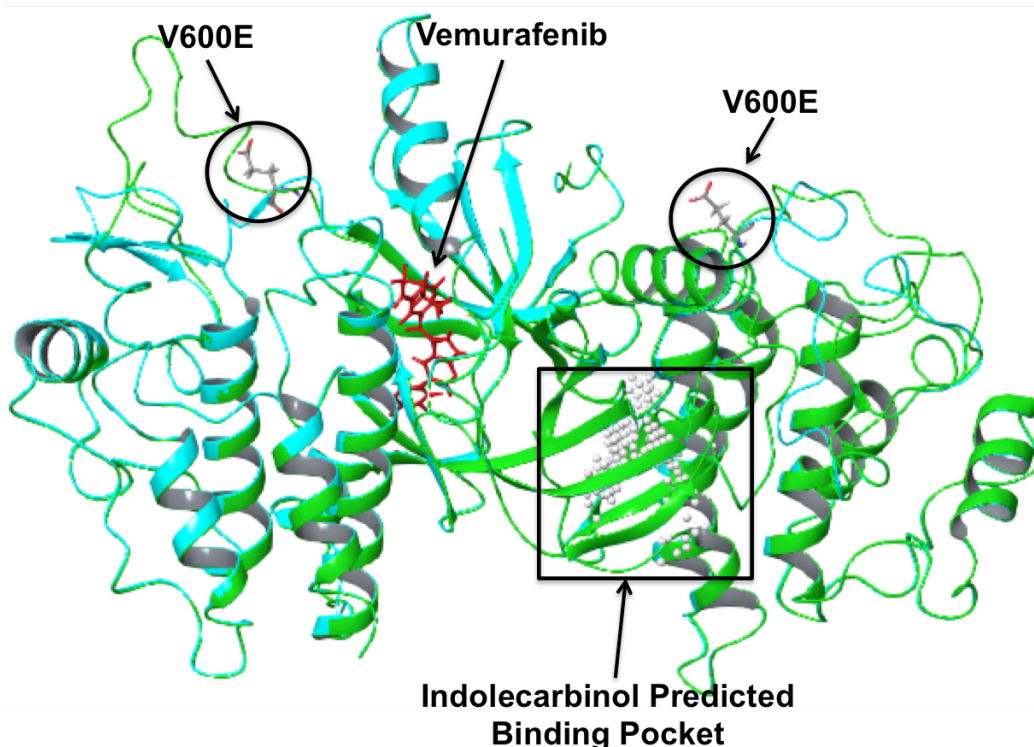


Figure 3.18A. Homology model of highly similar to RAF proto-oncogene serine/threonine protein kinase (Uniprot ID: B4E0X2). Ribbon diagram of homology model shown in teal, superimposed onto oncogenic B-Raf crystal structure (PDB accession code: 3OG7) shown in green. Oncogenic B-Raf inhibitor, Vemurafenib, is depicted in red. Mutated glutamate 600 residue is shown as sticks on each protomer. Box encircle SiteMap's best predicted indolecarbinol binding site on the homology model.

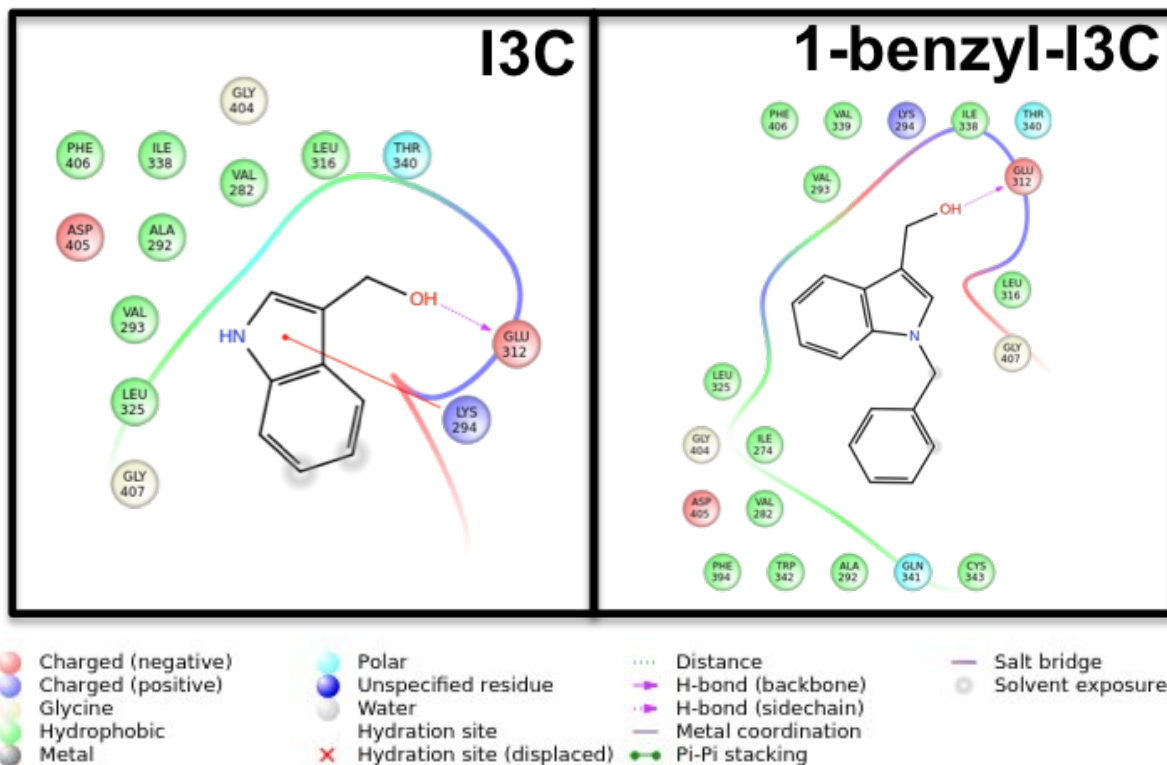
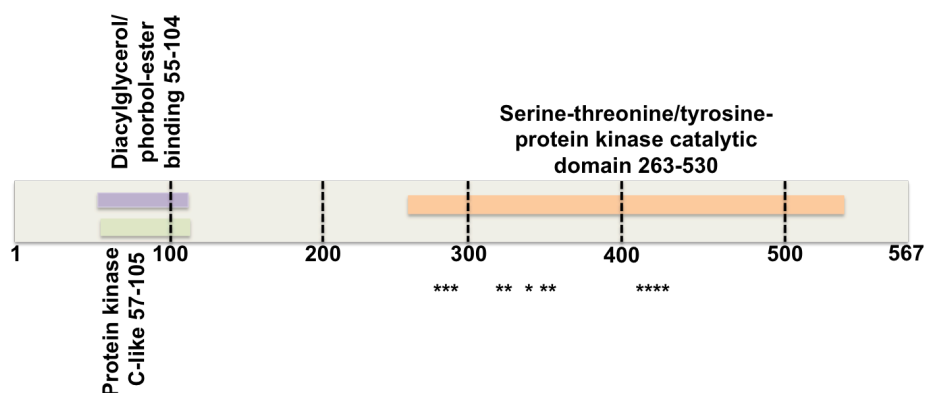


Figure 3.18B. Ligand interaction diagram of I3C and 1-benzyl-I3C in adducts with highly similar to RAF proto-oncogene serine/threonine protein kinase (Uniprot ID: B4E0X2) homology model. Predicted molecular docking interactions diagram depicting close contact residues within 4 Å of either I3C or 1-benzyl-I3C.

Indolecarbinol Compound	Structure	Predicted Interacting Amino Acid Residue	Amino Acid Biological Property
I3C		Leu 325 Val 293 Ala 292 Phe 406 Ile 338 Val 282 Leu 316	Hydrophobic
		Thr 340	Polar
		Glu 312 Asp 405	Charged (negative)
		Lys 294	Charged (positive)
		Gly 404 Gly 407	Unique
1-benzyl-I3C		Leu 316 Cys 343 Ala 292 Trp 342 Phe 394 Val 282 Le 274 Leu 325 Val 293 Phe 406 Val 339 Ile 338	Hydrophobic
		Thr 340	Polar
		Glu 312 Asp 405	Charged (negative)
		Lys 294	Charged (positive)
		Gly 404 Gly 407	Unique



I3C Predicted Contact Points: Val 282, Ala 292, Val 293, Lys 294, Glu 312, Leu 316, Leu 325, Ile 338, Thr 340, Gly 404, Asp 405, Phe 406, Gly 407, (denoted by *)

Figure 3.18C. Table and diagram of predicted molecular docking interactions of I3C and 1-benzyl-I3C with highly similar to RAF proto-oncogene serine/threonine protein kinase (Uniprot ID: B4E0X2) homology model. Table summarizes I3C and 1-benzyl-I3C interactions with their close contact residues and each residue's corresponding biological property. All interactions described are within 4 Å of each indolecarbinol analog. Figure below shows I3C predicted contact points within each subdomain of the homology model.

Underlining the key differences in the indolecarbinol binding sites between the Raf kinases of varying homology can illuminate our understanding of I3C target proteins and indolecarbinol binding modes. A homology model for the kinase suppressor of Ras 2 (Uniprot ID: E9PB13), which exhibits a 35.9% sequence similarity to oncogenic BRAF, was synthesized. The indole ring on I3C is positioned against hydrophobic residues Val 651, Ala 661, Val 694, Leu 712, Cys 713, Phe 764, Val 771, polar residues His 653, Thr 710, Ser 711, and Thr 773. The hydroxyl substituent is able to engage in hydrogen bond interactions with the side chain of Arg 663.

After examining the indolecarbinol binding sites between the homology models, it is evident there are patterns among the contact sites between oncogenic BRAF homologues of higher sequence similarity. From the three oncogenic BRAF homology models that exhibit greater than 75% sequence homology, they all appear to have I3C binding sites that are predominantly hydrophobic residues such as Leu, Val, Phe, polar residues Thr and Gln, positively charged residues like Lys, and negatively charged residues like Glu and Asp (Figure 3.19).


```

30G7_B 10 -----HGRDAA--DDWEIPDQITVGRIGSGSFGTV*
L7RRS6 329 -----KIRPRGORDSS--YYWEIEASEVMLSTRIGSGSFGTV
H7C155 208 -----KIRPRGORDSS--YYWEIEASEVMLSTRIGSGSFGTV
B4E0X2 248 -----KIRPRGORDSS--YYWEIEASEVMLSTRIGSGSFGTV
E9PB13 592 VHDEAESEDDFEEMNLSLLSARSEPRKASQTSIFLQEWIDIPFEQLEIGELIGKGRFGQV

30G7_B 41 YKGGKWHGDVAVKMLNVTAPTPOCLQAFKNEVGVLRKTRHVNILLFMGY-STAPQLAIVTQ**
L7RRS6 364 YKGGKWHGDVAVKILKVVDPTEQFQAFRNEVAVLRKTRHVNILLFMGY-MTKDNLAIVTQ
H7C155 243 YKGGKWHGDVAVKILKVVDPTEQFQAFRNEVAVLRKTRHVNILLFMGY-MTKDNLAIVTQ
B4E0X2 283 YKGGKWHGDVAVKILKVVDPTEQFQAFRNEVAVLRKTRHVNILLFMGY-MTKDNLAIVTQ
E9PB13 652 YHGRWHGEVAIRLIDIERDNEQLKAFKREVMAYRQTRHENVVLFMGACMSPPHLAIITS

30G7_B 100 WCEGSSLYHHLHASETKFEMKKLIDIAARQTARQMDYLHAKSIIHRDLKSNINIFLHEDNTV**
L7RRS6 423 WCEGSSLYKHLHVQETKFQMFQOLIDIAARQTAQGMDYLHAKNIIHRDMKSNINIFLHEGLTV
H7C155 302 WCEGSSLYKHLHVQETKFQMFQOLIDIAARQTAQGMDYLHAKNIIHRDMKSNINIFLHEGLTV
B4E0X2 342 WCEGSSLYKHLHVQETKFQMFQOLIDIAARQTAQGMDYLHAKNIIHRDMKSNINIFLHEGLTV
E9PB13 712 LCKGRITLYSVVRDAKIVLDVINKTRQIAQEIIVKGMGYLHAKGILHKDLKSKNVFYDNG-KV

30G7_B 160 KIGDFGLATEKSRWGSQH--FEQLSGSILWMAPEVIRM-----QDSNPYSFQSDVYA*
L7RRS6 483 KIGDFGLATVKSRWGSQQ--VEQPTGSVLWMAPEVIRM-----QDNNPFSFQSDVYS
H7C155 362 KIGDFGLATVKSRWGSQQ--VEQPTGSVLWMAPEVIRM-----QDNNPFSFQSDVYS
B4E0X2 402 KIGDFGLATVKSRWGSQQ--VEQPTGSVLWMAPEVIRM-----QDNNPFSFQSDVYS
E9PB13 771 VITDFGLFSISGVLOAGRREDKLRIQNGWICHIAPEIIRQLSPDTEEDKLPFSKHSDFVA

30G7_B 211 FGIVLYELMTGQLPYSNINNRDQIIFMVGGRGSLSPDLSKLYRNSNCPKRMKRLMAECLKKKR
L7RRS6 534 YGIVLYELMTGELPYSHINNRDQIIFMVGGRGYASPDLSKLYKNCPKAMKRLVADCVKVKV
H7C155 413 YGIVLYELMTGELPYSHINNRDQIIFMVGGRGYASPDLSKLYKNCPKAMKRLVADCVKVKV
B4E0X2 453 YGIVLYELMTGELPYSHINNRDQIIFMVGGRGYASPDLSKLYKNCPKAMKRLVADCVKVKV
E9PB13 831 LGTIHWYELHAREWPEKT-QPAEAIIMQMGTCM-KENLSQIG--MGKEISDILLFCWAFEQ

30G7_B 271 DERPSFPRILAEIEELARE-----
L7RRS6 594 EERPLFPQILSSIPELLQHS LPKINRSASEPSI-HRAAHTEDINACTLTTSRPLPVF
H7C155 473 EERPLFPQILSSIPELLQHS LPKINRSASEPSI-HRAAHTEDINACTLTTSRPLPVF
B4E0X2 513 EERPLFPQILSSIPELLQHS LPKINRSASEPSI-HRAAHTEDINACTLTTSRPLPVF
E9PB13 887 EERPTFTKLMDMLE----KLPKRNRLSHPGHFWKSAEL-----

```

Figure 3.19. Sequence alignment of oncogenic B-RAF V600E and corresponding homologous proteins. Homologous proteins are identified by their UniProt ID. * denotes predicted I3C amino acid contact points. Residues highlighted in black correspond to conserved sequences, regions in grey correspond to amino acid residues that are not conserved, but share the same biological property, while residues not highlighted show areas that are not conserved.

Wnt Homology Modeling

Wnt signaling is important in mediating embryogenesis and during carcinogenesis, it is deregulated [155, 173]. The wnt family consists of 19 secreted glycoproteins that can bind to various receptors or co-receptor complexes such as Frizzled (FZD), ROR1/2, RYK, and LRP, hence instigating a broad dimension of cellular responses [174]. In regards to transforming mouse mammary cells, wnt ligands have three primary classifications which are comprised of the highly transforming class (Wnt1, Wnt3A, Wnt7A, Wnt 5B and Wnt 2), a non-transforming class (Wnt4, Wnt5A and Wnt6), and an intermediate transforming class (Wnt7B). Previous studies have indicated that Wnt3A affects metabolism and alters mitochondrial behavior in melanoma [175]. It has also been demonstrated that all benign naevi express Wnt5A, Wnt7B, and Wnt10B [176]. Homology models were generated for Wnt3A, Wnt5B, Wnt7B, and Wnt10B using the 3D crystal structure of *Xenopus* Wnt8 (PDB accession number: 4F0A) as a structural template. Correspondingly, the homologues share roughly 40% in sequence similarity (Figure 3.20).

NCBI GenBank ID Code	Protein Name	Sequence Identity to protein Wnt-8 (PDB Accession Code: 4F0A) (%)
BAB61052.1	Wnt3A	34.072
BAB68399.1	Wnt7B	32.493
AAH74783.2	Wnt5A	31.061
AAB51685.1	Wnt10B	25.183
Uniprot ID: Q53TK5	Putative uncharacterized protein ADAM23	10.265

Figure 3.20. Wnt homologous proteins. Table depicts % sequence identity of each corresponding homologous protein to Wnt8. Proteins are identified by their NCBI GenBank ID code or UniProt ID.

Molecular Docking of Wnt homologous proteins and indolecarbinol compounds

In the *xwnt8*-*frizzled-8* cysteine rich domain (Fz8-CRD) crystal structure [162], *wnt* forms a two-domain structure that resembles a “hand” with a “thumb” and “index finger” unfurling out to bind to two points of contact on the Fz8-CRD. Docking studies using Glide, version 6.9 in Schrödinger revealed that I3C is predicted to bind to one of the two *wnt*/fz-CRD interfaces in *Wnt3A* (Figure 3.21A). One fz-CRD interaction site on *wnt* (from the *xwnt8*/fz-CRD crystal structure) is comprised of residues 301-338. These contact residues are highly conserved in all Wnts [162]. In *Wnt3A*, I3C is anchored to the apex of this contact region, engaging with hydrophobic residues Cys 81 and Cys 307 (Figure 3.21B & C). The binding of I3C to this region on *wnt* may prevent critical van der Waals interactions with main-chain and apolar residues on the Fz8-CRD. The predicted I3C binding site is also adjacent to a conserved patch of amino acid residues, particularly residues 216-219, 249-252, and 256-269. Because this region is away from the Fz binding site and is solvent exposed, this area serves as a potential binding site for additional co-receptors such as Lrp5/6 and/or Ryk which would allow the formation of a ternary complex with Fz. Ostensibly, the binding of I3C near this region may thus prevent the formation of a Wnt/Fz/LRP ternary complex.

