

Computational Tools for the High-Throughput Identification of Protein- targeted Drugs and Probes

By

Rhiannon Kamstra

**A thesis submitted in partial fulfillment of the requirements for the degree of
Master of Science in Chemistry
Faculty of Science and Environmental Studies
Department of Chemistry
Lakehead University
November 2013
© Rhiannon Kamstra 2013**

Abstract

This thesis is comprised of three projects that are driven by a common theme, which is the use of computational tools in aiding molecular probe and drug design. In the first project, the feasibility of using molecular docking and scoring to estimate binding affinity for small molecules labelled covalently with fluorophores was tested using several proof-of-concept experiments. The high-throughput nature of computational screening applications such as Hierarchical Virtual Ligand Screening (HierVLS) necessitate that, in order to screen these labelled compounds, there must be an automated way to generate the associated structures virtually from large databases of base compounds and fluorophores. A script was developed in MOE software using scientific vector language (SVL) that could identify key reactive functional groups in both reactive fluorophores and target base compounds, and create the appropriate labelled structures for screening. The final fluorescence-labelled database numbers 14,862 compounds, each tagged with the ATTO680 fluorophore.

In a subsequent project, the fluorescence-tagged library was screened against carbonic anhydrase 9 (CAIX), a protein implicated as a biomarker in several cancer types. This screening was accompanied by the screening of a validation set of known CAIX ligands and appropriately chosen decoys. The best scoring protocol according to our analyses was that which used principal components analysis. Ten of the top scoring candidates are suggested for future testing as probe candidates. CAIX binding sites were compared with equivalent residues in the sequences of 24 other CA isoforms to identify sites that might confer CAIX specificity, and the top scoring ligands were ranked according to this scheme.

Lastly, experimental characterization was performed on three previously identified potential ligands for a cancer-related receptor tyrosine kinase, EphB4. Two *in vitro* assay formats were used: a homogenous time-resolved fluorescence assay and an enzyme-coupled spectrophotometric assay. One candidate, DNP-L-Arg, was the only one of the three with some experimental evidence of affecting kinase activity. The first assays suggested that DNP-L-Arg may have an activating effect on EphB4. The plausibility of this effect discussed with respect to mechanisms found in the literature, and using predicted and experimental structures for docked ligands. The coupled assay format did not conclusively confirm this effect.

The research presented underscores the ability for computational tools to be incorporated into a variety of different areas within the fields of biochemistry and drug design. Future complementary experimental work will be crucial both in evaluating and refining the suggested probe candidates and in further validating and improving the computational techniques used.

Acknowledgements

I would personally like to thank my thesis supervisor and teacher, Dr. Wely Floriano, for her continued guidance and support throughout this process. I also extend that appreciation to my thesis committee members, Dr. Laura Curiel and Dr. Chris Phenix, and my external evaluator, Dr. Andrew Freywald, for their much valued constructive feedback and expertise. Work pertaining to EphB4 was completed in consultation with scientists Dr. Andrew Freywald and Dr. Suraj Abraham, who offered important insight into both concepts and experimental work. I would also like to thank the Lakehead University Chemistry Department specifically, along with all of the instructors, laboratory technicians, and staff in the Faculty of Science and Environmental Studies who have helped to shape my experience as both an undergraduate and graduate student at Lakehead University.

Funding for this research was generously offered through the Thunder Bay Regional Research Institute (TBRR) which, in addition to financial support and facilities, provided invaluable interdisciplinary research experience with a number of excellent scientists. This work was also funded in part by the RBC Royal Bank's Dr. Mark Poznansky Mentorship Development Award.

High-throughput computational work requires the availability of large scale computing power and technical support. The work in this project was made possible with resources and facilities provided through SHARCNET and Lakehead University (HPC). I thank Darryl Willick for his frequent and much-needed advice pertaining to all things computational.

The atmosphere within a workplace or research group can truly define how it is experienced. I thank all of my lab mates as well as my other LU and TBRR colleagues including

John Wigg, Saedeh Dadgar, and Tom Sitter, for creating an atmosphere where co-workers are also friends. On that note, I would finally like to thank all of the friends and family who have been understanding and who have provided some essential respite from academic life over the past few years.

Table of Contents

Abstract.....	I
Acknowledgements.....	III
Table of Contents.....	V
List of Figures.....	VIII
List of Tables.....	XV
List of Equations.....	XVI
List of Abbreviations and Symbols.....	XVIII
Chapter 1. Introduction.....	1
Chapter 2. Literature Review.....	2
2.1. Introduction.....	2
2.2. Computational Tools in Bioinformatics, Drug-Design, and Medicinal Chemistry.....	2
2.2.1. Overview.....	2
2.2.2. Sequence analyses and high-throughput data handling.....	3
2.2.3. General drug design approaches.....	4
2.2.4. Molecular docking and scoring for virtual ligand screening.....	7
2.3. Molecular Imaging and Probe Development.....	10
2.3.1. Overview.....	10
2.3.2. Molecular imaging in drug design.....	11
2.3.3. Clinical scope of molecular imaging.....	11
2.3.4. Molecular imaging modalities.....	12
2.4. Biomarkers for Cervical Cancer and Associated Hypoxia.....	16
2.4.1. Cervical cancer etiology and human papillomavirus.....	16
2.4.2. Tumour-related hypoxia.....	18
2.4.3. Carbonic anhydrase IX as a biomarker for cervical cancer.....	19
2.5. Receptor Tyrosine Kinases as Drug Targets.....	23
2.5.1. Structure and function of receptor tyrosine kinases.....	23
2.5.2. Eph receptors.....	24
2.5.3. EphB4 and ephrins-B2 signalling in cancer.....	27
2.6. Summary & Scope.....	29
2.7. References.....	30
Chapter 3. Methodology.....	1
3.1. Overview.....	1
3.2. Computational Methods.....	1
3.2.1. Force field-based methods.....	1
3.2.2. Hierarchical virtual ligand screening.....	5
3.3. Experimental Methods.....	6
3.3.1. Fluorescence-based binding assays.....	6
3.4. References.....	8
Chapter 4. Creating databases of fluorescently-labelled compounds for the discovery of target-specific molecular probes.....	1
4.1. Introduction.....	1
4.2. Materials and Methods.....	4
4.2.1. Preparation of the test libraries.....	4
4.2.2. Protein structure preparation, VLS, and hit selection.....	5

4.2.3. Source of chemicals for binding assays and fluorophore conjugation	6
4.2.4. Synthesis and purity of fluorescent conjugates.....	6
4.2.5. Fluorescence polarization overview	7
4.2.6. BoNT/A fluorescence polarization assay	7
4.2.7. Protein E6 intrinsic tryptophan fluorescence assay	8
4.2.8. Protein E6 fluorescence polarization assay.....	9
4.2.9. Fluorophore conjugation script and preparation of a labelled library	10
4.2.10. Base compound library for conjugation.....	11
4.2.11. Generating a library of fluorescent conjugates.	11
4.3. Results.....	13
4.3.1. Virtually screening fluorescence-labelled compound libraries is effective at identifying potential fluorescent probes	13
4.3.2. <i>In vitro</i> testing of BoNT/A ligands	17
4.3.3. <i>In vitro</i> testing of E6 ligands.....	19
4.3.4. An efficient methodology for creating large libraries of fluorescence-labelled compound enables the use of VLS for probe discovery	22
4.4. Discussion and Conclusions	22
4.5. References.....	26
Chapter 5. Identifying potential selective fluorescent probes for cancer-associated protein carbonic anhydrase IX using a computational approach	1
5.1. Introduction.....	1
5.2. Methods and Procedures	3
5.2.1. Preparation of the CAIX model structure	3
5.2.2. Multiple sequence alignment and binding site comparison.....	4
5.2.3. Known inhibitors library for structure validation	4
5.2.4. Decoy library for virtual screening.....	5
5.2.5. Validation Screening.....	6
5.2.6. CAIX Virtual Screening	7
5.3. Results and Discussion	8
5.3.1. Preparation of the CAIX model structure	8
5.3.2. Multiple sequence alignment and binding site comparison.....	10
5.3.3. Validation screening	12
5.3.4. CAIX Screening.....	17
5.4. Conclusions.....	23
5.5. References.....	24
Chapter 6. Identification and <i>in vitro</i> testing of potential EphB4 kinase modulators.....	1
6.1. Introduction.....	1
6.2. Materials and Methods.....	2
6.2.1. Computational screening of EphB4 kinase.....	2
6.2.2. Materials	3
6.2.3. Homogenous time-resolved FRET assay (HTRF).....	3
6.2.4. Spectrophotometric coupled kinase assay	6
6.3. Results and discussion	9
6.3.1. Compounds selected for experimental testing	9
6.3.2. Homogenous time-resolved FRET assay (HTRF).....	9
6.3.3. Possible mechanism of DNP-L-Arg interaction based on docked structure.....	11

6.3.4. Spectrophotometric coupled kinase assay	15
6.4. Conclusions.....	22
6.5. References.....	23
Chapter 7. Conclusions.....	28
Supplementary Information	i

List of Figures

- Figure 2.4.3-1**—Carbonic anhydrases catalyze the reversible hydration of carbon dioxide. Figure generated using ChemDraw Software (Perkin Elmer)..... 19
- Figure 2.4.3-2**—A basic cell representation depicting the relative localization of 13 different mammalian carbonic anhydrase (CA) isoforms. Isoforms IV, IX, XII, XIV, and XV are membrane-associated, VI is secreted, VA and VB are located in the mitochondria, and I-III, and VII, XIII are cytosolic.⁶³ Figure generated using ChemDraw Software (Perkin Elmer)..... 20
- Figure 2.4.3-3**—“Schematic representation of the catalytic mechanism for the α -CA catalyzed CO₂ hydration. The hydrophobic pocket for the binding of substrate(s) is shown schematically at step (B)”.⁶² 21
- Figure 2.5.3-1**—EphB4 receptors and their endogenous ligand, ephrin-B2, can interact between two neighboring cells and trigger a bidirectional signal. Stimulated activity and signalling that occurs in the EphB4-containing cell is termed “forward signalling” while activity that is stimulated in ephrin-B2 direction is termed “reverse signalling”. Figure generated using ChemDraw Software (Perkin Elmer)..... 28
- Figure 4.2.1-1**—N-hydroxysuccinimidyl ester (NHS ester) reactive fluorophores react with primary amines to form fluorophore/ligand conjugates that are linked via an amide bond. 5
- Figure 4.2.11-1**—Bodipy NHS ester (a) was the reactive fluorophore used for the experimental synthesis and testing of fluorescent small-molecule conjugates. ATTO680 NHS ester (b) is the reactive dye chosen for computational generation of tagged compound libraries..... 12
- Figure 4.3.1-1**—Molecular docking results for unlabelled ligands against binding region 21 in BoNT/A LC. The solid and dotted lines represent mean and \pm SD (standard deviation), respectively. Compounds scoring at or below the mean-SD threshold are expected to bind

experimentally to the target. Aspartame was one of the ligands showing high predicted binding affinity against BoNT/A with a score below the binding threshold. 14

Figure 4.3.1-2—Molecular docking results for Bodipy labelled ligands against binding region 21 in BoNT/A LC. The solid and dotted lines represent mean and \pm SD (standard deviation), respectively. Aspartame-Bodipy (APM-BDP) scored close to one SD below the mean value (about 0.5kCal/mol higher than SD), and it is expected to bind to BoNT/A..... 14

Figure 4.3.1-3—Molecular docking results for unlabelled ligands against protein E6. The solid and dotted lines represent mean and \pm 2SD (standard deviation), respectively. O-Succinyl-L-homoserine scored below the threshold of the mean-2SD, and was selected for experimental testing against the target..... 15

Figure 4.3.1-4—Molecular docking results for Bodipy-labelled ligands against protein E6. The solid and dotted lines represent mean and \pm 2SD (standard deviation), respectively. O-Succinyl-L-homoserine-Bodipy scored relatively poorly, with a predicted binding affinity almost 2SD worse than the average of all docked compounds and it is, hence, not expected to bind experimentally to E6. 15

Figure 4.3.1-5—The structures of aspartame (top) and aspartame-Bodipy (bottom) used in this work. Bodipy-labelled aspartame was synthesized by conjugating aspartame to the amine reactive 3-Bodipy-propanoylaminocaproic acid, N-hydroxysuccinimide ester (Bodipy). Both unlabelled compound and conjugate were predicted to bind to BoNT/A..... 16

Figure 4.3.1-6—The structures of O-succinyl-L-homoserine (top) and O-succinyl—L-homoserine-Bodipy (bottom) used in this work. Bodipy-labelled O-succinyl-L-homoserine was synthesized by conjugating O-succinyl-L-homoserine to the amine reactive 3-Bodipy-

propanoylaminocaproic acid, N-hydroxysuccinimide ester (Bodipy). In this case, the unlabelled compound was predicted to bind to protein E6, whereas its labelled counterpart was not. 17

Figure 4.3.2-1-Fluorescence polarization binding assay using aspartame-Bodipy against BoNT/A-LC. The data related to unlabelled aspartame (APM) is shown in black whereas unlabelled paclitaxel (PAC) results are presented in square pattern. The addition of aspartame at 1µM (left solid bar) reduces the fluorescence polarization associated with the binding of aspartame-Bodipy to BoNT/A (right solid bar), whereas the addition of paclitaxel (left checkered bar) increases the polarization of the corresponding aspartame-Bodipy reference (right checkered bar). Each labelled/unlabelled competition assay has its own reference because of the influence of DMSO on the fluorescence of labelled aspartame-Bodipy. The samples were run in triplicate at 37°C and backgrounds were subtracted for each sample. 18

Figure 4.3.3-1-Change in intrinsic tryptophan fluorescence of E6 after 30 minutes incubation at 37°C buffered to a pH of 7.3 (20mM HEPES 7.3, 0.01% tween (v/v). Protein was held at a constant concentration of 0.03 mg/mL. 6F4 antibody was used as a positive control at a concentration of 0.4ng/mL. O-succinyl-L-homoserine concentration was varied (300µM-0.01µM). The red line represents a 30% increase relative to baseline (no binding). 6F4 antibody, a positive control for binding, produces an increase in tryptophan fluorescence well over 30%. 20

Figure 4.3.3-2-O-Succinyl-L-homoserine-Bodipy (100, 10, and 1µM) and protein E6 (0.04mg/mL) at 60 minutes. The experiment was conducted at a temperature of 37°C and pH of 7.5. Our results indicate that the O-Succinyl-Homoserine-Bodipy is not binding at any concentrations tested. Control wells containing protein as well as the O-Succinyl-L-homoserine-Bodipy at all three concentrations dissolved in assay buffer were included in addition to the reaction wells. 21

Figure 5.3.1-1 —A Ramachandran plot for the CAIX model structure (generated using MOE software).	9
Figure 5.3.1-2 – A cartoon representation of the CAIX model structure with the zinc cation shown in the center of a large binding cavity.	10
Figure 5.3.2-1 – A surface representation of the CAIX model with the centers of sites 1, 3, and 4 (1 would be the most preferable for binding, 3 the second most, and 4 th , the least) shown as blue spheres. Surface colours represent hydrophilic (pink), hydrophobic (green) and neutral (white) regions.	11
Figure 5.3.2-2 – A surface representation of the CAIX model showing binding site 2 (blue sphere representing its center), which represents the catalytic site. The zinc cation is shown as a yellow sphere. Surface colours represent hydrophilic (pink), hydrophobic (green) and neutral (white) regions.	11
Figure 5.3.3-1 – The ROC curves for the PCA and raw (force field-based) scoring schemes used for the validation screen of CAIX. The area under each curve (AUC) is shown for each scoring scheme in parentheses.	13
Figure 5.3.3-2 – The PCA scoring scheme was selected for library screening. A frequency histogram of scores for decoys and true positives in the validation library are shown in (a). The specificity, sensitivity, and positive predictive value (PPV) of the PCA scheme determined using the validation library are shown in (b). The highest possible PPV value, assuming a 1% prevalence of hits in a database, occurs at a threshold of -3.317. Relatively few compounds in the validation library scored below the threshold value of -3.317 (c).	14
Figure 5.3.3-3 – The number of true positives found (N) using each indicated threshold (i.e., top x% of validation library screened) as a fraction of the total number of true positives present in	

the validation library (N=32). Enrichment factors (EFs) were calculated for each of the indicated threshold values and shown in tabulated format. The PC1 score threshold that was selected using ROC analysis gave the highest EF value, 6.92. 17

Figure 5.3.4-1 – A distribution of scores for molecules screened in the Atto680-tagged (blue) and Fluorescent MLSMR (red) databases. The threshold PC1 score -3.317 is marked. 19

Figure 6.2.3-1 – A basic flowchart indicating the steps associated with kinase reaction detection – EphB4 kinase activity phosphorylates Tyr on biotinylated peptide substrate followed by detection using an Eu-chelated anti-PTyr antibody (binding to PTyr) and Streptavidin-APC (binding to biotin) and subsequent FRET to produce signal. (Figure reproduced courtesy of EMD Millipore from the user guide for Cat#17-10015).²⁴ 4

Figure 6.2.4-1 – A spectrophotometric coupled assay was used to measure kinase activity. ADP produced by the kinase reaction would be measured through conversion of phosphoenolpyruvate (PEP) to pyruvate with pyruvate kinase (PK) followed by action by lactate dehydrogenase (LDH) to produce lactate through the oxidation of NADH. NADH absorbs at 340nm allowing the decrease in absorbance to be used to approximate ADP production by EphB4 kinase. ATP is regenerated during this process. Figure generated using ChemDraw Software (Perkin Elmer). ... 7

Figure 6.3.1-1 – Three hit compounds were selected for *in vitro* testing: N-(2,4)-dinitrophenyl-L-arginine (DNP-L-Arg), dihydrofolic acid, and aminopterin. 9

Figure 6.3.2-1 – The HTRF assay dose-response curve for staurosporine against EphB4 (16nM) with an IC50 estimate 18.59µM (95% confidence interval of 17.22-20.07µM). 10

Figure 6.3.2-2 – HTRF ratios are compared for the compounds tested at each concentration. Error bars represent standard deviation between duplicates. 11

Figure 6.3.2-3 – The HTRF assay dose-response curve for DNP-L-Arg against EphB4 (16nM) indicating a possible shift in effect direction relative to staurosporine, a known kinase inhibitor. 11

Figure 6.3.3-1 – Left: Docked or crystal structure conformations for three ligands are shown for EphB4 kinase relative to the N- and C-terminal lobes. Right: AMP-PNP (adenine moiety)(PDB:1JPA) (green) and 7X6 (inhibitor)⁸ docked using our protocol (orange) occupy similar sites in the structure while DNP-L-Arg (magenta) is predicted to bind further from the active site. 15

Figure 6.3.4-1 - Abs₃₄₀ at 10s with varying initial concentrations of ADP with the coupling system alone. The Abs₃₄₀ of the starting assay concentration of NADH alone, and in the presence of all reaction components except for PK/LDH (No PK/LDH), as well as a buffer control, are shown for reference..... 18

Figure 6.3.4-2 – Spectrophotometric coupled kinase assay indicating decrease in absorbance at 340nm over time where EphB4 kinase is present. 19

Figure 6.3.4-3 – The coupled assay data from two exemplary reaction wells with staurosporine and DNP-L-Arg each at 100μM with EphB4 (0.025μM). Linear regression was performed on time>1700s..... 20

Figure 6.3.4-4 – A comparison of individual reaction velocities ($\Delta\text{Abs}_{340}/\text{s}/\mu\text{mol EphB4}$) corrected for baseline rate, as a percent of the rate of normal EphB4 activity (100%) for the compounds tested..... 20

Figure 6.3.4-5 – A Hanes-Woolf plot is used to represent [Compound] (μM) as a function of [Compound] (μM)/V ($\Delta\text{Abs}_{340}/\text{s}/\mu\text{mol EphB4}$) using the coupling-assay. Reactions containing staurosporine, DNP-L-Arg, as well as a control reaction with varying peptide (polyGluTyr)

concentration are shown. All reactions were performed with the same concentration of ATP and
detection components..... 21

List of Tables

Table 4.3.3-1 -A table outlining the predicted binding scores, expected experimental result, and whether experimental binding was observed for four selected compounds.	21
Table 5.3.2-1 – The % sequence identity for residues within 5Å of each binding region to the sequence alignment of 23 CA isoforms is shown, along with the distance from the catalytic site (2).	12
Table 5.3.4-1 – Top 20 ranked compounds screened from both the Atto680-tagged and Fluorescent MLSMR databases.	19
Table 5.3.4-2 – Best 5 compounds with suggested binding preference for region 1 from the Fluorescent MLSMR database. Raw scores were examined relative to mean, and expressed as N*# of standard deviations from the mean score.	20
Table 5.3.4-3 – Best 5 compounds with suggested binding preference for region 1 from the Atto680-tagged MLSMR database. Raw scores were examined relative to mean, and expressed as N*# of standard deviations from the mean score.	21
Table 5.3.4-4 – Chemical structures of the top 5 scoring Fluorescent MLSMR compounds.	21
Table 5.3.4-5 – Base chemical structures of the top 5 scoring Atto680-tagged MLSMR compounds.	22
Table 6.3.4-1 – Regression parameters obtained from the Hanes-Woolf plot with associated constants for DNP-L-Arg, staurosporine, and varied concentrations of peptide substrate (no inhibitor). Note that reaction velocity is expressed as $\Delta\text{Abs}_{340}/\text{s}/\mu\text{mol EphB4}$	22

List of Equations

Equation 3.2.1-1 - Force field potential energy functions are summations of both bonded and non-bonded terms.....	2
Equation 3.2.1-2 - Simple harmonic oscillation function for bond stretching term. R denotes distance relative to equilibrium (R_e), k_e denotes the force-constant.	2
Equation 3.2.1-3 - Function describing the bond angle bending for two bonds between atoms IJ and JK (J atom shared). C_{IJK} is a function of the force constant and the equilibrium angle: $C_{IJK} = K_{IJK}/(\sin \theta_j^0)^2$	3
Equation 3.2.1-4 - Bond angle bending for molecules with linear equilibrium geometry is given by this potential.....	3
Equation 3.2.1-5 - Dihedral angle torsion is described about four atoms (IJKL) with respect to dihedral angle (ϕ), periodicity (n_{JK}), and rotation barrier (V_{JK}). The equilibrium angle is denoted by ϕ_{JK}^0	3
Equation 3.2.1-6 - Cases where one atom, I, is bonded to three other atoms, JKL, may require the inclusion of an inversion term which is given as a function of C_I , and the angle between the IL bond and the JIK plane (ψ). C is a function of the force constant (K) and the equilibrium bond-plane angle: $C_I = K_I/(\sin \psi_I^0)^2$	3
Equation 3.2.1-7 - A 12-6 Leonard-Jones potential is used to calculate van der Waals forces as a function of interatomic distance (r).....	4
Equation 3.2.1-8 - Electrostatic forces are calculated using a Coulombic interaction expression. $Q_{i,j}$ refer to atomic charges, ϵ , the dielectric constant, and R_{ij} the interatomic distance. 322.0637 is used as a conversion factor for kcal.	4

Equation 3.2.1-9– A potential used to describe hydrogen bonding in Dreiding as a function of the distance between the donor and acceptor atoms (R_{DA}) and the bond angle between donor and acceptor atoms (θ_{DA}). D_{hb} and R_{hb} values can be derived from quantum mechanical calculations or empirically. 4

Equation 3.2.2-1 – The binding affinities for the best conformers from Level 1 of HierVLS are calculated by subtracting both the energy of the solvated protein alone and the solvated ligand alone from the energy of the solvated complex. The AVGB solvation approach is used to calculate solvation energies. 6

Equation 3.3.1-1– Fluorescence polarization is calculated as the ratio of the difference between the fluorescence intensity parallel and perpendicular to the excitation plane to the sum of these intensities. 7

Equation 5.3.3-1 – The positive predictive value (PPV) is the fraction of all positives, true (TP) and false (FP), that are true positives. 14

Equation 5.3.3-2 – The enrichment factor (EF) is given by the ratio of the number of hits (n) found above the threshold (top x %) to the expected number of hits, the proportion (P) of overall hits in the database multiplied by the number of compounds overall (N) above the threshold (top x %). 15

Equation 6.2.3-1 – HTRF ratio calculation based on emission readings at 680nm and 620nm.... 6

Equation 6.2.3-2 – The relationship between the slope of the Hanes-Woolf regression (m) and V_{max} 6

Equation 6.2.3-3 – The relationship between the x-intercept of the Hanes-Woolf regression and the Michaelis-Menten constant (K_m). 6

List of Abbreviations and Symbols

^{18}F FDG – 2- ^{18}F -fluoro-2-deoxy-D-glucose

ADP – adenosine diphosphate

APC – allophycocyanin

ATP – adenosine triphosphate

AUC – area under the curve

Bodipy – 3-Bodipy-propanoylaminocaproic acid, N-hydroxysuccinimide ester

BoNT (LC) – Botulinum neurotoxin (light chain)

CA – carbonic anhydrase

CAIX – carbonic anhydrase 9

DHF – dihydrofolic acid

DNP-L-Arg – N-(2,4)-dinitrophenyl-L-arginine

EF – enrichment factor

EGFR – epidermal growth factor receptor

FRET – fluorescence resonance energy transfer

GPI – glycosylphosphatidylinositol

GUI – Graphical User Interface

GWAS – genome wide association study

HierVLS – Hierarchical Virtual Ligand Screening

HIF – hypoxia inducible factor

HPV – human papillomavirus

HTRF – homogenous time-resolved fluorescence

JMS – juxtamembrane segment

LC-MS – liquid chromatography mass spectrometry

LDH – lactate dehydrogenase

MLSMR - Molecular Libraries Small Molecule Repository

MRI – magnetic resonance imaging

NADH – β -Nicotinamide adenine dinucleotide (reduced)
NGS – next generation sequencing
NHS-ester - N-hydroxysuccinimidyl ester
NIR – near infrared
NMR – nuclear magnetic resonance
PAD – primary amines database
PCA – principal components analysis
PDB - Protein Data Bank
PEP – phosphoenolpyruvate
PET – positron emission tomography
PK – pyruvate kinase
PPV – positive predictive value
QSAR – quantitative structure-activity relationship
RCSB – Research Collaboratory for Structural Bioinformatics
RMSD – root mean squared deviation
ROC – receiver operating characteristic (curve)
RTKs – receptor tyrosine kinases
SMILES - simplified molecular-input line-entry system

Chapter 1.

Introduction

This thesis is comprised of several projects with a unifying theme – the utilization of computational techniques and predictions to assist in experimental biochemical research and drug-design. Two of these projects focus on screening biologically relevant protein targets, carbonic anhydrase 9 (CAIX) and erythropoietin producing hepatocellular carcinoma receptor B4 (EphB4), for small molecule ligands that could have applications in diagnostics, therapeutics, or research. Another one of these projects describes the development of a computational protocol to create libraries of fluorescently-labelled ligand structures that would be useful for computer assisted drug design efforts, including those employed in the CAIX project. Each of these projects is described in detail in a separate chapter, with specific methods, materials, and results. Preceding these chapters is a general literature overview which provides relevant background information pertaining to all of the projects, and contextualizes this thesis with respect to current scientific literature. The theory underlying the methods used in these projects is given in the Methodology chapter, and is explained further in the appropriate subsequent chapters' experimental sections. Finally, the results of these projects are summarized and directions for future research are suggested.

Chapter 2.

Literature Review

2.1. INTRODUCTION

A review of relevant literature is reported as it pertains to the topics discussed in this dissertation. Certain techniques, both computational and experimental, are outlined briefly in context with the current methods and applications being used in the field. More complete methodological theory is given in “Chapter 3. Methodology”, while detailed procedural steps are described for each method in the chapters discussing their use (Chapters 4-6).

2.2. COMPUTATIONAL TOOLS IN BIOINFORMATICS, DRUG-DESIGN, AND MEDICINAL CHEMISTRY

2.2.1. Overview

Computational tools have become increasingly used in fields like chemistry, biology, and medicine, often complementing traditional experimental techniques. The use of computers in these areas has facilitated the ability to process massive amounts of data that would otherwise be completely overwhelming. *Bioinformatics* is an emerging approach that focuses on processing and analyzing the wealth of information derived from biological systems, by interpreting it in the context of current molecular biology concepts through the application of computational and statistical techniques. Luscombe and co-authors define bioinformatics as “conceptualising biology in terms of molecules and applying ‘informatics techniques’ (derived from disciplines such as applied maths, computer science and statistics) to understand and organise the information associated with these molecules, on a large scale”.¹ Some of the commonly applied

methods of modern bioinformatics will be described briefly in the following sections, with particular emphasis on techniques used in computer-assisted drug design.

2.2.2. Sequence analyses and high-throughput data handling

As of August 2013, the U.S. National Center for Biotechnology Information (NCBI) “GenBank” database contains over 167 million sequence records containing more than 154 billion DNA bases, with the number of bases doubling roughly every 18 months since 1982.² Automated Sanger and next generation sequencing (NGS) technologies have made it increasingly possible to obtain large amounts of genomic data fast and cost effectively.³ This type of data can be used in a variety of applications including studies linking specific genetic variations or epigenetic markers to disease processes in genome-wide association studies (“GWAS”).³ However, as NGS technologies generate exponentially larger sets of data for a given assembled sequence than traditional methods, it has been suggested that data processing is likely to become the limiting factor for both the cost and time efficiency of NGS efforts.³

Information obtained from sequences can be used to identify specific protein targets, or even variants of a protein target, for use in drug design efforts or diagnostic tests. Many sequence-based techniques are available that can use a given sequence, along with other information, to provide predictions about the structure and function of a given nucleotide or amino-acid sequence. For example, chromatin structure, gene regulatory information, and epigenetic markers can be studied in this way.³ Protein amino-acid sequences, either obtained directly or through gene sequencing, can similarly be analyzed for patterns suggestive of specific three dimensional structures, post-transcriptional modification sites, and other functional motifs. Multiple available resources are designed to assist in sequence analysis. Sequence alignment tools are important for comparing different sequences and determining the degree of similarity,

and for identifying common features or relationships. Many algorithms are available to perform these tasks, and they are often specialized for certain applications or types of sequences.

Sequence alignment methods are normally categorized as either pairwise or multiple sequence alignment algorithms. Pairwise alignment tools are normally appropriate for aligning only two sequences, while multiple sequence alignments are suitable for aligning and comparing more than two sequences. Efficient use of computational time is an important consideration during the development of these software applications due to the very large size of these data sets.⁴

2.2.3. General drug design approaches

Computational tools are used to assist various stages of drug design. The physical properties of a drug or drug candidate can be estimated based on physical or empirical methods, and these can then be correlated with the molecule's observed biological effects. For example, rational drug design may utilize information about specific target protein amino acid sequences, which can be analyzed, compared, or visualized using different software to generate optimised ligands for particular sites in the protein structure. While applications of computational techniques in drug design are diverse, they tend to be sub-categorized based on whether they rely solely on information about ligands, or on information about both ligand(s) and the therapeutic target. These categories are often referred to as "ligand-based" and "structure-based" drug design methods, respectively.

2.2.3. a) Ligand-based drug design

Ligand-based drug design techniques analyze the relationships between structural features of a series of ligands, and their observed biological effects. For example, a series of structurally related molecular derivatives may be tested in an animal or cell-line model at different doses, and a particular effect measured. The graphical relationship between the dose

given and the degree of effect observed is referred to as the “dose-response curve”.

Computational techniques are often employed to describe quantitatively how the observed effects correlate with changes in ligand structure. This is known as a “quantitative structure-activity relationship” or QSAR. QSAR studies use various molecular descriptors, such as charge, molecular weight, and surface area of a ligand, either empirically determined or computationally derived. Current methods rely on a combination of 2D and 3D molecular descriptors in their calculations. Once a QSAR is determined, it can be applied to untested molecules to predict which ones are likely to have the desired activity. QSAR studies have been used as a part of successful drug design strategy for numerous clinical candidates. For example, QSAR methodology has been integral to the development of selective HIV protease inhibitors.⁵⁻⁸ While QSAR studies can provide valuable information leading to the identification of active molecules, they may be seriously limited in their effectiveness and versatility. QSAR studies are often limited by the number and the diversity of molecules that were experimentally tested to generate the initial training set. Another potential deficiency of QSAR is rooted in some commonly used analytical methodologies that cannot guarantee that all tested molecules are eliciting their effects via the same target protein. Therefore, this approach is best suited to improving the activity of existing lead compounds, but is not a robust method for identifying new, structurally unique, lead compounds.⁹⁻¹³

2.2.3. b) Structure-based drug design

Structure-based drug design methods require structural information about both the ligand and the target. These methods commonly make use of protein structures experimentally determined through X-ray crystallography or NMR, or computationally modelled using the experimental structure of another highly homologous protein, in the process known as

“homology modelling”.¹⁴⁻¹⁵ Modern molecular biology methods have made it possible to reliably produce and purify samples of proteins or protein-ligand complexes to obtain X-ray crystallography-based models of their structures.¹⁶ Structure-based methods can be used to analyze interactions that occur within a protein-ligand complex, and provide valuable insights into mechanisms by which ligands affect protein activity. A representation of the chemical features needed to elicit or inhibit biological activity can be created in the form of a *pharmacophore model*. This approach generates a model that represents molecular features and their relative orientations that are responsible for imparting a desired biological activity.¹⁷ This method may use 3D information about a given set of ligands with known activities, about a given active site and its properties, or about a ligand-protein complex.¹⁷ A pharmacophore model can be used to screen libraries of compounds for ones that possess a required arrangement of features, or to design chemical compounds *de novo*, based on possible arrangements of substructures producing optimal interactions with the target.¹⁷

Structure-based drug design efforts also make use of functions that can approximate the energy associated with interactions between the molecules of a system. Normally, molecular systems are described using either a classical mechanics or a quantum mechanics (QM) approach. Their reliance on first principles (without empirical assumptions) makes QM approaches highly appealing. These methods are considered to represent the most accurate analysis currently available. In addition to being used to describe structures and their associated potential energies, these methods can also be used to study dynamic intermolecular processes, such as reaction mechanisms. However, the use of quantum mechanical approaches, particularly for studying large biomolecules, remains heavily limited by the large amount of computational resources they require. QM methods are sometimes used to complement approaches where

classical potential energy functions are not adequate, such as in the case of describing catalytic mechanisms. Typically, however, protein-ligand systems are described through various classical approaches. Potential energy functions, in addition to being used to describe systems for the purposes of drug design, can also be used in combination with other techniques, such as molecular docking, to screen large libraries of chemical compounds to identify those that are likely to interact with a given target.^{14-15, 18}

2.2.4. Molecular docking and scoring for virtual ligand screening

2.2.4. a) Docking algorithms for protein/ligand complexes

Molecular docking involves the virtual placement of a small molecule (ligand) within a target protein binding site. This task is only seemingly simple, as both ligand and protein have varying conformational degrees of freedom, and varying conditions might affect conformations preferentially found in a natural system. These complications cannot be overlooked, as the appropriate sampling of each molecule's conformational space is crucial in ensuring that any subsequent scoring analysis is biologically relevant. While there are many available algorithms to handle conformational searches, each can be classified under one of three general approaches: systematic, stochastic, and simulation-based methods.¹⁸

Systematic approaches can provide very complete coverage of all reasonable conformations available to a particular ligand structure.¹⁸ Typically, these methods use strategies that either dock partial molecular fragments into a protein binding site and incrementally link them covalently to generate final conformers, or that dock rigid substructures and incrementally add flexible regions.¹⁸ Systematic searches can quickly become impractical even when moderately sized ligands (>5 rotatable bonds) are used, as the number of possible conformations

becomes too large.¹⁸ A systematic search that uses wide search increments may provide a subset of conformations that is very poorly representative of those actually available to the structure.

Stochastic or random search algorithms are also used in different types of docking software. Two popular stochastic techniques are Monte Carlo and genetic search algorithms.¹⁸ Using the Monte Carlo method, a random starting ligand conformer is scored and new conformations are randomly generated, scored and compared with the starting structure.¹⁸ Conformers are selected if they have an improved energy relative to other conformers, or alternatively, if they passed a probability function test.¹⁸

Simulation-based methods include those relying on molecular dynamics, where the time-dependent behaviour of a molecular system is modelled to theoretically observe all possible conformational states.¹⁸ To reach all energy minima, high-energy barriers may need to be overcome, which can require excessively long simulation times. However, molecular dynamics simulation is a useful way of probing conformational space of a molecular system, if sufficient computational resources are available.

2.2.4. b) Force field-based scoring methods for biological systems

Force field-based scoring functions are commonly employed for estimating the potential energy of protein-ligand interactions. These potential energy functions are based on classical mechanics, and represent a compromise between overall accuracy and efficiency. Force fields can provide reliable estimates of structural conformation energies using limited resources. A force field potential energy function describes intra- and intermolecular forces as a summation of bonded (depending on bonds between atoms) and non-bonded terms. Non-bonded and bonded interactions are each described by a number of functions to account for multiple contributions to

the potential energy of a system. A given set of parameters is required, which provides different constants describing specific atom types and their characteristics. Parameter values can be determined empirically or calculated using quantum mechanics. It is not uncommon for a force field to use a set of parameters whose origin may be both empirical and theoretical. Dreiding is the all-atoms force field that is applied extensively throughout this manuscript, and it will be described in more detail in Chapter 3 – Methodology.¹⁹

2.2.4. c) Virtual ligand screening and HierVLS

Virtual ligand screening consists of screening large libraries of chemical compounds using computational methods to identify candidates with probable biological activity or other desired properties. Various methods, including some of those previously described, can be used to screen databases in this way. QSAR and pharmacophore modelling studies can use information obtained from test sets or experimental data to screen libraries according to the structure-activity relationships or models identified. Molecular docking and scoring can also be used to dock small molecules into a target protein cavity in a structure-based approach. “Hit” compounds from screening are often carried through to extensive experimental validation and optimization to produce a viable drug candidate.¹⁸

Hierarchical virtual ligand screening (HierVLS) is a high-throughput protocol that docks different conformations of potential ligands into binding sites in a protein structure, and then scores them using a force field.²⁰ HierVLS systematically passes the best scoring conformers through different filters to limit the most computationally expensive calculations to most promising conformers.²⁰ A graphical user interface (GUI) for HierVLS has been developed to enhance user accessibility.²¹ HierVLS can be used in association with PASS to screen multiple binding sites within a protein's structure.²² PASS is a rapid and efficient program that uses a

geometry-based algorithm to identify putative binding pockets, basing their shape, size and depth relative to the protein surface.²² PASS is freely available and generates output that is generally compatible with applications used in later steps.

HierVLS itself is a hierarchical approach to virtual ligand screening that tests the largest number of bound conformations of each ligand using the least computationally expensive methods, and proceeds to use more computational resources to obtain more accurate predictions of binding energy for the most promising conformers.²⁰ A more detailed account of the HierVLS procedure is given in Chapter 3. HierVLS has been validated and used with several data sets and targets.^{20, 23-26} Most recently, HierVLS has been used as a part of a computational approach that successfully identified paclitaxel as an inhibitor of botulism neurotoxin A.²³

2.3. MOLECULAR IMAGING AND PROBE DEVELOPMENT

2.3.1. Overview

Molecular imaging is a modern medical imaging technique that can be applied both pre-clinically and clinically during the development and application of therapeutic drugs. Molecular imaging differs from traditional anatomical imaging techniques in that it is capable of imaging functional processes involving target molecules. Molecular imaging can also provide structural information, either inherently or when combined with anatomical imaging techniques. Pre-clinical molecular imaging is commonly used to assist in drug development, as it enables researchers to assess various drug characteristics *in vivo* using animal models prior to human clinical trials. Clinical molecular imaging is used in diagnostics, image-guided therapeutics, disease characterization, and assessment of response to therapy.²⁷

2.3.2. Molecular imaging in drug design

There are several advantages to using molecular imaging as a tool in pre-clinical phases of drug design and development. Molecular imaging can provide real-time information about biochemical processes as they are occurring *in vivo*. In early investigations, molecular imaging can be used as a tool to aid in understanding the functional relationships between different potential therapeutic targets.²⁷ Once a target is identified and a drug candidate produced, molecular imaging can also be used to experimentally evaluate drug performance *in vivo*, using animal models. Molecular imaging can be used to determine whether a drug candidate possesses the appropriate pharmacokinetic profile. This is possible because many molecular imaging techniques have a potential to be used quantitatively, and are inherently non or minimally invasive, so that functional processes can be imaged without disturbing the underlying physiology, and important information can be gleaned throughout the course of an experiment, instead of solely at the endpoint.²⁷

2.3.3. Clinical scope of molecular imaging

Several molecular imaging techniques, including functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) are routinely used clinically.²⁷⁻²⁸ These techniques are typically used for diagnostic purposes, and are designed to identify key molecular patterns or processes that are indicative of disease. Different molecular imaging techniques can be used either to image endogenous molecules or biomarkers, or to image processes involving exogenously administered reagents. Sometimes, contrast is achieved through fine tuning the inherent parameters of the imaging modality used to gain functional information, while in other instances, through the administration of an exogenous imaging agent or “probe”. The ability of different imaging modalities to image labelled reagents, which can be altered or targeted

specifically for the application, has sparked a great deal of interest in the development of target-specific probes for various disease biomarkers. Probe development works similarly to the drug development process so that selective molecules are designed to target specific proteins. However, the desired characteristics of probe candidates will be different from those of therapeutic drugs, as they should be able to generate contrast for a desired imaging modality, and thus, would require different chemical and pharmacokinetic properties. For instance, while therapeutic drugs are often desired to be orally administrable, this is less important for imaging applications. Molecular imaging fits well into the idealized concept of “individualized medicine”, where the biochemistry of an individual’s pathology is characterized through imaging or other techniques and therapy is directed in a manner that is determined by that characterization. Subsequently, because of its potential for molecular targeting and its non-invasiveness, molecular imaging can be used to predict response to targeted therapies and assess the response to therapy throughout the course of a selected treatment.²⁷⁻²⁸

2.3.4. Molecular imaging modalities

2.3.4. a) Functional MRI

Magnetic resonance imaging, or MRI, is a very powerful and adaptable technique that can be used for numerous types of anatomical and functional imaging studies. MRI relies on the same underlying principle as another powerful analytical technique, nuclear magnetic resonance spectroscopy (NMR). MRI uses specially designed gradient magnetic fields to spatially encode a given sample, allowing an image to be reproduced from a given signal. MRI can use a variety of techniques for applying radiofrequency pulses and magnetic fields in order to generate special types of contrast based on different molecular environments. Typical anatomical MRI detects ¹H nuclei that are found in water, which is highly abundant in the body. However, MRI can detect

various other nuclei with spin, as long as they are sufficiently abundant, meaning that it can also be used to provide information about chemical composition (spectroscopy). The versatility of MRI has been explored in many functional applications, relying on different techniques to generate contrast that enable the imaging of blood perfusion and oxygenation, as well as those that rely on exogenously administered contrast agents that can provide functional information. MRI is a highly sought-after technique because it does not require radioactive materials, and as such, is very safe for frequent or repeat imaging of patients.²⁹

2.3.4. b) Positron emission tomography

Positron emission tomography (PET) is a technique used in nuclear medicine that is strictly a molecular or functional technique. It requires a tracer or probe, radiolabelled with a positron emitting isotope such as ^{18}F , in order to generate contrast. Positron emitting isotopes can be generated in specialized facilities, such as cyclotrons, and must be prepared shortly in advance of administration and imaging due to their short half-lives (e.g., ^{18}F half-life is 110 minutes).²⁸ While this process does require specialized facilities, it is advantageous because the precursors used (such as H_2^{18}O) are more safely and readily available compared to materials used to generate isotopes for other nuclear medicine technologies. In PET, a minute amount of probe labelled with a positron emitter is administered into the body. When a positron is emitted from the probe upon isotope decay, it inevitably encounters an electron. Upon colliding with the electron, two coincident gamma rays are produced at directions 180° apart. These gamma rays can then be detected by a gamma camera to produce an image. There are many ways in which a PET probe or tracer can be used to get functional information about a patient. PET is typically used in cancer diagnostics employing the radiolabelled glucose analog, 2- ^{18}F -fluoro-2-deoxy-D-glucose (^{18}FDG). Actively growing tumours often exhibit enhanced glucose uptake and glycolysis

relative to surrounding tissue. FDG is used to probe relative glucose uptake of bodily tissues, exploiting this phenomenon to pinpoint tumour location and activity.²⁸ FDG's utility comes in part from the fact that it can be phosphorylated, like glucose, in preparation for glycolysis upon entering the cell.²⁸ However, because it lacks a key hydroxyl group at the 2- position occupied by ^{18}F , it cannot be carried further through the normal glycolytic process.²⁸ Thus, once phosphorylated, ^{18}F FDG becomes trapped in the cell, accumulating to generate an enhanced signal.²⁸ While ^{18}F FDG is the most common application of PET in clinical use, there are many other types of PET tracers that either are being used or have a potential for development as clinically useful molecular probes. Other isotopes used for PET include ^{15}O , ^{11}C , ^{13}N , and ^{64}Cu .²⁸ These isotopes can be incorporated into various types of tracers for different applications.²⁸ For example, water labelled with ^{15}O is used to examine blood perfusion in tumours.²⁸ PET has the advantage of being extremely sensitive (i.e., able to detect pM levels of injected probe) and depth-independent, meaning that developed tracers can be used in low enough concentrations so as to avoid possible toxic effects.²⁸ The limited half-life of positron emitting isotopes is usually in the range of hours and intrinsically limits the overall radiation exposure of the patient. Development of novel PET probes and their utility requires compounds that are creatively designed to provide physiologically relevant information and that exhibit ideal pharmacokinetic profiles. It is also important to consider that, because of their short half-life positron emitting isotopes must be easily incorporated into tracers before use. Therefore, only rapid and high yielding reactions can be used to synthesize PET tracers. Modern imaging techniques often use PET in combination with anatomical imaging modalities such as CT to provide more complete and clinically useful information.²⁸

2.3.4. c) Fluorescence imaging

Fluorescence imaging is an optical technique that has recently gained utility in the clinical setting, but that has long been useful for pre-clinical work and experimental biochemistry. There are various properties that can be exploited in optical and specifically fluorescence-based imaging in order to probe various characteristics of biological systems. Fluorescence occurs when certain molecules, known as fluorophores, are excited with a photon of a particular wavelength. Some of the absorbed energy is reemitted at a higher wavelength of light (lower energy) in the process known as fluorescence.³⁰ Simple excitation and emission are useful in the context of imaging or detection, but other optical properties can be also used to obtain valuable information. These properties include fluorescence anisotropy, polarization, and fluorescence-resonance energy transfer.³¹⁻³² Fluorescent molecules are commonly imaged using fluorescence microscopy. Specific probes, often antibodies, are labelled with a fluorescent dye and when bound to proteins or other molecules of interest provide useful molecular-level contrast.³³ Fluorescence-based imaging techniques have improved to a point that they may be used to obtain images through shallow tissue depths of up to a few centimeters.³⁴ Range and interference from autofluorescence are improved by utilizing fluorescent moieties that emit light in the near-infrared (NIR) range of the electromagnetic spectrum.³⁴ NIR wavelengths are poorly absorbed by water and hemoglobin, so use of NIR-emitting probes improves signal penetration and reduce changes in signal due to physiological composition.³⁴ While *in vitro* techniques typically use labelled antibodies, and fluorophores with varying optical properties, *in vivo* applications require differently specialized reagents. Antibodies generally have poor bioavailability and biodistribution, and have serious problems with immunogenicity and cost-effectiveness. Fluorophore-labelled small molecules that are specific for molecular targets would

be significantly more suitable for this type of application. Specialized fluorophores and vehicles for fluorophore delivery represent an area of ongoing research.³⁵⁻³⁸

Several types of probes can be developed for use in *in vivo* and cellular imaging. For example, probes can be intrinsically fluorescent and directly target a particular protein. These probes can be designed to either bind reversibly with very high affinity, or through covalent bonds linking the probe to the target protein. The probe can also be “activatable”, where a fluorescent moiety is covalently linked to a quencher, which limits fluorescence activity until it is covalently modified, such as through the action of an enzyme. Activatable probes could be used to quantify enzymatic activity if they can act as substrates for a particular enzyme that would remove the quencher and enable a fluorescent signal. This type of action would be advantageous because it amplifies the amount of signal produced, as multiple probes could be activated by a single enzyme molecule. This type of activated probe could be designed with properties that would trap the fluorescent moiety, once cleaved from the quencher, inside of the cell or an organelle of interest, so that the signal remains “trapped” and more efficiently localized to the area of interest.^{35, 39}

2.4. BIOMARKERS FOR CERVICAL CANCER AND ASSOCIATED HYPOXIA

2.4.1. Cervical cancer etiology and human papillomavirus

Cervical cancer is the second most common type of cancer affecting women worldwide, with an estimated 529,409 new cases developing in 2008.⁴⁰ Human papillomavirus (HPV) infection has been firmly established as the cause of nearly 100% of cervical cancer cases.⁴⁰ HPV infection also contributes to the development of a significant number of observed cases of anogenital and head & neck carcinomas.⁴⁰ While over 100 HPV subtypes have been identified, a specific subset of these are considered oncogenic or likely oncogenic. HPV-16 and HPV-18 are

two of the most common oncogenic subtypes and account for over 70% of cervical neoplasias.⁴⁰⁻
⁴² While the role of HPV in the etiology of cervical cancer is quite clear, it is also understood that there are other factors influencing the progression from oncogenic HPV infection to persistence and cervical neoplasia. Though at least half of sexually active women will develop an HPV infection during their life, most infections are cleared within a period of several months from the time of infection, with no apparent lasting pathology. However, in some cases, infection with an oncogenic HPV subtype becomes persistent and leads to the development of either a progressive low-grade intraepithelial lesion, or a high-grade intraepithelial lesion, either of which may develop into invasive cervical cancer over a period of one to several decades^{40-41, 43} if left untreated.^{40-41, 43}

In developed nations, the implementation of screening programs such as cytological Pap tests has enabled early detection and treatment of precancerous and cancerous lesions, lowering the incidence of invasive cervical cancer as well as its mortality rate.^{42, 44-46} The recent development, approval, and implementation of prophylactic vaccination programs for certain HPV subtypes is also expected to further decrease the burden of this disease.⁴⁷ Despite current successes in early diagnosis and treatment of cervical cancer, there remains a clinical application for rapid, specific, and selective diagnostic and prognostic tools. Current cytology-based methods are quite selective and specific in the detection of abnormal lesions, though there is still potential for improvement on this front. Cytology-based methods are intrinsically subjective, require complex laboratory infrastructure, and do not necessarily provide a large amount of information about the risk status of the HPV infection and the lesion, which would be useful in order to better direct therapy.^{42, 44-46} With the discovery of the role of HPV in cervical cancer, and subsequent research, many clinical diagnostic tests have been developed that are designed to

test specifically for HPV.^{42, 44-45} Despite the high specificity of some of these tests, many of them are poorly selective for high-risk HPV infections showing signs of precancerous lesion. This is likely due to cross-reactivity with other subtypes of HPV.⁴² The presence of an oncogenic subtype of HPV does not necessarily predict that an individual is developing or has developed a cancerous lesion. Current clinical recommendations do not include HPV-specific testing as part of the screening regimen for young women (ages 18-29) due to these concerns.⁴² However, combining screening with cytological tests is recommended for women between the ages of 30 and 40.⁴²

Current re-examination of the screening methodologies used for HPV-related cervical cancer highlights the need for specific markers able to detect clinically relevant cases of oncogenic HPV, providing prognostic information that can better direct therapy. Pre-clinical and clinical cervical cancer monitoring would also benefit from detection tools that are minimally invasive that can provide a measurable indication of the response of a tumour to a specific therapy or treatment.

2.4.2. Tumour-related hypoxia

Tissue hypoxia arises when a particular part of the body is deprived of an adequate supply of oxygen. Tumour-associated hypoxia is often characteristic of locally advanced solid tumours and is known to play an important role in tumour pathogenesis.⁴⁸ While it is a complex process, tumour hypoxia is generally arises as the result of an imbalance in the supply (through blood perfusion and diffusion) and demand of oxygen.⁴⁸⁻⁴⁹ A hypoxic tumour microenvironment can lead to gene-regulatory changes that alter the cell phenotype.⁴⁸ In patients with advanced cervical cancer, low tumour oxygenation has been repeatedly associated with an aggressive tumour phenotype and poor survival compared to patients whose tumours had normal oxygen

levels.⁵⁰⁻⁵³ Hypoxia can induce paradoxical cellular effects.⁴⁸⁻⁴⁹ While low oxygenation can induce apoptosis and retard cell growth, it can also enhance growth, the formation of metastases, and invasiveness in some cancer cells.^{48-49, 54-55} Hypoxia-inducible factor (HIF) is a family of transcription factors that is responsible for changing gene expression upon exposure to hypoxic conditions.⁴⁹ HIF-1 α is a transcription factor that controls expression of several enzymes, and whose presence is associated with an aggressive cancer phenotype and a poor prognosis for survival in patients.^{48, 54-58}

2.4.3. Carbonic anhydrase IX as a biomarker for cervical cancer

2.4.3. a) Carbonic anhydrase structure and function

Carbonic anhydrases (CAs) are enzymes that catalyze the reversible hydration of carbon dioxide to yield one bicarbonate anion and one proton (shown in Figure 2.4.3-1).⁵⁶⁻⁶²

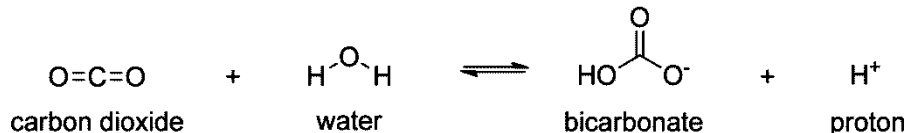


Figure 2.4.3-1—Carbonic anhydrases catalyze the reversible hydration of carbon dioxide. Figure generated using ChemDraw Software (Perkin Elmer).

CAs are crucial to many physiological processes including the maintenance of intra and extracellular pH, as well as gas exchange with red blood cells.⁵⁶⁻⁵⁹ CAs are encoded by 5 unrelated gene families α , β , γ , δ , ζ that have little or no sequence similarity.⁶³ Mammalian CAs belong to the α -class, and exist in many isoforms. α CAs invariably contain Zn^{2+} , which is essential for catalytic activity.⁵⁹⁻⁶² There are 16 mammalian α CA and CA-like isoforms, 13 of which are active.⁶²⁻⁶³ Active mammalian CA isoforms are differentially localized within the cell, with I-III, and VII, XIII isoforms being found predominantly in the cytosol, VA and VB in mitochondria, IV, IX, XII, XIV, and XV associated with the plasma membrane, and VI secreted

into the extracellular space.⁶²⁻⁶³ A complete representation of the localizations of mammalian CA isoforms is shown in Figure 2.4.3-2.

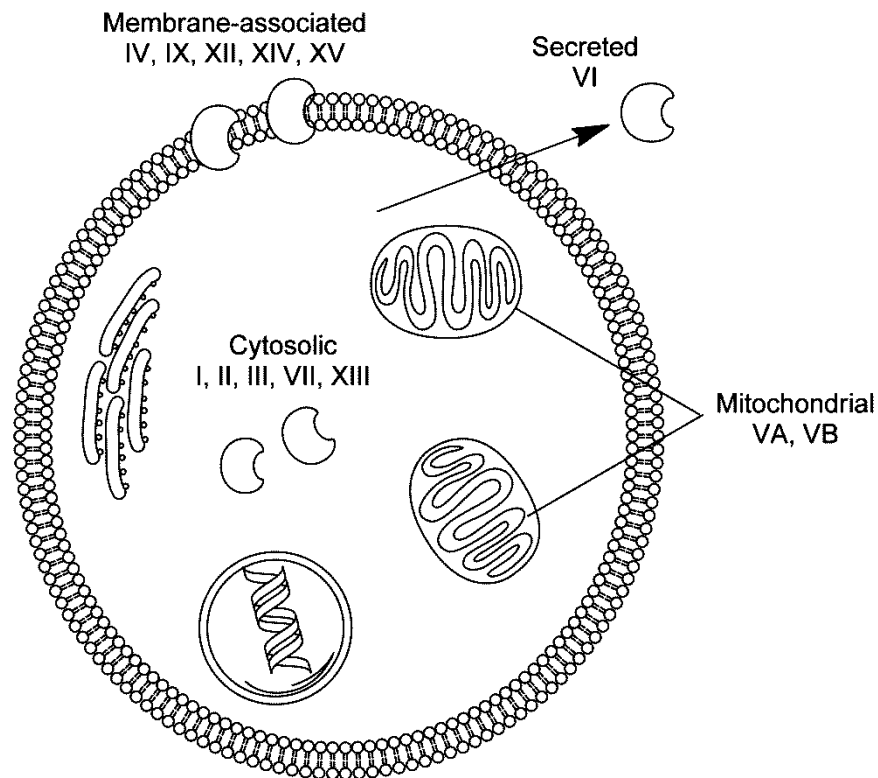


Figure 2.4.3-2-A basic cell representation depicting the relative localization of 13 different mammalian carbonic anhydrase (CA) isoforms. Isoforms IV, IX, XII, XIV, and XV are membrane-associated, VI is secreted, VA and VB are located in the mitochondria, and I-III, and VII, XIII are cytosolic.⁶³ Figure generated using ChemDraw Software (Perkin Elmer).

2.4.3. b) Carbonic anhydrase catalytic mechanism

The catalytic activity and mechanism of CAs has been well characterized, particularly with respect to the α family.^{59-62, 64} α CAs rely on a Zn^{2+} in the active site which is co-ordinated by three histidine residues.⁵⁹⁻⁶⁴ Isoforms that do not contain these key residues, such as VIII, X, and XI, are inactive.⁶³ The following description of carbonic anhydrase mechanism has been summarized from several more detailed reviews^{59-62, 64}, and refers to Figure 3 of a manuscript published in *Bioorganic & Medicinal Chemistry* by Supuran et al (2007) which is reproduced, for reference, as Figure 2.4.3-3 of this text.

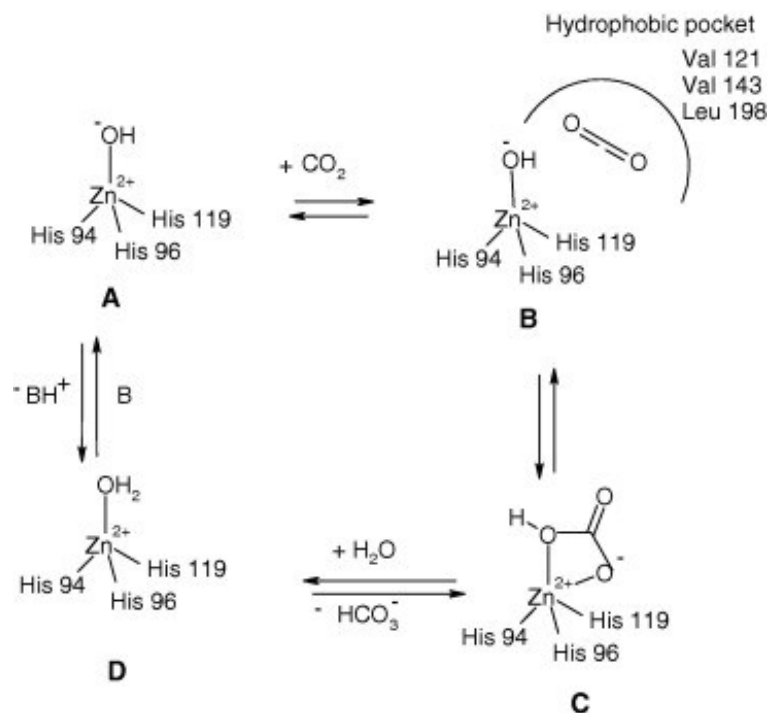


Figure 2.4.3-3—“Schematic representation of the catalytic mechanism for the α -CA catalyzed CO_2 hydration. The hydrophobic pocket for the binding of substrate(s) is shown schematically at step (B)”.⁶²

The zinc is located at the bottom of a deep cleft, and in addition to the three histidines, is coordinated by either a hydroxide (**A**) or water molecule (**D**) in the active and inactive form, respectively. The CO_2 substrate is held by a hydrophobic pocket of the active site (**B**). Hydrogen bonding interactions between the coordinated hydroxide and surrounding residues increase the nucleophilicity of the water and assist in positioning the CO_2 substrate. The nucleophilic hydroxide attacks the CO_2 substrate, forming bicarbonate, which remains coordinated to Zn^{2+} (**C**). A water molecule displaces the bicarbonate, which is released (**D**). The resultant form of the enzyme, which contains zinc coordinated to water (**D**), is inactive. The active, basic form is regenerated through a proton transfer reaction which is assisted by adjacent residues, such as histidine (in the case of isoforms II and IX, among others) which acts as a proton “shuttle”. This specific process, as well as other details of the catalytic mechanism, are described in a number of comprehensive reviews.^{59-62, 64}

2.4.3. c) Evidence for carbonic anhydrase IX as a biomarker

Carbonic anhydrase IX (CAIX) is a CA isoform that is expressed in several tumour types, including cervical cancer, and is rarely expressed in healthy tissues outside of the gastrointestinal tract.^{49, 58} CAIX is tethered to the plasma membrane by a transmembrane domain, with the catalytic domain located in the extracellular space.⁶³ CAIX activity and expression are associated with tumour-related hypoxia. CAIX transcription is induced when hypoxia-inducible transcription factor (HIF)-1 α binds to the promoter of the CAIX gene.⁴⁹ HIF-1 and other transcription factors also regulate other genes responsible for phenotype changes associated with hypoxia.⁴⁹ For example, in low oxygen conditions, cell metabolism shifts from oxidative phosphorylation to high levels of glycolysis, which requires increased expression of glycolytic enzymes and glucose transporters.⁴⁹

CAIX expression is thought to occur as a part of this adaptive response in advanced carcinomas for maintaining neutral intracellular pH while there is a high rate of aerobic glycolysis, with the by-product being extracellular acidification.^{49, 56, 58, 65} Acidification of the extracellular tumor microenvironment appears to contribute to an aggressive tumour phenotype.⁴⁹ There is increasing evidence that CAIX expression in tumours correlates significantly with an increased invasiveness, increased likelihood of metastasis, and a poor overall survival.^{50, 54, 66} A study by Loncaster et al (2001) demonstrated that in addition to being expressed in 79% of human cervical tumours sampled, CAIX is a significant prognostic indicator for disease-free and metastasis-free survival in a manner that is independent of disease stage.⁵⁰ The authors suggest that CAIX, in combination with information about disease stage could be predictive of factors like response to therapy and disease outcome. Using a much larger sample, Woelber et al (2011) further corroborated these findings, demonstrating that nearly 82% of

specimens tested expressed significant levels of CAIX, and that samples expressing moderate to high levels of CAIX were significantly correlated with advanced tumour stage, greater invasion depth, and undifferentiated tumour grade.⁶⁶ There was also a noted correlation between the number of metastatic lymph nodes present and expression of CAIX. These and other investigations have led CAIX to become the target of imaging probe and anticancer drug development.⁶¹ Many investigations have identified selective and potent inhibitors of CAIX.^{36, 61, 67-68} A 2011 investigation demonstrated that decreased CAIX expression and CAIX inhibition separately resulted in decreased metastasis and decreased tumour growth using cell culture and animal models.⁶⁸ Fluorescent inhibitors and other imaging agents have also been identified and experimentally tested using cell culture and animal models.^{36, 69-74} In 2005, Cecchi et al reported a series of fluorescent sulfonamides that act as dual imaging probes and therapeutic inhibitors of CAIX, and demonstrated that these compounds specifically targeted hypoxic cells expressing CAIX.³⁶ CAIX is clearly a high priority target for anti-cancer therapeutics. However, its expression and activity have also been shown to provide information about tumour pathology that is highly relevant to diagnostic and prognostic applications.^{49, 61} The development of probes for prognostic and diagnostic imaging and response to therapy, particularly for use in pre-clinical research, are important and ongoing areas of research.^{49, 61}

2.5. RECEPTOR TYROSINE KINASES AS DRUG TARGETS

2.5.1. Structure and function of receptor tyrosine kinases

Receptor tyrosine kinases (RTKs) are transmembrane proteins that respond to extracellular ligands by catalyzing intracellular autophosphorylation and transphosphorylation on tyrosine residues, leading to signal transduction across cell membrane and initiating a variety of biochemical pathways.⁷⁵

2.5.2. Eph receptors

The Eph family of receptor tyrosine kinases consists of sixteen members, fourteen of which have been identified in humans. Eph receptors belong to a subgroup, A (EphA receptors, EphA1-EphA10) or B (EphB receptors, EphB1-EphB6), based on their tendency to interact with ephrin-A or ephrin-B ligands, respectively.⁷⁶ Six type-A ephrins (ephrin-A - ephrin-A6) are tethered to the cell surface via a glycosylphosphatidylinositol (GPI) linker, while type-B ephrins (ephrin-B1 – ephrin-B3) possess a transmembrane domain linked to an intracellular domain, which can initiate downstream signaling events in a phenomenon known as “reverse signaling”(Figure 2.5.3-1).⁷⁷⁻⁸⁰ The term “reverse signaling” is used to refer specifically to the signal that is propagated in the direction of the ephrin ligand, differentiating it from the activity of the Eph receptor itself.

All Eph receptors share a conserved sequence of structural domains. The extracellular portion of these receptors consists of a globular ephrin-binding domain, a cysteine-rich region and two fibronectin type-III repeats. The intracellular portion is linked to the extracellular domain by one transmembrane domain, and consists of a juxtamembrane region followed sequentially by the catalytic (tyrosine kinase) domain, as well as SAM and PDZ domains.⁷⁷⁻⁸⁸ Kinase domains of Eph receptors share a high degree of structure and sequence homology with catalytic domains of other tyrosine kinases, including insulin receptor kinase (IRK) and epidermal growth factor receptor (EGFR).⁸⁹⁻⁹⁰

The juxtamembrane segment (JMS) is a short (approximate 10 amino-acid) sequence that is located N-terminally on the kinase domain of Eph receptors. The kinase domain contains one N-terminal and one C-terminal lobe. The N-terminal lobe generally consists of a 5-stranded β sheet that adopts a twisted structure (strands denoted by β 1 to β 5), along with one α -helix (α C).

α A and α B are short helical regions that are found in the JMS, and the nomenclature here reflects that used to previously describe the EphB2 structure (PDB: 1JPA).⁹¹ The C-terminal lobe consists mainly of helices, and is slightly larger than the N-terminal lobe. The N- and C-terminal lobes are connected by a short “hinge” linker, which forms a flexible inter-lobe cleft. Structural biology studies involving insulin receptor kinase (IRK), EphB2, and EphA4 have described the mechanistic details of kinase activation and catalysis in substantial detail.^{86, 91-93} Several key regions within the β -sheet and connecting loops are responsible for coordinating ATP within the active site, while the C-terminal lobe is associated with coordinating the peptide substrate. A flexible activation (A) loop structure spans the cleft between the two lobes, and plays a key role in regulating kinase activity. Several conserved tyrosine residues in the JMS and A-loop are known to be sites for trans- and auto-phosphorylation. Phosphotyrosines, which are the products of the autophosphorylation activity, can sometimes serve as key docking sites for downstream effector molecules possessing SH2 homology domains. The phosphorylation of specific tyrosine residues, including Tyr984, can also produce intramolecular conformational changes that are responsible for activating the kinase by promoting the catalytically competent conformation. In an autoinhibited kinase, the dephosphorylated A-loop acts as a pseudo-substrate inhibitor, blocking access of the peptide substrate to its binding site. Upon phosphorylation of A-loop tyrosines, however, the A-loop changes conformation and substrate access to the active site is restored. The EphB2 crystal structure (PDB:1JPA) demonstrates that in the autoinhibited state, the JMS associates closely with the N-terminal lobe, perturbing the position of catalytic residues and preventing the activation loop from assuming an “active” conformation.⁹¹ The JMS associates with both the N- and C-terminal lobes, likely restricting inter-lobe flexibility. Auto/trans phosphorylation was postulated to destabilize these interactions, promoting the active

conformation. NMR and X-ray crystal structure data obtained for EphB2 and EphA4 constructs support these hypotheses, and suggest that the kinase domain dynamically adopts a range of conformations in solution, and increased flexibility is associated with increased catalytic ability.⁹¹ Similar conformations have been observed for the insulin-like growth factor receptor 1, indicating a conserved functional mechanism. Li and co-authors postulate that the β -sheet/ α C cleft could be targeted by small molecules which would displace (or equivalent residue as several other RTKs, including Ephs, possess equivalent residues/interactions in these key regions) and activate the receptor.⁸⁹

There are several studies in the literature, which suggest that it is possible to use a small molecule to activate an isolated tyrosine kinase, possibly by affecting one of the mechanisms described above. In 2001, DMAQ-B1, a small molecule isolated from a tropical microbial extract, was found to specifically activate IRK in an insulin-independent manner.⁹⁴ It was proposed that the autoinhibitory conformation of the IR kinase was altered upon DMAQ-B1 binding. Protease sensitivity experiments suggested that DMAQ-B1 altered the conformation of IRK in regions proximal to the ATP-binding site. The same year, TLK16998, another small-molecule IRK activator, was found to sensitize cells to insulin, acting specifically on the kinase domain.⁹⁵ While these activities have been characterized *in vitro* and *in vivo*, the exact mechanism of activation remains unclear.

These previous investigations highlighted the potential for the development of allosteric tyrosine kinase activators. Due to their critical and somewhat ubiquitous role in cell signaling and development, receptor tyrosine kinases have been extensively studied, largely in the context of therapeutic inhibition. Though the inhibition of some specific tyrosine kinase-mediated signaling pathways seems to be desirable in the context of anti-cancer drug development, the

complex action of Eph receptors and other tyrosine kinases in cell regulation begs for the identification of both positive and negative modulators for these enzymes. The use of such modulators in research can provide valuable insights on the role of Ephs in specific cellular processes by enabling researchers to specifically activate these enzymes *in vitro* and *in vivo* for the purpose of investigating their roles in cellular processes. Novel therapeutic approaches could also make use of specific Eph activators, as the activity of these receptors is crucial for the positive or negative regulation of various biological processes, including bone resorption, blood vessel growth, cell differentiation, and cancer invasiveness.^{77-80, 82, 87-88, 96-97} To date, there are no specific small-molecule allosteric activators of the Eph receptor family that have been reported. The high degree of structural and functional conservation between tyrosine kinases makes it reasonable to expect that mechanisms exploited for drug-development in Ephs would also be useful in the context of related proteins.^{75, 86, 89, 91-92, 94, 98-99}

2.5.3. EphB4 and ephrins-B2 signalling in cancer

EphB4 is a receptor tyrosine kinase whose endogenous ligand is the transmembrane protein ephrin-B2.⁸² These two proteins interact in the extracellular space between adjacent cells, initiating a bidirectional signal in their host cells.⁷⁷ The signalling that occurs as a result of Eph receptor activation is normally termed “forward signalling” while ephrin signalling is referred to as “reverse signalling” (shown in Figure 2.5.3-1).