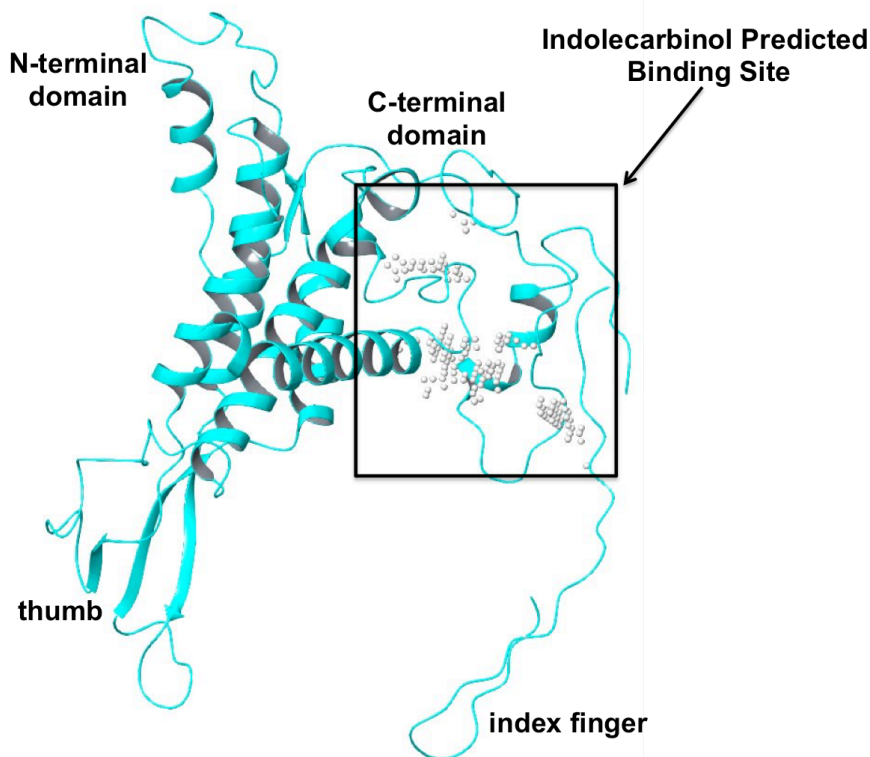


Figure 3.21A. Homology model of Wnt3A (NCBI GenBank ID: BAB61052.1). Ribbon diagram of homology model using the xenopus *Wnt8* X-ray crystal structure (PDB accession code: 4F0A) as a template. Box encircles SiteMap’s best predicted indolecarbinol binding site.

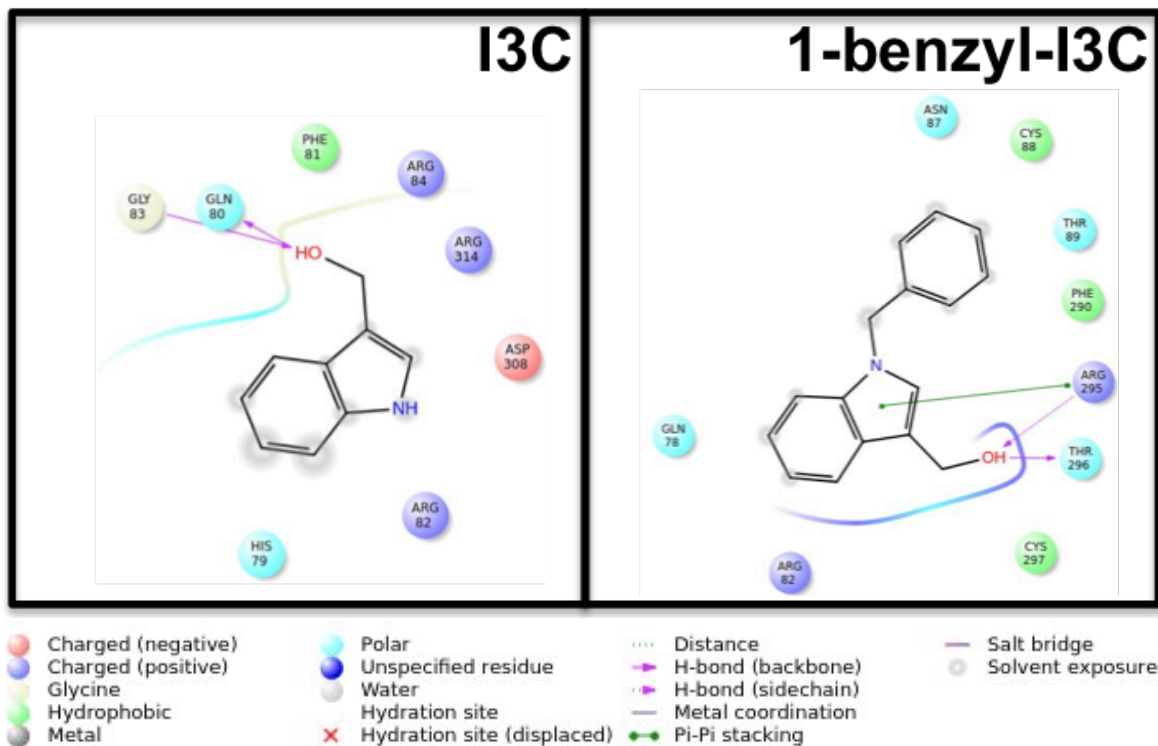
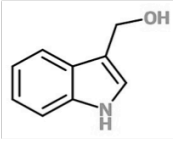
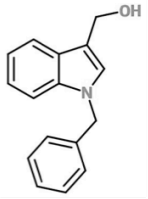
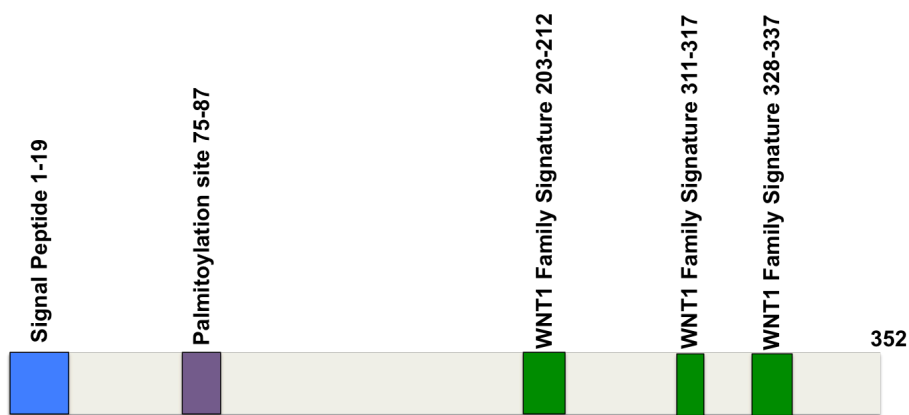


Figure 3.21B. Ligand interaction diagram of I3C and 1-benzyl-I3C in adducts with Wnt3A (NCBI GenBank ID: BAB61052.1) homology model. Predicted molecular docking interactions diagram depicting close contact residues within 4 Å of either I3C or 1-benzyl-I3C.

Indolecarbinol Compound	Structure	Predicted Interacting Amino Acid Residue	Amino Acid Biological Property
I3C		Phe 81	Hydrophobic
		His 79 Gln 80	Polar
		Asp 308	Charged (negative)
		Arg 82 Arg 84 Arg 314	Charged (positive)
1-benzyl-I3C		Cys 88 Phe 290 Cys 297	Hydrophobic
		Gln 78 Asn 87 Thr 89 Thr 296	Polar
		Arg 82 Arg 295	Charged (positive)



Predicted contact points: Gln 75, Gln 78, His 79, Arg 82, Gly 83, Asn 87, Phe 290, Asp 294, Arg 295, Thr 296, Cys 297, Cys 311 (denoted by *)

Figure 3.21C. Table and diagram of predicted molecular docking interactions of I3C and 1-benzyl-I3C with Wnt3A (NCBI GenBank ID: BAB61052.1) homology model. Table summarizes I3C and 1-benzyl-I3C interactions with their close contact residues and each residue's corresponding biological property. All interactions described are within 4 Å of each indolecarbinol analog. Figure below shows I3C predicted contact points within each subdomain of the homology model.

In the Wnt5A homology model, I3C is embedded in a region comprised of polar residues Asn 115, Ser 117, Thr 321, Gln 322, hydrophobic residue Leu 319, and positively charged residue 324 (Figure 3.21B).

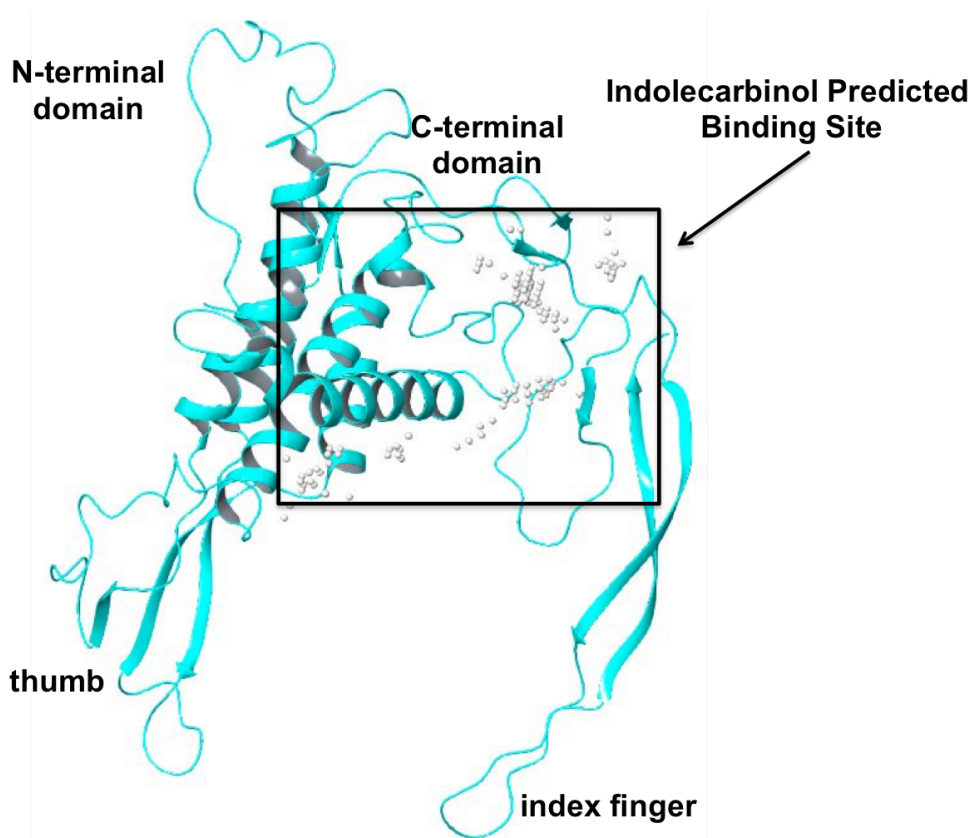


Figure 3.22A. Homology model of Wnt5A (NCBI GenBank ID: AAH74783.2). Ribbon diagram of homology model using the xenopus Wnt8 X-ray crystal structure (PDB accession code: 4F0A) as a template. Box encircles SiteMap's best predicted indolecarbinol binding site.

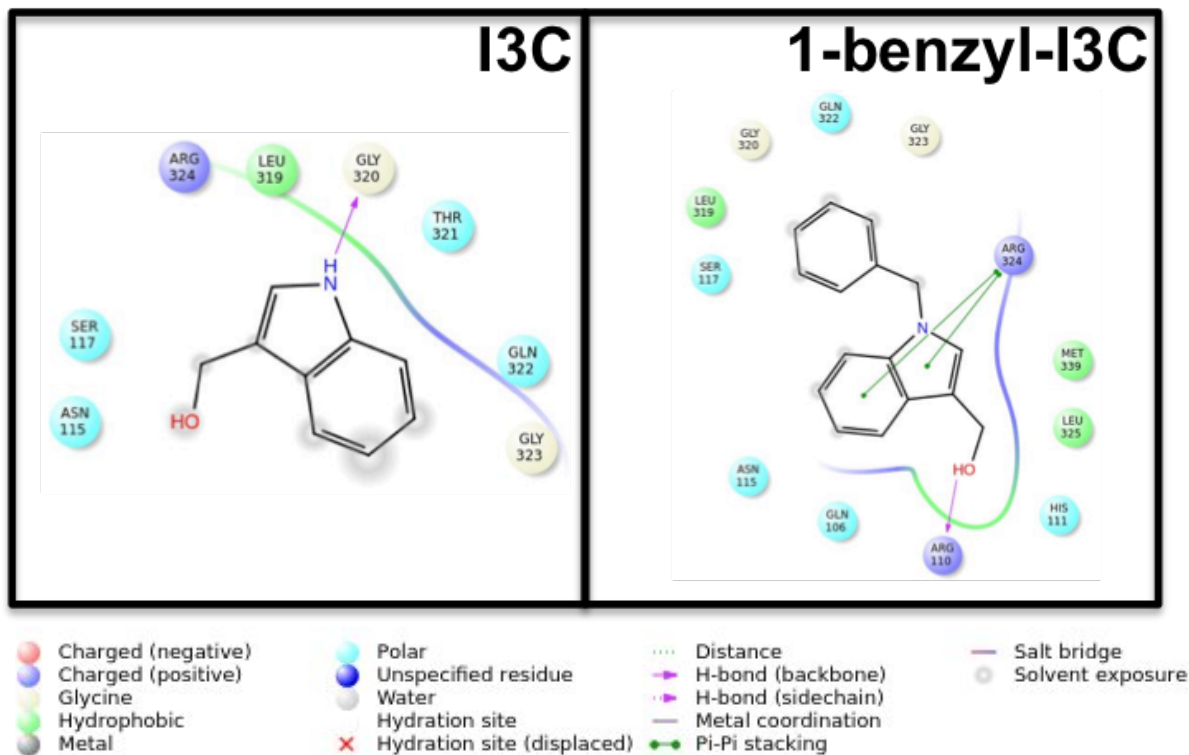
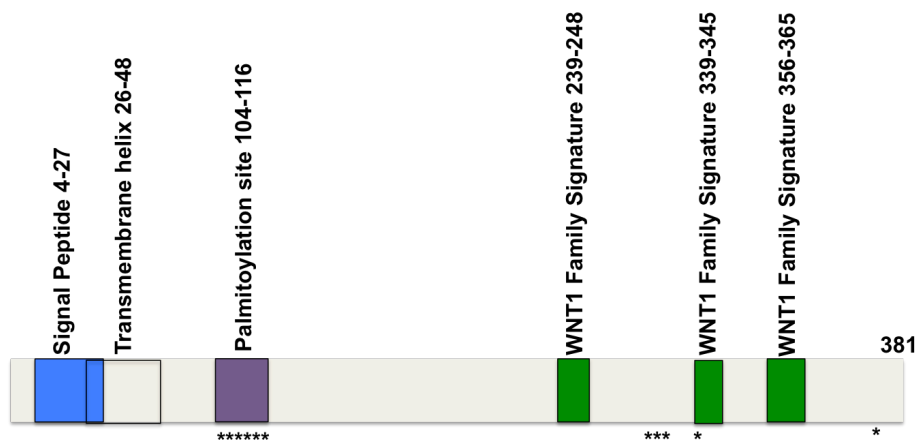


Figure 3.22B. Ligand interaction diagram of I3C and 1-benzyl-I3C in adducts with Wnt5A (NCBI GenBank ID: AAH74783.2) homology model. Predicted molecular docking interactions diagram depicting close contact residues within 4 Å of either I3C or 1-benzyl-I3C.

Indolecarbinol Compound	Structure	Predicted Interacting Amino Acid Residue	Amino Acid Biological Property
I3C		Leu 319	Hydrophobic
		Asn 115 Ser 117 Thr 321 Gln 322	Polar
		Arg 324	Charged (positive)
		Gly 320 Gly 323	Unique
1-benzyl-I3C		Leu 319 Leu 325 Met 339	Hydrophobic
		Gln 106 His 111 Asn 115 Ser 117 Gln 322	Polar
		Arg 110 Arg 324	Charged (positive)
		Gly 320 Gly 323	Unique



Predicted contact points: Lys 103, Gln 106, Tyr 107, Arg 110, Arg 112, Asn 115, Arg 324, Leu 325, Lys 326, Met 339, Val 379 (denoted by *)

Figure 3.22C. Table and diagram of predicted molecular docking interactions of I3C and 1-benzyl-I3C with Wnt5A (NCBI GenBank ID: AAH74783.2) homology model. Table summarizes I3C and 1-benzyl-I3C interactions with their close contact residues and each residue's corresponding biological property. All interactions described are within 4 Å of each indolecarbinol analog. Figure below shows I3C predicted contact points within each subdomain of the homology model.

It was observed in the Wnt7B homology model that I3C makes contacts with hydrophobic residues Phe 77, Trp 82, Tyr 232, Tyr 277, and Met 307. Cation- π interactions between the indole ring on I3C and positively charged residues Arg 81, Arg 102, and Lys 229 fasten I3C securely into place. The hydroxyl substituent also engages in hydrogen bond interactions with the sidechains of Glu 103 and Tyr 232 (Figure 3.23).

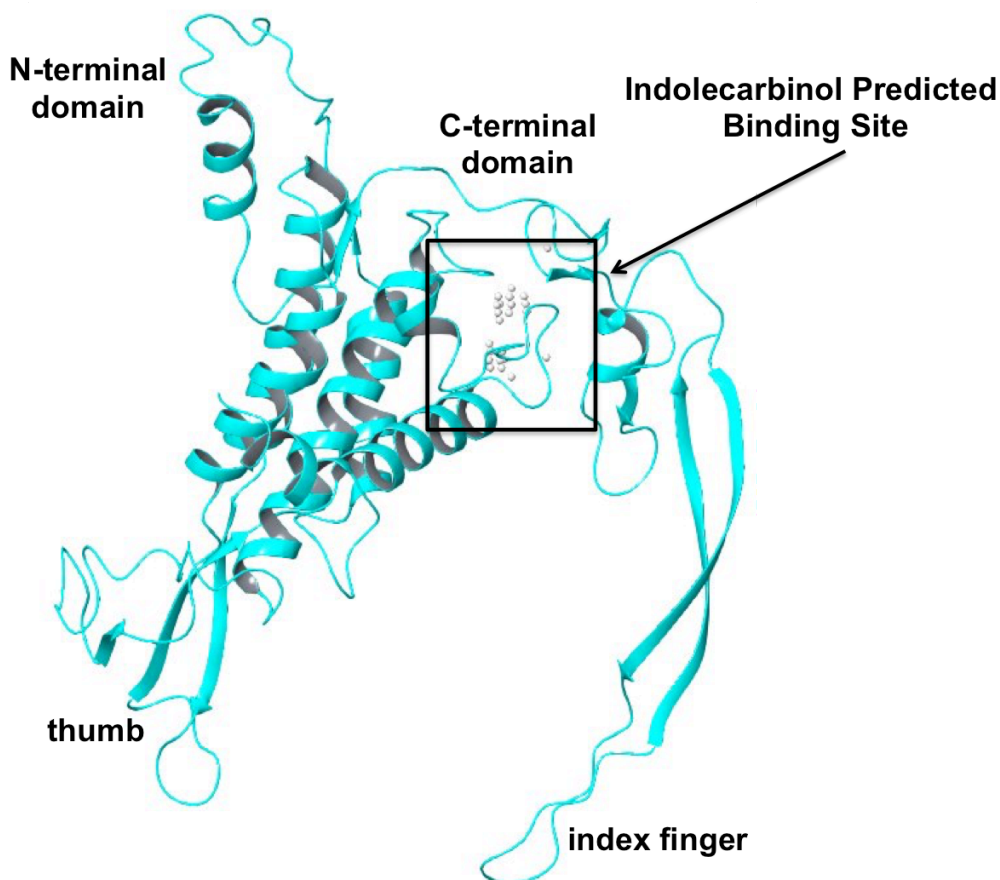


Figure 3.23A. Homology model of Wnt7B (NCBI GenBank ID: BAB68399.1). Ribbon diagram of homology model using the xenopus Wnt8 X-ray crystal structure (PDB accession code: 4F0A) as a template. Box encircles SiteMap's best predicted indolecarbinol binding site.

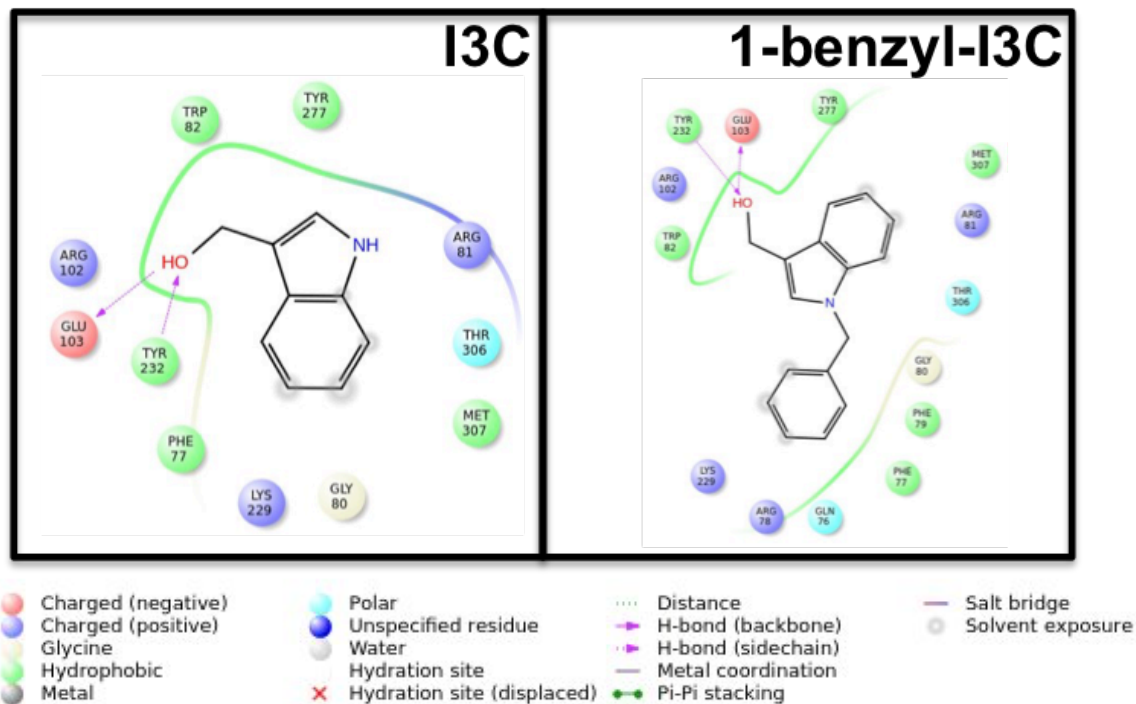
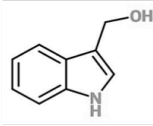
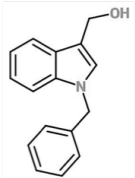
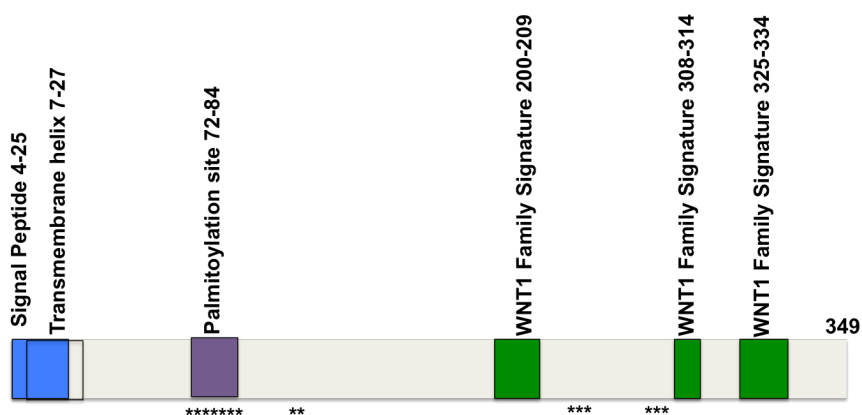


Figure 3.23B. Ligand interaction diagram of I3C and 1-benzyl-I3C in adducts with Wnt7B (NCBI GenBank ID: BAB68399.1) homology model. Predicted molecular docking interactions diagram depicting close contact residues within 4 Å of either I3C or 1-benzyl-I3C.

Indolecarbinol Compound	Structure	Predicted Interacting Amino Acid Residue	Amino Acid Biological Property
I3C		Phe 77 Trp 82 Tyr 232 Tyr 277 Met 307	Hydrophobic
		Thr 306	Polar
		Glu 103	Charged (negative)
		Arg 81 Arg 102 Lys 229	Charged (positive)
		Gly 80	Unique
1-benzyl-I3C		Phe 77 Phe 79 Trp 82 Tyr 277 Met 307	Hydrophobic
		Gln 76 Thr 306	Polar
		Glu 103	Charged (negative)
		Arg 78 Arg 81 Arg 102 Lys 229	Charged (positive)
		Gly 80	Unique



Predicted contact points: Tyr 75, Gln 76, Phe 77, Arg 78, Phe 79, Arg 81, Asn 83, Arg 102, Glu 103, Lys 229, Tyr 232, Asn 233, Asp 302, Gly 303, Thr 306 (denoted by *)

Figure 3.23C. Table and diagram of predicted molecular docking interactions of I3C and 1-benzyl-I3C with Wnt7B (NCBI GenBank ID: BAB68399.1) homology model. Table summarizes I3C and 1-benzyl-I3C interactions with their close contact residues and each residue's corresponding biological property. All interactions described are within 4 Å of each indolecarbinol analog. Figure below shows I3C predicted contact points within each subdomain of the homology model.

The indole ring on I3C is anchored by hydrophobic residues Cys 334, Trp 384, and Val 385, polar residues Ser 338 and Gln 358, and negatively charged residue Glu 383 in the Wnt10B homology model. The hydroxyl moiety is able to associate with the backbone of Thr 337 and the sidechain of Asn 386 via hydrogen bonding interactions (Figure 3.24).

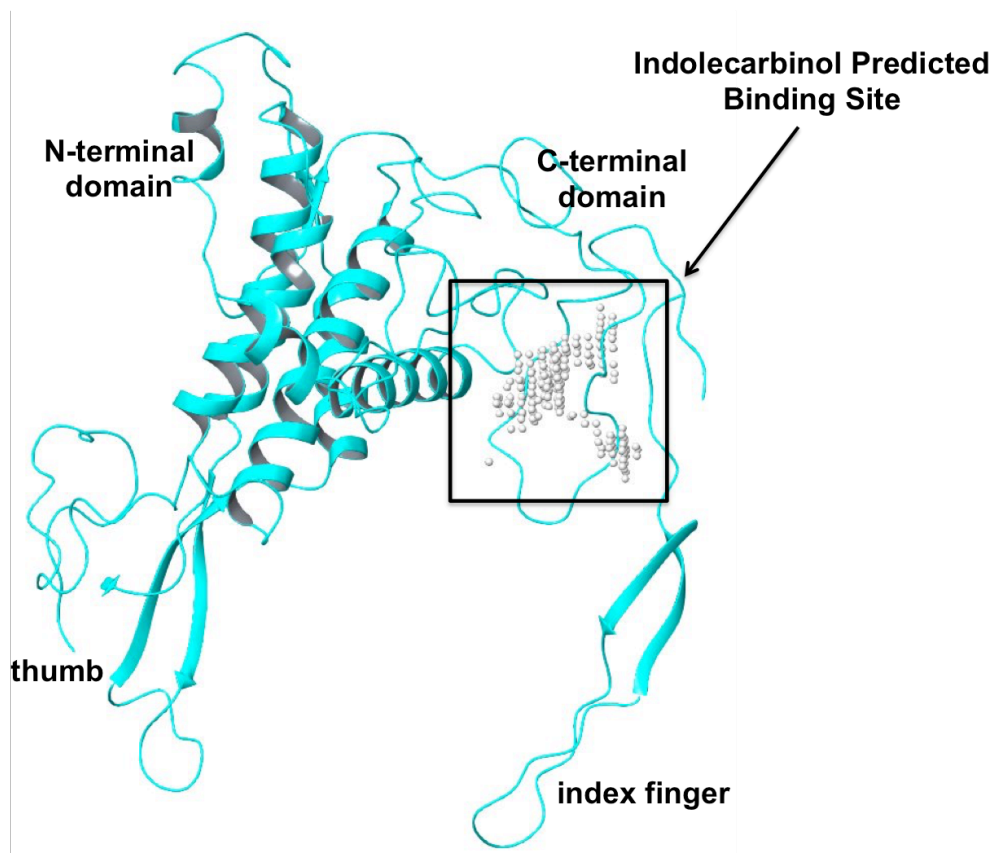


Figure 3.24A. Homology model of Wnt10B (NCBI GenBank ID: AAB51685.1). Ribbon diagram of homology model using the xenopus Wnt8 X-ray crystal structure (PDB accession code: 4F0A) as a template. Box encircles SiteMap's best predicted indolecarbinol binding site.

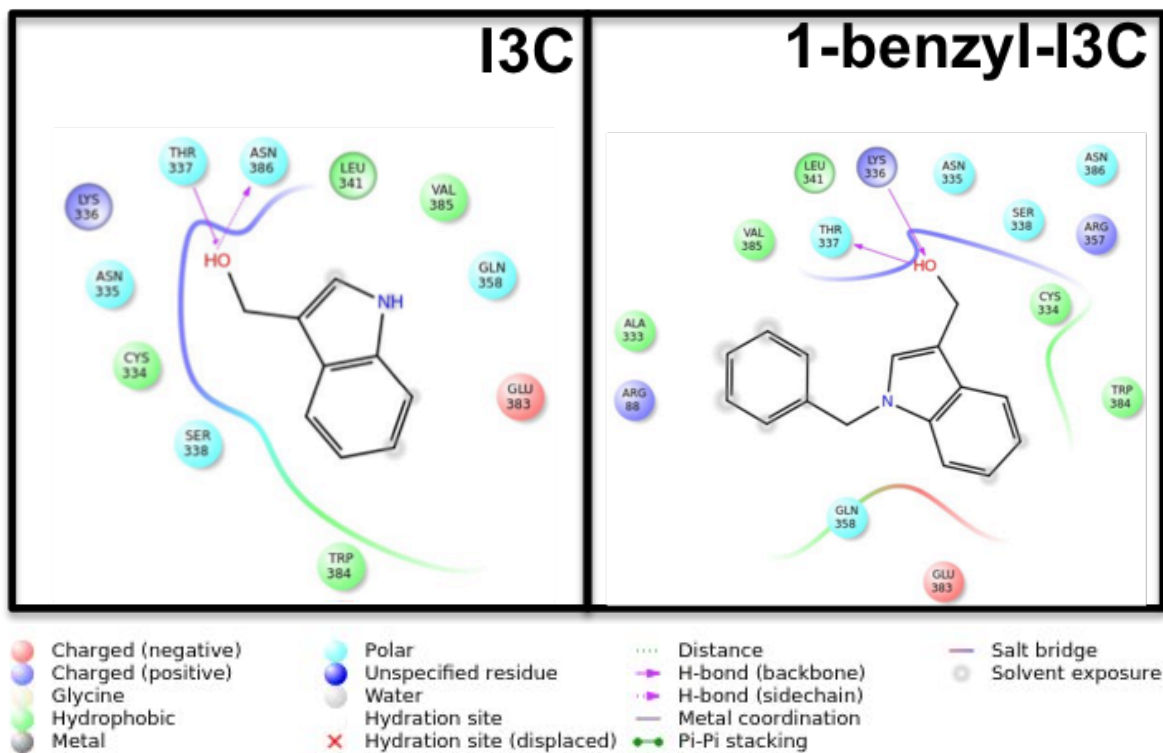
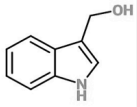
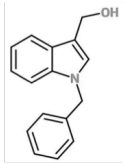
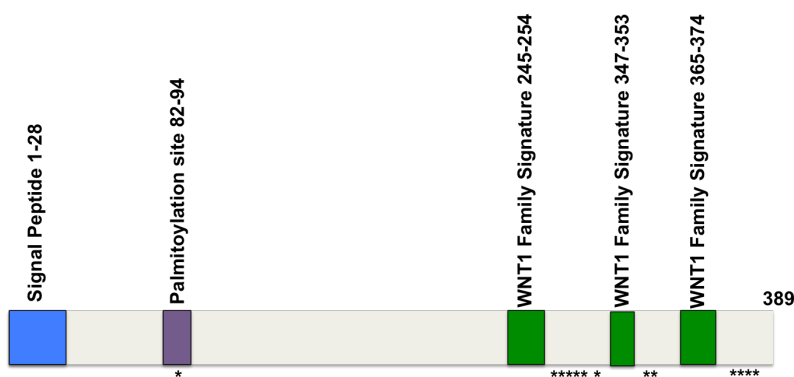


Figure 3.24B. Ligand interaction diagram of I3C and 1-benzyl-I3C in adducts with Wnt10B (NCBI GenBank ID: AAB51685.1) homology model. Predicted molecular docking interactions diagram depicting close contact residues within 4 Å of either I3C or 1-benzyl-I3C.

Indolecarbinol Compound	Structure	Predicted Interacting Amino Acid Residue	Amino Acid Biological Property
I3C		Trp 384 Cys 334 Leu 341 Val 385	Hydrophobic
		Ser 338 Asn 335 Thr 337 Asn 386 Gln 358	Polar
		Glu 383	Charged (negative)
		Lys 336	Charged (positive)
1-benzyl-I3C		Cys 334 Trp 384 Ala 333 Val 385 Leu 341	Hydrophobic
		Gln 358 Thr 337 Asn 335 Ser 338 Asn 386	Polar
		Glu 383	Charged (negative)
		Arg 88 Lys 336 Arg 357	Charged (positive)



Predicted contact points: Arg 88, Cys 334, Asn 335, Lys 336, Thr 337, Ser 338, Leu 341, Arg 357, Gln 358, Glu 383, Trp 384, Val 385, Asn 386 (denoted by *)

Figure 3.24C. Table and diagram of predicted molecular docking interactions of I3C and 1-benzyl-I3C with Wnt10B (NCBI GenBank ID: AAB51685.1) homology model. Table summarizes I3C and 1-benzyl-I3C interactions with their close contact residues and each residue's corresponding biological property. All interactions described are within 4 Å of each indolecarbinol analog. Figure below shows I3C predicted contact points within each subdomain of the homology model.

A homology model of the putative, uncharacterized ADAM23 protein which shares roughly 17% homology between the wnt homologues examined was generated to ascertain discernible differences in indolecarbinol binding (Figure 3.25). The best predicted binding site showed I3C binding to the thumb region on the NTD domain where the pocket is comprised of hydrophobic residues Tyr 257, Leu 278, Trp 280, Met 298 and positively charged residue Lys 259. The nitrogen on the indole ring interacts with the sidechain of Ser 258 and the hydroxyl moiety can associate with the side chain of Gln 260 both through hydrogen bond interactions.