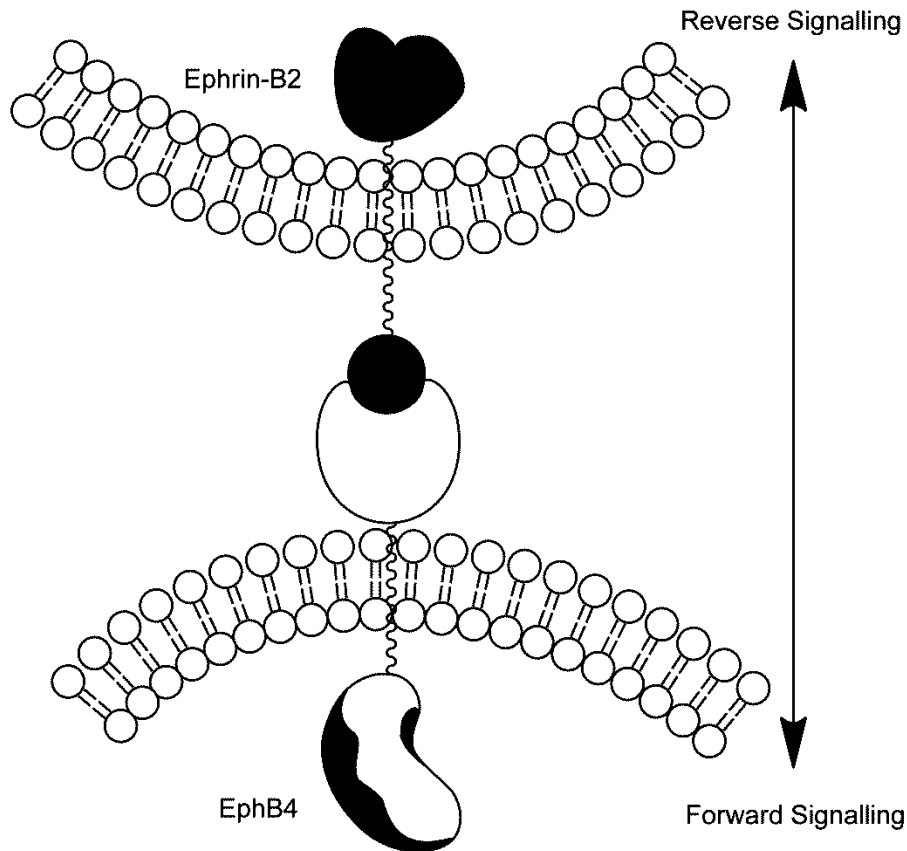


Figure 2.5.3-1-EphB4 receptors and their endogenous ligand, ephrin-B2, can interact between two neighboring cells and trigger a bidirectional signal. Stimulated activity and signalling that occurs in the EphB4-containing cell is termed “forward signalling” while activity that is stimulated in ephrin-B2 direction is termed “reverse signalling”. Figure generated using ChemDraw Software (Perkin Elmer).

EphB4 is able to dimerize and oligomerize with other EphB receptors, promoting the tyrosine kinase activity of these enzymes.⁷⁸ The EphB4 receptor and its ephrin-B2 ligand are critically involved in the development of new blood vessels from existing cardiovascular structures in a process known as *angiogenesis*.^{79-80, 82, 85, 87, 96, 100} EphB4 expression is significantly upregulated in some types of cancer including breast, gastric, and colon carcinoma.¹⁰¹⁻¹⁰³ Differentially high EphB4 expression has also been observed in cells lining vasculature.^{82, 101} Disrupting EphB4/Ephrin-B2 binding has been suggested as a viable chemotherapeutic approach that retards tumor growth by inhibiting vessel development, therefore depriving cancer cells of nutrients.^{85, 101, 104} Monoclonal antibodies that target the EphB4 ephrin-binding domain and impair EphB4-

ephrin-B2 signalling have been shown to inhibit angiogenesis and growth in human tumour xenografts in mice.¹⁰⁴ Several small-molecule inhibitors of the EphB4 kinase domain have been reported^{83, 105-112}, but there has yet to be a selective EphB4 kinase inhibitor approved for clinical treatment. Nevertheless, previous success in therapeutically targeting tyrosine kinases suggests that this approach is warranted in biologically relevant proteins of this family.⁷⁵ Novel selective EphB4 kinase inhibitors are desirable lead compounds to be further optimized for use as chemotherapeutic drugs. In addition, the EphB4 ephrin-binding domain can also be selectively targeted with the aim of disrupting EphB4-ephrin signalling. Currently, only peptides mimicking the ephrin-B2 interaction sequence have been identified as ligands for this domain.^{84, 113} Previous success in RTK-based therapies suggests that this approach is warranted in biologically relevant proteins of this family.⁷⁵ EphB4 specific molecules could also have potential roles in cancer diagnosis and treatment. EphB4 involvement in cancer development and prognosis makes it a potentially useful biomarker for functional imaging modalities like (PET).¹¹⁴ Probes labelled with positron-emitting isotopes could be used with PET to visualize tissues expressing high levels of EphB4, indicating the presence of cancerous cells. Information about the expression of EphB4 in a tumour could also provide prognostic information, based on the observation that high EphB4 expression in some tumours is associated with a higher potential for metastasis and fatal disease.⁸⁸

2.6. SUMMARY & SCOPE

The fields of drug design and medicinal chemistry have expanding horizons with the advent of novel computational techniques that can aid in the discovery and development process, and have an increased scope, as molecular imaging techniques expand the applications of targeted small molecules.

My project aims to use a series of computational techniques, including high-throughput molecular docking and scoring, against two biologically relevant targets, EphB4 and CAIX, with the ultimate aim of developing novel therapeutic or diagnostic agents. For CAIX probe development, ligand binding site preference will be estimated in order to predict whether the probe is likely to be isoform specific, and whether it is likely to interfere with enzymatic activity. Certain applications, such as evaluating response to an existing therapy in an experimental setting, may require that the probe does not interfere with activity of the molecule of interest (CAIX). In the process of attempting to discover selective molecules for these targets, this project also aims to develop a protocol for incorporating imaging agents into the screening process. Fluorescence imaging, in particular, will be utilized with an approach for producing large computational libraries of fluorescently labelled compounds, so that these can be screened and examined with the computational techniques available.

The overall scope of this project has been segmented into three distinct chapters. The first chapter outlines the development of a protocol for designing and computationally screening libraries of fluorescently labelled compounds. The second describes how a library of fluorescent compounds was used to identify potential optical probes for CAIX. The third describes the use of virtual ligand screening for identifying putative ligands for EphB4, and subsequent experimental validation of selected of candidates.

2.7. REFERENCES

1. Luscombe, N. M.; Greenbaum, D.; Gerstein, M., What is bioinformatics? A proposed definition and overview of the field. *Methods Inf. Med.* **2001**, *40* (4), 346-358.

2. National Center for Biotechnology Information: U. S. National Library of Medicine Growth of GenBank and WGS. <http://www.ncbi.nlm.nih.gov/genbank/statistics> (accessed August 2013).
3. Metzker, M. L., Sequencing technologies—the next generation. *Nat. Rev. Genet.* **2009**, *11* (1), 31-46.
4. McEntyre, J.; Ostell, J., The NCBI Handbook. National Center for Biotechnology Information (US): Bethesda (MD), 2002. <http://www.ncbi.nlm.nih.gov/books/NBK21101/>.
5. Waller, C. L.; Oprea, T. I.; Giolitti, A., et al., Three-dimensional QSAR of human immunodeficiency virus (I) protease inhibitors. 1. A CoMFA study employing experimentally-determined alignment rules. *J. Med. Chem.* **1993**, *36* (26), 4152-4160.
6. Kurup, A.; Mekapati, S. B.; Garg, R., et al., HIV-1 protease inhibitors: a comparative QSAR analysis. *Curr. Med. Chem.* **2003**, *10* (17), 1679-1688.
7. Katritzky, A. R.; Oliferenko, A.; Lomaka, A., et al., Six-membered cyclic ureas as HIV-1 protease inhibitors: a QSAR study based on CODESSA PRO approach. Quantitative structure-activity relationships. *Bioorg. Med. Chem. Lett.* **2002**, *12* (23), 3453-3457.
8. Di Santo, R.; Costi, R.; Artico, M., et al., Design, synthesis and QSAR studies on N-aryl heteroarylisopropanolamines, a new class of non-peptidic HIV-1 protease inhibitors. *Bioorg. Med. Chem.* **2002**, *10* (8), 2511-2526.
9. Verma, J.; Khedkar, V. M.; Coutinho, E. C., 3D-QSAR in drug design--a review. *Curr. Top. Med. Chem.* **2010**, *10* (1), 95-115.
10. Scior, T.; Medina-Franco, J. L.; Do, Q. T., et al., How to recognize and workaround pitfalls in QSAR studies: a critical review. *Curr. Med. Chem.* **2009**, *16* (32), 4297-4313.

11. Kubinyi, H., QSAR and 3D QSAR in drug design Part 1: methodology. *Drug Discov. Today* **1997**, 2 (11), 457-467.
12. Kubinyi, H., QSAR and 3D QSAR in drug design Part 2: applications and problems. *Drug Discov. Today* **1997**, 2 (12), 538-546.
13. Dudek, A. Z.; Arodz, T.; Galvez, J., Computational methods in developing quantitative structure-activity relationships (QSAR): a review. *Combinatorial Chem. High Throughput Screening* **2006**, 9 (3), 213-228.
14. Kindt, T.; Morse, S.; Gotschlich, E., et al., Structure-based strategies for drug design and discovery. *Nature* **1991**, 352, 581.
15. Verlinde, C. L. M. J.; Hol, W. G. J., Structure-based drug design: progress, results and challenges. *Structure* **1994**, 2 (7), 577-587.
16. Davis, A. M.; Teague, S. J.; Kleywegt, G. J., Application and Limitations of X-ray Crystallographic Data in Structure-Based Ligand and Drug Design. *Angew. Chem. Int. Ed.* **2003**, 42 (24), 2718-2736.
17. Yang, S.-Y., Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discov. Today* **2010**, 15 (11), 444-450.
18. Kitchen, D. B.; Decornez, H.; Furr, J. R., et al., Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* **2004**, 3 (11), 935-949.
19. Mayo, S. L.; Olafson, B. D.; Goddard, W. A., DREIDING: a generic force field for molecular simulations. *J. Phys. Chem.* **1990**, 94 (26), 8897-8909.
20. Floriano, W. B.; Vaidehi, N.; Zamanakos, G., et al., HierVLS hierarchical docking protocol for virtual ligand screening of large-molecule databases. *J. Med. Chem.* **2004**, 47 (1), 56-71.

21. Ramjan, Z. H.; Raheja, A.; Floriano, W. B. In *A cluster-aware graphical user interface for a virtual ligand screening tool*, Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE, IEEE: 2008; pp 4102-4105.
22. Brady, G. P.; Stouten, P. F., Fast prediction and visualization of protein binding pockets with PASS. *J. Comput. Aided Mol. Des.* **2000**, *14*, 383-401.
23. Dadgar, S.; Ramjan, Z.; Floriano, W. B., Paclitaxel Is an Inhibitor and Its Boron Dipyrromethene Derivative Is a Fluorescent Recognition Agent for Botulinum Neurotoxin Subtype A. *J. Med. Chem.* **2013**, *56* (7), 2791-2803.
24. Floriano, W. B.; Vaidehi, N.; Goddard, W. A., Making sense of olfaction through predictions of the 3-D structure and function of olfactory receptors. *Chem. Senses* **2004**, *29* (4), 269-290.
25. Li, X.; Bachmanov, A. A.; Maehashi, K., et al., Sweet taste receptor gene variation and aspartame taste in primates and other species. *Chem. Senses* **2011**, *36* (5), 453-475.
26. Vaidehi, N.; Schlyer, S.; Trabanino, R. J., et al., Predictions of CCR1 chemokine receptor structure and BX 471 antagonist binding followed by experimental validation. *J. Biol. Chem.* **2006**, *281* (37), 27613-27620.
27. Willmann, J. K.; van Bruggen, N.; Dinkelborg, L. M., et al., Molecular imaging in drug development. *Nat. Rev. Drug Discov.* **2008**, *7* (7), 591-607.
28. Gambhir, S. S., Molecular imaging of cancer with positron emission tomography. *Nat. Rev. Cancer* **2002**, *2* (9), 683-693.
29. Hornak, J. P., *The Basics of MRI. Interactive Learning Software*: Henrietta, New York, United States, 2011. <http://www.cis.rit.edu/htbooks/mri/> (accessed 2013).

30. Skoog, D. A., *Analytical Chemistry: An Introduction*. 7th ed.; Saunders College Pub.: Fort Worth, 1999.
31. Jameson, D. M.; Ross, J. A., Fluorescence polarization/anisotropy in diagnostics and imaging. *Chem. Rev.* **2010**, *110* (5), 2685-2708.
32. Jares-Erijman, E. A.; Jovin, T. M., FRET imaging. *Nat. Biotechnol.* **2003**, *21* (11), 1387-1395.
33. Alberts, B.; Johnson, A.; Lewis, J., et al., *Molecular Biology of the Cell*. 4 ed.; Garland Science: New York, 2002.
34. Ntziachristos, V.; Bremer, C.; Weissleder, R., Fluorescence imaging with near-infrared light: new technological advances that enable in vivo molecular imaging. *Eur. Radiol.* **2003**, *13* (1), 195-208.
35. Weissleder, R.; Tung, C. H.; Mahmood, U., et al., In vivo imaging of tumors with protease-activated near-infrared fluorescent probes. *Nat. Biotechnol.* **1999**, *17* (4), 375-378.
36. Cecchi, A.; Hulikova, A.; Pastorek, J., et al., Carbonic Anhydrase Inhibitors. Design of Fluorescent Sulfonamides as Probes of Tumor-Associated Carbonic Anhydrase IX That Inhibit Isozyme IX-Mediated Acidification of Hypoxic Tumors. *J. Med. Chem.* **2005**, *48* (15), 4834-4841.
37. Heilemann, M.; van de Linde, S.; Mukherjee, A., et al., Super-resolution imaging with small organic fluorophores. *Angew. Chem. Int. Ed.* **2009**, *48*, 6903-6908.
38. Messerli, S. M.; Prabhakar, S.; Tang, Y., et al., A Novel Method for Imaging Apoptosis Using a Caspase-1 Near-Infrared Fluorescent Probe. *Neoplasia* **2004**, *6* (2), 95-105.
39. Ntziachristos, V., Fluorescence molecular imaging. *Annu. Rev. Biomed. Eng.* **2006**, *8*, 1-33.

40. World Health Organization: Department of Reproductive Health and Research. Cervical cancer, human papillomavirus (HPV), and HPV vaccines 2008, p. 1-16.
http://whqlibdoc.who.int/hq/2008/WHO_RHR_08.14_eng.pdf.
41. Baseman, J. G.; Koutsky, L. A., The epidemiology of human papillomavirus infections. *J. Clin. Virol.* **2005**, *32 Suppl 1*, S16-24.
42. Luhn, P.; Wentzensen, N., HPV-based Tests for Cervical Cancer Screening and Management of Cervical Disease. *Curr. Obstet. Gynecol. Rep.* **2013**, *2*, 76-85.
43. Rana, M. M.; Huhtala, H.; Apter, D., et al., Understanding long-term protection of human papillomavirus vaccination against cervical carcinoma: Cancer registry-based follow-up. *Int. J. Cancer* **2013**, *132*, 2833-2838.
44. Carter, R. L.; Kang, L.; Darcy, K. M., et al., A modified Latent Class Model assessment of human papillomavirus-based screening tests for cervical lesions in women with atypical glandular cells: a Gynecologic Oncology Group study. *Cancer Causes Control* **2012**, *23*, 2013-2021.
45. Cuzick, J.; Bergeron, C.; von Knebel Doeberitz, M., et al., New technologies and procedures for cervical cancer screening. *Vaccine* **2012**, *30 Suppl 5*, F107-116.
46. Mayrand, M.-H.; Duarte-Franco, E.; Coutlée, F., et al., Randomized controlled trial of human papillomavirus testing versus Pap cytology in the primary screening for cervical cancer precursors: design, methods and preliminary accrual results of the Canadian cervical cancer screening trial (CCCaST). *Int. J. Cancer* **2006**, *119*, 615-623.
47. Markowitz, L. E.; Tsu, V.; Deeks, S. L., et al., Human papillomavirus vaccine introduction--the first five years. *Vaccine* **2012**, *30 Suppl 5*, F139-148.

48. Vaupel, P.; Mayer, A., Hypoxia in cancer: significance and impact on clinical outcome. *Cancer Metastasis Rev.* **2007**, *26*, 225-239.
49. Pastorekova, S.; Ratcliffe, P. J.; Pastorek, J., Molecular mechanisms of carbonic anhydrase IX-mediated pH regulation under hypoxia. *BJU Int.* **2008**, *101* (s4), 8-15.
50. Loncaster, J. A.; Harris, A. L.; Davidson, S. E., et al., Carbonic anhydrase (CA IX) expression, a potential new intrinsic marker of hypoxia: correlations with tumor oxygen measurements and prognosis in locally advanced carcinoma of the cervix. *Cancer Res.* **2001**, *61* (17), 6394-6399.
51. Hoekel, M.; Schlenger, K.; Höckel, S., et al., Tumor hypoxia in pelvic recurrences of cervical cancer. *Int. J. Cancer* **1998**, *79* (4), 365-369.
52. Höckel, M.; Vorndran, B.; Schlenger, K., et al., Tumor oxygenation: a new predictive parameter in locally advanced cancer of the uterine cervix. *Gynecol. Oncol.* **1993**, *51* (2), 141-149.
53. Höckel, M.; Schlenger, K.; Aral, B., et al., Association between tumor hypoxia and malignant progression in advanced cancer of the uterine cervix. *Cancer Res.* **1996**, *56* (19), 4509-4515.
54. Liao, S.; Rodgers, W. H.; Kauderer, J., et al., Carbonic anhydrase IX (CA-IX) and high-risk human papillomavirus (H-HPV) as diagnostic biomarkers of cervical dysplasia/neoplasia in Japanese women with a cytologic diagnosis of atypical glandular cells (AGC): a Gynecologic Oncology Group (GOG) Study. *Br. J. Cancer* **2011**, *104*, 353-360.
55. Shin, H.-J.; Rho, S. B.; Jung, D. C., et al., Carbonic anhydrase IX (CA9) modulates tumor-associated cell migration and invasion. *J. Cell Sci.* **2011**, *124*, 1077-1087.

56. Swietach, P.; Patiar, S.; Supuran, C. T., et al., The Role of Carbonic Anhydrase 9 in Regulating Extracellular and Intracellular pH in Three-dimensional Tumour Cell Growths. *J. Biol. Chem.* **2009**, *284* (30), 20299-20310.
57. Hulikoca, A.; Vaughan-Jones, R. D.; Swietach, P., Dual Role of CO₂/HCO₃⁻ Buffer in the Regulation of Intracellular pH of Three-dimensional Tumor Growths. *J. Biol. Chem.* **2011**, *286* (16), 13815-13826.
58. Swietach, P.; Vaughan-Jones, R. D., Regulation of tumor pH and the role of carbonic anhydrase 9. *Cancer Metastasis Rev.* **2007**, *26*, 299-310.
59. Khalifah, R. G., The Carbon Dioxide Hydration Activity of Carbonic Anhydrase. *J. Biol. Chem.* **1971**, *246* (8), 2561-2573.
60. Silverman, D. N.; Lindskog, S., The catalytic mechanism of carbonic anhydrase: implications of a rate-limiting protolysis of water. *Acc. Chem. Res.* **1988**, *21* (1), 30-36.
61. Supuran, C. T., Carbonic anhydrases: novel therapeutic applications for inhibitors and activators. *Nat. Rev. Drug Discov.* **2008**, *7* (2), 168-181.
62. Supuran, C. T.; Scozzafava, A., Carbonic anhydrases as targets for medicinal chemistry. *Biorg. Med. Chem.* **2007**, *15* (13), 4336-4350.
63. Aspatwar, A.; Tolvanen, M. E. E.; Parkkila, S., Phylogeny and expression of carbonic anhydrase-related proteins. *BMC Mol. Biol.* **2010**, *11* (25).
64. Lindskog, S., Structure and mechanism of carbonic anhydrase. *Pharmacol. Ther.* **1997**, *74* (1), 1-20.
65. Chiche, J.; Ilc, K.; Laferrière, J., et al., Hypoxia-inducible carbonic anhydrase IX and XII promote tumor cell growth by counteracting acidosis through the regulation of the intracellular pH. *Cancer Res.* **2009**, *69* (1), 358-368.

66. Woelber, L.; Kress, K.; Kersten, J. F., et al., Carbonic anhydrase IX in tumor tissue and sera of patients with primary cervical cancer. *BMC Cancer* **2011**, *11*, 12.
67. Vullo, D.; Franchi, M.; Gallori, E., et al., Carbonic anhydrase inhibitors: inhibition of the tumor-associated isozyme IX with aromatic and heterocyclic sulfonamides. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1005-1009.
68. Lou, Y.; McDonald, P. C.; Oloumi, A., et al., Targeting Tumor Hypoxia: Suppression of Breast Tumor Growth and Metastasis by Novel Carbonic Anhydrase IX Inhibitors. *Cancer Res.* **2011**, *71* (9), 3364-3376.
69. Tafreshi, N. K.; Bui, M. M.; Bishop, K., et al., Noninvasive detection of breast cancer lymph node metastasis using carbonic anhydrases IX and XII targeted imaging probes. *Clin. Cancer Res.* **2012**, *18*, 207-219.
70. Groves, K.; Bao, B.; Zhang, J., et al., Synthesis and evaluation of near-infrared fluorescent sulfonamide derivatives for imaging of hypoxia-induced carbonic anhydrase IX expression in tumors. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 653-657.
71. Dubois, L.; Lieuwes, N. G.; Maresca, A., et al., Imaging of CA IX with fluorescent labelled sulfonamides distinguishes hypoxic and (re)-oxygenated cells in a xenograft tumour model. *Radiother. Oncol.* **2009**, *92*, 423-428.
72. Dubois, L.; Douma, K.; Supuran, C. T., et al., Imaging the hypoxia surrogate marker CA IX requires expression and catalytic activity for binding fluorescent sulfonamide inhibitors. *Radiother. Oncol.* **2007**, *83*, 367-373.
73. Bao, B.; Groves, K.; Zhang, J., et al., In Vivo Imaging and Quantification of Carbonic Anhydrase IX Expression as an Endogenous Biomarker of Tumor Hypoxia. *PLoS ONE* **2012**, *7*, e50860.

74. Akurathi, V.; Dubois, L.; Lieuwes, N. G., et al., Synthesis and biological evaluation of a ^{99m}Tc-labelled sulfonamide conjugate for in vivo visualization of carbonic anhydrase IX expression in tumor hypoxia. *Nucl. Med. Biol.* **2010**, *37*, 557-564.
75. Levitzki, A.; Gazit, A., Tyrosine kinase inhibition: an approach to drug development. *Science* **1995**, *267* (5205), 1782-1789.
76. Lee, Y.-C.; Perren, J. R.; Douglas, E. L., et al., Investigation of the expression of the EphB4 receptor tyrosine kinase in prostate carcinoma. *BMC Cancer* **2005**, *5*, 119.
77. Surawska, H.; Ma, P. C.; Salgia, R., The role of ephrins and Eph receptors in cancer. *Cytokine Growth Factor Rev.* **2004**, *15* (6), 419-433.
78. Stein, E.; Lane, A.; Cerretti, D., et al., Eph receptors discriminate specific ligand oligomers to determine alternative signaling complexes, attachment, and assembly responses. *Genes Dev.* **1998**, *12*, 667-678.
79. Heroult, M.; Schaffner, F.; Augustin, H. G., Eph receptor and ephrin ligand-mediated interactions during angiogenesis and tumor progression. *Exp. Cell Res.* **2006**, *312* (5), 642-650.
80. Fuller, T.; Korff, T.; Kilian, A., et al., Forward EphB4 signaling in endothelial cells controls cellular repulsion and segregation from ephrinB2 positive cells. *J. Cell Sci.* **2003**, *116* (12), 2461-2470.
81. Shi, S.; Liu, J.; Joshi, S. B., et al., Biophysical characterization and stabilization of the recombinant albumin fusion protein sEphB4-HSA. *J. Pharm. Sci.* **2012**, *101*, 1969-1984.
82. Noren, N. K.; Lu, M.; Freeman, A. L., et al., Interplay between EphB4 on tumor cells and vascular ephrin-B2 regulates tumor growth. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (15), 5583-5588.

83. Martiny-Baron, G.; Holzer, P.; Billy, E., et al., The small molecule specific EphB4 kinase inhibitor NVP-BHG712 inhibits VEGF driven angiogenesis. *Angiogenesis* **2010**, *13* (3), 259-267.
84. Koolpe, M.; Burgess, R.; Dail, M., et al., EphB receptor-binding peptides identified by phage display enable design of an antagonist with ephrin-like affinity. *J. Biol. Chem.* **2005**, *280* (17), 17301-17311.
85. Kertesz, N.; Krasnoperov, V.; Reddy, R., et al., The soluble extracellular domain of EphB4 (sEphB4) antagonizes EphB4-EphrinB2 interaction, modulates angiogenesis, and inhibits tumor growth. *Blood* **2006**, *107* (6), 2330-2339.
86. Hubbard, S. R.; Mohammadi, M.; Schlessinger, J., Autoregulatory mechanisms in protein-tyrosine kinases. *J. Biol. Chem.* **1998**, *273* (20), 11987-11990.
87. Augustin, H. G.; Reiss, Y., EphB receptors and ephrinB ligands: regulators of vascular assembly and homeostasis. *Cell Tissue Res.* **2003**, *314* (1), 25-31.
88. Alam, S. M.; Fujimoto, J.; Jahan, I., et al., Coexpression of EphB4 and ephrinB2 in tumour advancement of ovarian cancers. *Br. J. Cancer* **2008**, *98* (4), 845-851.
89. Li, S.; Covino, N. D.; Stein, E. G., et al., Structural and biochemical evidence for an autoinhibitory role for tyrosine 984 in the juxtamembrane region of the insulin receptor. *J. Biol. Chem.* **2003**, *278*, 26007-26014.
90. Zhang, X.; Gureasko, J.; Shen, K., et al., An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell* **2006**, *125*, 1137-1149.
91. Wybenga-Groot, L. E.; Baskin, B.; Ong, S. H., et al., Structural basis for autoinhibition of the Ephb2 receptor tyrosine kinase by the unphosphorylated juxtamembrane region. *Cell* **2001**, *106*, 745-757.

92. Wiesner, S.; Wybenga-Groot, L. E.; Warner, N., et al., A change in conformational dynamics underlies the activation of Eph receptor tyrosine kinases. *EMBO J.* **2006**, *25*, 4686-4696.
93. Jura, N.; Endres, N. F.; Engel, K., et al., Mechanism for activation of the EGF receptor catalytic domain by the juxtamembrane segment. *Cell* **2010**, *137*, 1-25.
94. Salituro, G. M.; Pelaez, F.; Zhang, B. B., Discovery of a small molecule insulin receptor activator. *Recent Prog. Horm. Res.* **2001**, *56*, 107-126.
95. Manchem, V. P.; Goldfine, I. D.; Kohanski, R. a., et al., A novel small molecule that directly sensitizes the insulin receptor in vitro and in vivo. *Diabetes* **2001**, *50*, 824-830.
96. Adams, R.; Wilkinson, G.; Weiss, C., et al., Roles of ephrinB ligands and EphB receptors in cardiovascular development: demarcation of arterial/venous domains, vascular morphogenesis, and sprouting angiogenesis. *Genes Dev.* **1999**, *13*, 295-306.
97. Zhao, C.; Irie, N.; Takada, Y., et al., Bidirectional ephrinB2-EphB4 signaling controls bone homeostasis. *Cell Metab.* **2006**, *4*, 111-121.
98. Schindler, T.; Bornmann, W.; Pellicena, P., et al., Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science* **2000**, *289* (5486), 1938-1942.
99. Binns, K. L.; Taylor, P. P.; Sicheri, F., et al., Phosphorylation of Tyrosine Residues in the Kinase Domain and Juxtamembrane Region Regulates the Biological and Catalytic Activities of Eph Receptors. *Mol. Cell. Biol.* **2000**, *20*, 4791-4805.
100. Tonini, T.; Rossi, F.; Claudio, P. P., Molecular basis of angiogenesis and cancer. *Oncogene* **2003**, *22* (42), 6549-6556.
101. Kumar, S. R.; Singh, J.; Xia, G., et al., Receptor tyrosine kinase EphB4 is a survival factor in breast cancer. *Am. J. Pathol.* **2006**, *169* (1), 279-293.

102. Li, M.; Zhao, Z. W.; Zhang, Y., et al., Over-expression of Ephb4 is associated with carcinogenesis of gastric cancer. *Dig. Dis. Sci.* **2011**, *56* (3), 698-706.
103. Stephenson, S.; Slomka, S.; Douglas, E., et al., Receptor protein tyrosine kinase EphB4 is up-regulated in colon cancer. *BMC Mol. Biol.* **2001**, *2* (1), 15.
104. Krasnoperov, V.; Kumar, S. R.; Ley, E., et al., Novel EphB4 monoclonal antibodies modulate angiogenesis and inhibit tumor growth. *Am. J. Pathol.* **2010**, *176* (4), 2029-2038.
105. Lafleur, K.; Huang, D.; Zhou, T., et al., Structure-based optimization of potent and selective inhibitors of the tyrosine kinase erythropoietin producing human hepatocellular carcinoma receptor B4 (EphB4). *J. Med. Chem.* **2009**, *52* (20), 6433-6446.
106. Mitchell, S. A.; Danca, M. D.; Blomgren, P. A., et al., Imidazo[1,2-a]pyrazine diaryl ureas: Inhibitors of the receptor tyrosine kinase EphB4. *Bioorg. Med. Chem. Lett.* **2009**, *19* (24), 6991-6995.
107. Miyazaki, Y.; Nakano, M.; Sato, H., et al., Design and effective synthesis of novel templates, 3,7-diphenyl-4-amino-thieno and furo-[3,2-c]pyridines as protein kinase inhibitors and in vitro evaluation targeting angiogenetic kinases. *Bioorg. Med. Chem. Lett.* **2007**, *17* (1), 250-254.
108. Hu, S. X.; Soll, R.; Yee, S., et al., Metabolism and pharmacokinetics of a novel Src kinase inhibitor TG100435 ([7-(2,6-dichloro-phenyl)-5-methyl-benzo[1,2,4]triazin-3-yl]-[4-(2-pyrroli din-1-yl-ethoxy)-phenyl]-amine) and its active N-oxide metabolite TG100855 ([7-(2,6-dichloro-phenyl)-5-methylbenzo[1,2,4]triazin-3-yl]-{4-[2-(1-oxy-pyrrolidin-1-yl)-ethoxy]-phenyl}-amine). *Drug Metab. Dispos.* **2007**, *35* (6), 929-936.

109. Kolb, P.; Kipouros, C. B.; Huang, D., et al., Structure-based tailoring of compound libraries for high-throughput screening: discovery of novel EphB4 kinase inhibitors. *Proteins* **2008**, *73* (1), 11-18.
110. Bardelle, C.; Barlaam, B.; Brooks, N., et al., Inhibitors of the tyrosine kinase EphB4. Part 3: identification of non-benzodioxole-based kinase inhibitors. *Bioorg. Med. Chem. Lett.* **2010**, *20* (21), 6242-6245.
111. Bardelle, C.; Coleman, T.; Cross, D., et al., Inhibitors of the tyrosine kinase EphB4. Part 2: structure-based discovery and optimisation of 3,5-bis substituted anilinopyrimidines. *Bioorg. Med. Chem. Lett.* **2008**, *18* (21), 5717-5721.
112. Bardelle, C.; Cross, D.; Davenport, S., et al., Inhibitors of the tyrosine kinase EphB4. Part 1: Structure-based design and optimization of a series of 2,4-bis-anilinopyrimidines. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 2776-2780.
113. Xiong, C.; Huang, M.; Zhang, R., et al., In Vivo Small-Animal PET/CT of EphB4 Receptors Using ⁶⁴Cu-Labeled Peptide. *J. Nucl. Med.* **2011**, *52* (2), 241-248.
114. Lee, J. H.; Rosen, E. L.; Mankoff, D. A., The role of radiotracer imaging in the diagnosis and management of patients with breast cancer: part 1--overview, detection, and staging. *J. Nucl. Med.* **2009**, *50* (4), 569-581.

Chapter 3.

Methodology

3.1. OVERVIEW

This chapter will review the methods used in the subsequent chapters, focusing on the theory underlying each method. Methods will be described more briefly in their corresponding manuscripts, and specific procedural details will be provided there.

3.2. COMPUTATIONAL METHODS

3.2.1. Force field-based methods

Recall that force fields are potential energy functions that can be used to approximate the energy of protein-ligand interactions and to minimize the geometry of molecule structures. Force fields are particularly common and useful in applications, such as structural biology and computer-assisted drug design where the molecules of interest are very large. While quantum mechanics-based approaches are more accurate at describing molecular systems, they remain prohibitively expensive from a computational standpoint, for many applications, particularly those of a high-throughput nature. In these situations, force fields can be used to get informative results that require fewer resources. Force fields describe different forces within a molecular system using classical mechanics. The various intra and intermolecular forces are approximated individually in separate terms, and summed together to yield the final potential energy function. There are many available force fields, many of which have been designed and optimized for specific types of molecules or applications. Dreiding is the all-atoms force field used by the HierVLS scheme. Its versatility is useful for handling both protein, and ligand structures during the course of docking and scoring. A basic overview the Dreiding methodology is given in the

following paragraphs, summarized from the publication by Mayo et al, which completely describes the design and parameterization of this force field.¹ While the functions used to describe different atomic interactions may vary between different force fields, the general scheme described for Dreiding is representative of what is used by other force fields.²⁻³

In Dreiding, the potential function is given as the summation of the terms for bonded and non-bonded molecular interactions (Equation 3.2.1-1):

Equation 3.2.1-1 - Force field potential energy functions are summations of both bonded and non-bonded terms.

$$E_{Potential} = E_{Bonded} + E_{Non-Bonded}$$

Interactions between bonded atoms include bond stretching, bond angle bend, dihedral angle torsion, and inversion terms. Due to the tendency for starting structures to have very approximate geometries Dreiding by default describes bond stretching using simple harmonic oscillation (Equation 3.2.1-2):

Equation 3.2.1-2 - Simple harmonic oscillation function for bond stretching term. R denotes distance relative to equilibrium (R_e), k_e denotes the force-constant.

$$E = \frac{1}{2}k_e(R - R_e)^2$$

Alternatively, Dreiding can use a more accurate Morse function for further refinement of the initial simple harmonic oscillator function, if specified. Bond angle bending is described using a harmonic (cosine form) function (

Equation 3.2.1-3) with equilibrium angles (θ_j^0) obtained from standard reference structures.

Equation 3.2.1-3 - Function describing the bond angle bending for two bonds between atoms IJ and JK (J atom shared). C_{IJK} is a function of the force constant and the equilibrium angle: $C_{IJK} = K_{IJK}/(\sin \theta_j^0)^2$.

$$E_{IJK} = \frac{1}{2} C_{IJK} [\cos \theta_{IJK} - \cos \theta_j^0]^2$$

When the equilibrium geometry is linear (180°), this is instead approximated by Equation

3.2.1-4.

Equation 3.2.1-4 - Bond angle bending for molecules with linear equilibrium geometry is given by this potential.

$$E_{IJK} = K_{IJK} [1 + \cos \theta_{IJK}]^2$$

Dihedral angle torsion about four atoms (IJKL) is described as a function of the dihedral angle (φ), the equilibrium angle (φ_{JK}^0), the periodicity (n_{JK}), and the barrier to rotation (V_{JK}) (Equation 3.2.1-5).

Equation 3.2.1-5 - Dihedral angle torsion is described about four atoms (IJKL) with respect to dihedral angle (φ), periodicity (n_{JK}), and rotation barrier (V_{JK}). The equilibrium angle is denoted by φ_{JK}^0 .

$$E_{IJKL} = \frac{1}{2} V_{JK} (1 - \cos [n_{JK} (\varphi - \varphi_{JK}^0)])$$

The inversion term (for atoms bonded to three other atoms) is given by Equation 3.2.1-6.

Equation 3.2.1-6 - Cases where one atom, I, is bonded to three other atoms, JKL, may require the inclusion of an inversion term which is given as a function of C_I , and the angle between the IL bond and the JIK plane (ψ). C is a function of the force constant (K) and the equilibrium bond-plane angle: $C_I = K_I / (\sin \psi_I^0)^2$.

$$E_{IJKL} = \frac{1}{2} C_I (\cos \Psi - \cos \Psi_I^0)^2$$

This function differs from those used by certain force fields intended for biological simulations, which treat inversion as improper torsion. Non-bonded terms included describe van der Waals

forces, electrostatic interactions, and hydrogen bonding. Van der Waals and electrostatic interactions are only calculated between atoms that are not already described by a bonded interaction. Van der waals forces are, by default, described using a 12-6 Leonard-Jones potential given by Equation 3.2.1-7.

Equation 3.2.1-7 - A 12-6 Leonard-Jones potential is used to calculate van der Waals forces as a function of interatomic distance (r).

$$E_{vdw} = \frac{A}{r^{12}} - \frac{B}{r^6}$$

Electrostatic forces are calculated using the Coulombic expression shown in Equation 3.2.1-8.

Equation 3.2.1-8 - Electrostatic forces are calculated using a Coulombic interaction expression. $Q_{i,j}$ refer to atomic charges, ϵ , the dielectric constant, and R_{ij} the interatomic distance. 322.0637 is used as a conversion factor for kcal.

$$E_Q = \frac{(322.0637)Q_iQ_j}{\epsilon R_{ij}}$$

Hydrogen atoms that are bonded to very electronegative atoms and that might participate in hydrogen bond donation are specially typed. When these hydrogen atoms are in close proximity to a hydrogen bond acceptor, van der Waals and electrostatic forces are calculated for this set of atoms, and a specific hydrogen-bonding potential function (Equation 3.2.1-9) is applied.

Equation 3.2.1-9 - A potential used to describe hydrogen bonding in Dreiding as a function of the distance between the donor and acceptor atoms (R_{DA}) and the bond angle between donor and acceptor atoms (θ_{DA}). D_{hb} and R_{hb} values can be derived from quantum mechanical calculations or empirically.

$$E_{hb} = D_{hb} [5(R_{hb}/R_{DA})^{12} - 6(R_{hb}/R_{DA})^{10}] \cos^4(\theta_{DHA})$$

Generally, Dreiding is a relatively simple force field but has been designed to be versatile enough to be used with various types of molecules. Many commercial and non-commercial software packages include different force fields that are designed for use with particular types of

systems. For example, the MOE software used in different subsequent chapters features a set of Merck Molecular Force Fields which are specifically designed and parameterized for use with small molecules.⁴

3.2.2. Hierarchical virtual ligand screening

Hierarchical virtual ligand screening (HierVLS) is a structure-based method for high-throughput docking and scoring of large ligand libraries against 3D protein target structures.⁵ HierVLS identifies and filters promising binding conformations through a series of steps to ensure that the most computational time (and therefore, the best estimates of binding) are provided for the most promising conformers of a library.⁵

HierVLS is executable through a graphical, interactive user interface (GUI).⁶ The HierVLS protocol begins with an input defining the available binding sites for the given protein structure. These binding sites can be user-specified (such as might be the case for an experimentally determined structure with a well-defined site of interest) or can be generated using an algorithm that scans the surface of the protein structure for cavities. In the following chapters, HierVLS is implemented with such an algorithm, PASS, to rapidly identify potential sites within the protein structure using a geometric algorithm that characterizes available binding cavities by their shape, size, and depth with respect to the surface of the target structure.^{5, 7-8}

The user supplies molecule structure files for both the protein target and ligand library to be screened. Atoms must be appropriately typed for handling by the force field and solvation model. Gasteiger atomic charges are set for all ligands to be docked.⁹ The initial set of docked ligand conformations are generated using the Dock4.0 software package.¹⁰ Dock4.0 calculates an energy grid to determine the energy contribution to docked conformers from the protein

including any water molecules that are part of the protein structure provided. The initial conformations are evaluated based on the percentage of buried surface area and the number of hydrogen bonding interactions predicted between the protein and ligand structures. Conformers that do not have a user-defined percentage of buried surface area are eliminated. The remaining best conformers (a specified number, N) are carried through to the next step, Level 1. During Level 1, each ligand structure is energy minimized while keeping the protein structure fixed using the Dreiding force field without accounting for solvation effects. The lowest energy conformers from level 1 (usually a small number depending on available computational resources) are re-evaluated using the analytical volume generalized born solvation method (AVGB), which is applied both to the free ligand, and the protein-ligand complex.¹¹ Final binding scores are obtained by calculating the difference in energy between the protein-ligand complex and the solvated protein and ligand alone as shown in Equation 3.2.2-1.

Equation 3.2.2-1 – The binding affinities for the best conformers from Level 1 of HierVLS are calculated by subtracting both the energy of the solvated protein alone and the solvated ligand alone from the energy of the solvated complex. The AVGB solvation approach is used to calculate solvation energies.

$$Affinity = E_{solvated\ complex} - (E_{solvated\ protein\ alone} + E_{solvated\ ligand\ alone})$$

3.3. EXPERIMENTAL METHODS

3.3.1. Fluorescence-based binding assays

3.3.1. a) Fluorescence polarization

Fluorescence polarization is a phenomenon that occurs because fluorophores are excited maximally when their absorption transition vectors are aligned in parallel with respect to the excitation source.¹² Linearly polarized light can be used to selectively excite optimally aligned molecules.¹² The motion of the fluorophore prior to the emission (fluorescence) will affect the

degree to which this polarization is maintained. When a fluorescent molecule is free to rotate rapidly in solution and is excited by polarized light, the emitted light tends to be scattered in multiple planes, resulting in low polarization.¹³ However, if the fluorophore becomes bound into a larger, less mobile complex, there is less movement prior to light emission, and a high polarization value is obtained.¹²⁻¹³ This principle can be used to compare the fluorescence polarization of an unbound molecule with a bound complex to identify an intermolecular binding event.¹²⁻¹³ Fluorescence polarization can be calculated as a function of the fluorescence intensity parallel and perpendicular to the plane of excitation (Equation 3.3.1-1).¹²

Equation 3.3.1-1– Fluorescence polarization is calculated as the ratio of the difference between the fluorescence intensity parallel and perpendicular to the excitation plane to the sum of these intensities.

$$P = \frac{F_{\parallel} - F_{\perp}}{F_{\parallel} + F_{\perp}}$$

3.3.1. b) Fluorescence resonance energy transfer

Fluorescence resonance energy transfer (FRET) is another mechanism through which intermolecular interactions can be probed using fluorophores. FRET requires donor and acceptor moieties that are in close proximity. When excited, a donor molecule can transfer this energy to the acceptor molecule, which then emits a detectable photon.¹² FRET donors and acceptor pairs can consist of two different molecules, but the acceptor must be excited at emission wavelengths of the donor molecule.¹² The relatively close proximity required for FRET to occur (10-100Å) makes this a useful tool to detect binding interactions between two molecules, each containing a donor or acceptor pair.¹²

3.4. REFERENCES

1. Mayo, S. L.; Olafson, B. D.; Goddard, W. A., DREIDING: a generic force field for molecular simulations. *J. Phys. Chem.* **1990**, *94* (26), 8897-8909.
2. Kitchen, D. B.; Decornez, H.; Furr, J. R., et al., Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* **2004**, *3* (11), 935-949.
3. Meng, X.-Y.; Zhang, H.-X.; Mezei, M., et al., Molecular docking: a powerful approach for structure-based drug discovery. *Curr. Comput. Aided Drug Des.* **2011**, *7* (2), 146.
4. Halgren, T. A., Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490-519.
5. Floriano, W. B.; Vaidehi, N.; Zamanakos, G., et al., HierVLS hierarchical docking protocol for virtual ligand screening of large-molecule databases. *J. Med. Chem.* **2004**, *47* (1), 56-71.
6. Ramjan, Z. H.; Raheja, A.; Floriano, W. B. In *A cluster-aware graphical user interface for a virtual ligand screening tool*, Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE, IEEE: 2008; pp 4102-4105.
7. Dadgar, S.; Ramjan, Z.; Floriano, W. B., Paclitaxel Is an Inhibitor and Its Boron Dipyrromethene Derivative Is a Fluorescent Recognition Agent for Botulinum Neurotoxin Subtype A. *J. Med. Chem.* **2013**, *56* (7), 2791-2803.
8. Brady, G. P.; Stouten, P. F., Fast prediction and visualization of protein binding pockets with PASS. *J. Comput. Aided Mol. Des.* **2000**, *14*, 383-401.
9. Gasteiger, J.; Marsili, M., Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219-3228.

10. Ewing, T. J. A.; Kuntz, I. D., Critical evaluation of search algorithms for automated molecular docking and database screening. *J. Comput. Chem.* **1997**, *18* (9), 1175-1189.
11. Zamanakos, G. A fast and accurate analytical method for the computation of solvent effects in molecular simulations. California Institute of Technology, Pasadena, 2002.
12. Johnson, I.; Spence, M. T. Z., *Molecular Probes Handbook: A Guide to Fluorescent Probes and Labeling Technologies*. 11th ed.; Life Technologies Corporation: Carlsbad, California, United States, 2010.
13. Checovich, W. J.; Bolger, R. E.; Burke, T., Fluorescence polarization--a new tool for cell and molecular biology. *Nature* **1995**, *375* (6528), 254-256.

Chapter 4.

Creating databases of fluorescently-labelled compounds for the discovery of target-specific molecular probes*

4.1. INTRODUCTION

Many techniques in molecular biology laboratories demand the use of fluorescent reagents, which often function as optical tracers. These molecules may be intrinsically fluorescent or labelled with a fluorophore moiety. Chemically reactive fluorophores selective for biologically relevant functional groups are commonly used for labeling chemistry, often to tag large biomolecules like antibodies and nucleic acids.¹⁻² The most commonly used reactive fluorophores are designed to covalently label molecules containing a primary amine functional group such as lysine residues found in the peptides or protein structure. N-hydroxysuccinimidyl ester and isothiocyanate derivatives are among the most well characterized reactive fluorophores that are commercially available.¹ Relatively early applications for reactive fluorophores include identifying serum components using labelled antibodies and visualizing cell structures (e.g., actin filaments) using indirect immunofluorescent staining.²⁻⁴ Fluorescently-labelled or conjugated antibodies continue to be used for various direct and indirect immunofluorescence techniques that have applications in fundamental research and diagnostic pathology.²⁻⁶ Although antibodies can be used as selective probes, small molecules are often more desirable for certain applications, including *in vivo* imaging, where the production of antibodies is not feasible, or where size, immunogenicity, or poor pharmacokinetic properties of antibodies becomes

*This chapter represents a collection of work that is intended to be submitted for publication by multiple authors: Rhiannon L. Kamstra, Saedeh Dadgar, John Wigg, and Wely B. Floriano. Screening and experimental testing of candidates for BoNT/A and E6 were performed (and sections authored) by Saedeh Dadgar and John Wigg, respectively.

problematic.⁷ With the introduction of fluorophores that are active in the near-infrared spectral region, *in vivo* fluorescence imaging has become a viable functional technique for visualizing molecular composition of tissues and even dynamic intermolecular processes.⁸⁻⁹ Fluorescent probes can be designed that target specific receptors, that respond to environmental changes (e.g., pH), or that are activated by a specific enzyme.¹⁰⁻¹³

High-throughput virtual screening methods have been used to identify high-affinity ligands for a variety of protein targets.¹⁴⁻¹⁹ Hierarchical virtual ligand screening (HierVLS) is one protocol that uses a structure-based approach to screen large libraries of compounds against biologically relevant target structures, estimating a force-field binding energy for the complex. This method has been validated and used in previous investigations.^{14, 20} HierVLS, and other similar virtual screening methodologies have typically been used for identifying “hit” (lead) compounds intended for later development into therapeutic agents.²⁰⁻²²

In this paper, we demonstrate that the creation and computational screening of fluorescently-labelled small molecule libraries is an efficient strategy for discovering target-specific fluorescent molecular probes. We have created large virtual databases of easily synthesizable fluorescently-labelled small molecules and used virtual screening to identify ligands for two medically relevant protein targets. To demonstrate success, selected virtually identified ligands were synthesized and experimentally validated for predicted binding or non-binding against their respective protein targets.

Our validation screening and testing used a library of 1,153 primary amine-containing small molecules and its fluorescently labelled counterpart, which were screened against model structures of HPV protein E6 and botulinum neurotoxin subtype A (BoNT/A). Each protein

target has a publically available experimentally determined structure, is biologically relevant, and is the subject of ongoing investigation in our lab.

E6 is a cancer-associated protein encoded by human papillomavirus (HPV) that causes ubiquitin-mediated degradation of a key tumor suppressor, p53, in infected cells.²³⁻²⁵ There is evidence that E6, and E6 variants, play a role in conferring a high-risk phenotype to certain HPV strains.²⁴⁻²⁵ Specific detection agents for E6 variants could provide valuable insight into tumor cell pathology *in vivo*, and could also be developed to complement existing diagnostic tools. Botulinum neurotoxin (BoNT), produced by *Clostridium botulinum* bacteria, is one of the most potent toxins, causes food-borne illness (botulism) and has the potential to be used as a biological weapon.²⁶⁻²⁸ BoNT is also the key component of several available therapeutic and cosmetic treatments (e.g., “Botox” by Allergan, Inc.).²⁹⁻³⁰ While some *in vitro* BoNT identification platforms exist, the current standard for confirmatory testing remains the “mouse lethality assay” (live mice are exposed to different concentrations of samples and observed), which is inefficient, takes several days to complete, and is associated with ethical concerns.²⁸⁻²⁹ Particularly in light of its risk both in food products and as a bioterrorism threat, rapid *in vitro* tests for BoNT are currently in high demand.²⁷ In an earlier investigation, a small molecule, paclitaxel was identified as a novel inhibitor of BoNT serotype A using virtual ligand screening (VLS), giving confidence that this target is amenable to targeting with this technique.¹⁴

The experimental validation of our fluorescent-labelling and screening protocol provides the rationale and framework for developing a high-throughput method for generating virtual fluorophore-ligand conjugates, to facilitate screening of large libraries of such compounds. We describe the development and implementation of a software-based script for parsing databases of small molecules containing appropriate reactive functional groups, and generating structures that

are conjugated to user-specified reactive fluorophores. These databases of novel fluorescent conjugates can be virtually screened against protein biomarkers for the rapid identification of target-specific fluorescent molecular probes.

4.2. MATERIALS AND METHODS

4.2.1. Preparation of the test libraries

A library of 1,153 small molecules, each containing a single primary amine functional group, was used as a test virtual screening library. Compounds for this primary amines database (PAD) were downloaded from PubChem by drawing a basic primary amine structure and performing a structural similarity search. Lipinski's "rule of 5" was used as a filter for drug-likeness.³¹⁻³² Solvent and co-ion molecules were removed during processing, and entries containing atoms not parameterized by common force fields were excluded. The database was built and edited using MOE software.³³ The Gasteiger method was used to assign partial charges and an energy minimization was conducted using MMFF94x force field.³⁴⁻³⁵ The database was then saved in a Tripos MOL2 format, with a total of 1,153 entries (supplementary information).

In addition, the PAD was computationally tagged with a Bodipy fluorophore (3-Bodipy-propanoylaminocaproic acid, N-hydroxysuccinimide ester) by manually editing the structures. Conjugate structures were created according to the general reactivity scheme between primary amines and N-hydroxysuccinimidyl ester fluorophore derivatives that is described both in the literature³⁶⁻³⁷ and in the Molecular Probes Handbook.¹ This scheme is shown in Figure 4.2.1-1.

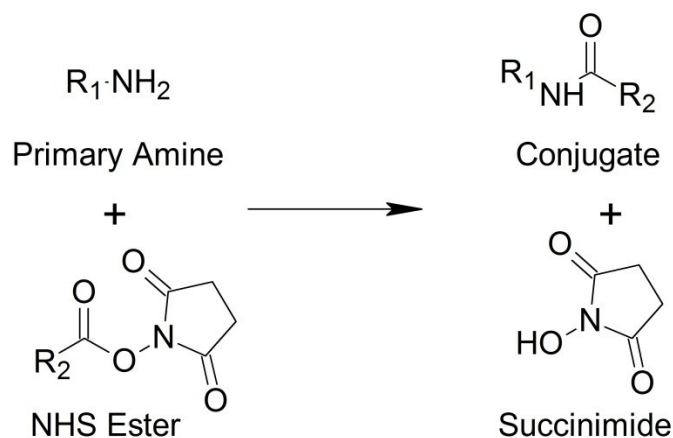


Figure 4.2.1-1-N-hydroxysuccinimidyl ester (NHS ester) reactive fluorophores react with primary amines to form fluorophore/ligand conjugates that are linked via an amide bond.

4.2.2. Protein structure preparation, VLS, and hit selection

4.2.2. a) Protein structure preparation

The experimental structure of the E6 protein (PDB Code: 2FK4) was obtained from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB). Energy minimization was performed on the structure in MOE using the MMFF94x force field.^{33, 35} For molecular docking, the protein structure was converted into Biograf format using Babel and Biograf.v400fsm files were created using an in-house algorithm. YASARA software was used to generate a PDB file with hydrogens removed.³⁸⁻⁴⁰ A model protein structure for botulinum neurotoxin A (BoNT/A) was prepared previously.¹⁴

4.2.2. b) Virtual ligand screening

The BoNT/A and E6 structures were scanned for docking regions using the PASS method.⁴¹ Both structures were screened against the PAD with molecular docking and scoring performed according to the HierVLS approach.²⁰ Docking and scoring of both PAD and Bodipy-tagged PAD against BoNT/A and against E6 was carried out using the same methodology detailed in Dadgar et al (2013) (manuscript in preparation).¹⁴ For the Bodipy-tagged PAD, an

adjusted buried surface area cutoff of 30% (from 70%) was used to accommodate the increased average size of the molecules being docked. Good binding candidates for BoNT/A were selected based on whether they had a calculated binding affinity of one standard deviation higher than the mean binding score for the entire library at particular binding regions believed to be subtype-selective (manuscript in preparation). Binding candidates for E6 were selected if they had a calculated binding energy below a threshold (-78.23kCal/mol) of two standard deviations below the mean binding score for the entire library at a particular binding region (-57.92kCal/mol). The scores of each selected binding candidate from the unlabelled ligand set were compared with those of their labelled counterparts.

4.2.3. Source of chemicals for binding assays and fluorophore conjugation

3-Bodipy-propanoylaminocaproic acid, N-hydroxysuccinimide ester (BODIPY® FL-X, hereafter abbreviated Bodipy-NHS; CAS 217190-09-5; Cat. B2527-40) and aspartame (CAS 22839-47-0; Cat. A7462) were purchased from US Biological and LKT Laboratories Inc., respectively. BoNT/A LC was obtained from US Biological. O-Succinyl-L-homoserine (CAS 1492-23-5) was purchased from Sigma Aldrich (Cat. S7129-25MG; 98% purity TLC).

4.2.4. Synthesis and purity of fluorescent conjugates

Detailed information on the synthesis and the purification of aspartame-Bodipy will be presented in a separate manuscript that is currently in preparation.⁴² Identity and purity of the isolated conjugate were confirmed using mass spectrometry and NMR respectively.⁴²

O-Succinyl-L-homoserine-Bodipy was synthesized by dissolving Bodipy-NHS (5 mg, x mmol), in 500 μ L of dimethylformamide (DMF). The reaction vial was placed, with stirring, in a water/ice bath at 0°C followed by addition of O-succinyl-homoserine (2.7 mg) and triethylamine.

The reaction was flushed with argon and left to proceed for 30 minutes at 0°C after which was allowed to stand at room temperature overnight. The reaction was quenched by addition of distilled water followed by extraction with ethyl acetate. The aqueous phase was frozen and sublimed using a freeze drier to yield the desired product (3 mg). The purity of the product was confirmed using TLC (5:1 dichloromethane:methanol) and the identity confirmed with liquid chromatography mass spectrometry(supplementary information).

4.2.5. Fluorescence polarization overview

Experimental validation of binding was performed using a fluorescence polarization assay. Fluorescence polarization assays rely on the principle that, when excited with polarized light, a static fluorophore will emit light that is polarized along the same plane. Depolarization of the light occurs in proportion to the amount of free rotation of the molecule in solution. Therefore, high polarization values are normally indicative of larger complexes that rotate less rapidly in solution, while low polarization indicates a relatively unrestrained fluorophore. Fluorescence polarization has been examined as a powerful and versatile method that can be used without complex separation or washing, does not require radiolabelled tracers and can be scaled for high-throughput applications.⁴³⁻⁴⁵

4.2.6. BoNT/A fluorescence polarization assay

Aspartame-Bodipy was tested in a fluorescence polarization assay competing with unlabelled aspartame and unlabelled paclitaxel against BoNT/A light chain (LC). Paclitaxel has been shown to inhibit BoNT/A LC with an estimated IC₅₀ of 5.2µM, with confirmed binding in a previous assay.¹⁴ In this experiment aspartame and paclitaxel were dissolved in assay buffer (20 mM HEPES buffer PH: 8.0, 0.1% Tween20, 0.3mM ZnCl₂) and DMSO respectively at the concentration of 1µM. The concentration of aspartame-Bodipy was fixed at its EC₅₀⁴² with a

fixed BoNT/ALC concentration of 2nM. Solvent (DMSO) controls were included. Samples were analyzed in triplicate using a 96-well black microplate (Costar) and a Synergy 4 Microplate Reader (Biotek, Inc) at 37°C, read every 15 minutes for 2.5 hours. A full size 510 nm dichroic mirror was used with excitation and emission ranges of 440-505nm and 515-640 nm, respectively. The excitation and emission filters used were 485/20 nm and 528/20 nm, respectively. The filters and mirrors were used to ensure selective excitation of and emission reading for the Bodipy fluorophore used. Data visualization and non-linear regression analyses were performed using Prism.⁴⁶

4.2.7. Protein E6 intrinsic tryptophan fluorescence assay

A change in intrinsic tryptophan fluorescence was used to measure the binding of O-succinyl-L-homoserine to E6. Protein E6 contains only one tryptophan which is partially solvent accessible in the experimental structure of the unbound protein. Preliminary experiments with an E6-specific antibody (6F4) indicated that binding to E6 significantly quenched the emission associated to E6's single tryptophan. Unlabelled O-succinyl-L-homoserine was tested at 8 concentrations (0.01, 0.1, 1, 3, 10, 25, 100, and 300 μ M) with a fixed concentration of protein E6 (0.03mg/mL). O-succinyl-L-homoserine concentrations were diluted from a 1mM stock solution dissolved in deionized water. E6 protein was accurately measured from the stock concentration (0.201mg/mL). The assay was buffered to a pH of 7.5 using 20mM HEPES, 0.01% tween (v/v) and incubated to a temperature of 37°C. Control wells were included for the unquenched protein using E6 (0.03mg/mL) in buffer alone and for each concentration of O-succinyl-L-homoserine. Samples were plated in triplicate and controls in duplicate in a 96-well black bottom microplate (Nunclon). The antibody 6F4 was used as a positive control at a concentration of 0.4ng/mL. Water, buffer, and sample were added first to all the wells in the microplate. E6 was added last to

sample or control wells, and the plate was incubated in the plate reader at 37°C. Emission spectra for each well were collected from 310nm-700nm at an excitation of 288nm. Fluorescence intensity was read using a probe height of 4mm, a 150µsec delay between wells, and 20 reads per well, every 30 minutes for 60 minutes using a sensitivity value of 100 with Gen5 software (Biotek).

4.2.8. Protein E6 fluorescence polarization assay

O-Succinyl-L-homoserine-Bodipy was tested in a fluorescence polarization assay at three concentrations (1, 10, and 100µM) with a fixed concentration of protein E6 (0.04mg/mL). The assay was buffered to a pH of 7.5 using 20mM HEPES, 0.01% tween (v/v) and incubated to a temperature of 37°C. Control wells were included for the E6 protein without ligand and O-Succinyl-L-homoserine-Bodipy in the absence of protein at each concentration tested. Samples were plated as single wells in a 96-well black bottom microplate (Nunclon). The experiment was conducted using a Synergy 4 Microplate Reader (Biotek) with a 485nm/20 excitation filter, 528nm/20 emission filter, and a 510nm dichromatic mirror. Excitation and emission maxima for the labelled O-succinyl-L-homoserine were determined in preliminary experiments.

Fluorescence polarization was read with a probe height of 4mm, a 350 µsec delay between wells, and 40 reads per well. Data was collected every 15 minutes for 75 minutes using a sensitivity value of 65 using Gen5 software (Biotek). Fluorescence polarization values were calculated with background polarization values discounted. Data analysis, including a 4-parameter non-linear regression, were performed using PRISM⁴⁶.

4.2.9. Fluorophore conjugation script and preparation of a labelled library

We designed a script able to virtually conjugate reactive fluorophores with ligands possessing the correct functional group, according to a pre-defined reactivity scheme. Literature review was used to identify common reactivity schemes for commercially available reactive fluorophores. MOE's scientific vector language (SVL) was chosen as the scripting environment due to the large number of available and relevant built-in functions and its suitability for large databases of small organic molecules.³³ A SMARTS-based (<http://www.daylight.com/>) functional group identification algorithm, included as part of the MOE package, was used to recognize the functional groups involved in the fluorophore-ligand reaction. We designed this first version of the script to identify N-hydroxysuccinimidyl ester derivatives, and to label small molecules containing one primary amine functional group. The small molecule becomes labelled with an amide bond according to the scheme shown in Figure 4.2.1-1.¹ The script accepts user-specified MOE molecular database (MDB) files as input for both fluorophores and ligands. Molecules (either fluorophore or ligand) that do not contain the appropriate functional groups required for the fluorescent labelling reaction are skipped. Conjugate structures are outputted as entries to a new MOE database and assigned molecule names. The script was tested using small databases of fluorophores and ligands to ensure that it correctly matched the specified molecular patterns targeted by the script, and skipped molecules containing "decoy" functional groups (supplementary information). Using this script, the same base compound library containing n primary amines can be conjugated with multiple fluorophores (m) to generate a library of $n \times m$ conjugates.

4.2.10. Base compound library for conjugation

The Molecular Libraries Small Molecule Repository (MLSMR) (<https://mli.nih.gov/mli/compound-repository/mlsmr-compounds/>) was selected as a source of molecular structures. There are three subsets each containing approximately 100,000 compounds, as well as some additional structures, totaling over 400,000 compounds as of June 2013.⁴⁷ The Molecular Libraries Initiative describes the selection criteria for these compounds on their website (<https://mli.nih.gov/mli/compound-repository/mlsmr-compounds/>). PubChem offers structures for download in both 2D and 3D formats. PubChem generates 3D structures for all compounds matching a specific set of criteria (detailed on PubChem3D).⁴⁸ One 3D conformer (if available) was downloaded in SDF format for each MLSMR compound from PubChem3D (<http://pubchem.ncbi.nlm.nih.gov/>), totaling 383,987 structures. The database was refined using MOE software (www.chemcomp.com/). SDF structures were loaded into MOE molecular databases and checked for counter ions and solvent molecules. Gasteiger atomic charges were calculated for each molecule.³⁴ An energy minimization was performed on each database using the MMFF94x force field.³⁵ Chirality was preserved for the downloaded structures.

4.2.11. Generating a library of fluorescent conjugates.

The fluorescent labeling script was used to parse the database of MLSMR compounds from PubChem3D and conjugate all compounds containing one primary amine to one NHS-ester functionalized fluorophore. ATTO680 NHS-ester (PubChemCID 16218508) was used as the amine-reactive fluorophore in this case (Figure 4.2.11-1b).

MOE software (www.chemcomp.com) was used for database preparation and refinement. Conjugates were outputted to a molecular database (MDB format) before being assigned Gasteiger atomic charges and energy minimized using MMFF94x force-field.³⁴⁻³⁵ Conjugates

were subjected to a conformational search using a built-in MOE function (“Conformational Import”), which is a high-throughput method for searching each molecule’s conformational space. Only one conformer was selected for each compound. Stochastic search failure and iteration limits of, respectively, 20 and 250 were used. Superposed conformations with an RMSD value of 0.15 or less were considered identical. Default constraints were used.

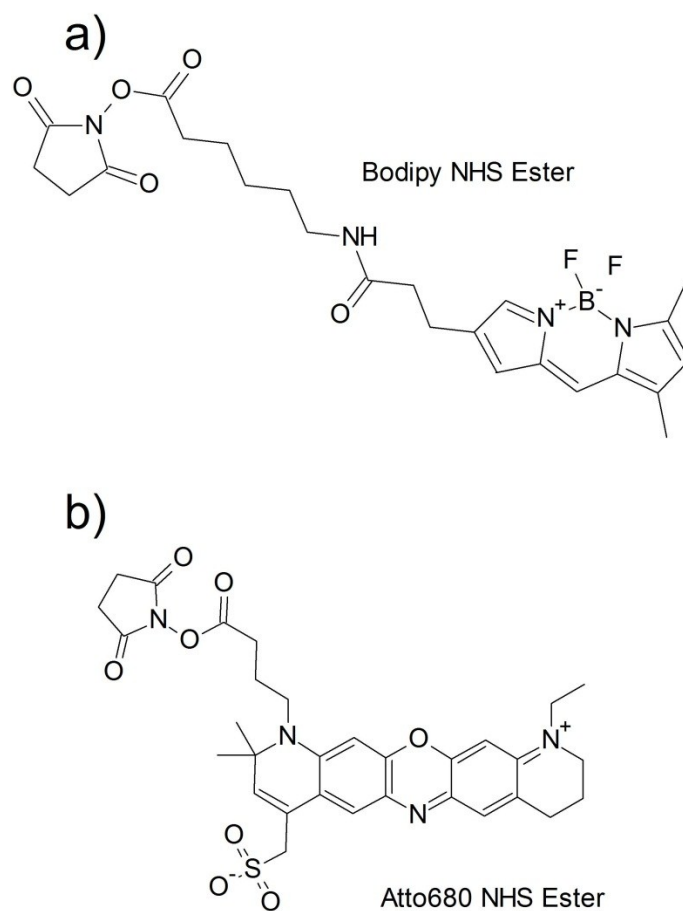


Figure 4.2.11-1–Bodipy NHS ester (a) was the reactive fluorophore used for the experimental synthesis and testing of fluorescent small-molecule conjugates. ATTO680 NHS ester (b) is the reactive dye chosen for computational generation of tagged compound libraries.

4.3. RESULTS

4.3.1. Virtually screening fluorescence-labelled compound libraries is effective at identifying potential fluorescent probes

A library of 1,153 primary amine-containing compounds (PAD) was screened using the HierVLS approach against two biologically relevant protein targets, botulinum toxin A (BoNT/A) and HPV 16 protein E6 (E6). Screening results for the unlabelled PAD against BoNT/A and E6 are shown in Figure 4.3.1-1 and Figure 4.3.1-3. The structures in this library were subsequently tagged with 3-Bodipy-propanoylaminocaproic acid, N-hydroxysuccinimide ester (Bodipy) by manually editing the structures to form an amide bond between primary amine nitrogen and the carbonyl carbon of the N-hydroxysuccinimidyl ester reactive group from the fluorophore (Figure 4.2.1-1).

The Bodipy-labelled PAD was rescreened against both targets (Figure 4.3.1-2 and Figure 4.3.1-4). The Bodipy fluorophore was chosen for virtual labeling because it is commercially available and suited for use with the available optical instruments in our laboratory. Aspartame and aspartame-Bodipy (Figure 4.3.1-5) both were predicted to bind to BoNT/A according to the binding threshold score used. O-Succinyl-L-homoserine was predicted to bind to E6 while its labelled counterpart, O-succinyl-L-homoserine-Bodipy (Figure 4.3.1-6) was not. These test cases were subject to experimental validation to confirm whether predicted binding or non-binding would be replicated in *in vitro* assays.

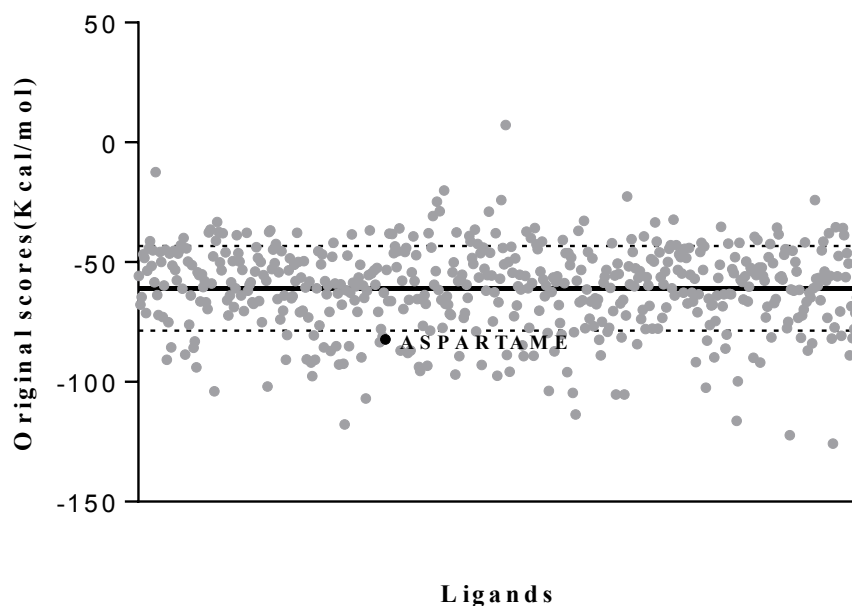


Figure 4.3.1-1—Molecular docking results for unlabelled ligands against binding region 21 in BoNT/A LC. The solid and dotted lines represent mean and \pm SD (standard deviation), respectively. Compounds scoring at or below the mean-SD threshold are expected to bind experimentally to the target. Aspartame was one of the ligands showing high predicted binding affinity against BoNT/A with a score below the binding threshold.

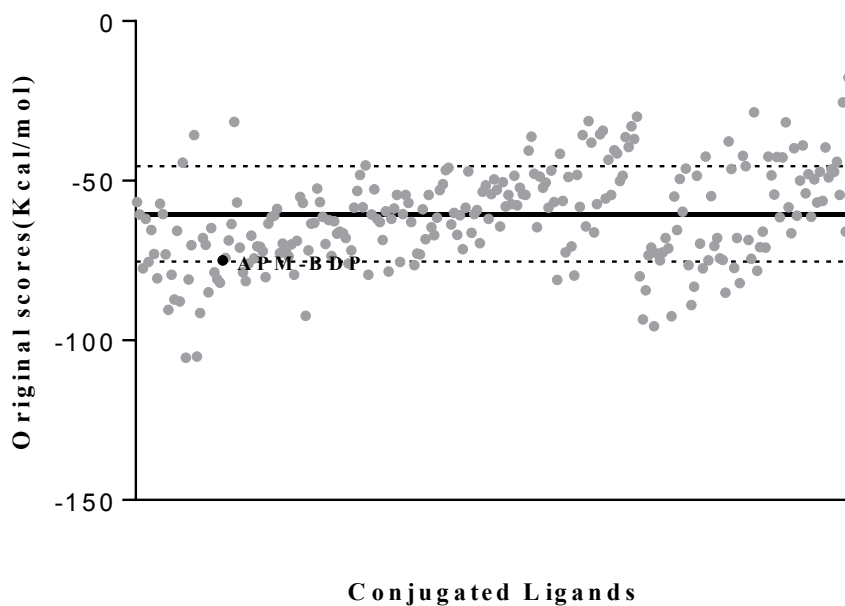


Figure 4.3.1-2—Molecular docking results for Bodipy labelled ligands against binding region 21 in BoNT/A LC. The solid and dotted lines represent mean and \pm SD (standard deviation),

respectively. Aspartame-Bodipy (APM-BDP) scored close to one SD below the mean value (about 0.5kCal/mol higher than SD), and it is expected to bind to BoNT/A.