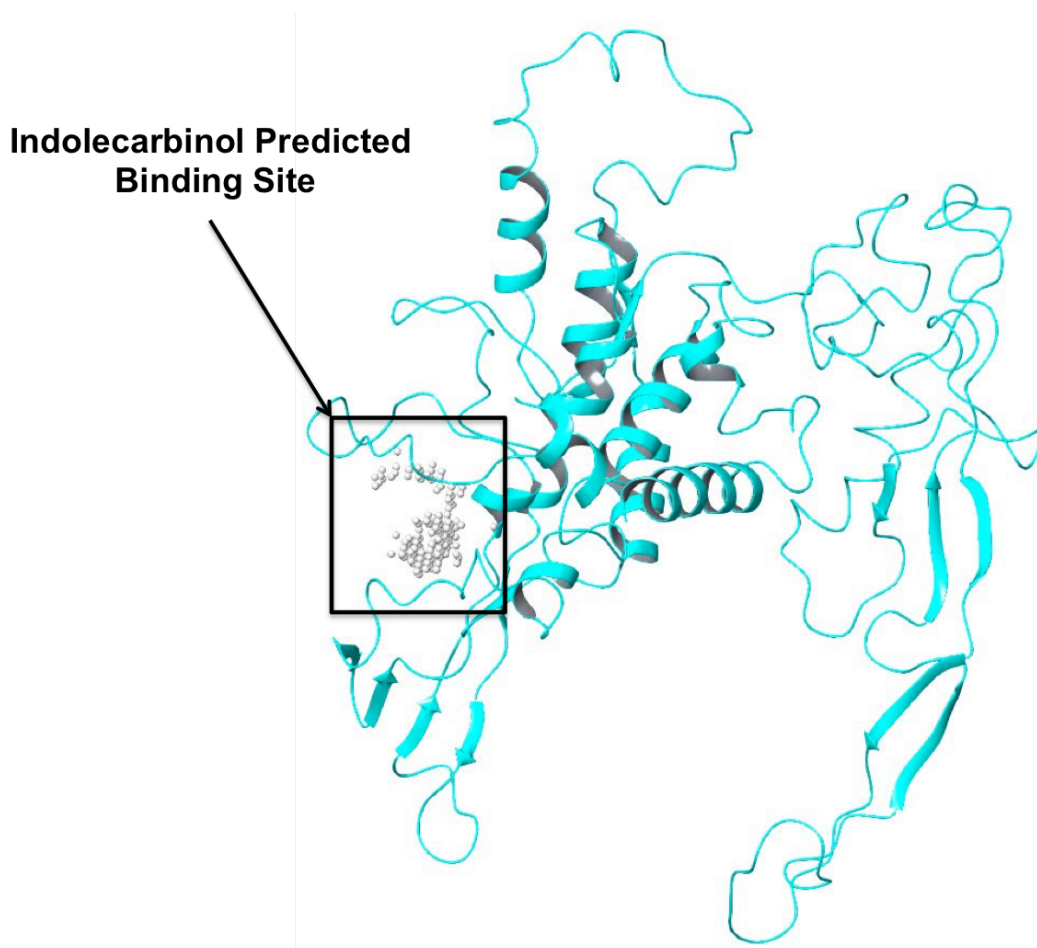
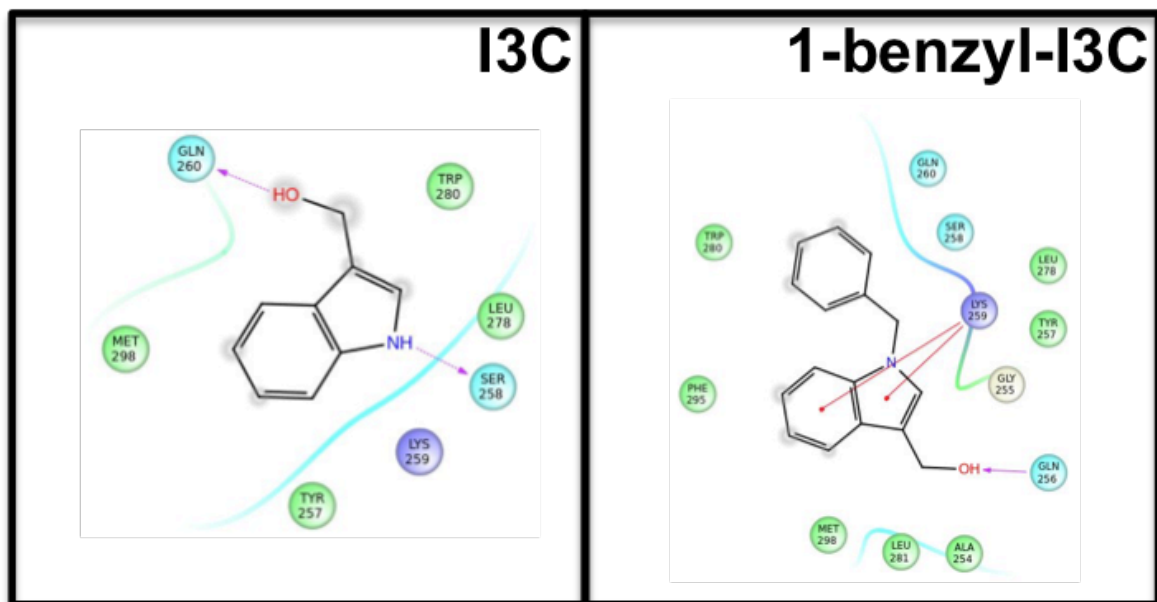
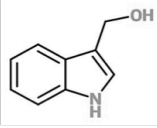
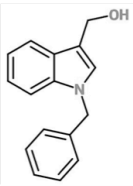


Figure 3.25A. Homology model of ADAM23 protein (Uniprot ID: Q53TK5). Ribbon diagram of homology model using the xenopus Wnt8 X-ray crystal structure (PDB accession code: 4F0A) as a template. Box encircles SiteMap's best predicted indolecarbinol binding site.



- Charged (negative)
- Charged (positive)
- Glycine
- Hydrophobic
- Metal
- Polar
- Unspecified residue
- Water
- Hydration site
- X Hydration site (displaced)
- ⋯ Distance
- H-bond (backbone)
- H-bond (sidechain)
- Metal coordination
- Pi-Pi stacking
- Salt bridge
- Solvent exposure

Figure 3.25B. Ligand interaction diagram of I3C and 1-benzyl-I3C in adducts with ADAM23 protein (Uniprot ID: Q53TK5) homology model. Predicted molecular docking interactions diagram depicting close contact residues within 4 Å of either I3C or 1-benzyl-I3C.

Indolecarbinol Compound	Structure	Predicted Interacting Amino Acid Residue	Amino Acid Biological Property
I3C		Tyr 257 Leu 278 Trp 280 Met 298	Hydrophobic
		Ser 258 Gln 260	Polar
		Lys 259	Charged (positive)
1-benzyl-I3C		Ala 254 Tyr 257 Leu 278 Trp 280 Leu 281 Phe 295 Met 298	Hydrophobic
		Gln 256 Ser 258 Gln 260	Polar
		Lys 259	Charged (positive)
		Gly 255	Unique

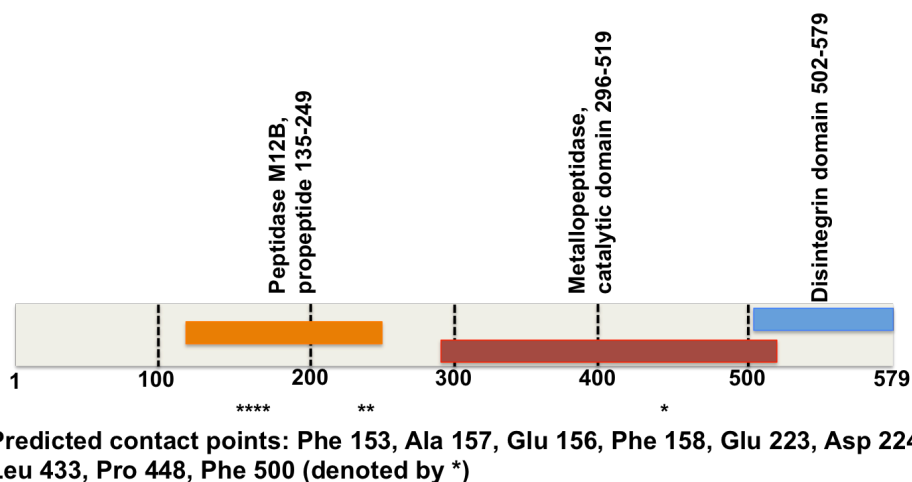


Figure 3.25C. Table and diagram of predicted molecular docking interactions of I3C and 1-benzyl-I3C with ADAM23 protein (Uniprot ID: Q53TK5) homology model. Table summarizes I3C and 1-benzyl-I3C interactions with their close contact residues and each residue's corresponding biological property. All interactions described are within 4 Å of each indolecarbinol analog. Figure below shows I3C predicted contact points within each subdomain of the homology model.

After examining the 4 wnt homology models (Wnt3A, Wnt5A, Wnt7B, Wnt10B), it is evident that the predicted indolecarbinol binding sites are all positioned within the same region of the apex of the index finger in the C-terminal domain (Figure 3.26). Interestingly, I3C is predicted to bind to a palmitoylation region that is conserved within the 4 wnt homologues.

Palmitoylation is required for the Wnt3a dependent accumulation of β -catenin. It is important for localizing to cellular membrane compartments. In Wnt3a, palmitoylation at Cys77 is essential for Wnt3a to bind to its receptors and to activate subsequent signaling. Moreover, it is required for internalization of LRP6 [177]. Palmitoylation and glycosylation are important for the actions and secretion of Wnt5A which induces the internalization of Fz5 and it promotes cell migration by stimulating focal adhesion turnover. Palmitoylation of Wnt-5a is necessary for inhibitory activity against the β -catenin pathway. It is also critical for its ability to stimulate cell migration. Palmitoylation of a conserved cysteine (Cys 104) was important for the ability of wnt-5a to inhibit Tcf-4 transcriptional activation and to stimulate cell migration and it is also important for the binding of wnt5a to Fz5 and the lipid-unmodified form could not induce the internalization of Fz5. Palmitoylation may be crucial for the anchoring of wnt-5a to the cell surface membrane. Palmitoylation may be necessary for the binding of wnt to lipoprotein particles (Kurayoshi *et al.*, 2007). Therefore, the importance of the conserved palmitoylation sites in the wnt proteins suggests that binding of the indolecarbinol compounds may hence impede important cell signaling where palmitoylation is involved.

```

WNT3A      1  KAPL-----S-----Y-----LCCSLKQALGSYPIWWSLAV--
WNT5A      1  PLOKSIGILSPGVALGMAGSAMSSKFFLVALA-----FFSF--QVIEANSWWSLGNNN
WNT7B      1  MHRN-----FRKWIIV-----ELCF--VQYVK-----LGA--
WNT10B     1  MLEPRPRPPP--SGLA-----S--LL-----LALCS-RAI--S---NEILG--
ADAM23     1  MKPP-----S-----SSSRQPPAGCSLGG-----ASC--

WNT3A      29  ---GPOYSSLGSRPIICASIPGLVETCLRF--RNYVE--LMPVAEGIKIGLCCCHOFFG
WNT5A      55  PVOSEVYIIGAC--PIGSOA--GLSQCKKICHLYQI--HVOYIGEGATIGRECCYQFRH
WNT7B      26  ---SSVVALGAN--IICNKIPGLAF--ORAK--QSRFD--AIVLIGEGAC--IGRECCYQFRF
WNT10B     35  --KIPGEPFLTAN--TICLIT--SGLSK--GLGCLRNED--VTASALQCLLAVRECCCHTRD
ADAM23     24  -----GPRGPAGSVEFSAFARTPPC--LLLVLL--LPP--ASS--PRAWGA-AAPSA

      * *
WNT3A      84  RRWNCSTVHDS--LALFGPVLQKATRE-----SAFVHAIASAGVAFV--RS--CAEGTA
WNT5A     112  RRWNCSTVNDN---TSVFGRVVQIGBRE-----IAFTYAV--AAGV--NAR--ACREGEL
WNT7B      80  GRWNCSAIAGE---KTVFGQERVGBRE-----AAFTYAI--AAGV--AHAV--AACCS--NL
WNT10B     90  CRWNCSAIEGGGR--PHHSA--LKR--BRE-----SAPFESS--AAGV--HAAV--ATACS--LKL
ADAM23     73  PRWNETAEKN---LGLMLDEDNTLQONSSSNISYSNA--MQKET--LPSR--IYYINQD--SESPY

WNT3A     134  AICGSSRHGGS--PG-----K--RWGGCSE
WNT5A     161  STCGSRAARPK--DLP-----K--RWGGCSD
WNT7B     129  SNCGDREKQGYNQA-----E--RWGGCSA
WNT10B    142  VSCCGWVGSSEQDRLRAKLLQLQALS--RGKSFPHSLPSPGPGSSPSPGPODTY--RWGGCNH
ADAM23    130  HVLDTKARHCQK--HN-----K--VLLQASF

WNT3A     158  DIEF-----GCMVSR--FADAR--ENR-----PDAR--SAM--RHN--NEAGRQA
WNT5A     186  NIDY-----GYRFA--EFVDAR--ERERI--HAKGSY--ESARILAM--RHN--NEAGRRT
WNT7B     155  DMRH-----GIDFSRRFVDAR--EIK-----SMAR--LMM--RHN--NEAGRKV
WNT10B    202  DMDF-----GCKFSRDF--LSR--DAP-----PDIQAR--RHN--NRV--GROV
ADAM23    154  QTEAFGSKFILDILNGLLSSDYV--IHYENG-----K--PQYSKGGEG--CYHYGSTRGV

WNT3A     195  IASHMH--KCKCHG--SGSCEV--KTCWNSQ--PFRAL-----
WNT5A     231  VYNLAD--ACKCHG--VSGSCL--KTCWLCQ--ADFRKV-----
WNT7B     192  LEDEMO--KCKCHG--VSGSCT--TKCWTT--LTKFREV-----
WNT10B    239  VTEM--KRKCKCHG--TSGSQP--KTCWRAA--PFRAV-----
ADAM23    206  KDSKV--ALST--N--H--MFED--TFVYM--EPELELVHDEKSTGRPHIIQKTLAGQYSKOMKNL

WNT3A     228  -----
WNT5A     264  -----
WNT7B     225  -----
WNT10B    272  -----
ADAM23    265  TMERGDQWPFLSELQWLKRRKRAVNP--SRGI--FEEMKYLELMIVNDHKTYKKHRSSHAHTNN

WNT3A     228  -GDFLK--KYDSASEV--VEKH--RESRWY--ETH--RP--RYTY--KVP--TERDLVY--Y--E--SF----
WNT5A     264  -GDALKEKYDSAAAR--N--SREKLV--Q--N--S--RNSPT--QDLVY--Y--E--SF----
WNT7B     225  -CHLLKEKYNAAVQ--E--V--RA--SLRQPTFLRIKQLRS--QK--E--TDLVY--Y--E--SF----
WNT10B    272  -SAALRE--LGRAIF--DT--HNRNS--G--FQPR--RP--RRLSG--LVY--Y--E--SF----
ADAM23    325  FKSVVNLVDS--TYKEC--N--R--VVV--VA--E--TWTEK--C--DITTNFVQMLH

      * * * * *
WNT3A     279  -----NFCBPNPBTGSE--STR--R--CNV--S-----R--
WNT5A     308  -----DVCV--N--ESTG--S--GTGR--CN--TS-----E--
WNT7B     276  -----NYCEDAATGSMGTGR--CN--TS-----E--
WNT10B    316  -----DFCE--DPT--G--S--TRGR--CN--TS-----R--L--
ADAM23    371  EFSKYRQRIKQHADAVHLISRVT--HYR--S--SYF--C--S--R--GVGVN--E--Y--LPMVAQA

WNT3A     304  -----ADGC-----
WNT5A     333  -----ADGC-----
WNT7B     301  -----ADGC-----

WNT10B    341  -----ADGC-----
ADAM23    428  VLSQSLAQLNLGIQWEPSSRKPKCDCTESWGC--IMEETGVSHSRKFSKCSILEYRDLFQRG

WNT3A     308  -----LLCCGRG--NARAER--REK--R--GV--FHWCCYVS--Q--E--C--R--VY--E--VHT--K--N
WNT5A     337  -----ELMCCGRGYDQFKTVQ--TER--C--R--FHWCCYVK--C--K--K--E--I--V--Q--F--V--C--K
WNT7B     305  -----LTMCCGRGYNTHQYT--VWCC--C--R--FHWCCYVK--C--N--T--C--S--E--R--T--E--V--T--E--K
WNT10B    345  -----GSLCCGRG--NVLQRT--V--E--R--C--R--FHWCCYVL--C--D--E--K--V--T--E--K--N--V--E--K
ADAM23    488  GGACLFNRPTKLF--EPT--C--G--V--E--A--G--E--C--G--F--H--E--Y--G--L--O--K--K--S--L--S--N--G--A--H--C--S--D--G

WNT3A     354  PGSRAGNSAHQP--PH--P--P--V--R--F--H--P--P--L--R--R-----A--G--K--V--P
WNT5A
WNT7B
WNT10B
ADAM23    544  PCCNNTSCLFQPRGYEC-----RDAVNECDITEYCTGDSGQ--

```

Figure 3.26. Sequence alignment of Wnt isoforms and corresponding homologous protein, ADAM23 protein. * denotes predicted I3C amino acid contact points. Residues highlighted in black correspond to conserved sequences, regions in grey correspond to amino acid residues that are not conserved, but share the same biological property, while residues not highlighted show areas that are not conserved.

Discussion

In ligand-protein docking, homology models have progressively been employed [2, 178, 179] and have been dramatically expanding the amount of protein targets that are feasible for synthesizing novel small molecule inhibitors. Not only are homology models advantageous in the drug design process, they have also been proven beneficial in other structural biology arenas as well which include the understanding of protein function and the recognition of druggable pockets [180, 181], identification of potential binding sites and binding modes on proteins [182, 183] and structure-based virtual screening [184-188].

New and accurate predictions can be made regarding interactions between proteins and small molecule ligands by delineating the properties of proteins across various orthologous species. In regards to interactions between proteins and ligands, previous studies have implied that there are a greater fraction of orthologous proteins that share ligands in comparison to paralogous proteins [189]. Additionally, small molecules are frequently developed to target specific members of a protein family like specific kinase inhibitors. A robust means of inference can be implemented to make predictions about protein-ligand function and interactions because previous studies have established that homologous proteins do in fact have similar function and often associate with their small molecule ligands in a similar fashion. Therefore, even if no crystal structure is available for a particular protein, it is possible to make predictions about protein-small molecule associations as long as there are structures that are sufficiently close in homology.

While there have been great lengths in ascertaining the inaccuracies in protein models that predict protein-ligand interactions and establishing general criteria for assessing the quality of these models, the studies presented here provide an initial framework for understanding the relationships between target proteins of high homology and for characterizing the details of ligand-protein interactions in regards to similarities in binding site. Presently, binding site prediction methods are classified into various groups which include procedures which employ evolutionary conservation of binding site motifs [190-193], methods which implement information about a structure of a complex [193-195], and docking [196, 197]. Structure-based methods rely on knowledge from a protein structure that yields binding site information based on physiochemical properties of individual residues, electrostatic characteristics, and binding site location in the 3D structure [198-203].

Since the crystal structures of many of the homologous proteins examined were not available in the Protein Data Bank (PDB), homology modeling studies were performed by using homologous high-resolution protein crystal structures as the template structure. The homology models were synthesized by employing Prime from the Schrödinger suite. Apart from comparing the target and template sequences between one another, Prime also compares the target's predicted secondary structure with the template's known secondary structure [204] and thereby improves sequence alignments by incorporating structural information. Previous studies conducted by Bernacki *et al.* [205] have demonstrated that when the sequence identity between a target and a template was

greater than 50%, the resulting docking poses were analogous to crystal structures and this can be attributed to the inference that when the active site is maintained between the target and template, ligand binding is not significantly altered. To corroborate this, various other studies have confirmed that when homology modeling is employed for docking applications, a minimum fraction of 40% sequence identity is necessary in order to garner reliable data [206, 207].

Here we focused on protein indolecarbinol binding sites and their corresponding biological relevance for protein function. In order to make predictions about additional indolecarbinol target proteins, we obtained the crystal structures of the four I3C target proteins to date (human neutrophil elastase, ubiquitin E3 ligase NEDD4-1, oncogenic BRAF V600E serine/threonine kinase, WNT) and utilized them as the starting template structure for synthesizing other homologous protein models.

The crystal structure for human neutrophil elastase (PDB accession code:5A8X) was implemented as the starting template for synthesizing the homology models for proteinase 3 (54.6% sequence similarity to HNE), azurocidin (45.7% sequence similarity to HNE), and PRSSL1 protein (35.7% sequence similarity to HNE). The homologous proteins of elastase exemplify the histidine active site of serine proteases and they are classified under the MEROPS peptidase family S1 (chymotrypsin family, clan PA(S)). A charge relay system implicating an aspartic acid residue hydrogen-bonded to a histidine, which itself is then hydrogen-bonded to a serine confers the catalytic activity of the serine proteases. Within this family of proteases, the sequences that are adjacent to the active site histidine residues are preserved. After superimposing the proteinase 3 and azurocidin homology models onto the elastase-I3C docking model, it is evident that there is a distinct region of conserved residues within the protein structures that confers binding to I3C. After aligning the elastase, proteinase 3, and azurocidin sequences, and delineating the predicted I3C binding site within each sequence, it is evident that there are specific conserved residues within the C-terminal region of elastase and its corresponding homolog that may be important for maintaining contact with I3C. Notably, residues such as valine, threonine, cysteine, and arginine, appear to be essential for securely anchoring I3C in place. As the sequence similarity for elastase and the homologous proteins continued to decrease, it was apparent that critical amino acid residues, like valine and cysteine, for maintaining contact with I3C were lost. Notably, in the PRSSL1 protein which exhibits a 35.7% sequence similarity to elastase, the cysteine residue (Cys 152 in elastase, Cys 141 in proteinase 3, and Cys 180 in azurocidin) was replaced with a tryptophan. This cysteine to tryptophan substitution that's observed in the PRSSL1 protein may contribute to a less favorable interaction with additional I3C or 1-benzyl-I3C derived compounds.

The crystal structure of the E3 ubiquitin ligase NEDD4-1 Hect domain (PDB accession code: 2XBF) was used as the starting template structure for generating the homology models for the neural cell expressed, developmentally down-regulated 4like, isoform CRA_b (64.1% sequence similarity to NEDD4-1), the highly similar to E3 ubiquitin-protein ligase NEDD-4like protein (62.6% sequence similarity to NEDD4-1), the E3 ubiquitin-protein ligase NEDD4-like (60.0% sequence similarity to NEDD4-1), the E3

ubiquitin-protein ligase HUWE1 (54.7% sequence similarity to NEDD4-1), and the E3 ubiquitin-protein ligase HERC2 (26.5% sequence similarity to NEDD4-1). The homologous NEDD4-1 proteins that were examined here represent the HECT E3 ubiquitin ligase family of proteins. While the proteins exhibit varying degrees of sequence similarity, they all harbor a conserved catalytic HECT domain which is responsible for regulating the association with cognate E2s and generating a thioester adduct with ubiquitin through an evolutionally conserved cysteine residue [43, 57, 58, 208]. Since it was previously reported that the structure of NEDD4-1^{HECT} was co-crystallized with an indole-based small molecule (PDB Accession code: 5C91) [169], the location of this small molecule inhibitor was used as a starting point for docking the indolecarbinol compounds onto the homology models. After aligning the homology models with the NEDD4-1 HECT domain sequence, there are some discernible residues that are conserved across the homologous proteins that appear to be important for maintaining contact with I3C. The I3C binding pockets within the homologous proteins that share greater than 50% homology are predominantly comprised of residues such as glutamate, tyrosine, leucine, cysteine, and arginine. The E3 ubiquitin ligase HERC2 which exhibits a 26.5% sequence similarity to NEDD4-1, appears to lose essential residues that appear to be important for maintaining contact with I3C. In particular, in the HERC2 sequence, at position 4498, where a tyrosine residue which is present in the NEDD4-1 and remaining 4 homologous proteins examined in this study, has been replaced with a threonine residue. At positions 4525, 4526, and 4532, the cysteine, arginine, and tyrosine residues which are present in NEDD4-1 and the other high sequence similarity proteins have been substituted with an arginine, alanine, and methionine respectively.

Since oncogenic B-RAF V600E serine/threonine kinase has been implicated as an I3C target protein, the crystal structure (PDB accession code: 3OG7) was implemented as the template structure for modeling other homologous proteins. This family of protein kinases is characterized by an evolutionally conserved serine-threonine/tyrosine protein kinase catalytic domain at the C-terminal domain, followed by various ubiquitin related domains and diacylglycerol/phorbol-ester binding domains interspersed throughout the N-terminal region. Oncogenic BRAF harbors a valine to glutamate mutation at position 600, which confers its constitutive activity. In order to understand the I3C specificity for oncogenic B-RAF, molecular docking analysis was initially performed on BRAF V600E and wild-type BRAF to compare the I3C binding pockets within each protein. Previous co-crystallization studies of the BRAF V600E protein in adducts with vemurafenib have demonstrated that BRAF forms a dimer. As generally observed among various kinase domains, each BRAF protomer is comprised of two varied-sized lobes that are connected by a hinge region and a neighboring cleft that becomes the ATP binding region [209]. Both protomers conform to the 'DFG-in' orientation where the phenylalanine side chain of the DFG motif within the activation loops is planted within the interior and away from the ATP-binding pocket. In the mutated BRAF V600E, the salt bridge that links Glu 600 and Lys 507 keeps the activation loop in the 'DFG-in' conformation and hence renders the protein constitutively active. Our lab has previously demonstrated that I3C cooperatively works with Vemurafenib, a well-established, FDA clinically approved BRAF inhibitor, to mitigate MIT-M expression and

suppress melanoma cell proliferation. Docking analysis was performed on the oncogenic BRAF V600E protein in complex with Vemurafenib to ascertain where I3C would bind relative to the known BRAF inhibitor. In the oncogenic BRAF model in complex with both Vemurafenib and I3C, molecular docking simulations had predicted I3C to bind to the opposing apo-protomer not bound to Vemurafenib and was embedded in a pocket near the DFG motif and is comprised of residues Ile 463, Val 471, Ala 481, Lys 483, Leu 514, Thr 529, Gln 530, Trp 531, Cys 532, Phe 583, Asp 594 (chain B). In the wild-type BRAF model in complex with I3C, *in silico* analysis showed that I3C was predicted to bind to the protomer bound to Vemurafenib, and is surrounded by residues Ile 462, Val 470, Ala 480, Lys 482, Leu 513, Thr 528, Gln 529, Trp 530, Cys 531, and Phe 594 (chain A).

The oncogenic BRAF V600E crystal structure (PDB accession code: 3OG7) was used as a template for generating homology models for V-Raf-1 murine leukemia viral oncogene-like protein 1 isoform 3 (75.7% sequence similarity to BRAF V600E), RAF proto-oncogene serine/threonine-protein kinase (75.7% sequence similarity to BRAF V600E), highly similar to RAF proto-oncogene serine/threonine-protein kinase (75.5% sequence similarity to BRAF V600E), and kinase suppressor of Ras 2 to determine whether or not there were similarities in the I3C binding pocket across proteins that share a high sequence homology.

SiteMap from the Schrödinger suite [170] was employed to make predictions about potential ligand binding sites on the protein. After superimposing the homology models that exhibit approximately 75% sequence homology onto the oncogenic B-RAF V600E model in complex with Vemurafenib and I3C, it is evident that the predicted I3C binding site that is present on BRAF V600E is also observed in the other homology models examined as well. When the sequences of BRAF V600E and the remaining 4 homologous proteins were aligned, and the I3C binding sites were mapped onto the sequences, it is apparent that there are evolutionally conserved amino acid residues within each protein that may be important for anchoring I3C into its appropriate pocket. The predicted I3C binding sites for BRAF V600E and the homologous proteins that harbor ~75% sequence homology are positioned at the N-terminal region. Notably, residues such as lysine, leucine, glutamine, tryptophan, cysteine, and phenylalanine, and aspartate appear to be important for I3C binding. When the sequence of the kinase suppressor of Ras2, which shares 35.9% sequence homology, was aligned with BRAF V600E and the other proteins that exhibit relatively high sequence similarity, it was observed that many of the critical residues that appear to be significant for I3C binding were lost. At positions 663, 694, 711, 712, the lysine, leucine, glutamine, and tryptophan which occupy these sites in BRAF V600E and the other highly homologous proteins of greater than 75% sequence similarity, were substituted with arginine, valine, serine, and leucine respectively in the kinase suppressor of Ras2 sequence.

Wnt3A, Wnt5A, Wnt7B, and Wnt10B are all proteins expressed in melanoma and have been regarded as I3C target proteins. Generally, the wnt proteins that have been examined here are characterized by an evolutionarily conserved WNT1 family signature located at the C-terminal region and a signal peptide and a palmitoylation site

positioned at the N-terminal domain. Homology models were constructed by using the only crystallized Wnt structure, Xwnt8-Fz (PDB accession code: 4F0A) as the template structure. Between the 4 wnt proteins studied here, they all share approximately 40% sequence similarity. Despite the relatively low sequence similarity that each corresponding wnt protein shares with Wnt-8, derived from the xenopus species, (Wnt3A: 34.072%; Wnt7B: 32.493%; Wnt5A: 31.061%; Wnt10B: 25.183%) molecular docking simulations were still performed to gain a preliminary understanding of how I3C interacts with each protein and to compare similarities in I3C binding pockets within each wnt protein. In the xwnt8-frizzled-8 cysteine rich domain (Fz8-CRD) crystal structure (Janda *et al.*, 2012), wnt forms a two-domain structure that resembles a “hand” with a “thumb” and “index finger” unfurling out to bind to two points of contact on the Fz8-CRD. Docking studies using Glide, version 6.9 in Schrödinger revealed that I3C is predicted to bind to one of the two wnt/fz-CRD interfaces. One fz-CRD interaction site on wnt (from the xwnt8/fz-CRD crystal structure) is comprised of residues 301-338. These contact residues are highly conserved in all Wnts (Janda *et al.*, 2012). I3C is anchored to the apex of this contact region, engaging with hydrophobic residues Cys 297, Cys 311, and Phe 290. The binding of I3C to this region on wnt may prevent critical van der Waals interactions with main-chain and apolar residues on the Fz8-CRD. The predicted I3C binding site is also adjacent to a conserved patch of amino acid residues, particularly residues 216-219, 249-252, and 256-269. Because this region is away from the Fz binding site and is solvent exposed, this area serves as a potential binding site for additional co-receptors such as Lrp5/6 and/or Ryk which would allow the formation of a ternary complex with Fz. Ostensibly, the binding of 1-benzyl-I3C near this region may thus prevent the formation of a Wnt/Fz/LRP ternary complex.

After superimposing the wnt proteins onto one another, it is evident that they all share a common I3C binding site. Upon delineating the precise amino acid contact points on each wnt sequence, and then subsequently aligning the corresponding sequence between each other, it is evident that there are critical contact points that appear to be preserved throughout each wnt protein examined here. In particular, arginine in the N-terminal region, and glycine, threonine, and arginine in the C-terminal region are common residues that appear to anchor I3C securely into its pocket. However, when a protein of lower sequence similarity, namely, the putative uncharacterized protein (which exhibits 22.87% sequence identity to human wnt5a), was assessed for similarities in I3C binding pockets, it was apparent that critical I3C contact point residues that were present in the wnt proteins were lost in a protein of lesser homology. In particular, at positions 402, 403, and 406, a glycine, threonine, arginine were present in the wnt proteins, but in the ADAM23 protein, a serine, tyrosine, and glycine respectively substituted these corresponding residues.

The results presented here have provided invaluable insight regarding various I3C target proteins and have unlocked the potential for structure-based drug design. The quest to locate additional I3C target proteins has been facilitated through scrupulous structural comparisons of other homologous proteins and their corresponding indolecarbinol binding sites. The conservation of specific structural domains between homologous proteins can be subdivided into classes of similar folds, and as it would

appear, the fraction of protein folds employed by nature is narrow [210]. Because extensive efforts have been made to procure experimental structure data for a substantial portion of these possible protein folds, this has made homology-based modeling of protein structures feasible and it has progressively become a relevant equivalent to experimental structure determination. Understanding the specificities of indolecarbinol binding sites and gleaning the importance of particular amino acid residues through the comparisons of homologous proteins with various degrees of sequence similarity will improve the manner in which new I3C-derived compounds are synthesized.

References

1. Anfinsen CB: Principles That Govern Folding of Protein Chains. *Science* 1973, 181(4096):223-230.
2. Cavasotto CN, Phatak SS: Homology modeling in drug discovery: current trends and applications. *Drug Discov Today* 2009, 14(13-14):676-683.
3. Westbrook J, Feng ZK, Chen L, Yang HW, Berman HM: The Protein Data Bank and structural genomics. *Nucleic Acids Res* 2003, 31(1):489-491.
4. Tramontano A, Morea V: Assessment of homology-based predictions in CASP5 (vol 53, 352, 2003). *Proteins* 2004, 55(3):782-782.
5. Lesk AM, Chothia CH: The Response of Protein Structures to Amino-Acid-Sequence Changes. *Philos T R Soc A* 1986, 317(1540):345-356.
6. Venclovas C, Zemla A, Fidelis K, Moulton J: Assessment of progress over the CASP experiments. *Proteins* 2003, 53:585-595.
7. Xiang ZX: Advances in homology protein structure modeling. *Curr Protein Pept Sc* 2006, 7(3):217-227.
8. Fiser A, Sali A: ModLoop: automated modeling of loops in protein structures. *Bioinformatics* 2003, 19(18):2500-2501.
9. Sanchez R, Sali A: Comparative protein structure modeling as an optimization problem. *J Mol Struct-Theochem* 1997, 398:489-496.
10. Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T: Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protoc* 2009, 4(1):1-13.
11. Topham CM, Thomas P, Overington JP, Johnson MS, Eisenmenger F, Blundell TL: An Assessment of Composer - a Rule-Based Approach to Modeling Protein-Structure. *Biochem Soc Symp* 1990, 57:1-9.
12. Sali A, Blundell TL: Comparative Protein Modeling by Satisfaction of Spatial Restraints. *J Mol Biol* 1993, 234(3):779-815.
13. Peitsch MC: Protein Modeling by E-Mail. *Bio-Technol* 1995, 13(7):658-660.
14. Harrison RW, Chatterjee D, Weber IT: Analysis of six protein structures predicted by comparative modeling techniques. *Proteins-Structure Function and Genetics* 1995, 23(4):463-471.
15. Mosimann S, Meleshko R, James MNG: A Critical-Assessment of Comparative Molecular Modeling of Tertiary Structures of Proteins. *Proteins* 1995, 23(3):301-317.
16. Yang AS, Honig B: An integrated approach to the analysis and modeling of protein sequences and structures. III. A comparative study of sequence conservation in protein structural families using multiple structural alignments. *J Mol Biol* 2000, 301(3):691-711.
17. Sauder JM, Arthur JW, Dunbrack RL: Large-scale comparison of protein sequence alignment algorithms with structure alignments. *Proteins* 2000, 40(1):6-22.
18. Pieper U, Eswar N, Stuart AC, Ilyin VA, Sali A: MODBASE, a database of annotated comparative protein structure models. *Nucleic Acids Res* 2002, 30(1):255-259.