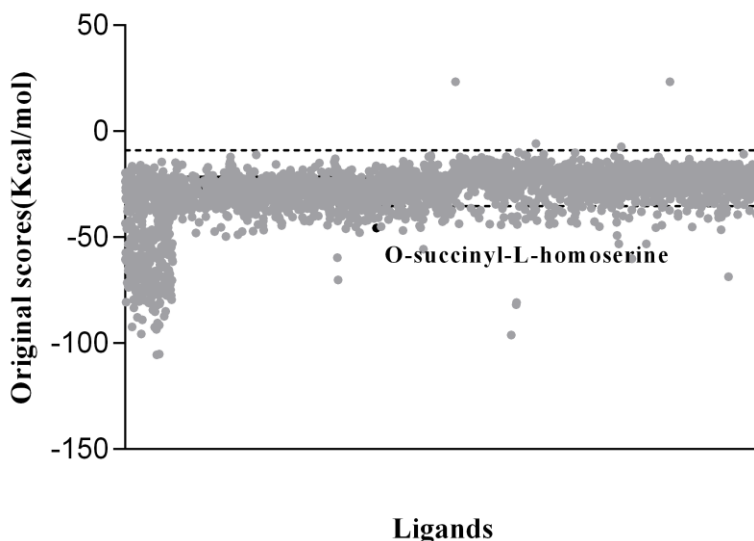


Figure 4.3.1-3—Molecular docking results for unlabelled ligands against protein E6. The solid and dotted lines represent mean and $\pm 2SD$ (standard deviation), respectively. O-Succinyl-L-homoserine scored below the threshold of the mean-2SD, and was selected for experimental testing against the target.

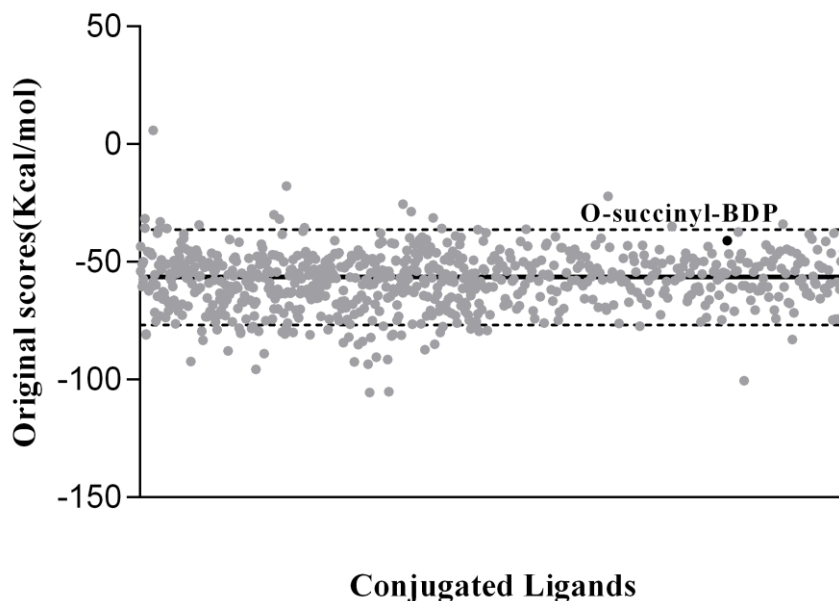


Figure 4.3.1-4—Molecular docking results for Bodipy-labelled ligands against protein E6. The solid and dotted lines represent mean and $\pm 2SD$ (standard deviation), respectively. O-Succinyl-L-homoserine-Bodipy scored relatively poorly, with a predicted binding affinity almost 2SD

worse than the average of all docked compounds and it is, hence, not expected to bind experimentally to E6.

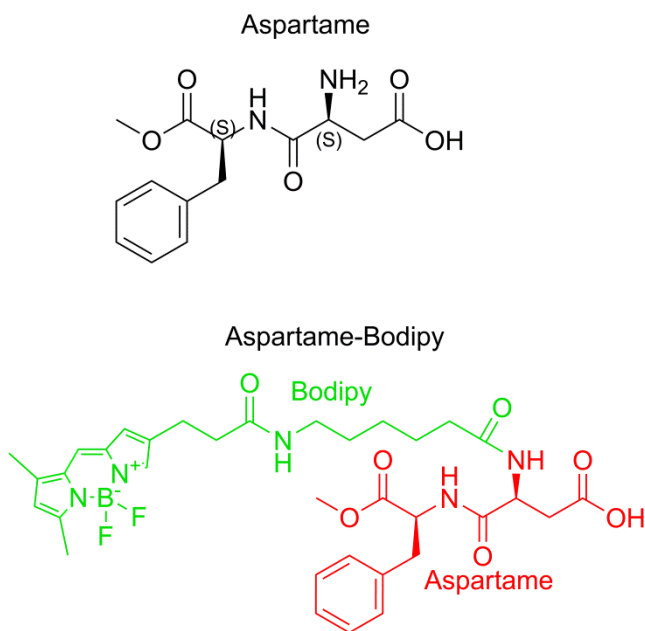


Figure 4.3.1-5—The structures of aspartame (top) and aspartame-Bodipy (bottom) used in this work. Bodipy-labelled aspartame was synthesized by conjugating aspartame to the amine reactive 3-Bodipy-propanoylaminocaproic acid, N-hydroxysuccinimide ester (Bodipy). Both unlabelled compound and conjugate were predicted to bind to BoNT/A.

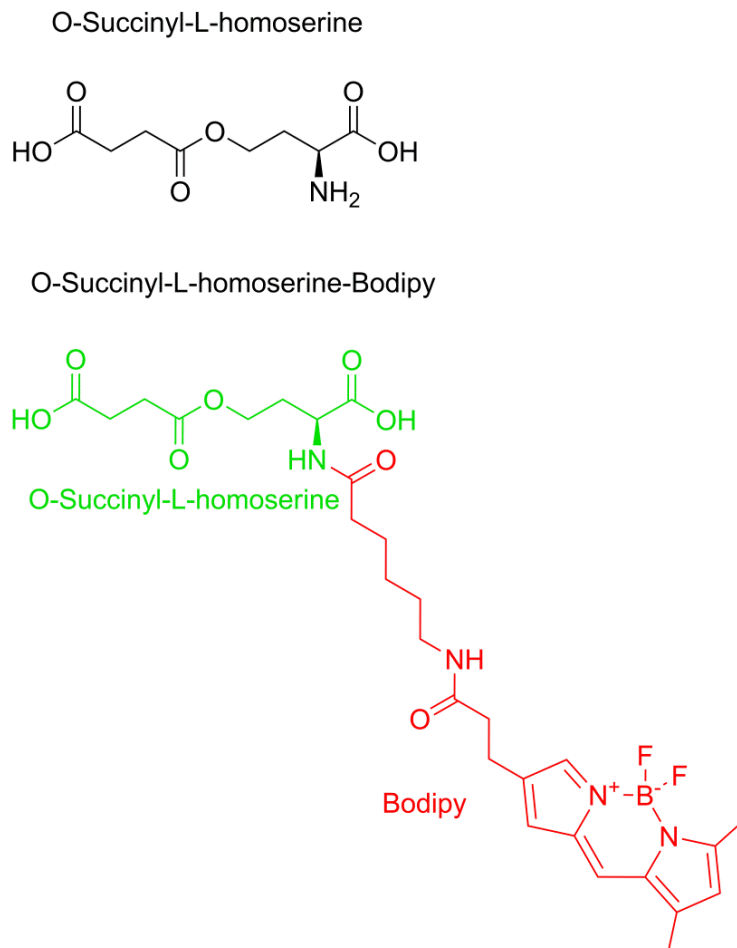


Figure 4.3.1-6—The structures of O-succinyl-L-homoserine (top) and O-succinyl—L-homoserine-Bodipy (bottom) used in this work. Bodipy-labelled O-succinyl-L-homoserine was synthesized by conjugating O-succinyl-L-homoserine to the amine reactive 3-Bodipy-propanoylaminocaproic acid, N-hydroxysuccinimide ester (Bodipy). In this case, the unlabelled compound was predicted to bind to protein E6, whereas its labelled counterpart was not.

4.3.2. *In vitro* testing of BoNT/A ligands

The synthesized aspartame-Bodipy was tested in a fluorescence polarization binding assay to determine its binding affinity experimentally for the target, BoNT/A light chain (LC). It was also tested in competition with unlabelled paclitaxel, a confirmed BoNT/A inhibitor, and unlabelled aspartame. The data provided in Figure 4.3.2-1 shows that aspartame-Bodipy binds to BoNT/A-LC as the polarization value for aspartame-Bodipy in the presence of BoNT/A is significantly higher than the value associated to free aspartame-Bodipy (data not shown). The

data also indicates that aspartame and aspartame-Bodipy are competing for the same binding region on the BoNT/A-LC because the polarization value is decreased when experiments were performed with both aspartame and aspartame-Bodipy, compared to aspartame-Bodipy complexed to BoNTA/A.

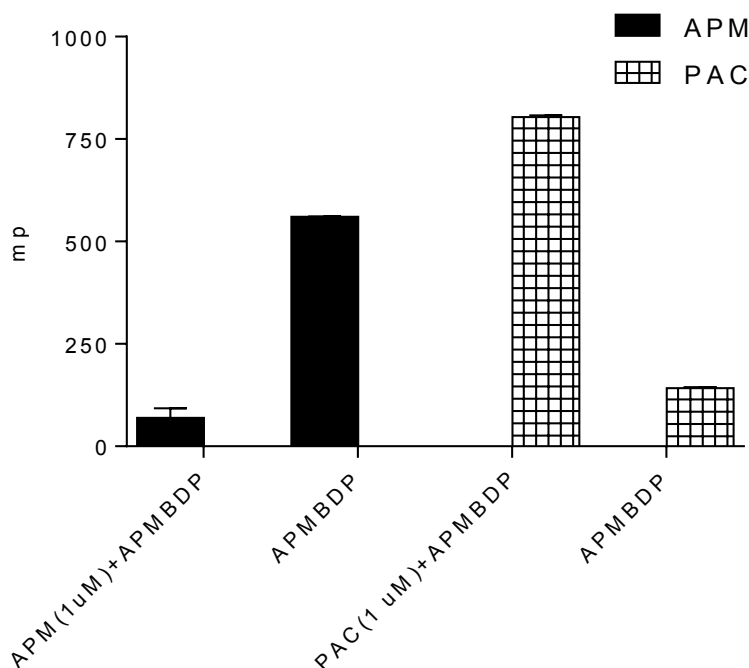


Figure 4.3.2-1-Fluorescence polarization binding assay using aspartame-Bodipy against BoNT/A-LC. The data related to unlabelled aspartame (APM) is shown in black whereas unlabelled paclitaxel (PAC) results are presented in square pattern. The addition of aspartame at 1 μ M (left solid bar) reduces the fluorescence polarization associated with the binding of aspartame-Bodipy to BoNT/A (right solid bar), whereas the addition of paclitaxel (left checked bar) increases the polarization of the corresponding aspartame-Bodipy reference (right checked bar). Each labelled/unlabelled competition assay has its own reference because of the influence of DMSO on the fluorescence of labelled aspartame-Bodipy. The samples were run in triplicate at 37°C and backgrounds were subtracted for each sample.

Testing against unlabelled paclitaxel confirms the binding of aspartame-Bodipy to BoNT/A and indicates that aspartame-Bodipy and paclitaxel (a confirmed experimental inhibitor¹⁴ of BoNT/A-LC) bind to two distinct binding sites in BoNT/AL-C. Evidence for this hypothesis is provided by the increased polarization value in the sample containing both aspartame-Bodipy

and paclitaxel compared to the control containing aspartame-Bodipy alone. Binding of two ligands to distinct sites of BoNT/A-LC would enhance the size of the BoNT/A-LC complex and consequently increase the Fluorescence polarization value relative to that of BoNT/A-LC complexed with a single ligand. Each labelled/unlabelled competition assay had its own reference because of the influence of DMSO (necessary to dissolve paclitaxel) on the fluorescence of labelled aspartame-Bodipy. More details about the binding of aspartame-Bodipy to BoNT/A will be provided in a separate manuscript.⁴²

4.3.3. *In vitro* testing of E6 ligands

O-Succinyl-L-homoserine was tested for binding to E6 using an intrinsic tryptophan fluorescence assay. O-Succinyl-L-homoserine-Bodipy was synthesized and tested for binding against protein E6 using a fluorescence polarization-based assay. The unlabelled ligand was tested at 8 concentrations (0.01, 0.1, 1, 3, 10, 25, 100, and 300 μM) while the labelled ligand was tested at three (1, 10, 100 μM) against a fixed concentration of protein. The results for the unlabelled and labelled compound are shown in Figure 4.3.3-1 and Figure 4.3.3-2 respectively. Unlabelled O-succinyl-L-homoserine exhibits a distinctive increase in tryptophan fluorescence relative to baseline control. An increase in signal is consistent with binding, as indicated by the positive (E6 antibody) control. No significant dose-dependency is observed at concentrations tested, however further characterization will be completed as part of a separate manuscript authored by John Wigg.

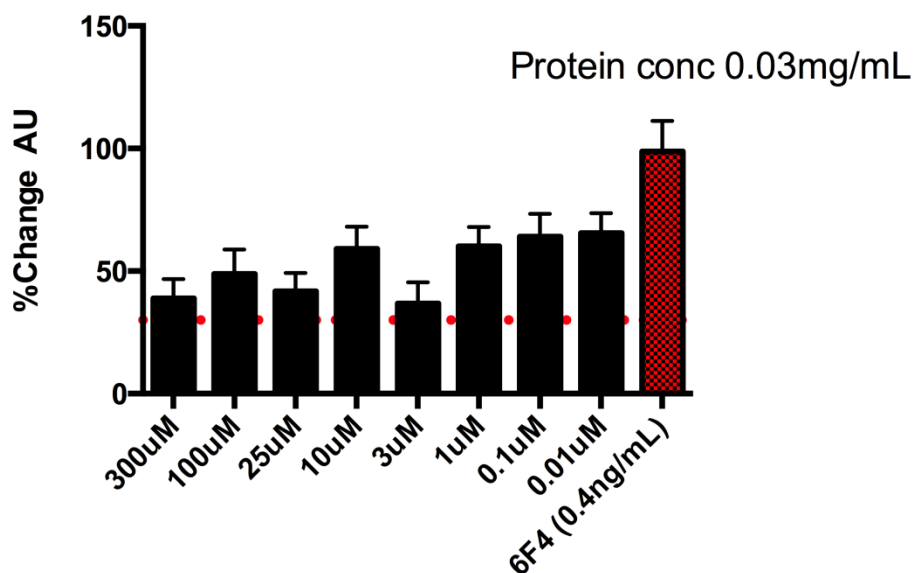


Figure 4.3.3-1-Change in intrinsic tryptophan fluorescence of E6 after 30 minutes incubation at 37°C buffered to a pH of 7.3 (20mM HEPES 7.3, 0.01% tween (v/v)). Protein was held at a constant concentration of 0.03 mg/mL. 6F4 antibody was used as a positive control at a concentration of 0.4ng/mL. O-succinyl-L-homoserine concentration was varied (300μM-0.01μM). The red line represents a 30% increase relative to baseline (no binding). 6F4 antibody, a positive control for binding, produces an increase in tryptophan fluorescence well over 30%.

Labelled O-succinyl-L-homoserine-Bodipy polarization values after one hour do not differ substantially from control. This binding assay yielded no evidence that O-succinyl-L-homoserine-Bodipy binds to protein E6, predicted by the low binding score of the docked compound relative to the threshold (Table 4.3.3-1).

The experimental results (summarized in Table 4.3.3-1) for BoNT/A and E6 indicate that our VLS strategy accurately predicts whether a fluorescently-labelled compound will bind or not bind to the intended target protein. Moreover, the competition assay results suggest that this strategy can identify probes that bind to distinct sites within the protein structure.

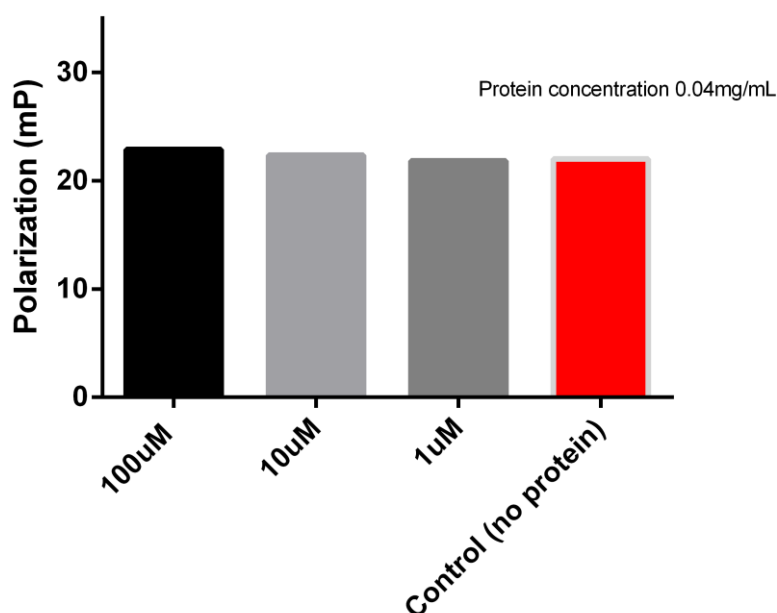


Figure 4.3.3-2-O-Succinyl-L-homoserine-Bodipy (100, 10, and 1µM) and protein E6 (0.04mg/mL) at 60 minutes. The experiment was conducted at a temperature of 37°C and pH of 7.5. Our results indicate that the O-Succinyl-Homoserine-Bodipy is not binding at any concentrations tested. Control wells containing protein as well as the O-Succinyl-L-homoserine-Bodipy at all three concentrations dissolved in assay buffer were included in addition to the reaction wells.

Table 4.3.3-1-A table outlining the predicted binding scores, expected experimental result, and whether experimental binding was observed for four selected compounds.

Compound	Binding Score (kcal/mol)	Binding Threshold (kcal/mol)	Predicted Binding (yes/no)	Experimental Binding (yes/no)
Aspartame	-82.20	-78.50	Yes	Yes
Aspartame-Bodipy	-74.90	-75.31	Yes*	Yes
O-Succinyl-L-homoserine	-45.62	-35.30	Yes	Yes
O-Succinyl-L-homoserine-Bodipy	-40.85	-56.50	No	No

*Aspartame-Bodipy scored close to the binding threshold. The labelled compound was expected to have some affinity based on previous tests that indicated that aspartame had high affinity, and based on an analysis of the predicted molecular interactions. Aspartame-Bodipy was selected as a candidate for testing based on these, and other practical criteria.

4.3.4. An efficient methodology for creating large libraries of fluorescence-labelled compound enables the use of VLS for probe discovery

A script for conjugating reactive fluorophores to databases of small molecules (supporting information) was prepared using MOE's scientific vector language (SVL). This script was able to distinguish between true candidates for conjugation and decoys (decoy database in supporting information). This script was used to virtually label a publically available database of base compounds with ATTO680 NHS ester (ATTO-TEC, Siegen, Germany; PubChemCID 16218508) according to the reactivity scheme shown in Figure 4.2.1-1. The resulting database contains 14,862 fluorescently-labelled compounds (supporting information) that would be suitable for VLS applications aiming at identifying targeted optical probes. The script can be easily modified to work with a different amine-reactive fluorophore for labelling.

4.4. DISCUSSION AND CONCLUSIONS

Significantly more extensive validation procedures have been previously described, which have given us confidence that HierVLS is a valid approach for screening small molecules for binding affinity.^{20, 24-27} However, this approach has not been used specifically for estimating the affinity of larger fluorophores and their small molecule conjugates. Our experimental results are in agreement with our computational predictions, which suggest that both unlabelled and labelled aspartame (aspartame and aspartame-Bodipy) have significant binding affinity for BoNT/A LC, and that O-succinyl-L-homoserine and its Bodipy conjugate will and will not bind to E6, respectively (summarized in Table 4.3.3-1).

This demonstrates that a VLS approach, such as HierVLS, can be an appropriate tool for estimating the binding of conjugates of reactive fluorophores and appropriate small molecules to protein targets. As these screening approaches are designed to be high-throughput methods,

allowing for thousands of molecules to be screened against targets of interest, it is necessary to devise a method for generating fluorophore/ligand conjugates rapidly, from large databases of substructures. To realize this idea, we designed a script that recognizes the appropriate reactive groups associated with conjugating an NHS-ester fluorophore to ligands containing primary amines. The script was written using a vector-based programming language known as scientific vector language (SVL) used by MOE (www.chemcomp.com) software. This specific language and the additional pre-defined applications and functions found in MOE are particularly suited for high-throughput applications involving large molecular databases. The script is capable of handling a fluorophore database containing multiple entries, and will create conjugates for each fluorophore containing an NHS-ester functional group. The script will only create conjugates for NHS-ester/primary amine combinations, and will output all conjugates with a name based on the concatenated ligand and fluorophore names or IDs. In cases where either the fluorophore or ligand is missing the appropriate reactive group, the conjugation is skipped, and a message is written to a log file. A SMARTS-based connectivity algorithm, built into MOE, is used for recognizing the appropriate functional groups. A database of test compounds containing various functional groups was prepared and tested to ensure that functional groups other than primary amines were not being inadvertently recognized by the algorithm used. Particular attention was given to ensure that other nitrogen-containing functional groups, including nitriles, amides, nitrates, and secondary or tertiary amines were being correctly discriminated from primary amines. Compounds containing multiple primary amines were completely avoided by the script, with a message written to a log file to alert the user. While these compounds would likely react with NHS-esters, it would be necessary to determine which amine would be preferentially labelled. Moreover, experimental synthesis of these conjugates would likely require additional

purification or separation steps to ensure that the appropriate conjugate was being selected and used for testing. These practical considerations led to the decision to avoid compounds containing multiple reactive groups. This version of the script is designed to be a basic implementation of our idea, which can be expanded in later versions to include other types of reactive fluorophores, and exclude entries containing undesirable molecular features such as groups which could confer toxicity or that commonly participate in side reactions with reactive fluorophores.

Once tested, the script was deemed adequate for the virtual labeling of large databases of small molecules with reactive NHS-ester fluorophores. We generated a large labelled database using publically available small molecule structures and one NHS-ester fluorophore, for later use in virtual screening applications. [1-(3-Carboxypropyl)-11-ethyl-2,2-dimethyl-1,8,9,10-tetrahydro-2H-dipyrido[3,2-b:2',3'-i]phenoxazin-11-ium-4-yl]methanesulfonate, marketed as ATTO680 (ATTO-TEC, Siegen, Germany) is the fluorophore moiety that was chosen for this database. ATTO680 is a relatively hydrophilic commercial fluorophore that fluoresces in the red region, which is more appropriate for cell-based *in vitro* applications, and possibly *in vivo* applications.⁴⁹ In comparisons with other popular fluorophore series including AlexaFluor® ((Molecular Probes Inc.) and BODIPY® (Molecular Probes Inc.), ATTO680 exhibits among the best signal to noise ratios for aqueous applications.⁴⁹ ATTO680 has been shown to be an appropriate choice for both steady-state and time-resolved fluorescence applications, which would enhance the experimental utility and versatility of any fluorescent conjugates bearing this moiety.⁴⁹ ATTO680 NHS-ester (PubChemCID 16218508) was used as the amine-reactive fluorophore (Figure 4.2.11-1b). Its structure is publically available and does not contain exotic

atoms that could interfere with VLS. While ATTO680 was chosen for generating an in-house tagged library, the script can easily work other NHS-ester fluorophores.

The Molecular Libraries Small Molecule Repository (MLSMR) (<https://mli.nih.gov/mli/compound-repository/mlsmr-compounds/>) was selected as a source of small molecule structures due to its public availability and its suitability for probe development. MLSMR is a collection of small molecules accessible for the purposes of high-throughput screening for biomedical applications, maintained as a part of the Molecular Libraries Initiative (MLI). The MLI is an NIH-funded programs which “offers public sector biomedical researchers access to the large-scale screening capacity necessary to identify small molecules that can be optimized as chemical probes to study the functions of genes, cells, and biochemical pathways” (<http://commonfund.nih.gov/Molecularlibraries/overview.aspx>). The MLSMR restricts entries based on numerous criteria with specific considerations given to the suitability of these compounds for high-throughput experimental screening. For example, minimum water solubility thresholds and minimum vendor amounts and purity were considered. The database is designed to retain a high level of diversity while implementing minimal restrictions on the types of chemical structures.⁴⁷ These characteristics may facilitate the translation of future screening efforts to experimental work.

The development of a high-throughput methodology for virtually generating structures of putative fluorescently labelled conjugates is an important step in facilitating the computational identification and/or design of these types of molecules as potential molecular probes. However, there are several limitations and future challenges that must be considered. Firstly, our implementation of this idea is quite rudimentary, and it's suited for one specific reactive fluorophore class. It would be feasible to expand this simple script in the future such that a more

complex piece of software could be developed that would enable screening of other reactive fluorophores, such as isothiocyanates and maleimides, thus expanding the scope of possible conjugation reactions. Additional criteria could also be applied such that the conjugates produced are most likely to be feasible experimentally, by ignoring molecules containing functional groups that might interfere with the labeling process, for example. While both small-molecule conjugates tested in this case were readily synthesized in the laboratory using standard techniques, it would be ideal to come up with a range of molecular descriptors that would prioritize molecules based on their predicted reactivity for labeling. Fluorophore-ligand conjugates are also typically quite large, and thus, have many conformational degrees of freedom. It would be important that, in any screening efforts using these molecules, this issue would be accounted for by using one or more conformational sampling or simulation techniques, such as molecular dynamics simulations.

The final fluorescently-labelled database numbers 14,862 compounds, each tagged with the ATTO680 fluorophore. This library is highly suitable for VLS efforts seeking to identify fluorescent probes for biomarkers. While the use of fluorescent labeling technology has been highlighted by large biomolecule labeling, there is the demand for labelled small molecules for specific *in vitro* and *in vivo* applications, the discovery of which can be computationally aided through such methods.

4.5. REFERENCES

1. Johnson, I.; Spence, M. T. Z., *Molecular Probes Handbook: A Guide to Fluorescent Probes and Labeling Technologies*. 11th ed.; Life Technologies Corporation: Carlsbad, California, United States, 2010.

2. Giepmans, B. N.; Adams, S. R.; Ellisman, M. H., et al., The fluorescent toolbox for assessing protein location and function. *Science* **2006**, *312* (5771), 217-224.
3. Sack, U.; Conrad, K.; Csernok, E., et al., Autoantibody detection using indirect immunofluorescence on HEp-2 cells. *Ann. N. Y. Acad. Sci.* **2009**, *1173*, 166-173.
4. Hoxha, E.; Harendza, S.; Zahner, G., et al., An immunofluorescence test for phospholipase-A2-receptor antibodies and its clinical usefulness in patients with membranous glomerulonephritis. *Nephrol. Dial. Transplant.* **2011**, *26* (8), 2526-2532.
5. Kinders, R. J.; Hollingshead, M.; Lawrence, S., et al., Development of a Validated Immunofluorescence Assay for γ H2AX as a Pharmacodynamic Marker of Topoisomerase I Inhibitor Activity. *Clin. Cancer Res.* **2010**, *16* (22), 5447-5457.
6. Schmidt, E.; Zillikens, D., Modern diagnosis of autoimmune blistering skin diseases. *Autoimmun. Rev.* **2010**, *10* (2), 84-89.
7. Chames, P.; Van Regenmortel, M.; Weiss, E., et al., Therapeutic antibodies: successes, limitations and hopes for the future. *Br. J. Pharmacol.* **2009**, *157* (2), 220-233.
8. Ntziachristos, V.; Bremer, C.; Weissleder, R., Fluorescence imaging with near-infrared light: new technological advances that enable in vivo molecular imaging. *Eur. Radiol.* **2003**, *13* (1), 195-208.
9. Hilderbrand, S. A.; Weissleder, R., Near-infrared fluorescence: application to in vivo molecular imaging. *Curr. Opin. Chem. Biol.* **2010**, *14* (1), 71-79.
10. Tung, C.-H.; Lin, Y.; Moon, W. K., et al., A Receptor-Targeted Near-Infrared Fluorescence Probe for In Vivo Tumor Imaging. *ChemBiochem* **2002**, *3* (8), 784-786.
11. Zhang, Z.; Achilefu, S., Design, synthesis and evaluation of near-infrared fluorescent pH indicators in a physiologically relevant range. *Chem. Commun.* **2005**, (47), 5887-5889.

12. Weissleder, R.; Tung, C. H.; Mahmood, U., et al., In vivo imaging of tumors with protease-activated near-infrared fluorescent probes. *Nat. Biotechnol.* **1999**, *17* (4), 375-378.
13. Messerli, S. M.; Prabhakar, S.; Tang, Y., et al., A Novel Method for Imaging Apoptosis Using a Caspase-1 Near-Infrared Fluorescent Probe. *Neoplasia* **2004**, *6* (2), 95-105.
14. Dadgar, S.; Ramjan, Z.; Floriano, W. B., Paclitaxel Is an Inhibitor and Its Boron Dipyrromethene Derivative Is a Fluorescent Recognition Agent for Botulinum Neurotoxin Subtype A. *J. Med. Chem.* **2013**, *56* (7), 2791-2803.
15. Shah, F.; Mukherjee, P.; Gut, J., et al., Identification of Novel Malarial Cysteine Protease Inhibitors Using Structure-Based Virtual Screening of a Focused Cysteine Protease Inhibitor Library. *J. Chem. Inf. Model.* **2011**, *51* (4), 852-864.
16. Naylor, E.; Arredouani, A.; Vasudevan, S. R., et al., Identification of a chemical probe for NAADP by virtual screening. *Nat. Chem. Biol.* **2009**, *5* (4), 220-226.
17. Doman, T. N.; McGovern, S. L.; Witherbee, B. J., et al., Molecular Docking and High-Throughput Screening for Novel Inhibitors of Protein Tyrosine Phosphatase-1B. *J. Med. Chem.* **2002**, *45* (11), 2213-2221.
18. Kolb, P.; Rosenbaum, D. M.; Irwin, J. J., et al., Structure-based discovery of β 2-adrenergic receptor ligands. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (16), 6843-6848.
19. De Graaf, C.; Kooistra, A. J.; Vischer, H. F., et al., Crystal structure-based virtual screening for fragment-like ligands of the human histamine H₁ receptor. *J. Med. Chem.* **2011**, *54* (23), 8195-8206.
20. Floriano, W. B.; Vaidehi, N.; Zamanakos, G., et al., HierVLS hierarchical docking protocol for virtual ligand screening of large-molecule databases. *J. Med. Chem.* **2004**, *47* (1), 56-71.

21. Kitchen, D. B.; Decornez, H.; Furr, J. R., et al., Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* **2004**, *3* (11), 935-949.
22. Meng, X.-Y.; Zhang, H.-X.; Mezei, M., et al., Molecular docking: a powerful approach for structure-based drug discovery. *Curr. Comput. Aided Drug Des.* **2011**, *7* (2), 146.
23. Scheffner, M.; Huibregtse, J. M.; Vierstra, R. D., et al., The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. *Cell* **1993**, *75* (3), 495-505.
24. Huibregtse, J. M.; Scheffner, M.; Howley, P. M., A cellular protein mediates association of p53 with the E6 oncoprotein of human papillomavirus types 16 or 18. *EMBO J.* **1991**, *10* (13), 4129.
25. Scheffner, M.; Werness, B. A.; Huibregtse, J. M., et al., The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell* **1990**, *63* (6), 1129-1136.
26. Sobel, J.; Tucker, N.; Sulka, A., et al., Foodborne botulism in the United States, 1990–2000. *Emerg. Infect. Dis.* **2004**, *10* (9), 1606.
27. Cai, S.; Singh, B. R.; Sharma, S., Botulism diagnostics: from clinical symptoms to in vitro assays. *Crit. Rev. Microbiol.* **2007**, *33* (2), 109-125.
28. Arnon, S. S.; Schechter, R.; Inglesby, T. V., et al., Botulinum toxin as a biological weapon. *J. Am. Med. Assoc.* **2001**, *285* (8), 1059-1070.
29. Jankovic, J., Botulinum toxin in clinical practice. *J. Neurol. Neurosurg. Psychiatry* **2004**, *75* (7), 951-957.
30. Münchau, A.; Bhatia, K. P., Regular review: Uses of botulinum toxin injection in medicine today. *BMJ* **2000**, *320* (7228), 161.

31. Bolton, E.; Wang, Y.; Thiessen, P. A., et al., PubChem: Integrated Platform of Small Molecules and Biological Activities. *Annu. Rep. Comput. Chem.* **2008**, *4*, 217-241.
32. Lipinski, C. A., Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* **2000**, *44* (1), 235-249.
33. Chemical Computing Group Inc. *Molecular Operating Environment (MOE)*, 2010.10; Chemical Computing Group, Inc.: Montreal, Quebec, Canada, 2010.
34. Gasteiger, J.; Marsili, M., Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219-3228.
35. Halgren, T. A., Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490-519.
36. Mujumdar, R. B.; Ernst, L. A.; Mujumdar, S. R., et al., Cyanine dye labeling reagents: sulfoindocyanine succinimidyl esters. *Bioconjug. Chem.* **1993**, *4*, 105-111.
37. Southwick, P. L.; Ernst, L. A.; Tauriello, E. W., Cyanine dye labeling reagents—carboxymethylindocyanine succinimidyl esters. *Cytometry* **1990**, *430*, 418-430.
38. OpenBabel *OpenBabel*, 1.3.
39. Yasara Biosciences Inc. *Yet Another Scientific Artificial Reality Application (YASARA)*, 10.8.2; 2010.
40. Krieger, E.; Koraimann, G.; Vriend, G., Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. *Proteins: Struct. Funct. Bioinform.* **2002**, *47* (3), 393-402.
41. Brady, G. P.; Stouten, P. F., Fast prediction and visualization of protein binding pockets with PASS. *J. Comput. Aided Mol. Des.* **2000**, *14*, 383-401.

42. Dadgar, S.; Floriano, W. B.; et al, Aspartame and aspartame-bodipy as ligands for BoNT/A. *Manuscript in preparation* **2014**.
43. Allen, M.; Reeves, J.; Mellor, G., High Throughput Fluorescence Polarization: A Homogeneous Alternative to Radioligand Binding for Cell Surface Receptors. *J. Biomol. Screen.* **2000**, *5* (2), 63-69.
44. Jolley, M. E., Fluorescence polarization assays for the detection of proteases and their inhibitors. *J. Biomol. Screen.* **1996**, *1* (1), 33-38.
45. Checovich, W. J.; Bolger, R. E.; Burke, T., Fluorescence polarization--a new tool for cell and molecular biology. *Nature* **1995**, *375* (6528), 254-256.
46. GraphPad Software *PRISM*, 6.02 (Windows); GraphPad Software: La Jolla, California, United States, 2013.
47. National Institutes of Health: Molecular Libraries Program MLSMR Compounds. <https://mli.nih.gov/mli/compound-repository/mlsmr-compounds/> (accessed June 2013).
48. National Center for Biotechnology Information: PubChem PubChem3D Release Notes. <http://pubchem.ncbi.nlm.nih.gov/release3d.html> (accessed June 2013).
49. Buschmann, V.; Weston, K. D.; Sauer, M., Spectroscopic study and evaluation of red-absorbing fluorescent dyes. *Bioconjug. Chem.* **2003**, *14*, 195-204.

Chapter 5.

Identifying potential selective fluorescent probes for cancer-associated protein carbonic anhydrase IX using a computational approach

5.1. INTRODUCTION

Carbonic anhydrase IX (CAIX) is a protein that is expressed in several tumour types, including cervical cancer, and is rarely expressed in healthy tissues outside of the gastrointestinal tract.¹ Carbonic anhydrases are a family of proteins that catalyze the reversible hydration of carbon dioxide. CAIX activity and expression are associated with tumour-related hypoxia. There is increasing evidence that CAIX activity in tumours correlates significantly with increased invasiveness, increased likelihood of metastasis, and poor overall survival, suggesting that CAIX expression is an important indicator for prognostically relevant tumour pathology.²⁻³

A reliable method of measuring of CAIX expression *in vivo* through imaging could become an important tool for assessing the cellular and organism-level response to novel tumour treatments during pre-clinical research. Optically active probes could be used to selectively image changes in CAIX expression throughout a course of treatment, either *in vitro* or *in vivo*. However, CAIX-based assessments of treatment efficacy should not be biased by any effects the probe may have on the activity of CAIX or other CA isoforms. Thus, an ideal probe for tumour monitoring should have high affinity and specificity for the CAIX isoform and should not interfere with its normal function.