19. Kelley LA, MacCallum RM, Sternberg MJE: Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J Mol Biol* 2000, 299(2):499-520.
20. Teichmann SA, Chothia C, Gerstein M: Advances in structural genomics. *Curr Opin Struc Biol* 1999, 9(3):390-399.
21. Brenner SE, Chothia C, Hubbard TJP: Population statistics of protein structures: Lessons from structural classifications. *Curr Opin Struc Biol* 1997, 7(3):369-376.
22. Francoijs CJJ, Klomp JPG, Knegt RMA: Sequence annotation of nuclear receptor ligand-binding domains by automated homology modeling. *Protein Eng* 2000, 13(6):391-394.
23. Zhou YS, Johnson ME: Comparative molecular modeling analysis of 5-amidinoindole and benzimidazole binding to thrombin and trypsin: specific H-bond formation contributes to high 5-amidinoindole potency and selectivity for thrombin and factor Xa. *J Mol Recognit* 1999, 12(4):235-241.
24. Takeda-Shitaka M, Takaya D, Chiba C, Tanaka H, Umeyama H: Protein structure prediction in structure based drug design. *Curr Med Chem* 2004, 11(5):551-558.
25. Needleman SB, Wunsch CD: A General Method Applicable to Search for Similarities in Amino Acid Sequence of 2 Proteins. *J Mol Biol* 1970, 48(3):443-+.
26. Smith TF, Waterman MS: Identification of Common Molecular Subsequences. *J Mol Biol* 1981, 147(1):195-197.
27. Gotoh O: An Improved Algorithm for Matching Biological Sequences. *J Mol Biol* 1982, 162(3):705-708.
28. Taylor WR: Identification of Protein-Sequence Homology by Consensus Template Alignment. *J Mol Biol* 1986, 188(2):233-258.
29. Chappay C, Danckaert A, Dessen P, Hazout S: Mash - an Interactive Program for Multiple Alignment and Consensus Sequence Construction for Biological Sequences. *Comput Appl Biosci* 1991, 7(2):195-202.
30. Suyama M, Matsuo Y, Nishikawa K: Comparison of protein structures using 3D profile alignment. *J Mol Evol* 1997, 44:S163-S173.
31. Lolkema JS, Slotboom DJ: Estimation of structural similarity of membrane proteins by hydropathy profile alignment. *Mol Membr Biol* 1998, 15(1):33-42.
32. Barton GJ, Sternberg MJE: Flexible Protein-Sequence Patterns - a Sensitive Method to Detect Weak Structural Similarities. *J Mol Biol* 1990, 212(2):389-402.
33. Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997, 25(17):3389-3402.
34. Krogh A, Brown M, Mian IS, Sjolander K, Haussler D: Hidden Markov-Models in Computational Biology - Applications to Protein Modeling. *J Mol Biol* 1994, 235(5):1501-1531.
35. Mittelman D, Sadreyev R, Grishin N: Probabilistic scoring measures for profile-profile comparison yield more accurate short seed alignments. *Bioinformatics* 2003, 19(12):1531-1539.
36. Rychlewski L, Jaroszewski L, Li WZ, Godzik A: Comparison of sequence profiles. Strategies for structural predictions using sequence information. *Protein Sci* 2000, 9(2):232-241.

37. Yona G, Levitt M: Within the twilight zone: A sensitive profile-profile comparison tool based on information theory. *J Mol Biol* 2002, 315(5):1257-1275.
38. Ohlson T, Wallner B, Elofsson A: Profile-profile methods provide improved fold-recognition: A study of different profile-profile alignment methods. *Proteins* 2004, 57(1):188-197.
39. Matsuo Y, Bryant SH: Identification of homologous core structures. *Proteins-Structure Function and Genetics* 1999, 35(1):70-79.
40. Perola E, Walters WP, Charifson PS: A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins* 2004, 56(2):235-249.
41. Pickart CM: Back to the future with ubiquitin. *Cell* 2004, 116(2):181-190.
42. Petroski MD, Deshaies RJ: Mechanism of lysine 48-linked ubiquitin-chain synthesis by the cullin-RING ubiquitin-ligase complex SCF-Cdc34. *Cell* 2005, 123(6):1107-1120.
43. Huibregtse JM, Scheffner M, Beaudenon S, Howley PM: A Family of Proteins Structurally and Functionally Related to the E6-Ap Ubiquitin-Protein Ligase (Vol 92, Pg 2563, 1995). *P Natl Acad Sci USA* 1995, 92(11):5249-5249.
44. Joazeiro CAP, Weissman AM: RING finger proteins: Mediators of ubiquitin ligase activity. *Cell* 2000, 102(5):549-552.
45. Huibregtse JM, Scheffner M, Howley PM: Cloning and Expression of the Cdna for E6-Ap, a Protein That Mediates the Interaction of the Human Papillomavirus E6 Oncoprotein with P53. *Mol Cell Biol* 1993, 13(2):775-784.
46. Scheffner M, Nuber U, Huibregtse JM: Protein Ubiquitination Involving an E1-E2-E3 Enzyme Ubiquitin Thioester Cascade. *Nature* 1995, 373(6509):81-83.
47. Hatakeyama S, Yada M, Matsumoto M, Ishida N, Nakayama KI: U box proteins as a new family of ubiquitin-protein ligases. *J Biol Chem* 2001, 276(35):33111-33120.
48. Jiang JH, Ballinger CA, Wu YX, Dai Q, Cyr DM, Hohfeld J, Patterson C: CHIP is a U-box-dependent E3 ubiquitin ligase - Identification of Hsc70 as a target for ubiquitylation. *J Biol Chem* 2001, 276(46):42938-42944.
49. Boname JM, Stevenson PG: MHC class I ubiquitination by a viral PHD/LAP finger protein. *Immunity* 2001, 15(4):627-636.
50. Coscoy L, Sanchez DJ, Ganem D: A novel class of herpesvirus-encoded membrane-bound E3 ubiquitin ligases regulates endocytosis of proteins involved in immune recognition. *J Cell Biol* 2001, 155(7):1265-1273.
51. Lu ZM, Xu SC, Joazeiro C, Cobb MH, Hunter T: The PHD domain of MEKK1 acts as an E3 ubiquitin ligase and mediates ubiquitination and degradation of ERK1/2. *Mol Cell* 2002, 9(5):945-956.
52. Huang L, Kinnucan E, Wang GL, Beaudenon S, Howley PM, Huibregtse JM, Pavletich NP: Structure of an E6AP-UbcH7 complex: Insights into ubiquitination by the E2-E3 enzyme cascade. *Science* 1999, 286(5443):1321-1326.
53. Zheng N, Wang P, Jeffrey PD, Pavletich NP: Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases. *Cell* 2000, 102(4):533-539.

54. Harvey KF, Kumar S: Nedd4-like proteins: an emerging family of ubiquitin-protein ligases implicated in diverse cellular functions. *Trends Cell Biol* 1999, 9(5):166-169.
55. Plant PJ, Lafont F, Lecat S, Verkade P, Simons K, Rotin D: Apical membrane targeting of Nedd4 is mediated by an association of its C2 domain with annexin XIIIb. *J Cell Biol* 2000, 149(7):1473-1483.
56. Metzger MB, Hristova VA, Weissman AM: HECT and RING finger families of E3 ubiquitin ligases at a glance. *J Cell Sci* 2012, 125(3):531-537.
57. Scheffner M, Staub O: HECT E3s and human disease. *Bmc Biochem* 2007, 8.
58. Rotin D, Kumar S: Physiological functions of the HECT family of ubiquitin ligases. *Nat Rev Mol Cell Bio* 2009, 10(6):398-409.
59. Lin DYW, Diao JB, Chen J: Crystal structures of two bacterial HECT-like E3 ligases in complex with a human E2 reveal atomic details of pathogen-host interactions. *P Natl Acad Sci USA* 2012, 109(6):1925-1930.
60. Lin DYW, Diao JB, Zhou DG, Chen J: Biochemical and Structural Studies of a HECT-like Ubiquitin Ligase from *Escherichia coli* O157:H7. *J Biol Chem* 2011, 286(1):441-449.
61. Zhang Y, Higashide WM, McCormick BA, Chen J, Zhou DG: The inflammation-associated *Salmonella* SopA is a HECT-like E3 ubiquitin ligase. *Mol Microbiol* 2006, 62(3):786-793.
62. Verdecia MA, Joazeiro CAP, Wells NJ, Ferrer JL, Bowman ME, Hunter T, Noel JP: Conformational flexibility underlies ubiquitin ligation mediated by the WWP1HECT domain E3 ligase. *Mol Cell* 2003, 11(1):249-259.
63. Ogunjimi AA, Briant DJ, Pece-Barbara N, Le Roy C, Di Guglielmo GM, Kavsak P, Rasmussen RK, Seet BT, Sicheri F, Wrana JL: Regulation of Smurf2 ubiquitin ligase activity by anchoring the E2 to the HECT domain. *Mol Cell* 2005, 19(3):297-308.
64. Kamadurai HB, Souphron J, Scott DC, Duda DM, Miller DJ, Stringer D, Piper RC, Schulman BA: Insights into Ubiquitin Transfer Cascades from a Structure of a Ubch5B similar to Ubiquitin-HECTNEDD4L Complex. *Mol Cell* 2009, 36(6):1095-1102.
65. Shearwin-Whyatt L, Dalton HE, Foot N, Kumar S: Regulation of functional diversity within the Nedd4 family by accessory and adaptor proteins. *Bioessays* 2006, 28(6):617-628.
66. Shea FF, Rowell JL, Li YCW, Chang TH, Alvarez CE: Mammalian Alpha Arrestins Link Activated Seven Transmembrane Receptors to Nedd4 Family E3 Ubiquitin Ligases and Interact with Beta Arrestins. *Plos One* 2012, 7(12).
67. Han SO, Kommaddi RP, Venkataramanan V, Shenoy SK: beta-arrestin2 and ARRDC proteins have distinct roles in beta(2)AR trafficking and signaling. *Faseb J* 2012, 26.
68. Wiesner S, Ogunjimi AA, Wang HR, Rotin D, Sicheri F, Wrana JL, Forman-Kay JD: Autoinhibition of the HECT-Type ubiquitin ligase smurf2 through its c2 domain. *Cell* 2007, 130(4):651-662.
69. Gallagher E, Gao M, Liu YC, Karin M: Activation of the E3 ubiquitin ligase Itch through a phosphorylation-induced conformational change. *P Natl Acad Sci USA* 2006, 103(6):1717-1722.

70. Behrends C, Harper JW: Constructing and decoding unconventional ubiquitin chains. *Nat Struct Mol Biol* 2011, 18(5):520-528.
71. Komander D, Rape M: The Ubiquitin Code. *Annu Rev Biochem* 2012, 81:203-229.
72. Kulathu Y, Komander D: Atypical ubiquitylation - the unexplored world of polyubiquitin beyond Lys48 and Lys63 linkages. *Nat Rev Mol Cell Bio* 2012, 13(8):508-523.
73. Husnjak K, Dikic I: Ubiquitin-Binding Proteins: Decoders of Ubiquitin-Mediated Cellular Functions. *Annu Rev Biochem* 2012, 81:291-322.
74. Kumar S, Tomooka Y, Noda M: Identification of a Set of Genes with Developmentally down-Regulated Expression in the Mouse-Brain. *Biochem Bioph Res Co* 1992, 185(3):1155-1161.
75. Kumar S, Harvey KF, Kinoshita M, Copeland NG, Noda M, Jenkins NA: cDNA cloning, expression analysis, and mapping of the mouse Nedd4 gene (vol 40, pg 435, 1997). *Genomics* 1997, 44(1):156-156.
76. Yang B, Kumar S: Nedd4 and Nedd4-2: closely related ubiquitin-protein ligases with distinct physiological functions. *Cell Death Differ* 2010, 17(1):68-77.
77. Katz M, Shtiegman K, Tal-Or P, Yakir L, Mosesson Y, Harari D, Machluf Y, Asao H, Jovin T, Sugamura K *et al*: Ligand-independent degradation of epidermal growth factor receptor involves receptor ubiquitylation and hgs, an adaptor whose ubiquitin-interacting motif targets ubiquitylation by Nedd4. *Traffic* 2002, 3(10):740-751.
78. Cao XR, Lill NL, Boase N, Shi PP, Croucher DR, Shan HB, Qu J, Sweezer EM, Place T, Kirby PA *et al*: Nedd4 Controls Animal Growth by Regulating IGF-1 Signaling. *Sci Signal* 2008, 1(38).
79. Yang BL, Gay DL, MacLeod MKL, Cao X, Hala T, Sweezer EM, Kappler J, Marrack P, Oliver PM: Nedd4 augments the adaptive immune response by promoting ubiquitin-mediated degradation of Cbl-b in activated T cells. *Nat Immunol* 2008, 9(12):1356-1363.
80. Kawabe H, Neeb A, Dimova K, Young SM, Takeda M, Katsurabayashi S, Mitkovski M, Malakhova OA, Zhang DE, Umikawa M *et al*: Regulation of Rap2A by the ubiquitin ligase Nedd4-1 controls neurite development. *Neuroforum* 2010, 16(2):193-195.
81. Drinjakovic J, Jung HS, Campbell DS, Strohlic L, Dwivedy A, Holt CE: E3 Ligase Nedd4 Promotes Axon Branching by Downregulating PTEN. *Neuron* 2010, 65(3):341-357.
82. Liu Y, Oppenheim RW, Sugiura Y, Lin WC: Abnormal development of the neuromuscular junction in Nedd4-deficient mice. *Dev Biol* 2009, 330(1):153-166.
83. Nagpal P, Plant PJ, Correa J, Bain A, Takeda M, Kawabe H, Rotin D, Bain JR, Batt JAE: The Ubiquitin Ligase Nedd4-1 Participates in Denervation-Induced Skeletal Muscle Atrophy in Mice. *Plos One* 2012, 7(10).
84. Fouladkou F, Lu C, Jiang C, Zhou LM, She YM, Walls JR, Kawabe H, Brose N, Henkelman RM, Huang A *et al*: The Ubiquitin Ligase Nedd4-1 Is Required for Heart Development and Is a Suppressor of Thrombospondin-1. *J Biol Chem* 2010, 285(9):6770-6780.

85. Liu PY, Xu N, Malyukova A, Scarlett CJ, Sun YT, Zhang XD, Ling D, Su SP, Nelson C, Chang DK *et al*: The histone deacetylase SIRT2 stabilizes Myc oncoproteins. *Cell Death Differ* 2013, 20(3):503-514.
86. Yasuda J, Nakao M, Kawaoka Y, Shida H: Nedd4 regulates egress of Ebola virus-like particles from host cells. *J Virol* 2003, 77(18):9987-9992.
87. Timmins J, Schoehn G, Ricard-Blum S, Scianimanico S, Vernet T, Ruigrok RWH, Weissenhorn W: Ebola virus matrix protein VP40 interaction with human cellular factors Tsg101 and Nedd4. *J Mol Biol* 2003, 326(2):493-502.
88. Blot V, Perugi F, Gay B, Prevost MC, Briant L, Tangy F, Abriel H, Staub O, Dokhelar MC, Pique C: Nedd4.1-mediated ubiquitination and subsequent recruitment of Tsg101 ensure HTLV-1 Gag trafficking towards the multivesicular body pathway prior to virus budding. *J Cell Sci* 2004, 117(11):2357-2367.
89. Kikonyogo A, Bouamr F, Vana ML, Xiang Y, Aiyar A, Carter C, Leis J: Proteins related to the Nedd4 family of ubiquitin protein ligases interact with the L domain of Rous sarcoma virus and are required for gag budding from cells. *P Natl Acad Sci USA* 2001, 98(20):11199-11204.
90. Yasuda J, Hunter E, Nakao M, Shida H: Functional involvement of a novel Nedd4-like ubiquitin ligase on retrovirus budding. *Embo Rep* 2002, 3(7):636-640.
91. Bouamr F, Melillo JA, Wang MQ, Nagashima K, Los Santos MD, Rein A, Goff SP: PPPYEPTAP motif is the late domain of human T-Cell leukemia virus type 1 GaG and mediates its functional interaction with cellular proteins Nedd4 and Tsg101. *J Virol* 2003, 77(22):11882-11895.
92. Vana ML, Tang Y, Chen AP, Medina G, Carter C, Leis J: Role of Nedd4 and ubiquitination of Rous sarcoma virus gag in budding of virus-like particles from cells. *J Virol* 2004, 78(24):13943-13953.
93. Segura-Morales C, Pescia C, Chatellard-Causse C, Sadoul R, Bertrand E, Basyuk E: Tsg101 and Alix interact with murine leukemia virus Gag and cooperate with Nedd4 ubiquitin ligases during budding. *J Biol Chem* 2005, 280(29):27004-27012.
94. Ikeda A, Caldwell RG, Longnecker R, Ikeda M: Itchy, a Nedd4 ubiquitin ligase, downregulates latent membrane protein 2A activity in B-cell signaling. *J Virol* 2003, 77(9):5529-5534.
95. Borregaard N, Theilgaard-Monch K, Cowland JB, Stahle M, Sorensen OE: Neutrophils and keratinocytes in innate immunity - cooperative actions to provide antimicrobial defense at the right time and place. *J Leukocyte Biol* 2005, 77(4):439-443.
96. Korkmaz B, Moreau T, Gauthier F: Neutrophil elastase, proteinase 3 and cathepsin G: Physicochemical properties, activity and physiopathological functions. *Biochimie* 2008, 90(2):227-242.
97. Segal AW: How neutrophils kill microbes. *Annu Rev Immunol* 2005, 23:197-223.
98. Brinkmann V, Reichard U, Goosmann C, Fauler B, Uhlemann Y, Weiss DS, Weinrauch Y, Zychlinsky A: Neutrophil extracellular traps kill bacteria. *Science* 2004, 303(5663):1532-1535.
99. Papayannopoulos V, Zychlinsky A: NETs: a new strategy for using old weapons. *Trends Immunol* 2009, 30(11):513-521.

100. Lee WL, Downey GP: Leukocyte elastase - Physiological functions and role in acute lung injury. *Am J Resp Crit Care* 2001, 164(5):896-904.
101. Moraes TJ, Chow CW, Downey GP: Proteases and lung injury. *Crit Care Med* 2003, 31(4):S189-S194.
102. Horwitz M, Benson KF, Person RE, Aprikyan AG, Dale DC: Mutations in ELA2, encoding neutrophil elastase, define a 21-day biological clock in cyclic haematopoiesis. *Nat Genet* 1999, 23(4):433-436.
103. Horwitz MS, Duan ZJ, Korkmaz B, Lee HH, Mealiffe ME, Salipante SJ: Neutrophil elastase in cyclic and severe congenital neutropenia. *Blood* 2007, 109(5):1817-1824.
104. Jenne DE, Tschopp J, Ludemann J, Utecht B, Gross WL: Wegener Autoantigen Decoded. *Nature* 1990, 346(6284):520-520.
105. Pham CTN, Ley TJ: Dipeptidyl peptidase I is required for the processing and activation of granzymes A and B in vivo. *P Natl Acad Sci USA* 1999, 96(15):8627-8632.
106. Adkison AM, Raptis SZ, Kelley DG, Pham CTN: Dipeptidyl peptidase I activates neutrophil-derived serine proteases and regulates the development of acute experimental arthritis. *J Clin Invest* 2002, 109(3):363-371.
107. Hart TC, Hart PS, Bowden DW, Michalec MD, Callison SA, Walker SJ, Zhang YZ, Firatli E: Mutations of the cathepsin C gene are responsible for Papillon-Lefevre syndrome. *J Med Genet* 1999, 36(12):881-887.
108. Toomes C, James J, Wood AJ, Wu CL, McCormick D, Lench N, Hewitt C, Moynihan L, Roberts E, Woods CG *et al*: Loss-of-function mutations in the cathepsin C gene result in periodontal disease and palmoplantar keratosis. *Nat Genet* 1999, 23(4):421-424.
109. Kobayashi SD, Voyich JM, Burlak C, DeLeo FR: Neutrophils in the innate immune response. *Arch Immunol Ther Ex* 2005, 53(6):505-517.
110. Belaouaj AA, Kim KS, Shapiro SD: Degradation of outer membrane protein A in *Escherichia coli* killing by neutrophil elastase. *Science* 2000, 289(5482):1185-1187.
111. Weinrauch Y, Drujan D, Shapiro SD, Weiss J, Zychlinsky A: Neutrophil elastase targets virulence factors of enterobacteria. *Nature* 2002, 417(6884):91-94.
112. Zasloff M: Antimicrobial peptides in health and disease. *New Engl J Med* 2002, 347(15):1199-1200.
113. Ossovskaya VS, Bunnett NW: Protease-activated receptors: Contribution to physiology and disease. *Physiol Rev* 2004, 84(2):579-621.
114. Vergnolle N: Protease-activated receptors as drug targets in inflammation and pain. *Pharmacol Therapeut* 2009, 123(3):292-309.
115. Renesto P, Si-Tahar M, Moniatte M, Balloy V, VanDorselaer A, Pidard D, Chignard M: Specific inhibition of thrombin-induced cell activation by the neutrophil proteinases elastase, cathepsin G, and proteinase 3: Evidence for distinct cleavage sites within the aminoterminal domain of the thrombin receptor. *Blood* 1997, 89(6):1944-1953.
116. Bode W, Wei AZ, Huber R, Meyer E, Travis J, Neumann S: X-Ray Crystal-Structure of the Complex of Human-Leukocyte Elastase (Pmn Elastase) and the 3rd Domain of the Turkey Ovomuroid Inhibitor. *Embo J* 1986, 5(10):2453-2458.

117. Fujinaga M, Chernaia MM, Halenbeck R, Koths K, James MNG: The crystal structure of PR3, a neutrophil serine proteinase antigen of Wegener's granulomatosis antibodies. *J Mol Biol* 1996, 261(2):267-278.
118. Hof P, Mayr I, Huber R, Korzus E, Potempa J, Travis J, Powers JC, Bode W: The 1.8 angstrom crystal structure of human cathepsin G in complex with Suc-Val-Pro-Phe(P)-(OPh)(2): A Janus-faced proteinase with two opposite specificities. *Embo J* 1996, 15(20):5481-5491.
119. Schechter I, Berger A: On Size of Active Site in Proteases .I. Papain. *Biochem Biophys Res Co* 1967, 27(2):157-+.
120. Almeida RP, Melchior M, Campanelli D, Nathan C, Gabay JE: Complementary-DNA Sequence of Human Neutrophil Azurocidin, an Antibiotic with Extensive Homology to Serine Proteases. *Biochem Biophys Res Co* 1991, 177(2):688-695.
121. Morgan JG, Sukiennicki T, Pereira HA, Spitznagel JK, Guerra ME, Larrick JW: Cloning of the Cdna for the Serine Protease Homolog Cap37 Azurocidin, a Microbicidal and Chemotactic Protein from Human Granulocytes. *J Immunol* 1991, 147(9):3210-3214.
122. Watorek W: Azurocidin - inactive serine proteinase homolog acting as a multifunctional inflammatory mediator. *Acta Biochim Pol* 2003, 50(3):743-752.
123. Iversen PO, Lewis ID, Turczynowicz S, Hasle H, Niemeyer C, Schmiegelow K, Bastiras S, Biondi A, Hughes TP, Lopez AF: Inhibition of granulocyte-macrophage colony-stimulating factor prevents dissemination and induces remission of juvenile myelomonocytic leukemia in engrafted immunodeficient mice. *Blood* 1997, 90(12):4910-4917.
124. Lindmark A, Garwicz D, Rasmussen PB, Flodgaard H, Gullberg U: Characterization of the biosynthesis, processing, and sorting of human HBP/CAP37/azurocidin. *J Leukocyte Biol* 1999, 66(4):634-643.
125. Witko-Sarsat V, Halbwachs-Mecarelli L, Schuster A, Nusbaum P, Ueki I, Canteloup S, Lenoir G, Descamps-Latscha B, Nadel JA: Proteinase 3, a potent secretagogue in airways, is present in cystic fibrosis sputum. *Am J Resp Cell Mol* 1999, 20(4):729-736.
126. Tapper H, Furuya W, Grinstein S: Localized exocytosis of primary (lysosomal) granules during phagocytosis: Role of Ca²⁺-dependent tyrosine phosphorylation and microtubules. *J Immunol* 2002, 168(10):5287-5296.
127. Mayet WJ, Csernok E, Szymkowiak C, Gross WL, Zumbuschfeld KHM: Human Endothelial-Cells Express Proteinase-3, the Target Antigen of Anticytoplasmic Antibodies in Wegeners Granulomatosis. *Blood* 1993, 82(4):1221-1229.
128. Lee TD, Gonzalez ML, Kumar P, Chary-Reddy S, Grammas P, Pereira HA: CAP37, a novel inflammatory mediator - Its expression in endothelial cells and localization to atherosclerotic lesions. *Am J Pathol* 2002, 160(3):841-848.
129. Ruan X, Chodosh J, Callegan MC, Booth MC, Lee TD, Kumar P, Gilmore MS, Pereira HA: Corneal expression of the inflammatory mediator CAP37. *Invest Ophth Vis Sci* 2002, 43(5):1414-1421.
130. Kirkwood JM, Strawderman MH, Ernstoff MS, Smith TJ, Borden EC, Blum RH: Interferon alfa-2b adjuvant therapy of high-risk resected cutaneous melanoma:

- The Eastern Cooperative Oncology Group trial EST 1684. *J Clin Oncol* 1996, 14(1):7-17.
131. Garnett MJ, Rana S, Paterson H, Barford D, Marais R: Wild-type and mutant B-RAF activate C-RAF through distinct mechanisms involving heterodimerization. *Mol Cell* 2005, 20(6):963-969.
 132. Rushworth LK, Hindley AD, O'Neill E, Kolch W: Regulation and role of Raf-1/B-Raf heterodimerization. *Mol Cell Biol* 2006, 26(6):2262-2272.
 133. Rajakulendran T, Sahmi M, Lefrancois M, Sicheri F, Therrien M: A dimerization-dependent mechanism drives RAF catalytic activation. *Nature* 2009, 461(7263):542-U114.
 134. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, Garnett MJ, Bottomley W *et al*: Mutations of the BRAF gene in human cancer. *Nature* 2002, 417(6892):949-954.
 135. Maurer G, Tarkowski B, Baccarini M: Raf kinases in cancer-roles and therapeutic opportunities. *Oncogene* 2011, 30(32):3477-3488.
 136. Cantwell-Dorris ER, O'Leary JJ, Sheils OM: BRAF(V600E): Implications for Carcinogenesis and Molecular Therapy. *Mol Cancer Ther* 2011, 10(3):385-394.
 137. Wan PTC, Garnett MJ, Roe SM, Lee S, Niculescu-Duvaz D, Good VM, Jones CM, Marshall CJ, Springer CJ, Barford D *et al*: Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. *Cell* 2004, 116(6):855-867.
 138. Garnett MJ, Marais R: Guilty as charged: B-RAF is a human oncogene. *Cancer Cell* 2004, 6(4):313-319.
 139. Eggermont AMM, Robert C: New drugs in melanoma: It's a whole new world. *Eur J Cancer* 2011, 47(14):2150-2157.
 140. Wilhelm SM, Carter C, Tang LY, Wilkie D, McNabola A, Rong H, Chen C, Zhang XM, Vincent P, McHugh M *et al*: BAY 43-9006 exhibits broad spectrum oral antitumor activity and targets the RAF/MEK/ERK pathway and receptor tyrosine kinases involved in tumor progression and angiogenesis. *Cancer Res* 2004, 64(19):7099-7109.
 141. Flaherty KT: BRAF Inhibitors and Melanoma. *Cancer J* 2011, 17(6):505-511.
 142. Sosman JA, Kim KB, Schuchter L, Gonzalez R, Pavlick AC, Weber JS, McArthur GA, Hutson TE, Moschos SJ, Flaherty KT *et al*: Survival in BRAF V600-Mutant Advanced Melanoma Treated with Vemurafenib. *New Engl J Med* 2012, 366(8):707-714.
 143. Ascierto PA, Kirkwood JM, Grob JJ, Simeone E, Grimaldi AM, Maio M, Palmieri G, Testori A, Marincola FM, Mozzillo N: The role of BRAF V600 mutation in melanoma. *J Transl Med* 2012, 10.
 144. Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, Dummer R, Garbe C, Testori A, Maio M *et al*: Improved Survival with Vemurafenib in Melanoma with BRAF V600E Mutation. *New Engl J Med* 2011, 364(26):2507-2516.
 145. Kim KB, Flaherty KT, Chapman PB, Sosman JA, Ribas A, McArthur GA, Amaravadi RK, Lee RJ, Nolop KB, Puzanov I: Pattern and outcome of disease progression in phase I study of vemurafenib in patients with metastatic melanoma (MM). *J Clin Oncol* 2011, 29(15).

146. McArthur GA, Ribas A, Chapman PB, Flaherty KT, Kim KB, Puzanov I, Nathanson KL, Lee RJ, Koehler A, Spleiss O *et al*: Molecular analyses from a phase I trial of vemurafenib to study mechanism of action (MOA) and resistance in repeated biopsies from BRAF mutation-positive metastatic melanoma patients (pts). *J Clin Oncol* 2011, 29(15).
147. Villanueva J, Vultur A, Lee JT, Somasundaram R, Fukunaga-Kalabis M, Cipolla AK, Wubbenhorst B, Xu XW, Gimotty PA, Kee D *et al*: Acquired Resistance to BRAF Inhibitors Mediated by a RAF Kinase Switch in Melanoma Can Be Overcome by Cotargeting MEK and IGF-1R/PI3K. *Cancer Cell* 2010, 18(6):683-695.
148. Lacouture ME, O'Reilly K, Rosen N, Solit DB: Induction of Cutaneous Squamous Cell Carcinomas by RAF Inhibitors: Cause for Concern? *J Clin Oncol* 2012, 30(3):329-330.
149. Polakis P: Wnt Signaling in Cancer. *Csh Perspect Biol* 2012, 4(5).
150. Clements WM, Groden J, Lowy AM: Wnt signaling increases invasiveness of gastric and pancreatic cancer cells independent of MMP-7. *Ann Surg Oncol* 2003, 10(1):S26-S26.
151. Lammi L, Arte S, Somer M, Jarvinen H, Lahermo P, Thesleff I, Pirinen S, Nieminen P: Mutations in AXIN2 cause familial tooth agenesis and predispose to colorectal cancer. *Am J Hum Genet* 2004, 74(5):1043-1050.
152. Polakis P: The many ways of Wnt in cancer. *Curr Opin Genet Dev* 2007, 17(1):45-51.
153. Anastas JN, Moon RT: WNT signalling pathways as therapeutic targets in cancer. *Nat Rev Cancer* 2013, 13(1):11-26.
154. Pinson KI, Brennan J, Monkley S, Avery BJ, Skarnes WC: An LDL-receptor-related protein mediates Wnt signalling in mice. *Nature* 2000, 407(6803):535-538.
155. Clevers H: Wnt/beta-catenin signaling in development and disease. *Cell* 2006, 127(3):469-480.
156. Weeraratna AT: A wnt-er wonderland - The complexity of wnt signaling in melanoma. *Cancer Metast Rev* 2005, 24(2):237-250.
157. Kuhl M, Sheldahl LC, Park M, Miller JR, Moon RT: The Wnt/Ca²⁺ pathway - a new vertebrate Wnt signaling pathway takes shape. *Trends Genet* 2000, 16(7):279-283.
158. Dissanayake SK, Wade M, Johnson CE, O'Connell MP, Leotlela PD, French AD, Shah KV, Hewitt KJ, Rosenthal DT, Indig FE *et al*: The Wnt5A/protein kinase C pathway mediates motility in melanoma cells via the inhibition of metastasis suppressors and initiation of an epithelial to mesenchymal transition. *J Biol Chem* 2007, 282(23):17259-17271.
159. Bazan JF, de Sauvage FJ: Structural Ties between Cholesterol Transport and Morphogen Signaling. *Cell* 2009, 138(6):1055-1056.
160. Hausmann G, Basler K: Wnt lipid modifications: Not as saturated as we thought. *Dev Cell* 2006, 11(6):751-752.
161. Mulligan KA, Fuerer C, Ching W, Fish M, Willert K, Nusse R: Secreted Wingless-interacting molecule (Swim) promotes long-range signaling by maintaining Wingless solubility. *P Natl Acad Sci USA* 2012, 109(2):370-377.

162. Janda CY, Waghray D, Levin AM, Thomas C, Garcia KC: Structural Basis of Wnt Recognition by Frizzled. *Science* 2012, 337(6090):59-64.
163. Kawano Y, Kypta R: Secreted antagonists of the Wnt signalling pathway. *J Cell Sci* 2003, 116(13):2627-2634.
164. Hsieh JC, Kodjabachian L, Rebbert ML, Rattner A, Smallwood PM, Samos CH, Nusse R, Dawid IB, Nathans J: A new secreted protein that binds to Wnt proteins and inhibits their activities. *Nature* 1999, 398(6726):431-436.
165. Apweiler R, Martin MJ, O'Donovan C, Magrane M, Alam-Faruque Y, Antunes R, Casanova EB, Bely B, Bingley M, Bower L *et al*: Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res* 2012, 40(D1):D71-D75.
166. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic Local Alignment Search Tool. *J Mol Biol* 1990, 215(3):403-410.
167. Rost B: Twilight zone of protein sequence alignments. *Protein Eng* 1999, 12(2):85-94.
168. Aronchik I, Kundu A, Quirrit JG, Firestone GL: The Antiproliferative Response of Indole-3-Carbinol in Human Melanoma Cells Is Triggered by an Interaction with NEDD4-1 and Disruption of Wild-Type PTEN Degradation. *Mol Cancer Res* 2014, 12(11):1621-1634.
169. Kathman SG, Span I, Smith AT, Xu ZY, Zhan J, Rosenzweig AC, Statsyuk AV: A Small Molecule That Switches a Ubiquitin Ligase From a Processive to a Distributive Enzymatic Mechanism. *J Am Chem Soc* 2015, 137(39):12442-12445.
170. Halgren T: New method for fast and accurate binding-site identification and analysis. *Chem Biol Drug Des* 2007, 69(2):146-148.
171. Nguyen HH, Aronchik I, Brar GA, Nguyen DHH, Bjeldanes LF, Firestone GL: The dietary phytochemical indole-3-carbinol is a natural elastase enzymatic inhibitor that disrupts cyclin E protein processing. *P Natl Acad Sci USA* 2008, 105(50):19750-19755.
172. Kundu A QJ, Khouri MG, Firestone GL: Inhibition of oncogenic BRAF activity by indole-3-carbinol disrupts microphthalmia-associated transcription factor expression and arrests melanoma cell proliferation. *Mol Carcinog* 2016.
173. Klaus A, Birchmeier W: Wnt signalling and its impact on development and cancer. *Nat Rev Cancer* 2008, 8(5):387-398.
174. O'Connell MP WA: Hear the Wnt Ror: how melanoma cells adjust to changes in Wnt. *Pigment Cell Melanoma Res* 2009, 6:724-739.
175. Larue L, Luciani F, Kumasaka M, Champeval D, Demirkan N, Bonaventure J, Delmas V: Bypassing melanocyte senescence by beta-catenin: A novel way to promote melanoma. *Pathol Biol* 2009, 57(7-8):543-547.
176. Pham K, Milovanovic T, Barr RJ, Truong T, Holcombe RF: Wnt ligand expression in malignant melanoma: pilot study indicating correlation with histopathological features. *J Clin Pathol-Mol Pa* 2003, 56(5):280-285.
177. Komekado H, Yamamoto H, Chiba T, Kikuchi A: Glycosylation and palmitoylation of Wnt-3a are coupled to produce an active form of Wnt-3a. *Genes Cells* 2007, 12(4):521-534.
178. Hillisch A, Pineda LF, Hilgenfeld R: Utility of homology models in the drug discovery process. *Drug Discov Today* 2004, 9(15):659-669.