Fluorescence-based imaging techniques are improving and may be used to obtain images through shallow tissue depths of up to a few centimeters.⁴ Range is improved and interference from autofluorescence is minimized by utilizing fluorescent moieties that emit light in the near-infrared (NIR) range of the electromagnetic spectrum.⁴ NIR wavelengths are poorly absorbed by water and hemoglobin, and therefore, use of NIR-emitting probes improves signal penetration and reduce changes in signal due to physiological composition.⁴ Relatively few available molecules possess the desired optical qualities, making it unlikely to identify a small molecule intrinsically suitable for imaging, while also conferring strong affinity and selectivity for the desired target, CAIX. However, it is possible to conjugate a suitable reactive dye to a small molecule containing specific chemical functional groups to create a target-specific fluorescent probe.

In this paper, we propose 20 fluorescent CAIX probe candidates for experimental validation. Candidates were selected from a virtual database of 20,860 compounds, which were either fluorescently-labelled or had published evidence of intrinsic fluorescence. The database was screened against a CAIX model using hierarchical virtual ligand screening (HierVLS), which combines a force field-based approach with a hierarchical framework to maximize efficiency for high-throughput screening applications.⁵ Previously published experimental data on CAIX inhibitors was used to validate the methodology used.⁶ Available sequence information for other CA isoforms was used to prioritize candidates based on the estimated “uniqueness” of each binding site. The extensively studied structure and mechanism of carbonic anhydrases, in addition to the availability of a crystal structure for CAIX⁷, influenced the decision to adopt a computational approach.

5.2. METHODS AND PROCEDURES

5.2.1. Preparation of the CAIX model structure

The only available experimental structure of the target protein, carbonic anhydrase IX (CAIX), was downloaded from the RCSB Protein Data Bank (www.rcsb.com). This structure (PDB Code: 3IAI) was obtained using X-ray diffraction at a resolution of 2.20Å.⁷ The biological assemblies associated with PDB:3IAI are proposed homodimers. Chain A, corresponding to one CAIX monomer, was selected for modeling. Editing and structure quality checking of the protein structure were performed using the programs YASARA (www.yasara.org) and/or MOE (www.chemcomp.com), unless otherwise noted. Solvent molecules, co-crystallized ligands (except for Zn²⁺) and covalently-linked carbohydrate moieties were removed from the structure. A mutated residue in the crystal structure (Cys41Ser) was replaced with its wild-type amino acid (Cys41). Missing atoms and side chains were identified and corrected. The CAIX model structure was then assigned CHARMM22 atomic charges and subjected to 2,000 steps of conjugate-gradient minimization using the Dreiding force field.⁸⁻⁹ Model coordinates are included as supplementary material.

Structural validation tools including PROCHECK¹⁰ and WHATCHECK¹¹ were used to check the quality of the resulting modeled structure (complete results detailed in supplementary material). MOE software was used to generate a Ramachandran plot (Figure 5.3.1-1) and figures of the model structure (Figure 5.3.1-2 to Figure 5.3.2-2). The CAIX model structure was compared to the experimental structure (PDB: 3IAI) to ensure that significant conformational changes were not induced by the energy minimization procedure. The model structure was aligned and superposed with 3IAI (Chain A) and all-atom and C- α RMSDs were calculated. This

procedure provided a visual confirmation that residues proximal to the active site had retained their relative positions after modeling.

5.2.2. Multiple sequence alignment and binding site comparison

The amino acid sequences of 23 known CA isoforms (supplementary material) were obtained from NCBI GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>). A multiple sequence alignment was constructed using ClustalX 2.1¹² using the Gonnet matrix series, gap opening: 10, gap extension: 0.2, delay divergent sequences: 30%.

Binding pockets within the CAIX model structure were identified using PASS, which is a freely available algorithm that relies on geometry to identify potential sites.¹³ Binding site centers were used to construct a graphical representation of each binding site within the model structure using MOE software. Residues with atoms between 0-5Å of each predicted site center were identified, and their position was annotated with respect to the multiple sequence alignment of CA isoforms. Percentage of amino acid identity to CAIX at each marked position was determined, and each binding site was scored according to its uniqueness to the CAIX isoform (Table 5.3.2-1).

5.2.3. Known inhibitors library for structure validation

A literature review was conducted to identify 32 small molecule inhibitors of CAIX with reported K_i values in the 14-305 nM range.⁶ Structures for these molecules were obtained either from PubChem (<http://pubchem.ncbi.nlm.nih.gov/>), BindingDB (www.bindingDB.org), or were constructed using MOE software (www.chemcomp.com) according to the structure reported in the literature. Inhibitor structures were assigned Gasteiger atomic charges and energy-minimized using the MMFF94x force-field.¹⁴⁻¹⁵

5.2.4. Decoy library for virtual screening

A decoy library of chemical compounds was created for validation purposes. No available decoy databases were found for carbonic anhydrase IX in two commonly used decoy repositories, DUD (<http://dud.docking.org/>) and DUDE (<http://dude.docking.org/>). Several target proteins that are structurally dissimilar to each other and that are also unrelated in function and structure to carbonic anhydrase were selected and the structures of known high-affinity ligands for these targets were downloaded from a publicly available repository of ligand structures and binding data known as BindingDB (www.bindingDB.org). Our assumption is that high affinity ligands for targets that are dissimilar to CAIX, will be unlikely to have any significant affinity for CAIX, and would therefore be suitable as decoy molecules. Ligands for several different “targets” were used as decoys to help ensure some chemical diversity in the decoy library, and not bias the library toward a particular chemical substructure. Specific inhibitors with K_i or IC_{50} values below 10nM were obtained for the following targets (number of inhibitors in parentheses): neuraminidase (25), metabotropic glutamate receptor isoforms 1-8 (213), monoamine oxidase A and B (199), VEGFR (44), and EphB4 (39).¹⁶⁻¹⁹ 3D structures for these decoy ligands were generated using MOE (www.chemcomp.com) from available 2D SDF structures or canonical SMILES (simplified molecular-input line-entry system) strings. Redundant molecules were eliminated yielding a decoy library of 478 unique entries. Decoys were cross-referenced for any documented activity against any carbonic anhydrase isoform in secondary BindingDB queries. Various chemical properties were calculated for the decoy set and compared to those of the inhibitor library in order to evaluate whether the decoys sufficiently represent the properties of the inhibitor library. Molecule structure files were prepared for screening by removing counter ions and solvent molecules and eliminating molecules containing

unparameterized atoms. Atoms were assigned Gasteiger charges and ligand structures were energy-minimized using the MMFF94x force-field.¹⁴⁻¹⁵ Various molecular descriptors were calculated for both the inhibitor and decoy set using MOE software in order to compare the distribution of properties between the groups.

5.2.5. Validation Screening

The combined library of 460 decoys and 32 known CAIX inhibitors was computationally screened against the 3D model of CAIX using PASS for binding site identification and HierVLS for molecular docking and force field-based scoring. PASS is a rapid and efficient program that uses a geometry-based algorithm to identify putative binding pockets based their shape, size, and depth relative to the protein surface.¹³ It is freely available and generates output that is generally compatible with applications used in later steps. HierVLS is a hierarchical approach to virtual ligand screening that tests the largest number of bound conformations of each ligand using the least computationally expensive methods, and proceeds to use more computational resources to obtain more accurate predictions of binding energy for the most promising conformers.⁵ Screening was initiated using a graphical user interface (GUI) known as Cassandra, and calculations were run using the SHARCNET high-performance computing network (www.sharcnet.ca).²⁰ For each ligand, a maximum number of 10,000 docked conformations were generated in level 0 (coarse-grain conformational search), with the 1500 best conformers passed to the filter step. Conformations were eliminated by the filter if they had less than 70% buried surface area. 150 conformations per ligand were passed from the filter step through to level 1 (protein-fixed minimization). The best 3 docked conformations for each ligand were subject to an all-atoms minimization (level 2). Final raw binding scores (kcal/mol) consider the desolvation effects of the protein/ligand complex by applying the analytical volume generalized Borne

solvation model to the complex, and the protein and ligand alone. The steps and parameters of the HierVLS docking and scoring scheme have been described previously.⁵ A complete list of HierVLS parameters used for this experiment is included in the supplementary material. Raw binding scores were collected for each ligand docked to each of the 4 identified binding sites. Three separate score-adjustment schemes were applied to the raw scores. Raw binding scores were adjusted to give site-specific propensity relative to other binding sites using two separate in-house algorithms. In a separate step, raw binding scores for each ligand in all regions were subjected to principal components analysis (PCA) using PAST statistical software (<http://folk.uio.no/ohammer/past/>). Ligands without binding data for any of the 4 sites (due to failing the buried surface cut-off filter) were excluded from analysis (18 decoys), leaving 460 and 32 scored decoys and true positives, respectively. PCA was performed using a correlation matrix type with singular value decomposition (SVD) enabled and iterative imputation to handle missing values with the following other parameters: Boot N=0, Jolliffe cut-off=0.7. An analysis of the test characteristics for each score-adjustment scheme (raw scores, in-house algorithms, and PCA) was performed using PRISM²¹ and Microsoft Excel 2010.

5.2.6. CAIX Virtual Screening

The CAIX model was screened against two compound libraries. One library contains 14,862 primary-amine containing compounds that were identified from Molecular Libraries Initiative Small Molecule Repository (MLSMR) and tagged virtually with Atto680 NHS-ester (database available with supplementary material). Atto680 is a near-IR fluorophore that is commercially available in a reactive N-hydroxysuccinimidyl ester form.²² The preparation of this database is described in another manuscript (Chapter 4). An additional subset of the MLSMR, was identified using PubChem BioAssay data. Molecules with reported activity in a published

series of high-throughput fluorescence assays¹⁹ were selected, totaling 5,998 compounds. The overall number of compounds screened against CAIX was 20,860. These databases and associated PubChemBioAssayIDs are included with supplementary material. The parameters used for screening were identical to those used during the validation procedure with the exception of the buried surface area cut-off for docked ligands, which was adjusted from 70% to 40% when screening the fluorescently-labelled compounds to reflect their larger overall size.

“Hit” candidates were identified based on the scoring scheme and binding thresholds that best performed during the validation step. Best conformers for the binding candidates were examined for favourable protein-ligand interactions (using MOE software). Candidates were also assessed for binding site preference by comparing raw scores between specific sites relative to each binding site’s mean scores and score standard deviation for that database.

5.3. RESULTS AND DISCUSSION

5.3.1. Preparation of the CAIX model structure

A CAIX model structure was prepared from the 3IAI PDB structure. A Ramachandran plot was generated (Figure 5.3.1-1), and PROCHECK/WHATCHECK quality analyses performed (supplementary material).

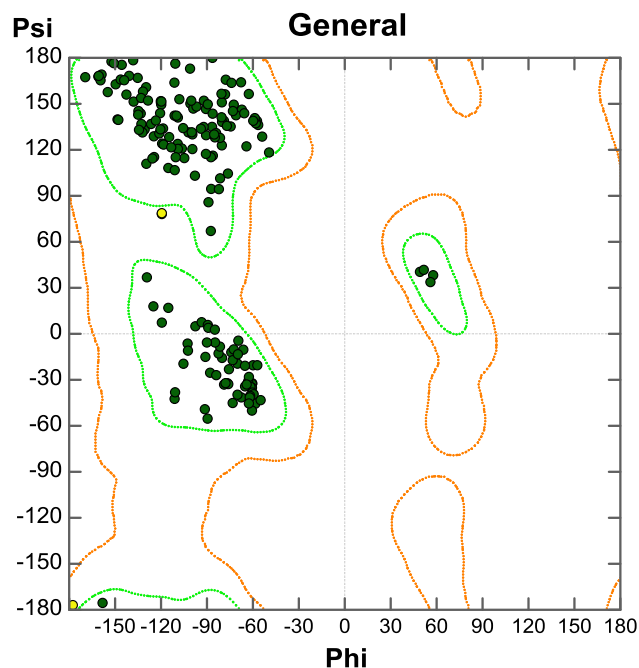


Figure 5.3.1-1—A Ramachandran plot for the CAIX model structure (generated using MOE software).

Psi-phi bond angles (Ramachandran) appeared to be consistent with a reasonable structure.

Quality analyses were used to check to ensure that missing sidechains (and atoms) had been added during the modelling process, and that the starting and final structures did not have any serious structural abnormalities. The model was compared to the crystal structure by calculating the all-atoms RMSD which was found to be 0.354Å. When a c-alpha RMSD was performed, results were similar, and showed that less than 5 residues had α C-alpha RMSD greater than 2 standard deviations from the mean RMSD (0.48Å was the maximum). Combined with the quality results, these comparisons gave us confidence that our model would be representative of the actual structure. A cartoon representation of the CAIX model is shown in Figure 5.3.1-2.

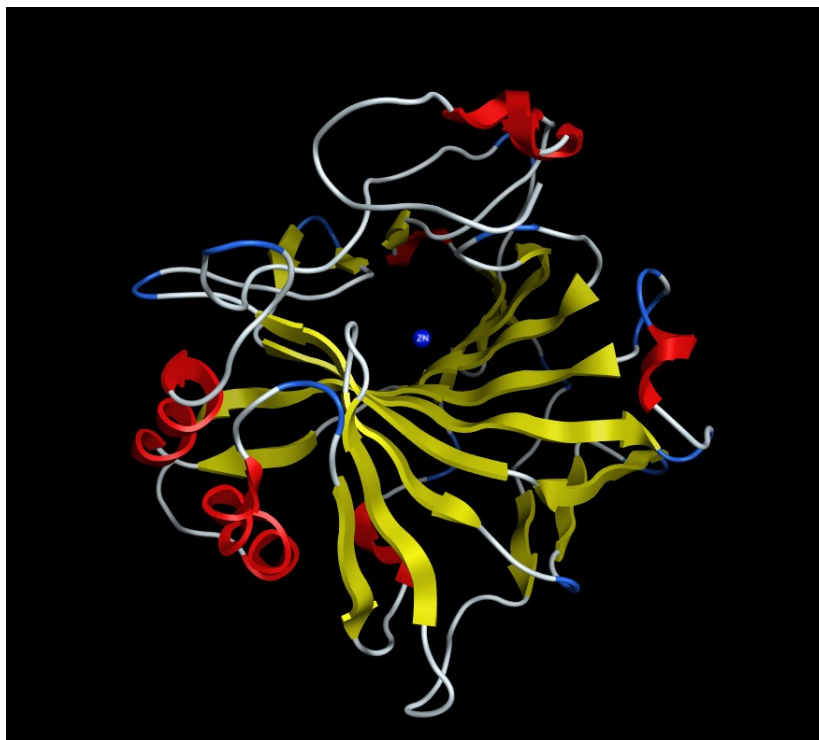


Figure 5.3.1-2 – A cartoon representation of the CAIX model structure with the zinc cation shown in the center of a large binding cavity.

Four binding sites were identified using PASS. Sites 1, 3, and 4 are shown as surfaces Figure 5.3.2-1. Site 2 corresponds to the catalytic site (Zn^{2+} shown) and has the most buried surface of the four sites (Figure 5.3.2-2).

5.3.2. Multiple sequence alignment and binding site comparison

A multiple sequence alignment of 23 human isoforms was performed (provided in supplemental material). All residues within 5\AA of each binding site of CAIX were compared for identity using the sequence alignment. The distances from the center of each binding site to the catalytic site (site 2) were also calculated. A summary of % homology and distance to the catalytic site for all binding sites is shown in Table 5.3.2-1. Based on this analysis, hits will be prioritized for experimental testing if they show selectivity for sites 1 or 3, which show the least homology to other CAIX isoforms, and largest distance from the catalytic site.

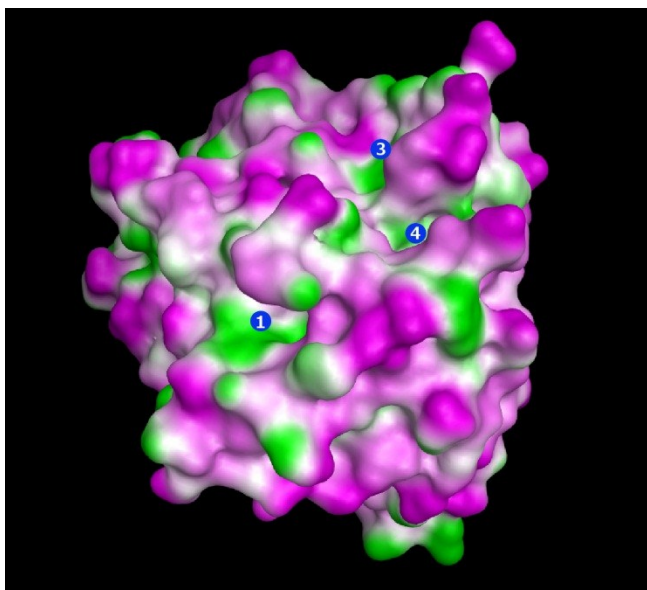


Figure 5.3.2-1 – A surface representation of the CAIX model with the centers of sites 1, 3, and 4 (1 would be the most preferable for binding, 3 the second most, and 4th, the least) shown as blue spheres. Surface colours represent hydrophilic (pink), hydrophobic (green) and neutral (white) regions.

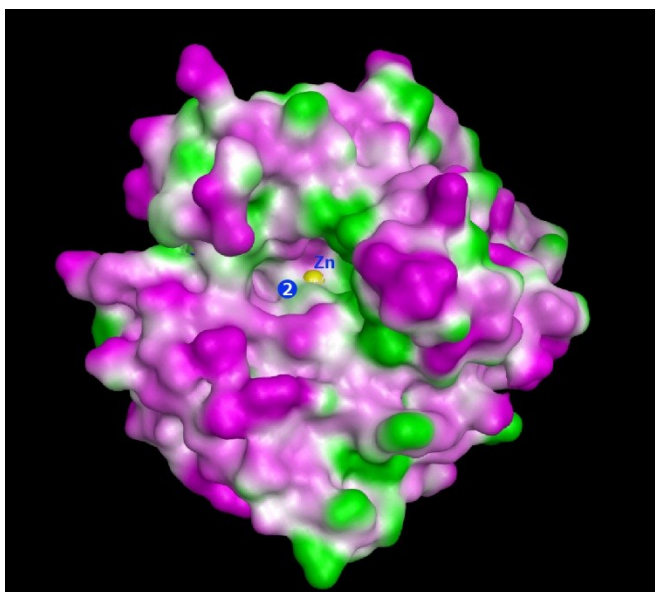


Figure 5.3.2-2 – A surface representation of the CAIX model showing binding site 2 (blue sphere representing its center), which represents the catalytic site. The zinc cation is shown as a yellow sphere. Surface colours represent hydrophilic (pink), hydrophobic (green) and neutral (white) regions.

Table 5.3.2-1 – The % sequence identity for residues within 5Å of each binding region to the sequence alignment of 23 CA isoforms is shown, along with the distance from the catalytic site (2).

Region	Identity (%)	Distance from Catalytic Site (Å)
1	14.6	21.28
2	67.2	n/a
3	17.4	13.76
4	29.2	11.68

5.3.3. Validation screening

Four scoring schemes were assessed using a test set of 32 known CAIX inhibitors⁶ and 460 decoys. Raw (force field-based) scores, two different holistic scores (adjusted for site-specific propensity using different in-house algorithms), and scores given by a principal components analysis (PCA) were analyzed. Raw scores from region 2, which corresponds to the active site, were used because 32 inhibitors in the validation site bind to the active-site of the protein and, hence, their binding affinities for that site may correlate with experimental values. For the PCA scoring scheme, the first principal component (PC1) was used because it carries the most variance (65.96%) in the calculated binding affinity it represents. PCA was the only scheme to improve the discrimination between true binders and decoys relative to the raw scores.

Figure 5.3.3-1 represents the sensitivity and specificity of the PCA scheme relative to the raw scoring scheme using a receiver operating curve (ROC). The area under the curve (AUC) was calculated for all scoring methods. AUC is a commonly used metric for evaluating diagnostic test performance, and signifies the probability that a random true positive will be

ranked better than a random true negative.²³ The PCA scoring scheme performed the best, with an AUC of 92%.

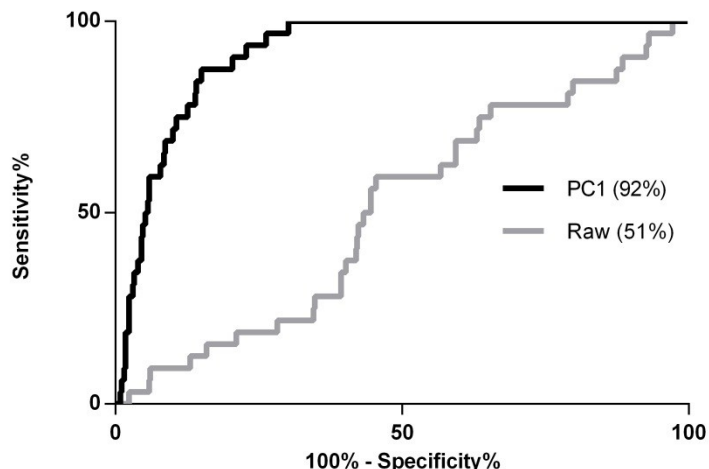


Figure 5.3.3-1 – The ROC curves for the PCA and raw (force field-based) scoring schemes used for the validation screen of CAIX. The area under each curve (AUC) is shown for each scoring scheme in parentheses.

A frequency histogram of PC1 scores for both decoys and true positives is shown in Figure 5.3.3-2a. It is evident from this distribution that the PC1 scores for true positives lay toward the negative side of the distribution. In order to choose a threshold below which candidates will be selected for experimental testing, the positive predictive value (PPV) of the PCA scheme was examined. The PPV signifies the probability that a compound with a score that is better than the threshold will be a true binder (Equation 5.3.3-1). The PPV depends on the sensitivity and specificity of the selected threshold, as well as the prevalence of true positives in the database being screened. With respect to the ROC curve, increasing sensitivity and specificity values are associated with a shift in the curve toward the topmost and leftmost parts of the graph, respectively. For our purposes, we were most interested in enhancing the PPV, because we would like to increase the probability that compounds selected for testing will have true affinity for the target.

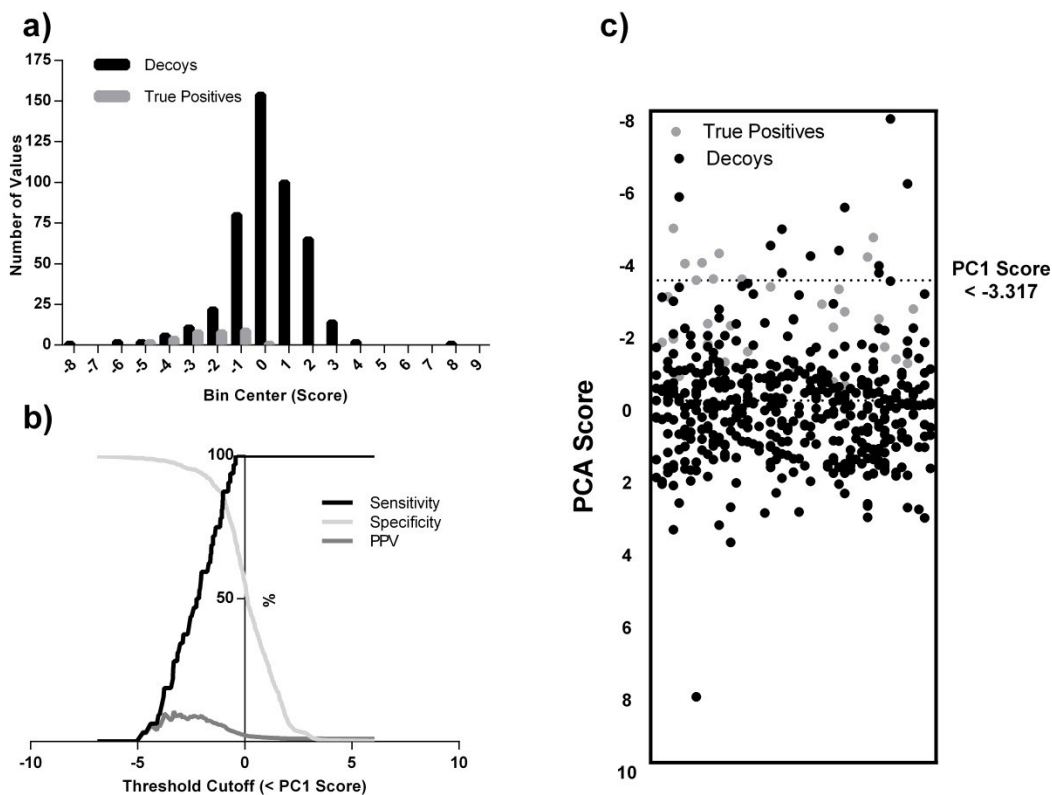


Figure 5.3.3-2 – The PCA scoring scheme was selected for library screening. A frequency histogram of scores for decoys and true positives in the validation library are shown in (a). The specificity, sensitivity, and positive predictive value (PPV) of the PCA scheme determined using the validation library are shown in (b). The highest possible PPV value, assuming a 1% prevalence of hits in a database, occurs at a threshold of -3.317. Relatively few compounds in the validation library scored below the threshold value of -3.317 (c).

Equation 5.3.3-1 – The positive predictive value (PPV) is the fraction of all positives, true (TP) and false (FP), that are true positives.

$$PPV = TP / (TP + FP)$$

It is acceptable if our selected threshold has a relatively low sensitivity, as it is unnecessary for us to successfully identify *all* true positives within a database. In this context it was ideal to improve the discriminatory ability of our threshold such that few false negatives were chosen. Alternatively, if we believed that there were very few true positives in a database, or if we were using this type of screening for another application where identifying all actives were a priority, we could adjust the threshold for better sensitivity. Assuming a prevalence of

true positives of 1% in a hypothetical database, we graphed the PPV across a range of scoring thresholds, also including associated specificity and sensitivity values (Figure 5.3.3-2b). The PC1 score threshold that gave the best PPV was -3.317. At this threshold, the PPV would be 10.63% for a database containing 1% hits. The corresponding sensitivity and specificity for this threshold are 28.13% and 97.61% respectively. Figure 5.3.3-2c shows a distribution of the PC1 scores for the validation set, with decoys and true positives shown with respect to the chosen threshold. The PPV for this scoring scheme suggests that in future screening experiments at least 10-15 top candidates should be identified for experimental testing to increase the likelihood of identifying a true binder. The enrichment factor (EF) is a metric commonly used to describe the performance of virtual screening methods. EF gives the ratio of true positives detected in the top x% of scored compounds to the number that would have been expected based on random selection:

Equation 5.3.3-2 – The enrichment factor (EF) is given by the ratio of the number of hits (*n*) found above the threshold (top x %) to the expected number of hits, the proportion (*P*) of overall hits in the database multiplied by the number of compounds overall (*N*) above the threshold (top x %).

$$EF = \frac{n_{hits\ in\ top\ x\ \%}}{P_{hits} N_{in\ top\ x\ \%}}$$

An enrichment factor of 1 would indicate that the method is no better at selecting true positives than random selection. The enrichment factors are given for the chosen threshold (PC1 score < -3.317) as well as for the top 2, 5, 10, and 20% of scored compounds in the validation library in Figure 5.3.3-3. This figure also displays the true positives (hits) identified at each threshold as a proportion of the total number of true positives (N=32) in the validation library. The threshold score chosen using ROC analyses (PC1 score < -3.317) gave the best enrichment factor, 6.92. The interpretation of this EF is that the probability of a true positive occurring

among compounds selected using the chosen threshold is increased relative to random selection within the overall database by a factor of almost 7. When systematically assessed for performance with test ligand sets across multiple targets, 11 different docking programs were shown to have enrichment factors between 0.0 and 10.0 for the top 10% of their databases.²⁴ For the PCA scoring scheme, EF=5.65 for the top 10% of PCA scores, indicating that our observed EFs agree reasonably well with values reported for other screening protocols.

These analyses should be considered with some uncertainty, as they are based on a relatively small validation library. This library was also necessarily composed of compounds targeting only one site, the active site, of the protein. Nonetheless, when binding scores vary continuously on a normal distribution, it can be very difficult to decide upon a binary threshold. The use of a validation set to estimate the predictive value of our scoring scheme is important for justifying a threshold value in subsequent tests, despite this uncertainty. Comparison of enrichment factors with those obtained in the literature for other docking and scoring methods gives us confidence that our scoring scheme has significant predictive value for selecting binding candidates for CAIX.

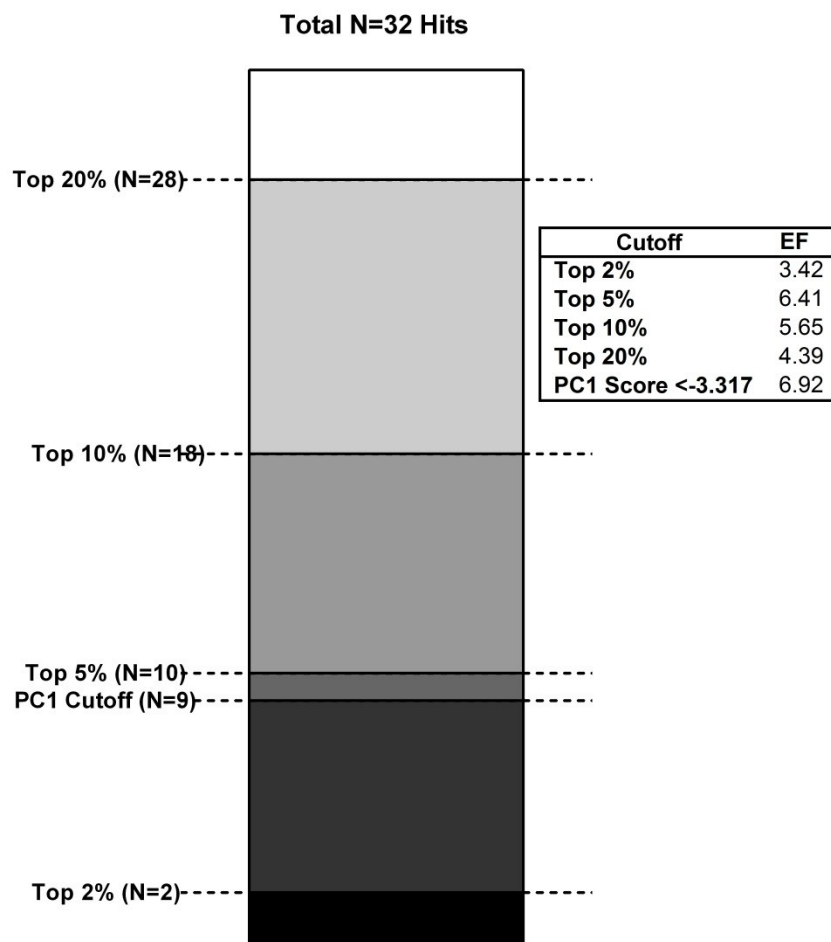


Figure 5.3.3-3 – The number of true positives found (N) using each indicated threshold (i.e., top x% of validation library screened) as a fraction of the total number of true positives present in the validation library (N=32). Enrichment factors (EFs) were calculated for each of the indicated threshold values and shown in tabulated format. The PC1 score threshold that was selected using ROC analysis gave the highest EF value, 6.92.

5.3.4. CAIX Screening

A total of 20,860 compounds were screened against the CAIX model. Scores were adjusted using principal components analysis (PCA) which was determined in the previous step to be the preferred scoring scheme. A PC1 score threshold of -3.317 was used, below which molecules were considered candidates for binding. A distribution of the PC1 scores for molecules screened is shown in Figure 5.3.4-1.

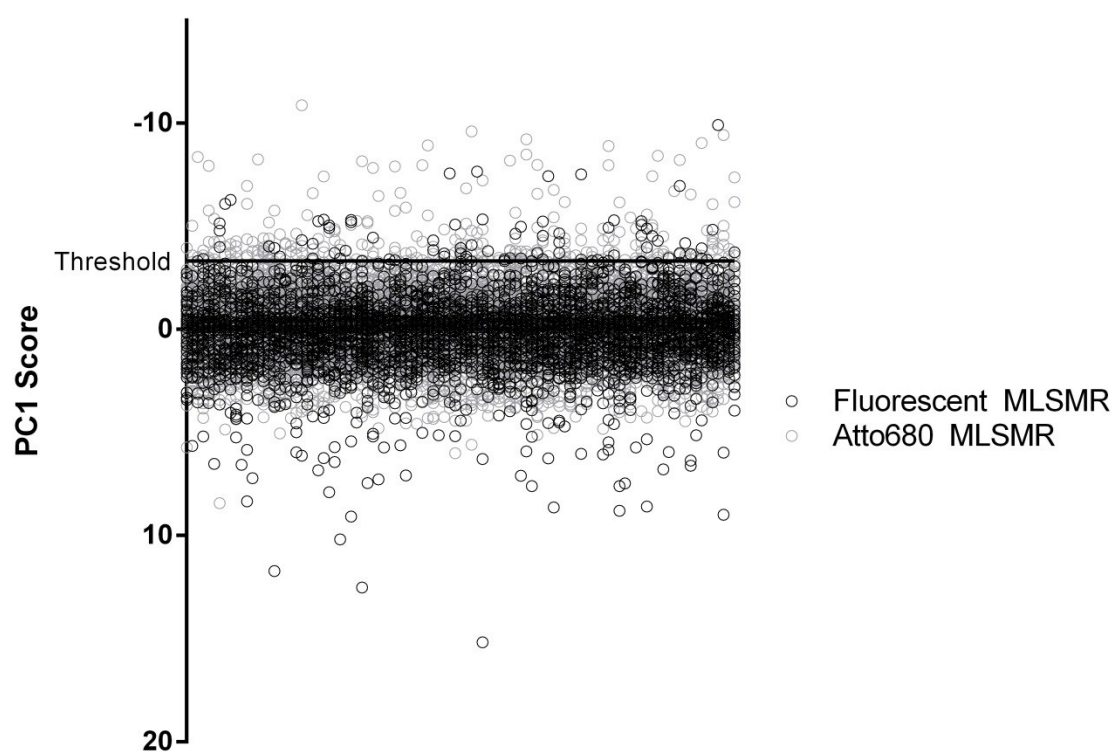


Figure 5.3.4-1 – A distribution of scores for molecules screened in the Atto680-tagged (blue) and Fluorescent MLSMR (red) databases. The threshold PC1 score -3.317 is marked.

Table 5.3.4-1 summarizes 20 top scoring binding candidates for both the Atto680-tagged and Fluorescent MLSMR libraries. For each of these candidates, best scoring conformers from each binding site were examined, and site-specific raw scores compared. Scores were examined for number of standard deviations above the mean score for that site (in a particular library). Site 1 had been selected previously as the most preferred binding site based on lack of homology

between other isoforms, and distance from catalytic site. The site specific scores relative to the mean are summarized in Table 5.3.4-2 and Table 5.3.4-3 for the best 5 compounds in each of the libraries. Each of these 5 ligands (or in the case of the Atto680-tagged library, the parent unlabelled primary amine) is shown in Table 5.3.4-4 and Table 5.3.4-5.

Table 5.3.4-1 – Top 20 ranked compounds screened from both the Atto680-tagged and Fluorescent MLSMR databases.

Fluorescent MLSMR		Atto680 MLSMR	
PubChem CID	PC1 Score	PubChem CID*	PC1 Score
CID652931	-9.927	CID16746148	-10.887
CID5154	-7.674	CID5939529	-9.614
CID1256741	-7.576	CID666466	-9.445
CID2999504	-7.538	CID5939530	-9.232
CID665212	-7.441	CID20846957	-9.059
CID854591	-6.975	CID666309	-8.933
CID446849	-6.293	CID667134	-8.915
CID649288	-6.104	CID665941	-8.498
CID2368420	-5.351	CID535684	-8.437
CID441975	-5.330	CID666821	-8.378
CID656158	-5.323	CID3153996	-8.259
CID648233	-5.310	CID1939105	-8.229
CID5389398	-5.261	CID3104972	-8.201
CID662206	-5.255	CID25163910	-8.165
CID3243178	-5.243	CID6916824	-7.989
CID1301431	-5.217	CID8795	-7.982
CID666072	-5.152	CID666424	-7.976
CID4849	-5.097	CID1491128	-7.952

CID16231	-5.091	CID667344	-7.913
CID5546	-5.057	CID1853629	-7.852

*PubChemCID given for parent compound of Atto680-tagged conjugate.