179. Ferrara P, Jacoby E: Evaluation of the utility of homology models in high throughput docking. *J Mol Model* 2007, 13(8):897-905.
180. Autin L, Steen M, Dahlback B, Villoutreix BO: Proposed structural models of the prothrombinase (FXa-FVa) complex. *Proteins* 2006, 63(3):440-450.
181. Song N, Sedgewick RD, Durand D: Domain architecture comparison for multidomain homology identification. *J Comput Biol* 2007, 14(4):496-516.
182. Sheng YH, Sali A, Herzog H, Lahnstein J, Krilis SA: Site-directed mutagenesis of recombinant human beta(2)-glycoprotein I identifies a cluster of lysine residues that are critical for phospholipid binding and anti-cardiolipin antibody activity. *J Immunol* 1996, 157(8):3744-3751.
183. Luther KB, Schindelin H, Haltiwanger RS: Structural and Mechanistic Insights into Lunatic Fringe from a Kinetic Analysis of Enzyme Mutants. *J Biol Chem* 2009, 284(5):3294-3305.
184. Clem B, Telang S, Clem A, Yalcin A, Meier J, Simmons A, Rasku MA, Arumugam S, Dean WL, Eaton J *et al*: Small-molecule inhibition of 6-phosphofructo-2-kinase activity suppresses glycolytic flux and tumor growth. *Mol Cancer Ther* 2008, 7(1):110-120.
185. Cozza G, Gianoncelli A, Montopoli M, Caparrotta L, Venerando A, Meggio F, Pinna LA, Zagotto G, Moro S: Identification of novel protein kinase CK1 delta (CK1 delta) inhibitors through structure-based virtual screening. *Bioorg Med Chem Lett* 2008, 18(20):5672-5675.
186. Hellmuth K, Grosskopf S, Lum CT, Wurtele M, Roder N, von Kries JP, Rosario M, Rademann J, Birchmeier W: Specific inhibitors of the protein tyrosine phosphatase Shp2 identified by high-throughput docking. *P Natl Acad Sci USA* 2008, 105(20):7275-7280.
187. Nguyen TL, Gussio R, Smith JA, Lannigan DA, Hecht SM, Scudiero DA, Shoemaker RH, Zaharevitz DW: Homology model of RSK2 N-terminal kinase domain, structure-based identification of novel RSK2 inhibitors, and preliminary common pharmacophore. *Bioorgan Med Chem* 2006, 14(17):6097-6105.
188. Park H, Bahn YJ, Jung SK, Jeong DG, Lee SH, Yoon TS, Kim SJ, Ryu SE: Discovery of novel Cdc25 phosphatase inhibitors with micromolar activity based on the structure-based virtual screening. *J Med Chem* 2008, 51(18):5533-5541.
189. Kruger FA, Overington JP: Global Analysis of Small Molecule Binding to Related Protein Targets. *Plos Comput Biol* 2012, 8(1).
190. Capra JA, Singh M: Predicting functionally important residues from sequence conservation. *Bioinformatics* 2007, 23(15):1875-1882.
191. Zhang T, Zhang H, Chen K, Shen S, Ruan J, Kurgan L: Accurate sequence-based prediction of catalytic residues. *Bioinformatics* 2008, 24(20):2329-2338.
192. Fischer JD, Mayer CE, Soding J: Prediction of protein functional residues from sequence by probability density estimation. *Bioinformatics* 2008, 24(5):613-620.
193. Ota M, Kinoshita K, Nishikawa K: Prediction of catalytic residues in enzymes based on known tertiary structure, stability profile, and sequence conservation. *J Mol Biol* 2003, 327(5):1053-1064.
194. Burgoyne NJ, Jackson RM: Predicting protein interaction sites: binding hot-spots in protein-protein and protein-ligand interfaces. *Bioinformatics* 2006, 22(11):1335-1342.

195. Liang SD, Zhang C, Liu S, Zhou YQ: Protein binding site prediction using an empirical scoring function. *Nucleic Acids Res* 2006, 34(13):3698-3707.
196. Campbell SJ, Gold ND, Jackson RM, Westhead DR: Ligand binding: functional site location, similarity and docking. *Curr Opin Struc Biol* 2003, 13(3):389-395.
197. Thibert B, Bredesen DE, del Rio G: Improved prediction of critical residues for protein function based on network and phylogenetic analyses. *Bmc Bioinformatics* 2005, 6.
198. Bray T, Chan P, Bougouffa S, Greaves R, Doig AJ, Warwicker J: SitesIdentify: a protein functional site prediction tool. *Bmc Bioinformatics* 2009, 10.
199. Brylinski M, Prymula K, Jurkowski W, Kochanczyk M, Stawowczyk E, Konieczny L, Roterman I: Prediction of functional sites based on the fuzzy oil drop model. *Plos Comput Biol* 2007, 3(5):909-923.
200. Jones S, Thornton JM: Analysis of protein-protein interaction sites using surface patches. *J Mol Biol* 1997, 272(1):121-132.
201. Landgraf R, Xenarios I, Eisenberg D: Three-dimensional cluster analysis identifies interfaces and functional residue clusters in proteins. *J Mol Biol* 2001, 307(5):1487-1502.
202. Pazos F, Valencia A: In silico two-hybrid system for the selection of physically interacting protein pairs. *Proteins-Structure Function and Genetics* 2002, 47(2):219-227.
203. Teichmann SA, Murzin AG, Chothia C: Determination of protein function, evolution and interactions by structural genomics. *Curr Opin Struc Biol* 2001, 11(3):354-363.
204. Nayeem A, Sitkoff D, Krystek S: A comparative study of available software for high-accuracy homology modeling: From sequence alignments to structural models. *Protein Sci* 2006, 15(4):808-824.
205. Bernacki K, Kalyanaraman C, Chorny I, Jacobson MP: Improving the quality of virtual ligand screening against homology models. *Abstr Pap Am Chem S* 2005, 229:U803-U803.
206. Ring CS, Sun E, Mckerrow JH, Lee GK, Rosenthal PJ, Kuntz ID, Cohen FE: Structure-Based Inhibitor Design by Using Protein Models for the Development of Antiparasitic Agents. *P Natl Acad Sci USA* 1993, 90(8):3583-3587.
207. Schafferhans A, Klebe G: Docking ligands onto binding site representations derived from proteins built by homology modelling. *J Mol Biol* 2001, 307(1):407-427.
208. Schwarz SE, Rosa JL, Scheffner M: Characterization of human hect domain family members and their interaction with Ubch5 and Ubch7. *J Biol Chem* 1998, 273(20):12148-12154.
209. Bollag G, Tsai J, Zhang JZ, Zhang C, Ibrahim P, Nolop K, Hirth P: Vemurafenib: the first drug approved for BRAF-mutant cancer. *Nat Rev Drug Discov* 2012, 11(11):873-886.
210. Govindarajan S, Recabarren R, Goldstein RK: Estimating the total number of protein folds. *Proteins-Structure Function and Genetics* 1999, 35(4):408-414.

Chapter 4

Conclusions and Future Directions

Despite comprehensive biomedical research endeavors, the treatment alternatives for a variety of pernicious cancers still remain constricted and the flow through of the current drug development pipeline has been receding which in part can be attributed to ongoing drug discovery efforts aimed at targeting previously validated 'druggable' protein families [1]. Conceivably, this leaves an ample fraction of proteins unexploited by cancer drugs. Hence, this ultimately calls for the identification and substantiation of new cancer-relevant targets.

Intravenous cytotoxic chemotherapy has for decades been the hallmark for treating cancer. Rapidly dividing cells, which include both cancer cells and certain normal tissues, have been the target for these chemotherapeutic drugs and consequently, many patients must endure the classic toxicities of alopecia, gastrointestinal symptoms, and myelosuppression [2]. However, within the past decade, a dramatic transition in cancer therapies transpired as targeted therapies emerged as an alternative treatment for various types of cancers which include breast, colorectal, lung, pancreatic, lymphoma leukemia, and multiple myeloma.

The two primary modes of targeted therapy include monoclonal antibodies and small molecule inhibitors. Because of their distinct mechanisms of actions and toxicities, targeted therapies have broadened the approach of individually personalized cancer treatment. The expression of specific molecular targets in different cancers has paved the way for customizing treatment for each patient. Understanding the toxicities and potential drug interactions that are linked to targeted cancer therapies has become increasingly crucial.

Molecular targeted therapies work by suppressing critical biochemical pathways or mutant proteins that are necessary for tumor cell growth and survival [3]. Within a molecularly defined cluster of patients, these drugs have the capacity to arrest tumor progression and can impel dramatic regressions.

Identifying novel classes of highly potent therapeutic agents that specifically act on molecular targets with diminished side effects after continued treatment has been a challenging obstacle to overcome in the treatment of cancer.

Among the array of molecular targeted agents employed, indole-3-carbinol (I3C) has emerged as a promising natural compound that has been demonstrated to inhibit tumor cell proliferation in many various cancer cell types including, breast, prostate, endometrial, colon and leukemia cells [4], can engender a G1/S cell cycle arrest, and can promote apoptosis *in vitro*. Moreover, studies where human reproductive cancer cells were treated with I3C revealed that it can induce or attenuate distinct transcriptional, signal transduction, and metabolic cascades that instigate a cell cycle arrest, apoptosis, constrict cancer cell migration, and modulation of hormone receptor signaling [4-9]. Previously, our lab was the first to establish the direct inhibition of elastase enzymatic activity, which causes the indole-dependent disruption of cyclin E processing in human breast cancer cells [10]. In spite of the cogent cellular and physiological evidence that showcase the anti-cancer effects of I3C, there are still

complications that restrict I3C's potential as a therapeutic alternative for treating cancer. I3C lacks stability in cells and is promptly transformed into a variety of products *in vivo* [11] which have been shown to have specific biological effects on breast cancer cells despite some shared targets that I3C and its corresponding bi-products may exhibit [5, 8]. It has been estimated that approximately 20% of I3C is transformed into DIM, the predominant condensation product, and effectively a fraction of these acid-catalyzed reaction bi-products themselves confer specific anti-carcinogenic and anti-proliferative responses. Because both the conditions under which I3C can maintain stability and the precise ratio of conversion of I3C into its condensation products are not clear, estimating an appropriate dose for I3C *in vivo* will be difficult to reconcile. Moreover, in an effort to inhibit the growth of human breast cancer cells, treatment required relatively high doses within 50-100 μM range [12-16]. In spite of this, however, I3C's efficacy in exerting control over the expression and activity of distinct cell cycle components in various cancer cell types alludes to the possibility that this phytochemical would be an instrumental compound for synthesizing novel, more potent anti-cancer compounds. Thorough analysis of the I3C structure-activity highlighted the significance of adding hydrophobic alkyl substituents to the nitrogen position 1 in the indole ring whereby the growth arrest response was enhanced by several hundred fold [17]. With this understanding, our lab has reported 1-benzyl-I3C as the most effective I3C-derived synthetic analog to date with a 1000-fold elevated effectiveness for growth arrest of human estrogen responsive and estrogen-independent human breast cancer cells. For *in vivo* growth of breast cancer cell xenografts in athymic mice, 1-benzyl-I3C also exhibits stronger effects. Given these findings, this prompted the question as to whether or not we can use I3C and 1-benzyl-I3C as a scaffold for generating stronger, more efficacious, cancer therapeutic agents.

Cementing human neutrophil elastase as an I3C target protein and demonstrating that 1-benzyl-I3C acts as a more potent inhibitor in suppressing elastase enzymatic activity incited the pursuit for additional target proteins in human cancer cells that do not express elastase and require it for cellular proliferation. Recently our lab has observed that I3C instigates an apoptotic response in human melanoma cells and through a series of cellular and biochemical studies, has proven that I3C binds and inhibits the E3 ubiquitin ligase which is responsible for mediating AKT-1-dependent cell survival cascades in human melanoma cells. Given this information, we have inaugurated a novel set of I3C and 1-benzyl-I3C synthetic analogs and have demonstrated that these compounds are effective inhibitors of both NEDD4-1 ubiquitin ligase and elastase activity and the I3C indolecarbinol structure serves as a novel arena for small molecule inhibitors for HECT-domain containing ubiquitin ligases and elastase. Collectively, our *in vitro* binding and inhibition of enzymatic activity of purified NEDD4-1 and elastase by I3C, 1-benzyl-I3C, and of the five additional synthetic indolecarbinol analogs exhibit selective modifications in the physical and chemical properties of the resulting chemical structure. The increased effectiveness of the indolecarbinol compounds to inhibit both NEDD4-1 and elastase activity required an optimal nucleophilic π system within the benzene ring. A loss of flexibility of the chemical scaffold had subsequently diminished the efficacy of the indolecarbinol compounds. For each of the synthetic indolecarbinol analogs, it was essential to maintain the C-3 hydroxy methyl substituent for the

biological activity of the I3C parental compound [17]. Moreover, the effectiveness of the anti-proliferative effects in melanoma cells and breast cancer cells by the examined indolecarbinol compounds generally correlated with their *in vitro* inhibition of NEDD4-1 ubiquitin ligase and elastase activity, respectively, thus corroborating the significance of the indolecarbinol structure scaffold in synthesizing stable and highly potent anti-cancer compounds.

Examining the binding modes of I3C, 1-benzyl-I3C, and their corresponding synthetic derivatives in adducts with their target proteins, namely, elastase, NEDD4-1, oncogenic BRAF V600E serine/threonine kinase, and wnt, can illuminate various patterns in binding sites and can hence provide valuable information for locating additional indolecarbinol target proteins. Homology modeling of the four target proteins were employed to highlight similarities in protein folds. Subsequently, SiteMap was used to scan each protein model and make predictions based on shape, electrostatics, and hydrophobicity, about where each indolecarbinol compound would bind. After delineating the precise binding sites on each homology model, the sequences of the homologous proteins were aligned to assess patterns in binding sites. It became evident that proteins of higher homology (generally greater than 50%) to their respective target protein, exhibited a greater conservation of I3C contact points than proteins of lower homology (generally below 30%). Conceivably, information garnered from these studies will be useful in making predictions about identifying additional indolecarbinol compound target proteins.

Through a systematic structure-function approach, we were able to demonstrate that I3C, 1-benzyl-I3C, and their synthetic analogs are able to bind, inhibit, and hence trigger multiple anti-proliferative cascades in both melanoma and breast cancer. With the similarities in binding sites between the homologous proteins of their corresponding target proteins, there is still a comprehensive array of studies that need to be executed to unlock the potential of these indolecarbinol compounds. It would be advantageous to test the effects of these compounds on additional highly homologous proteins and determine if they also have the capacity to not only bind, but suppress the enzymatic activity of each newly examined target protein. The results presented here provide the fundamental framework for understanding the mode of inhibition of natural and synthetic indolecarbinol compounds and the proteins they effectively target.

References

1. Rask-Andersen M, Almen MS, Schioth HB: Trends in the exploitation of novel drug targets. *Nat Rev Drug Discov* 2011, 10(8):579-590.
2. Gerber DE: Targeted therapies: A new generation of cancer treatments. *Am Fam Physician* 2008, 77(3):311-319.
3. Druker BJ: Imatinib as a paradigm of targeted therapies. *J Clin Oncol* 2003, 21(23):239s-245s.
4. Aggarwal BB, Ichikawa H: Molecular targets and anticancer potential of indole-3-carbinol and its derivatives. *Cell Cycle* 2005, 4(9):1201-1215.
5. Safe S, Papineni S, Chintharlapalli S: Cancer chemotherapy with indole-3-carbinol, bis(3'-indolyl)methane and synthetic analogs. *Cancer Lett* 2008, 269(2):326-338.
6. Kim YS, Milner JA: Targets for indole-3-carbinol in cancer prevention. *J Nutr Biochem* 2005, 16(2):65-73.
7. Firestone GL, Sundar SN: Minireview: Modulation of Hormone Receptor Signaling by Dietary Anticancer Indoles. *Mol Endocrinol* 2009, 23(12):1940-1947.
8. Firestone GL, Bjeldanes LF: Indole-3-carbinol and 3-3'-diindolylmethane antiproliferative signaling pathways control cell-cycle gene transcription in human breast cancer cells by regulating promoter-Sp1 transcription factor interactions. *J Nutr* 2003, 133(7):2448s-2455s.
9. Sundar SN, Kerekatte V, Equinozio CN, Doan VB, Bjeldanes LF, Firestone GL: Indole-3-carbinol selectively uncouples expression and activity of estrogen receptor subtypes in human breast cancer cells. *Mol Endocrinol* 2006, 20(12):3070-3082.
10. Nguyen HH, Aronchik I, Brar GA, Nguyen DHH, Bjeldanes LF, Firestone GL: The dietary phytochemical indole-3-carbinol is a natural elastase enzymatic inhibitor that disrupts cyclin E protein processing. *P Natl Acad Sci USA* 2008, 105(50):19750-19755.
11. Riby JE, Feng CL, Chang YC, Schaldach CM, Firestone GL, Bjeldanes LF: The major cyclic trimeric product of indole-3-carbinol is a strong agonist of the estrogen receptor signaling pathway. *Biochemistry-Us* 2000, 39(5):910-918.
12. Cover CM, Hsieh SJ, Tran SH, Hallden G, Kim GS, Bjeldanes LF, Firestone GL: Indole-3-carbinol inhibits the expression of cyclin-dependent kinase-6 and induces a G(1) cell cycle arrest of human breast cancer cells independent of estrogen receptor signaling. *J Biol Chem* 1998, 273(7):3838-3847.
13. Garcia HH, Brar GA, Nguyen DHH, Bjeldanes LF, Firestone GL: Indole-3-carbinol (I3C) inhibits cyclin-dependent kinase-2 function in human breast cancer cells by regulating the size distribution, associated cyclin E forms, and subcellular localization of the CDK2 protein complex. *J Biol Chem* 2005, 280(10):8756-8764.
14. Meng QH, Qi M, Chen DZ, Yuan RQ, Goldberg ID, Rosen EM, Auburn K, Fan SJ: Suppression of breast cancer invasion and migration by indole-3-carbinol: associated with up-regulation of BRCA1 and E-cadherin/catenin complexes. *J Mol Med* 2000, 78(3):155-165.
15. Cram EJ, Liu BD, Bjeldanes LF, Firestone GL: Indole-3-carbinol inhibits CDK6 expression in human MCF-7 breast cancer cells by disrupting Sp1 transcription

- factor interactions with a composite element in the CDK6 gene promoter. *J Biol Chem* 2001, 276(25):22332-22340.
16. Cover CM, Hsieh SJ, Cram EJ, Hong CB, Riby JE, Bjeldanes LF, Firestone GL: Indole-3-carbinol and tamoxifen cooperate to arrest the cell cycle of MCF-7 human breast cancer cells. *Cancer Res* 1999, 59(6):1244-1251.
 17. Jump SM, Kung J, Staub R, Kinseth MA, Cram EJ, Yudina LN, Preobrazhenskaya MN, Bjeldanes LF, Firestone GL: N-Alkoxy derivatization of indole-3-carbinol increases the efficacy of the G1 cell cycle arrest and of 13C-specific regulation of cell cycle gene transcription and activity in human breast cancer cells. *Biochem Pharmacol* 2008, 75(3):713-724.