Table 5.3.4-2 – Best 5 compounds with suggested binding preference for region 1 from the Fluorescent MLSMR database. Raw scores were examined relative to mean, and expressed as N*# of standard deviations from the mean score.

PubChem CID	Raw Scores	N x Std. Dev. Above Mean Raw Site 1 Score	N x Std. Dev. Above Mean Raw Site2 Score
CID652931	-60.8793	3.7	0.3
CID5154	-56.8195	3.1	0.5
CID1256741	-51.6917	2.3	-0.5
CID2999504	-50.746	2.2	-0.7
CID665212	-51.671	2.3	-0.4

Table 5.3.4-3 – Best 5 compounds with suggested binding preference for region 1 from the Atto680-tagged MLSMR database. Raw scores were examined relative to mean, and expressed as N*# of standard deviations from the mean score.

PubChem CID*	Raw Scores	N x Std. Dev. Above Mean Raw Site 1 Score*	N x Std. Dev. Above Mean Raw Site 2 Score
CID667134	-8.915	6.6	5.3
CID535684	-8.437	7.8	4.6
CID666821	-8.378	5.6	4.4
CID3104972	-8.201	5.9	2.8
CID8795	-7.982	5.1	4.4

Table 5.3.4-4 – Chemical structures of the top 5 scoring Fluorescent MLSMR compounds.

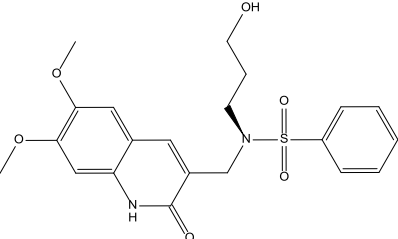
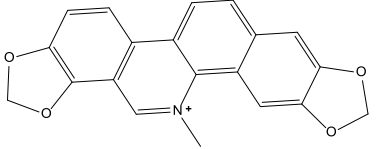
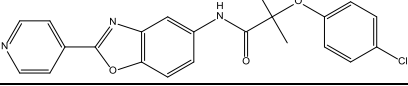
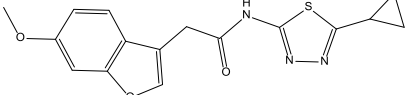
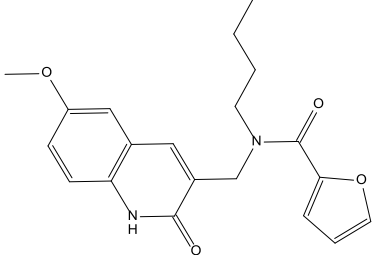
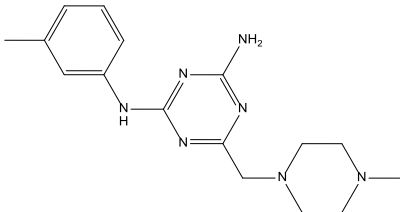
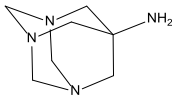
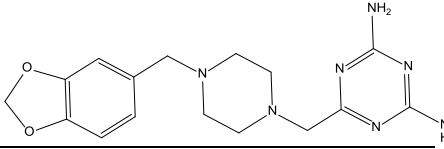
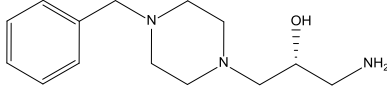
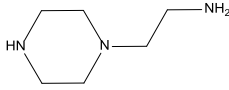
PubChem CID	Structure
CID652931	
CID5154	
CID1256741	
CID2999504	
CID665212	

Table 5.3.4-5 – Base chemical structures of the top 5 scoring Atto680-tagged MLSMR compounds.

PubChem CID	Structure
CID667134	
CID535684	
CID666821	
CID3104972	
CID8795	

5.4. CONCLUSIONS

By using a test-set validated scoring scheme, and additional ranking criteria, we have identified several promising probe candidates for experimental testing against CAIX. We predict that the post-screening criteria used for ranking of these compounds will impart additional selectivity against CAIX isoforms than would have otherwise been present.

We are very confident about the ability of our methodology to predict the true binding of fluorescent probes to CAIX. However, there remain several challenges in our study. CAIX activity is Zn-dependent, and the orientation of the active site relative to the Zn^{2+} ion is crucial both for catalysis and inhibition. While we ensured that Zn orientation within the active site was not significantly altered, it is still possible that small changes in the active site orientation due to

minimization and modeling procedures may have affected the accuracy of our model. Zinc and other metal ions are typically handled poorly by force field-based approaches, which likely affected the accuracy of our scores. We provided estimates of binding site homology based on a multiple sequence alignment of all human CA isoforms and predicted the preferential site of binding for each ligand based on site-specific scores from HierVLS. The scoring method for ranking compounds, PCA of raw scores, does not retain information about binding site selectivity. This method is useful as a rough approximation of the amount of conservation between isoforms, however, it does not account for 3D structural variation. An alternate approach could include developing homology models for the most common isoforms, and cross-screening promising candidates against them in order to estimate isoform specificity. Such an approach would require additional time and computational resources, but would be recommended as part of a subsequent investigation prior to experimental testing.

5.5. REFERENCES

1. Swietach, P.; Vaughan-Jones, R. D., Regulation of tumor pH and the role of carbonic anhydrase 9. *Cancer Metastasis Rev.* **2007**, *26*, 299-310.
2. Loncaster, J. A.; Harris, A. L.; Davidson, S. E., et al., Carbonic anhydrase (CA IX) expression, a potential new intrinsic marker of hypoxia: correlations with tumor oxygen measurements and prognosis in locally advanced carcinoma of the cervix. *Cancer Res.* **2001**, *61* (17), 6394-6399.
3. Woelber, L.; Kress, K.; Kersten, J. F., et al., Carbonic anhydrase IX in tumor tissue and sera of patients with primary cervical cancer. *BMC Cancer* **2011**, *11*, 12.

4. Ntziachristos, V.; Bremer, C.; Weissleder, R., Fluorescence imaging with near-infrared light: new technological advances that enable in vivo molecular imaging. *Eur. Radiol.* **2003**, *13* (1), 195-208.
5. Floriano, W. B.; Vaidehi, N.; Zamanakos, G., et al., HierVLS hierarchical docking protocol for virtual ligand screening of large-molecule databases. *J. Med. Chem.* **2004**, *47* (1), 56-71.
6. Vullo, D.; Franchi, M.; Gallori, E., et al., Carbonic anhydrase inhibitors: inhibition of the tumor-associated isozyme IX with aromatic and heterocyclic sulfonamides. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1005-1009.
7. Alterio, V.; Hilvo, M.; Di Fiore, A., et al., Crystal structure of the catalytic domain of the tumor-associated human carbonic anhydrase IX. *Proceedings of the National Academy of Sciences* **2009**, *106* (38), 16233-16238.
8. MacKerell, A. D.; Bashford, D.; Dunbrack, R. L., et al., All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *The Journal of Physical Chemistry B* **1998**, *102*, 3586-3616.
9. Mayo, S. L.; Olafson, B. D.; Goddard, W. A., DREIDING: a generic force field for molecular simulations. *J. Phys. Chem.* **1990**, *94* (26), 8897-8909.
10. Laskowski, R. a.; MacArthur, M. W.; Moss, D. S., et al., PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283-291.
11. Hooft, R.; Vriend, G.; Sander, C., et al., Errors in protein structures. *Nature* **1996**.
12. Larkin, M. a.; Blackshields, G.; Brown, N. P., et al., Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)* **2007**, *23*, 2947-2948.

13. Brady, G. P.; Stouten, P. F., Fast prediction and visualization of protein binding pockets with PASS. *J. Comput. Aided Mol. Des.* **2000**, *14*, 383-401.
14. Gasteiger, J.; Marsili, M., Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219-3228.
15. Halgren, T. A., Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490-519.
16. Bardelle, C.; Barlaam, B.; Brooks, N., et al., Inhibitors of the tyrosine kinase EphB4. Part 3: identification of non-benzodioxole-based kinase inhibitors. *Bioorg. Med. Chem. Lett.* **2010**, *20* (21), 6242-6245.
17. Bardelle, C.; Coleman, T.; Cross, D., et al., Inhibitors of the tyrosine kinase EphB4. Part 2: structure-based discovery and optimisation of 3,5-bis substituted anilinopyrimidines. *Bioorg. Med. Chem. Lett.* **2008**, *18* (21), 5717-5721.
18. Bardelle, C.; Cross, D.; Davenport, S., et al., Inhibitors of the tyrosine kinase EphB4. Part 1: Structure-based design and optimization of a series of 2,4-bis-anilinopyrimidines. *Bioorg. Med. Chem. Lett.* **2008**, *18* (9), 2776-2780.
19. Simeonov, A.; Jadhav, A.; Thomas, C. J., et al., Fluorescence spectroscopic profiling of compound libraries. *J. Med. Chem.* **2008**, *51*, 2363-2371.
20. Ramjan, Z. H.; Raheja, A.; Floriano, W. B. In *A cluster-aware graphical user interface for a virtual ligand screening tool*, Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE, IEEE: 2008; pp 4102-4105.
21. GraphPad Software *PRISM*, 6.02 (Windows); GraphPad Software: La Jolla, California, United States, 2013.

22. Buschmann, V.; Weston, K. D.; Sauer, M., Spectroscopic study and evaluation of red-absorbing fluorescent dyes. *Bioconjug. Chem.* **2003**, *14*, 195-204.
23. Kirchmair, J.; Markt, P.; Distinto, S., et al., Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection—What can we learn from earlier mistakes? *J. Comput. Aided Mol. Des.* **2008**, *22* (3-4), 213-228.
24. Warren, G. L.; Andrews, C. W.; Capelli, A.-M., et al., A Critical Assessment of Docking Programs and Scoring Functions. *J. Med. Chem.* **2005**, *49* (20), 5912-5931.

Chapter 6.

Identification and *in vitro* testing of potential EphB4 kinase modulators

6.1. INTRODUCTION

EphB4 receptor tyrosine kinase is frequently overexpressed in many types of cancer, including breast, gastric, and colon carcinoma.¹⁻³ Disrupting EphB4 signalling may be a chemotherapeutic approach that retards tumor growth by inhibiting vessel development, therefore depriving cancer cells of nutrients.^{1, 4-5} Several small-molecule inhibitors of the EphB4 kinase domain have been reported⁶⁻¹⁴, but there has yet to be a selective EphB4 kinase inhibitor approved for clinical treatment. However, previous success in targeting tyrosine kinases therapeutically suggests that this approach is warranted in biologically relevant proteins of this family.¹⁵ Novel selective EphB4 kinase inhibitors could represent important lead compounds to be further optimized for use as chemotherapeutic drugs. EphB4-specific molecules could also have potential roles in cancer diagnosis and treatment. The involvement of EphB4 expression in tumour growth and cancer prognosis makes it a potentially useful biomarker for functional imaging modalities like positron-emission tomography (PET).¹⁶ Probes labelled with positron-emitting isotopes could be used with PET to visualize patient tissues expressing high levels of EphB4 indicating the presence of cancerous cells. Information about the expression of EphB4 in a tumour could also provide prognostic information, based on the observation that high EphB4 expression in some tumours is associated with a higher potential for metastasis and fatal disease.¹⁷ In addition to its role in disease, EphB4 plays several crucial functions related to

normal development and physiology, including angiogenesis and neuronal development. Any EphB4 specific modulators could be useful for examining these processes in a research setting.

In previous work, Hierarchical Virtual Ligand Screening (HierVLS) was used to screen a model of the EphB4 kinase domain against 1,153 small molecules and identify three potential ligands.¹⁸ HierVLS is a high-throughput technique that is used to dock different conformations of potential ligands into binding sites of a known protein structure, and then score them using a force-field based approach.¹⁹ Several X-ray crystal structures of the EphB4 kinase domain are available from the Protein Data Bank, and 2VWZ was used as the target structure for virtual screening. In this investigation, we sought to experimentally characterize the effect of three selected hits, aminopterin, dihydrofolic acid (DHF), and N-(2,4)-dinitrophenyl-L-arginine (DNP-L-Arg) on EphB4 kinase activity using two *in vitro* kinase assays: a homogenous time-resolved FRET assay for kinase activity as well as a spectrophotometric coupled kinase assay. Here we describe the results of these assays and discuss potential challenges and future research directions.

6.2. MATERIALS AND METHODS

6.2.1. Computational screening of EphB4 kinase

Virtual modelling and screening of EphB4 kinase are described in full detail in a previous work.¹⁸ EphB4 kinase structure PDB:2VWZ was used for modeling. Missing side chains and atoms were added into the protein structure using YASARA software.²⁰ A missing loop segment was completed by homology modelling using the entire EphB4 sequence. Waters, ligands, and all cofactors (Mn^{2+} , Mg^{2+}) were removed from the structure. The EphB4 model was energy minimized using the Dreiding force field prior to docking.²¹ A database of 2,153 compounds (obtained from PubChem) containing a primary amine functional group was prepared for

docking by removing counter-ions and performing an energy minimization using the MMFF94x force field in MOE.²²⁻²³ This database was prepared in collaboration with John Wigg, and has been described in more detail in his Master's thesis document. The database was screened against the target structure using HierVLS, and three hit compounds were selected for testing based on raw predicted binding score and commercial availability.

6.2.2. Materials

All reagents were obtained from Sigma Aldrich, Inc unless otherwise noted.

6.2.3. Homogenous time-resolved FRET assay (HTRF)

Kinase activity was measured using a homogenous time-resolved fluorescence resonance energy transfer (HTRF) assay kit according to the manufacturer's recommendations (Millipore, Inc. Cat#17-10015).²⁴ The basic scheme for this assay is shown in Figure 6.2.3-1. During the reaction, EphB4 phosphorylates a commercial biotinylated peptide substrate. A europium-chelated (Eu) anti-phosphotyrosine (PTyr) antibody binds to phosphorylated tyrosine residues produced by EphB4 kinase activity. Streptavidin-allophycocyanin complex (Strep-APC) is added after the reaction. Strep-APC will form a strong non-covalent bond with biotin on the peptide substrate. When the substrate has been phosphorylated (from EphB4 kinase activity), the Eu-antibody will bind to PTyr. This will bring the fluorescence resonance energy transfer donor (Eu) into close enough proximity to the acceptor (APC) for the energy transfer (and fluorescence from APC) to occur. Eu, when excited, will fluoresce at 680nm, while APC, when acting as an acceptor, will fluoresce at 620nm. The "HTRF ratio", a relative measure of fluorescence at the two wavelengths, indicates whether this interaction is taking place. The ratio helps to correct for background emission.

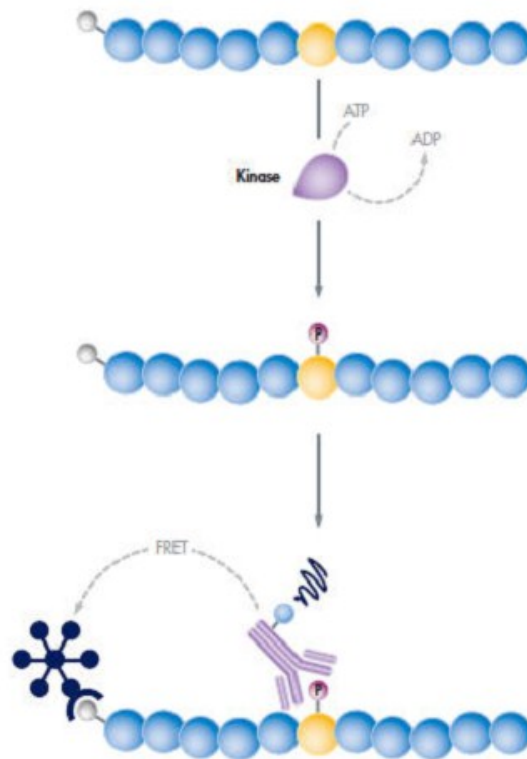


Figure 6.2.3-1 – A basic flowchart indicating the steps associated with kinase reaction detection – EphB4 kinase activity phosphorylates Tyr on biotinylated peptide substrate followed by detection using an Eu-chelated anti-PTyr antibody (binding to PTyr) and Streptavidin-APC (binding to biotin) and subsequent FRET to produce signal. (Figure reproduced courtesy of EMD Millipore from the user guide for Cat#17-10015).²⁴

Active recombinant EphB4 kinase (a.a. 563-987) was obtained from GenWay Biotech, Inc. 1X KinEASE (Millipore, Inc) buffer was prepared following manufacturer’s instructions with components at the following reaction concentrations: stock KinEASE buffer (Part# CS206508) – 1X MnCl₂ (Part# CS206510, 1 mM), MgCl₂ (Part# CS206511) – 5mM, sodium orthovanadate (Part# CS206465) – 0.1mM, and dithiothreitol (DTT) (Part# 90499) – 1mM. 1X KinEASE buffer was used to dilute reaction components from stock unless otherwise specified. Kinase reactions were carried out for 1 hour with light shaking at 37°C in a 96 well black Costar half-area microplate at a 25µL reaction volume with the following reagent concentrations: EphB4 Kinase– 0.8ng/µL (20ng/well), ATP (Part# CS206509) - 100µM, biotinylated TK2 Peptide

(Part# CS206505) – 1 μ M. Detection reagent mix was prepared in detection buffer (Part# CS206478) and added following the kinase reaction, with the final concentrations (in a total well volume of 50 μ L) as follows: Eu-4G10 anti-phosphotyrosine antibody (Part# CS206507) – 0.175 μ g/mL and streptavidin-APC (Part# CS206501) – 0.0625 μ M. Note that the relative detection concentrations of biotin/streptavidin-APC have a fixed 1:8 ratio (biotinylated TK2 peptide –0.5 μ M, streptavidin-APC –0.0625 μ M). Fixing the biotin/Strep-APC ratio is necessary to ensure consistent signal if the concentration of biotinylated peptide changes.

Optimal EphB4 kinase concentration was determined to be 20ng/well (16nM) by performing a standard curve according to the manufacturer's suggested protocol (supplementary material). 1mM stock ligand solutions were prepared for aminopterin, DNP-L-Arg (Cat# D8129), and dihydrofolic acid (DHF) (Cat# D7006) in 20%, 20%, and 0% DMSO (respectively) in water for a maximum reaction DMSO concentration of 2%. A 500 μ M staurosporine (kinase inhibitor control) stock was prepared from the 2.5mM solution provided with the kit (Part# CS201728) for a maximum DMSO reaction concentration of 3.3%. Two-fold serial dilutions of the test compound working solutions were prepared in 1X KinEASE buffer for reaction concentrations of 100, 50, 25, 12.5, and 6.25 μ M. Kinase only (kinase with 3% DMSO in absence of test-compound), buffer only, kinase negative, and Eu-4G10 only control wells were included according to the manufacturer's recommendation. Each reaction was performed in duplicate.

A dose-response curve for biotinylated phosphotyrosine peptide standard was also prepared in tandem with the compound testing assay on the same plate. Lyophilized biotinylated phosphotyrosine standard (Part# CS206507) was dissolved in 109.6 μ L of distilled water for a 50 μ M working solution. A phosphotyrosine peptide standard solution was then prepared using 4.8 μ L of the 50 μ M stock, combined with 30 μ L of streptavidin-APC (stock) and 85.2 μ L of

detection buffer (stock). Plate was read 1 hour after detection reagent addition with the following read parameters: excitation - 360/40 nm, emission - 680-30 nm and 620-40nm, counting delay - 50µsec, integration time - 400µsec. HTRF ratios were calculated according to:

Equation 6.2.3-1 – HTRF ratio calculation based on emission readings at 680nm and 620nm.

$$Ratio = \frac{Emission_{680nm}}{Emission_{620nm}} \times 10,000$$

Curve fitting was performed using PRISM software using the non-linear sigmoidal dose-response curve (variable slope) or linear regression unless otherwise noted²⁵. Hanes-Woolf plots were used to roughly approximate kinetic parameters according to the following relationships:

Equation 6.2.3-2 – The relationship between the slope of the Hanes-Woolf regression (m) and V_{max} .

$$m = 1/v_{max}$$

Equation 6.2.3-3 – The relationship between the x-intercept of the Hanes-Woolf regression and the Michaelis-Menten constant (K_m).

$$x - intercept = -K_m$$

6.2.4. Spectrophotometric coupled kinase assay

The EphB4 kinase reaction was coupled to the enzymatic oxidation of NADH via pyruvate kinase (PK) and lactate dehydrogenase (LDH) enzymes in the presence of phosphoenolpyruvate (PEP) and ATP (Figure 6.2.4-1).

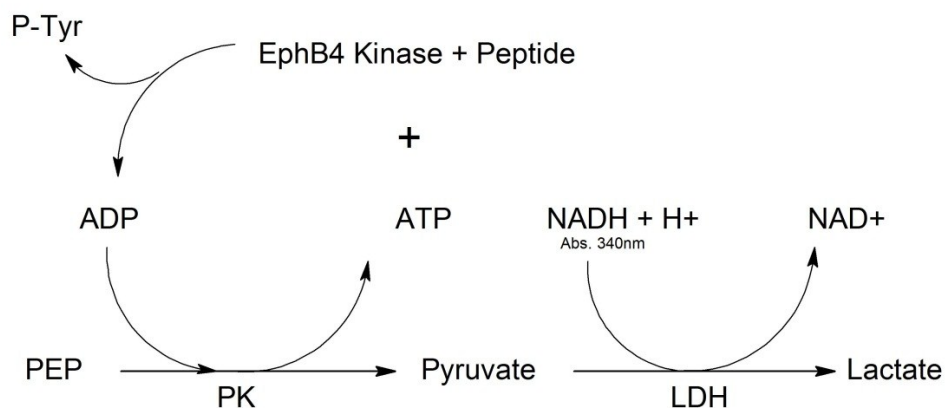


Figure 6.2.4-1 – A spectrophotometric coupled assay was used to measure kinase activity. ADP produced by the kinase reaction would be measured through conversion of phosphoenolpyruvate (PEP) to pyruvate with pyruvate kinase (PK) followed by action by lactate dehydrogenase (LDH) to produce lactate through the oxidation of NADH. NADH absorbs at 340nm allowing the decrease in absorbance to be used to approximate ADP production by EphB4 kinase. ATP is regenerated during this process. Figure generated using ChemDraw Software (Perkin Elmer).

This coupling scheme has been employed in numerous kinetic studies of kinases²⁶⁻³³, including tyrosine kinases such as EGFR^{28, 33} and EphA4 (an EphB4 relative)^{27, 32}. Enzyme, substrate, and cofactor concentrations were selected according to ranges found in the literature.^{27, 32}

Recombinant EphB4 kinase (Cat# SRP0347) was obtained from Sigma Aldrich. Reactions were carried out in 60mM HEPES pH 7.5 buffer containing 20mM MgCl₂, 10mM MnCl₂ and 0.1mM dithiothreitol (DTT). Coupling reaction component concentrations were as follows (unless otherwise stated): PEP (Cat# P7127) – 1mM, PK/LDH (Cat# P0294) – PK – 10-20U & LDH – 20-30U, NADH (Cat# N0786) – 0.3mM). Preliminary reactions were carried out at a fixed EphB4 concentration of 0.05μM with varying Poly-Glu-Tyr (4:1) substrate (Cat# P0275) concentrations and/or varying ATP (Cat# A3377) concentrations. Reaction ATP concentration was otherwise 2mM. NADH standard curves (5 points between 1.5mM – 0.09375mM) were prepared on each plate. Several concentrations of ADP (Cat# A5285) were tested with the coupling system alone (no EphB4), and in the presence of test compound (DNP-L-Arg) to evaluate baseline NADH turnover (alone and in the presence of test compound). To test

compounds for their effect on EphB4 activity, varying DNP-L-Arg (Cat# D8129) and staurosporine (Millipore Cat# CS201728) reaction concentrations were used (200, 100, 50, 25, 12.5, 6.25 μ M for DNP-L-Arg and 100, 50, 25, 12.5, 6.25 μ M for staurosporine). Separate control reactions containing vehicle control (3% DMSO) instead of test compound with and without EphB4 were performed. Test compound reactions and associated controls were run with an EphB4 kinase concentration of 0.025 μ M, Poly-Glu-Tyr concentration of 122 μ M and an ATP concentration of 2mM, in triplicate. All reaction volumes were 100 μ L in a 96 well clear bottom microplate. Reactions were initiated with the addition of EphB4 kinase and proceeded at room temperature. Absorbance at 340nm was read at minimum possible time intervals (2-30seconds) using a Synergy 4 microplate reader. Reactions were monitored for 20-45 minutes depending on whether a steady state appeared to be reached (whether well values continued to change significantly over time). Reaction velocities ($\Delta_{\text{abs}340\text{nm}}/\text{s}$) were determined by selecting a time threshold beyond which reactions exhibited a “steady-state” of absorption decrease over time (generally between 500-1700s from reaction initiation) and performing a linear regression (where $x > \text{threshold}$). For subsequent analysis, reaction wells with a linear regression coefficient (R^2) poorer than 0.65 were excluded to a maximum of 1 out of 3 replicates.

6.3. RESULTS AND DISCUSSION

6.3.1. Compounds selected for experimental testing

Based on the virtual screening of EphB4 and preliminary *in vitro* testing in a previous investigation¹⁸, three compounds were selected for further experimental testing to determine whether they could affect EphB4 activity: N-(2,4)-dinitrophenyl-L-arginine (DNP-L-Arg), dihydrofolic acid, and aminopterin (Figure 6.3.1-1).

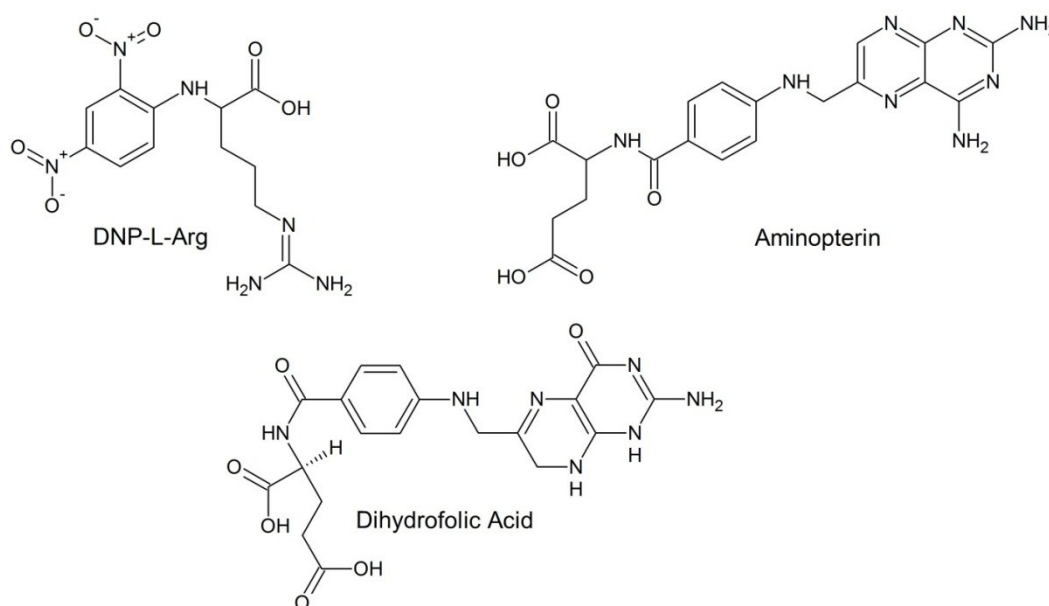


Figure 6.3.1-1 – Three hit compounds were selected for *in vitro* testing: N-(2,4)-dinitrophenyl-L-arginine (DNP-L-Arg), dihydrofolic acid, and aminopterin.

6.3.2. Homogenous time-resolved FRET assay (HTRF)

A kit-based homogenous time-resolved fluorescence assay (HTRF) was used to test all three compounds as well as staurosporine, a known general kinase inhibitor used as a positive control. The dose-response curve for staurosporine against EphB4 obtained using the HTRF assay is shown in Figure 6.3.2-1. There is evidence of an inhibitory effect, and non-linear curve fitting using PRISM yielded an IC₅₀ estimate of 18.59 μ M (95% confidence interval of 17.22-20.07 μ M). The dose-response of staurosporine is also shown in Figure 6.3.2-1 relative to the

other compounds tested. At each concentration tested, DHF and aminopterin showed little change relative to control (EphB4 alone), and did not exhibit a clear dose-dependency. DNP-L-Arg was the only other compound that appeared to have some evidence of effect – at higher DNP-L-Arg concentrations, product formation (evidenced by HTRF ratio) was possibly elevated, and appeared to decrease with decreasing concentration of DNP-L-Arg (Figure 6.3.2-2).

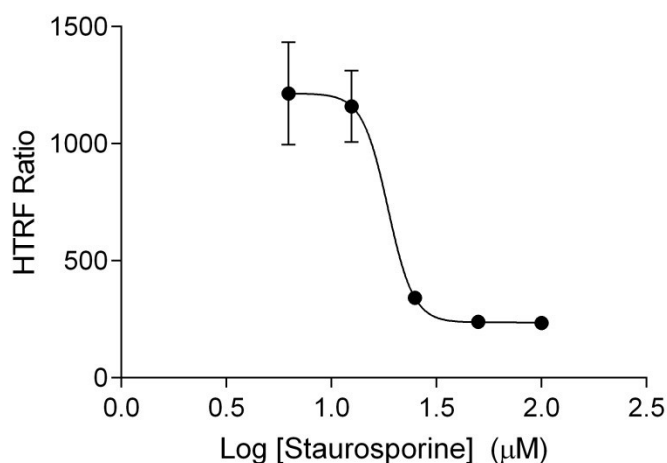


Figure 6.3.2-1 – The HTRF assay dose-response curve for staurosporine against EphB4 (16nM) with an IC₅₀ estimate 18.59µM (95% confidence interval of 17.22-20.07µM).

The dose response curve for DNP-L-Arg is shown in Figure 6.3.2-3. The EC₅₀ estimate from non-linear regression was 26.78µM however a confidence interval could not be approximated given the high level of standard deviation between replicates and poor curve fit. Though there was significant signal variation between replicates, DNP-L-Arg appeared to exhibit an effect in the opposite direction to that of the known kinase inhibitor, staurosporine. This was unexpected, as we anticipated either an inhibitory effect or no effect (such as was the case with the other two compounds tested). Due to the limited ability for this assay format to show significant increase in activity relative to baseline, and overall poor repeatability, we decided to retest DNP-L-Arg against EphB4 kinase again using a separate assay format.

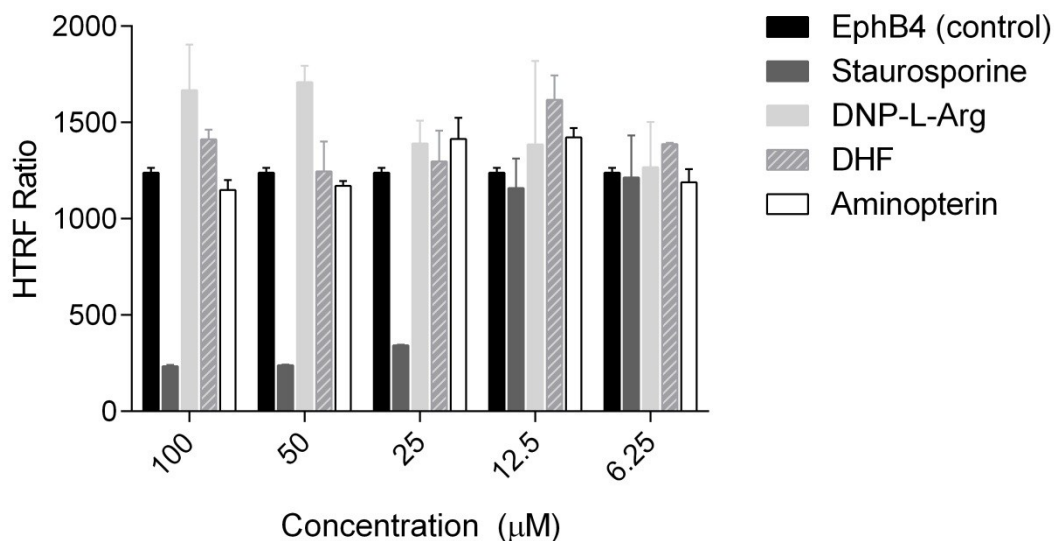


Figure 6.3.2-2 – HTRF ratios are compared for the compounds tested at each concentration. Error bars represent standard deviation between duplicates.

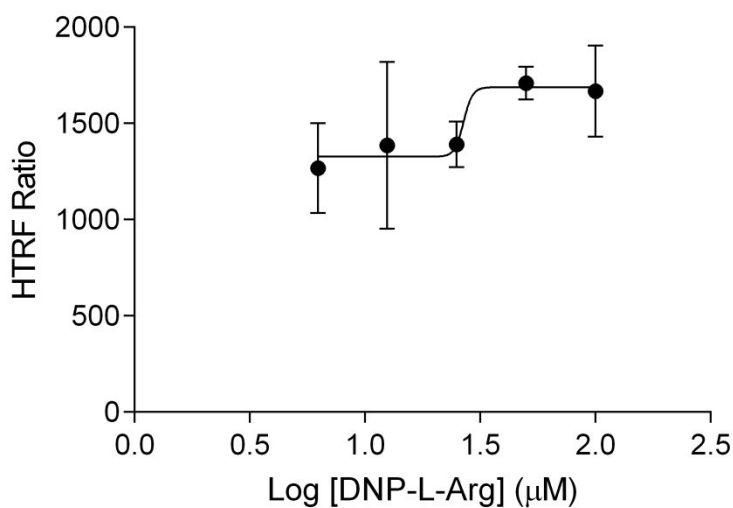


Figure 6.3.2-3 – The HTRF assay dose-response curve for DNP-L-Arg against EphB4 (16nM) indicating a possible shift in effect direction relative to staurosporine, a known kinase inhibitor.

6.3.3. Possible mechanism of DNP-L-Arg interaction based on docked structure

Initial screening had aimed primarily to identify kinase inhibitors. As mentioned in the introduction, many studies in the literature examine kinase inhibition for its therapeutic relevance, and no alternative types of small modulators have been described for the EphB4

kinase domain. Based on the results from the HTRF assay that were possibly counter indicative of inhibition, we revisited the computational docked structure and proposed alternative mechanisms of action. Based on a literature review of tyrosine kinase modulation, we suggest some plausible mechanisms through which a small molecule such as DNP-L-Arg could activate EphB4 kinase.

The kinase domain contains one N-terminal and one C-terminal lobe. The N-terminal lobe generally consists of a 5-stranded β sheet that adopts a twisted structure (strands denoted by $\beta 1$ to $\beta 5$), along with one α -helix (αC). The juxtamembrane segment (JMS) is a short (approximate 10 amino-acid) sequence near the N-terminus of kinase domains. Nomenclature adopted here follows the one previously used to describe the EphB2 structure (PDB: 1JPA).³⁴ The larger C-terminal lobe consists mainly of helices and is connected to the N-terminal lobe by a short “hinge” linker, which forms a flexible inter-lobe cleft. Structural studies of insulin receptor kinase (IRK), EphB2, and EphA4 have described the mechanism of kinase activation and catalysis in detail.^{28, 32, 34-35} Several regions within the β -sheet and connecting loops are responsible for coordinating ATP in the active site, while the C-terminal lobe coordinates the peptide substrate. A flexible activation (A-) loop structure spans the cleft between the two lobes, and plays a key role in regulating activity. Several conserved tyrosine residues in the JMS and A-loop are sites for trans- and auto-phosphorylation. The phosphorylation of these tyrosine residues can activate the kinase by promoting a conformation favorable for catalysis. In an autoinhibited kinase, the dephosphorylated A-loop acts as a pseudo-substrate inhibitor, blocking access of the peptide substrate to its binding site. Upon phosphorylation of A-loop tyrosines, however, the A-loop changes conformation and restores substrate access to the active site. The EphB2 crystal structure (PDB: 1JPA) demonstrates that when autoinhibited, the JMS associates closely with the

N-terminal lobe, altering the position of catalytic residues and preventing an “active” conformation likely by restricting inter-lobe flexibility. This auto/trans phosphorylation activation mechanism is supported by studies on EphB2, EphA4 and insulin-like growth factor receptor.^{27, 34, 36} Li et al postulate that the cleft anchoring autoinhibitory tyrosine residues could also be targeted by small molecules which would displace tyrosine and activate the receptor.³⁶

Several studies also suggest that it is possible to activate tyrosine kinases using small molecules, by affecting one of the mechanisms described above. In 2001, DMAQ-B1, a small molecule microbial isolate, was found to activate IRK in an insulin-independent manner.³⁷ It was proposed that the autoinhibitory conformation of IRK is altered upon DMAQ-B1 binding. Protease sensitivity experiments suggested that DMAQ-B1 altered the conformation of IRK in regions proximal to the ATP-binding site which would be consistent with mechanisms described. The same year, TLK16998, another small-molecule IRK activator, was found to sensitize cells to insulin, acting through the kinase domain.³⁸ While these compounds’ activities have been characterized *in vitro* and *in vivo*, the exact activation mechanism remains unclear.

These previous investigations highlighted the potential for the development of allosteric tyrosine kinase activators. Due to their critical and somewhat ubiquitous role in cell signaling, and development, receptor tyrosine kinases have been extensively studied, largely in the context of therapeutic inhibition. Though the inhibition of some specific tyrosine kinase-mediated signaling pathways seems to be desirable in the context of anti-cancer drug development, the complex roles of Ephs and other tyrosine kinases would likely be better understood if these activation mechanisms were better characterized. The development of alternative (non-inhibitory) modulators could provide valuable insight for research. To date, there are no specific small-molecule allosteric activators of the Eph receptor family that have been reported. The high

degree of structural and functional conservation between these families of enzymes makes it reasonable to expect that mechanisms exploited for drug-development in Ephs would also be useful in the context of related proteins. With this in mind, we examined docked structures for DNP-L-Arg compared to structures containing docked or crystallized active site inhibitors.

Figure 6.3.3-1 displays the superposed orientation of AMP-PNP(green), an ATP-competitive inhibitor co-crystallized with EphB2³⁴, along with the orientation of a known EphB4 inhibitor, 7X6 (orange)(which has also been previously co-crystallized with EphB4⁸) in the orientation that was predicted using our docking protocol, as well as the docked structure for DNP-L-Arg (magenta). Note that both AMP-PNP and 7X6 have overlapping regions that correspond to the ATP binding site. The predicted binding site for DNP-L-Arg, however, does not overlap with the ATP binding location. The predicted orientation of DNP-L-Arg is adjacent to the ATP binding site, associating with the N-terminal domain and, more generally, the inter-lobe cleft region. Two effects could arise from this binding location: because the inter-lobe cleft plays a role in substrate binding, DNP-L-Arg could prevent activity by interfering with the interactions of this lobe and the peptide substrate, or by interfering with the accessibility of the ATP-binding site. DNP-L-Arg could also interfere with normal kinase autoinhibition, in which case its effect could be less predictable. It is reasonable to suggest that DNP-L-Arg could increase the favorability of a catalytically competent conformation, increasing the number of enzyme molecules participating in catalysis. Alternatively, affecting inter-lobe flexibility may alter the substrate binding rates of catalytically competent enzymes that may not be in a fully activated state.

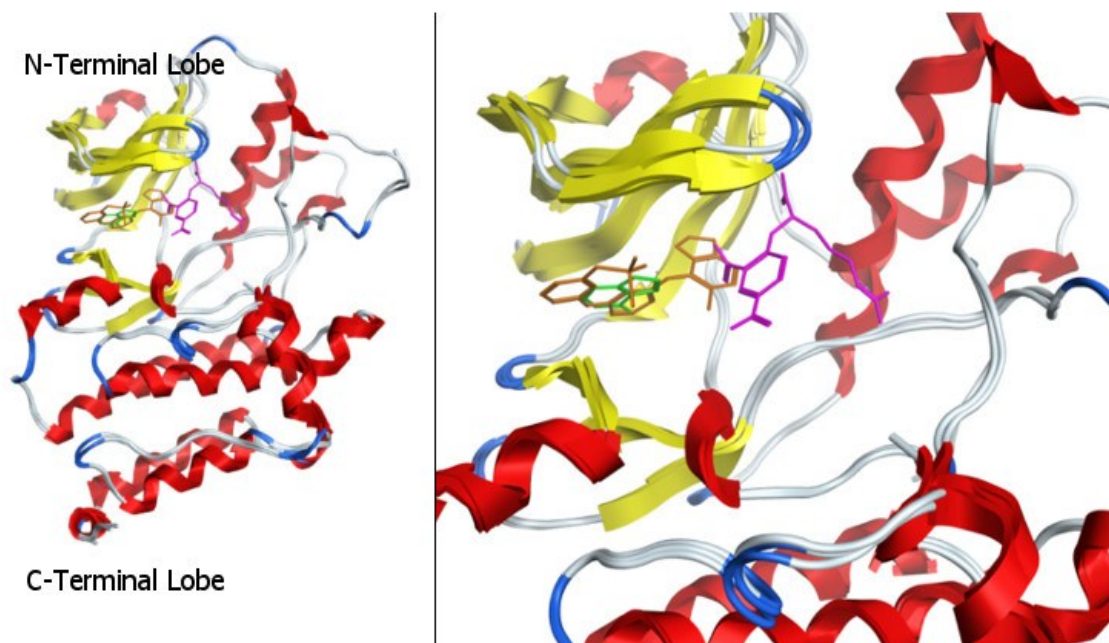


Figure 6.3.3-1 – Left: Docked or crystal structure conformations for three ligands are shown for EphB4 kinase relative to the N- and C-terminal lobes. Right: AMP-PNP (adenine moiety)(PDB:1JPA) (green) and 7X6 (inhibitor)⁸ docked using our protocol (orange) occupy similar sites in the structure while DNP-L-Arg (magenta) is predicted to bind further from the active site.

6.3.4. Spectrophotometric coupled kinase assay

A spectrophotometric coupled assay, as described in the literature, was chosen due to its non-proprietary detection components (in order to facilitate future work), its kinetic format, and its previous use in examining allosteric activation mechanisms in other tyrosine kinases. To test the function of the coupling system, we incubated varying concentrations of ADP with the coupling system alone to ensure rapid turnover of large amounts of ADP (relative to what would likely be produced by EphB4 kinase). The coupling system performed as expected, with all of the observed decrease in absorbance at 340nm (Abs₃₄₀) occurring within the first 10 seconds after the addition of lactate dehydrogenase and pyruvate kinase (PK and LDH). Though these reactions were monitored over time, because the turnover occurred rapidly, we used absorbance at 340nm after 10s to signify overall response (Figure 6.3.4-1). It is clear from this preliminary experiment that even when completely saturated with ADP substrate, the coupling system is

capable of complete turnover in a very short (10s) time window, giving us confidence that it would be appropriately responsive for kinetic tests measuring EphB4 kinase activity.

EphB4 kinase was tested alone (without inhibitor or test ligand) and a constant decrease in absorbance (corresponding to the coupled oxidation of NADH) was observed relative to a control where all components were included except for EphB4 kinase (Figure 6.3.4-2). An initial, very rapid, decrease in Abs_{340} was invariably observed in our tests whenever ATP and the coupling components were present (even in the absence of EphB4), and we suggest that this is likely due to small amounts of ADP present with the ATP, which are rapidly converted by the coupling system near the start of the reaction, and are unrelated to EphB4 kinase activity. Overall, regression coefficients for analyses of the linear regions of each reaction were high (>0.85). Some reactions exhibited a delay in apparent reaction initiation, such that few data points were captured in the linear region. This could perhaps be due to temperature variation or inconsistency in the timing of addition of reagents and measurement. A “lag phase” has been observed in other tyrosine kinase studies²⁷, and could be shortened by pre-incubating the kinase with ATP prior reaction initiation to facilitate phosphorylation-mediated autoactivation. The overall steady-state rates observed ranged between 4-8 $\Delta Abs_{340}/s/\mu\text{mol}$ EphB4, which are significantly higher than those observed by Binns et al using the same assay format for EphA4, who observed a maximum steady-state velocity of 2.12 $\Delta Abs_{340}/s/\mu\text{mol}$ of kinase.²⁷ This difference may be accounted for by the fact that a different Eph isoform, EphA4 was used. DNP-L-Arg and staurosporine were both tested at multiple concentrations (6 and 5 respectively) between 200 μM and 6.25 μM . Selected reactions for DNP-L-Arg and staurosporine at 100 μM are shown in Figure 6.3.4-3.

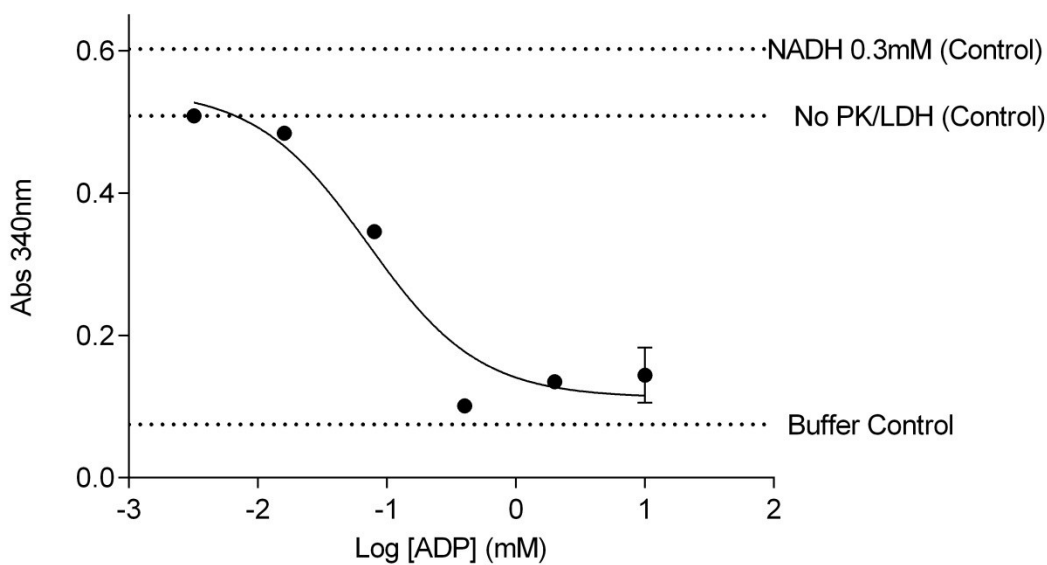


Figure 6.3.4-1 - Abs₃₄₀ at 10s with varying initial concentrations of ADP with the coupling system alone. The Abs₃₄₀ of the starting assay concentration of NADH alone, and in the presence of all reaction components except for PK/LDH (No PK/LDH), as well as a buffer control, are shown for reference.

It is difficult to clearly discern effects here given the high level of variability between assay replicates. In all cases except for at the 100 μ M concentration, staurosporine either exhibited no significant change, or some decrease in velocity relative to control. Testing at higher concentrations, with additional replicates could be used to determine whether the data at 100 μ M was anomalous. DNP-L-Arg data is similarly ambiguous given the standard deviation between replicates. Relative reaction velocities for DNP-L-Arg and staurosporine wells are given in Figure 6.3.4-4. A Hanes-Woolf plot of [Compound]/V as a function of [Compound] is shown in Figure 6.3.4-5.

Regression parameters and associated constants are compiled in Table 6.3.4-1. R^2 values are not given as they cannot be used as a reliable measure of the goodness of the fitting for Hanes-Woolf plots (as plotted variables are not independent). The x-intercept results in a negative K_m determination for the peptide substrate. This could be a result of both the limited

number and range of concentrations tested as well as the limitations of linear approximations for kinetic analyses. Note that apparent V_{\max} values obtained in the presence of DNP-L-Arg and staurosporine differ. Given that staurosporine is known to be an ATP-competitive inhibitor, dissimilar V_{\max} values for DNP-L-Arg, staurosporine, and peptide could indicate different or mixed site of action. However, generally, definitive conclusions based on these data would be unwise. Future experimental work should aim to ensure optimization of reaction components and repeatability with this assay format. A better understanding of compound effects on EphB4 would be possible if staurosporine and DNP-L-Arg were tested at multiple concentrations while varying substrate concentrations (ATP and peptide).

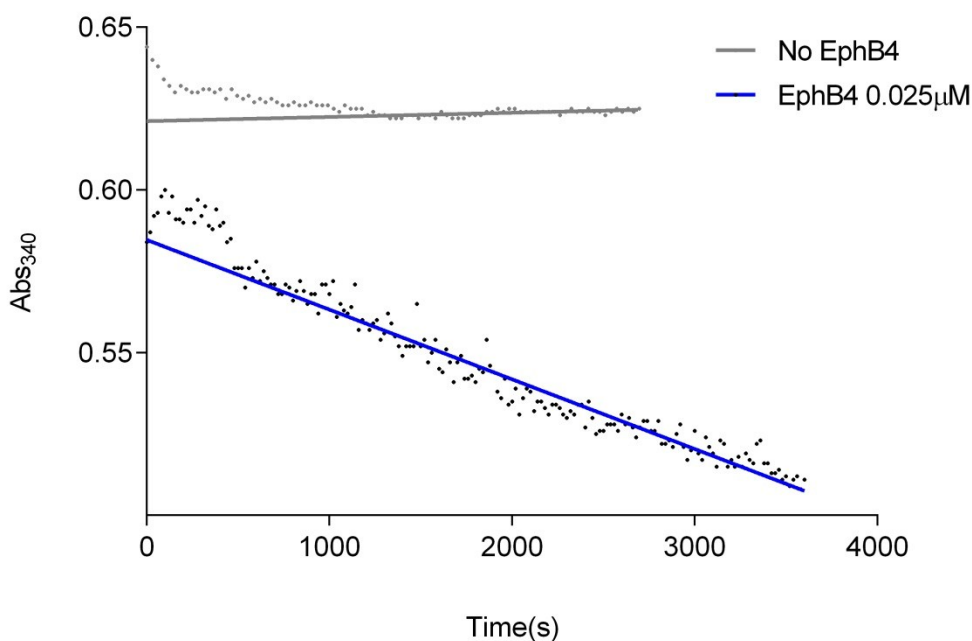


Figure 6.3.4-2 – Spectrophotometric coupled kinase assay indicating decrease in absorbance at 340nm over time where EphB4 kinase is present.

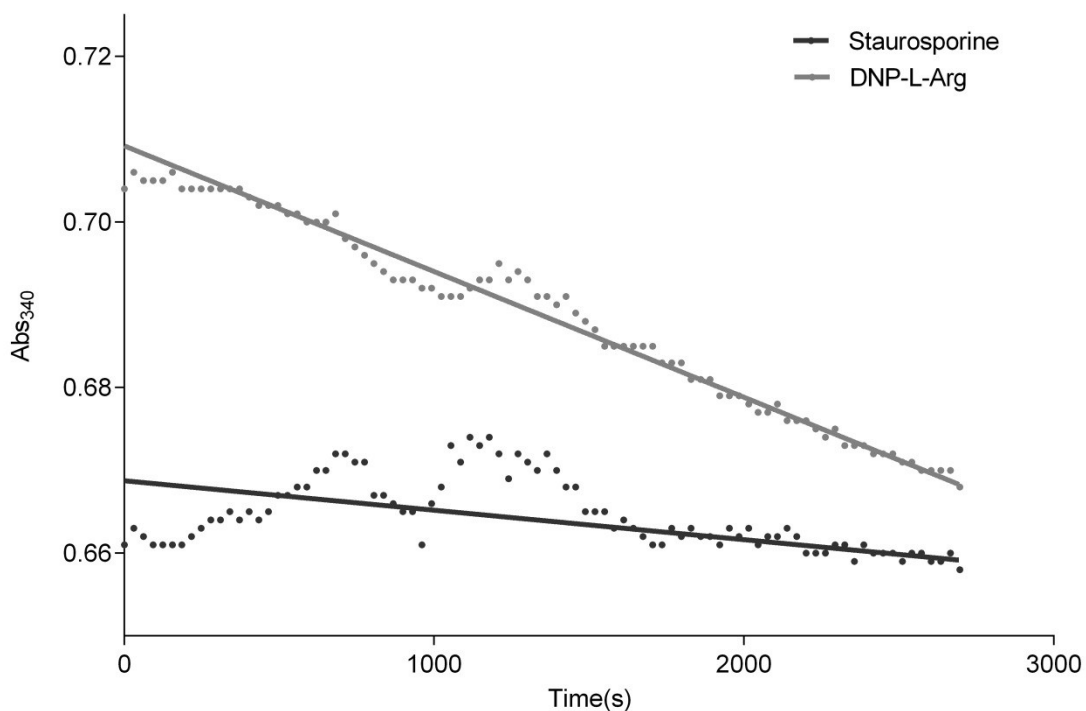


Figure 6.3.4-3 – The coupled assay data from two exemplary reaction wells with staurosporine and DNP-L-Arg each at 100 μ M with EphB4 (0.025 μ M). Linear regression was performed on time>1700s.

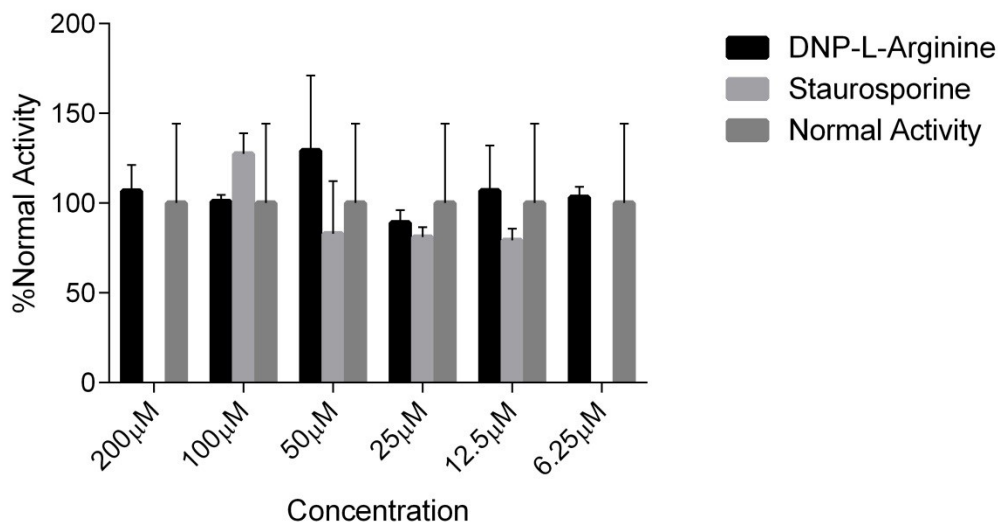


Figure 6.3.4-4 – A comparison of individual reaction velocities (Δ Abs₃₄₀/s/ μ mol EphB4) corrected for baseline rate, as a percent of the rate of normal EphB4 activity (100%) for the compounds tested.

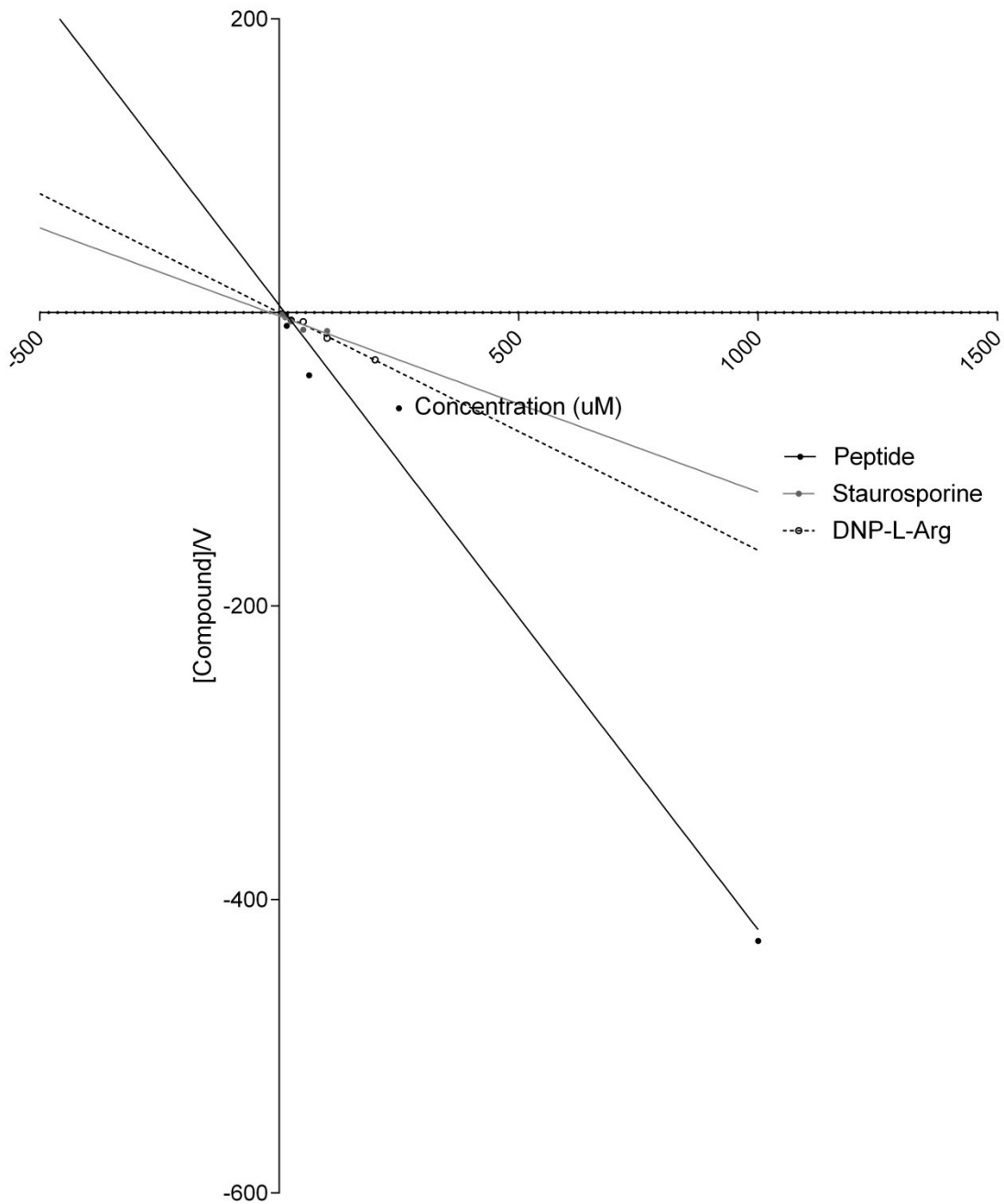


Figure 6.3.4-5 – A Hanes-Woolf plot is used to represent $[Compound]$ (μM) as a function of $[Compound]$ (μM)/ V ($\Delta Abs_{340}/s/\mu mol$ EphB4) using the coupling-assay. Reactions containing staurosporine, DNP-L-Arg, as well as a control reaction with varying peptide (polyGluTyr) concentration are shown. All reactions were performed with the same concentration of ATP and detection components.

Table 6.3.4-1 – Regression parameters obtained from the Hanes-Woolf plot with associated constants for DNP-L-Arg, staurosporine, and varied concentrations of peptide substrate (no inhibitor). Note that reaction velocity is expressed as $\Delta\text{Abs}_{340}/\mu\text{mol EphB4}$.

Compound	Slope (m)	Y-Intercept (b)	X-Intercept (x)	K_m (μM)	V_{max}
DNP-L-Arg	-0.162	-0.0235	-0.145	0.145	-6.18
Staurosporine	-0.120	-2.34	-19.5	19.5	-8.34
Peptide Substrate	-0.426	5.00	11.8	-11.8	-2.35

6.4. CONCLUSIONS

Through high-throughput virtual screening we identified three potential candidates for testing as EphB4 kinase ligands. One *in vitro* assay identified DNP-L-Arg as the most promising candidate for affecting EphB4 kinase activity; however the observed response was that of possible activation rather than inhibition. A second spectrophotometric coupled assay gave inconclusive results. A review of the literature supports the hypothesis that a small molecule such as DNP-L-Arg could activate kinase activity through interfering with normal autoinhibitory mechanisms. The docked structure of DNP-L-Arg further suggests that if DNP-L-Arg is affecting EphB4 activity, it may be doing so through a different mechanism than by competing for the active site, and the possible effects could include activation or inhibition.

The spectrophotometric assay format has the advantages of being rapid, flexible, non-proprietary (uses fairly ubiquitous detection components), and non-radioactive. While it has been used to characterize other, similar enzymes (such as EphA4), no examples of this assay being used with EphB4 have been published as of yet. The data obtained using this assay requires caution. For example, it is possible that the addition of untested components could affect the coupling system. We tested this in a preliminary experiment to ensure that turnover remained rapid in the presence of our test compounds, however this would have to be repeated if different compounds or conditions were to be tested.

The relative uncertainty surrounding the mechanisms by which kinases may be activated or regulated underscores the importance of continuing this research. Additional experimentation could be used to better characterize the effect of DNP-L-Arg on EphB4 kinase activity with varying compound and substrate concentrations. It would be valuable to compare this with the action of known small molecule kinase inhibitors and activation mechanisms using the same assay format.

6.5. REFERENCES

1. Kumar, S. R.; Singh, J.; Xia, G., et al., Receptor tyrosine kinase EphB4 is a survival factor in breast cancer. *Am. J. Pathol.* **2006**, *169* (1), 279-293.
2. Li, M.; Zhao, Z. W.; Zhang, Y., et al., Over-expression of Ephb4 is associated with carcinogenesis of gastric cancer. *Dig. Dis. Sci.* **2011**, *56* (3), 698-706.
3. Stephenson, S.; Slomka, S.; Douglas, E., et al., Receptor protein tyrosine kinase EphB4 is up-regulated in colon cancer. *BMC Mol. Biol.* **2001**, *2* (1), 15.
4. Kertesz, N.; Krasnoperov, V.; Reddy, R., et al., The soluble extracellular domain of EphB4 (sEphB4) antagonizes EphB4-EphrinB2 interaction, modulates angiogenesis, and inhibits tumor growth. *Blood* **2006**, *107* (6), 2330-2339.
5. Krasnoperov, V.; Kumar, S. R.; Ley, E., et al., Novel EphB4 monoclonal antibodies modulate angiogenesis and inhibit tumor growth. *Am. J. Pathol.* **2010**, *176* (4), 2029-2038.
6. Bardelle, C.; Barlaam, B.; Brooks, N., et al., Inhibitors of the tyrosine kinase EphB4. Part 3: identification of non-benzodioxole-based kinase inhibitors. *Bioorg Med Chem Lett* **2010**, *20* (21), 6242-6245.

7. Bardelle, C.; Coleman, T.; Cross, D., et al., Inhibitors of the tyrosine kinase EphB4. Part 2: structure-based discovery and optimisation of 3,5-bis substituted anilinopyrimidines. *Bioorg. Med. Chem. Lett.* **2008**, *18* (21), 5717-5721.
8. Bardelle, C.; Cross, D.; Davenport, S., et al., Inhibitors of the tyrosine kinase EphB4. Part 1: Structure-based design and optimization of a series of 2,4-bis-anilinopyrimidines. *Bioorg. Med. Chem. Lett.* **2008**, *18* (9), 2776-2780.
9. Lafleur, K.; Huang, D.; Zhou, T., et al., Structure-based optimization of potent and selective inhibitors of the tyrosine kinase erythropoietin producing human hepatocellular carcinoma receptor B4 (EphB4). *J. Med. Chem.* **2009**, *52* (20), 6433-6446.
10. Martiny-Baron, G.; Holzer, P.; Billy, E., et al., The small molecule specific EphB4 kinase inhibitor NVP-BHG712 inhibits VEGF driven angiogenesis. *Angiogenesis* **2010**, *13* (3), 259-267.
11. Mitchell, S. A.; Danca, M. D.; Blomgren, P. A., et al., Imidazo[1,2-a]pyrazine diaryl ureas: Inhibitors of the receptor tyrosine kinase EphB4. *Bioorg. Med. Chem. Lett.* **2009**, *19* (24), 6991-6995.
12. Miyazaki, Y.; Nakano, M.; Sato, H., et al., Design and effective synthesis of novel templates, 3,7-diphenyl-4-amino-thieno and furo-[3,2-c]pyridines as protein kinase inhibitors and in vitro evaluation targeting angiogenetic kinases. *Bioorg. Med. Chem. Lett.* **2007**, *17* (1), 250-254.
13. Hu, S. X.; Soll, R.; Yee, S., et al., Metabolism and pharmacokinetics of a novel Src kinase inhibitor TG100435 ([7-(2,6-dichloro-phenyl)-5-methyl-benzo[1,2,4]triazin-3-yl]-[4-(2-pyrroli din-1-yl-ethoxy)-phenyl]-amine) and its active N-oxide metabolite TG100855 ([7-(2,6-

- dichloro-phenyl)-5-methylbenzo[1,2,4]triazin-3-yl]-{4-[2-(1-oxy-pyrrolidin-1-yl)-ethoxy]-phenyl}-amine). *Drug Metab. Dispos.* **2007**, *35* (6), 929-936.
14. Kolb, P.; Kipouros, C. B.; Huang, D., et al., Structure-based tailoring of compound libraries for high-throughput screening: discovery of novel EphB4 kinase inhibitors. *Proteins* **2008**, *73* (1), 11-18.
 15. Levitzki, A.; Gazit, A., Tyrosine kinase inhibition: an approach to drug development. *Science* **1995**, *267* (5205), 1782-1789.
 16. Lee, J. H.; Rosen, E. L.; Mankoff, D. A., The role of radiotracer imaging in the diagnosis and management of patients with breast cancer: part 1--overview, detection, and staging. *J. Nucl. Med.* **2009**, *50* (4), 569-581.
 17. Alam, S. M.; Fujimoto, J.; Jahan, I., et al., Coexpression of EphB4 and ephrinB2 in tumour advancement of ovarian cancers. *Br. J. Cancer* **2008**, *98* (4), 845-851.
 18. Kamstra, R. L. The identification and testing of potential EphB4 ligands using a combination of computational and experimental methods. Honour's Thesis, Lakehead University, Thunder Bay, ON, 2011.
 19. Floriano, W. B.; Vaidehi, N.; Zamanakos, G., et al., HierVLS hierarchical docking protocol for virtual ligand screening of large-molecule databases. *J. Med. Chem.* **2004**, *47* (1), 56-71.
 20. Yasara Biosciences Inc. *Yet Another Scientific Artificial Reality Application (YASARA)*, 10.8.2; 2010.
 21. Mayo, S. L.; Olafson, B. D.; Goddard, W. A., DREIDING: a generic force field for molecular simulations. *J. Phys. Chem.* **1990**, *94* (26), 8897-8909.

22. Halgren, T. A., Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490-519.
23. Chemical Computing Group Inc. *Molecular Operating Environment (MOE)*, 2010.10; Chemical Computing Group, Inc.: Montreal, Quebec, Canada, 2010.
24. Millipore Inc, Phosphotyrosine HTRF ® Assay. 2012.
25. GraphPad Software *PRISM*, 6.02 (Windows); GraphPad Software: La Jolla, California, United States, 2013.
26. Barker, S. C.; Kassel, D. B.; Weigl, D., et al., Characterization of pp60c-src tyrosine kinase activities using a continuous assay: autoactivation of the enzyme is an intermolecular autophosphorylation process. *Biochemistry (Mosc.)* **1995**, *34*, 14843-14851.
27. Binns, K. L.; Taylor, P. P.; Sicheri, F., et al., Phosphorylation of Tyrosine Residues in the Kinase Domain and Juxtamembrane Region Regulates the Biological and Catalytic Activities of Eph Receptors. *Mol. Cell. Biol.* **2000**, *20*, 4791-4805.
28. Jura, N.; Endres, N. F.; Engel, K., et al., Mechanism for activation of the EGF receptor catalytic domain by the juxtamembrane segment. *Cell* **2010**, *137*, 1-25.
29. Nagar, B.; Bornmann, W. G.; Pellicena, P., et al., Crystal Structures of the Kinase Domain of c-Abl in Complex with the Small Molecule Inhibitors PD173955 and Imatinib (STI-571). *Cancer Res.* **2002**, *62* (15), 4236-4243.
30. Schindler, T.; Bornmann, W.; Pellicena, P., et al., Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science* **2000**, *289* (5486), 1938-1942.
31. Technikova-Dobrova, Z.; Sardanelli, A. M.; Papa, S., Spectrophotometric determination of functional characteristics of protein kinases with coupled enzymatic assay. *FEBS Lett.* **1991**, *292*, 69-72.

32. Wiesner, S.; Wybenga-Groot, L. E.; Warner, N., et al., A change in conformational dynamics underlies the activation of Eph receptor tyrosine kinases. *EMBO J.* **2006**, *25*, 4686-4696.
33. Zhang, X.; Gureasko, J.; Shen, K., et al., An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell* **2006**, *125*, 1137-1149.
34. Wybenga-Groot, L. E.; Baskin, B.; Ong, S. H., et al., Structural basis for autoinhibition of the Ephb2 receptor tyrosine kinase by the unphosphorylated juxtamembrane region. *Cell* **2001**, *106*, 745-757.
35. Hubbard, S. R.; Mohammadi, M.; Schlessinger, J., Autoregulatory mechanisms in protein-tyrosine kinases. *J. Biol. Chem.* **1998**, *273* (20), 11987-11990.
36. Li, S.; Covino, N. D.; Stein, E. G., et al., Structural and biochemical evidence for an autoinhibitory role for tyrosine 984 in the juxtamembrane region of the insulin receptor. *J. Biol. Chem.* **2003**, *278*, 26007-26014.
37. Salituro, G. M.; Pelaez, F.; Zhang, B. B., Discovery of a small molecule insulin receptor activator. *Recent Prog. Horm. Res.* **2001**, *56*, 107-126.
38. Manchem, V. P.; Goldfine, I. D.; Kohanski, R. a., et al., A novel small molecule that directly sensitizes the insulin receptor in vitro and in vivo. *Diabetes* **2001**, *50*, 824-830.

Chapter 7.

Conclusions

The focus of these projects was the utilization of computational tools as an integral part of the drug design process. Using virtual ligand screening we identified a series of probe candidates for CAIX that, if confirmed on experimental testing, could be further developed into agents for diagnosis or response to therapy either *in vitro* or *in vivo*. Virtual ligand screening was also used as a proof-of-concept tool to validate an approach that could create computational structures of ligands bound to fluorescent dye moieties, which would be useful for future molecular docking studies against any number of potential targets. Lastly, additional experimental characterization was performed on an EphB4 kinase ligand, DNP-L-Arg, which was previously identified through virtual ligand screening. These results indicate that DNP-L-Arg may act on EphB4 through an activation mechanism, which is supported by the literature and can be explained by software-assisted visualizations of the predicted binding site. The number of techniques and applications explored throughout these projects highlights the versatility of computational methods in these fields and exemplifies the ways in which these techniques can be integrated into different areas of multidisciplinary research. A major advantage of computational techniques is the ability to explore molecular systems theoretically, in a manner that is less costly, prior to undertaking significant experimental work. However, there are always limitations in the ability of simulations to accurately predict conformations and interactions. In each of the projects described here, additional experimental work will be required to translate the computational results into relevant solutions for biochemical applications.

Supplementary Information

Due to the electronic nature of much of the supplementary data and material referred to in this text, it has been organized in an online repository that is freely available at the following URL:

<http://oxala.lakeheadu.ca/bin/view/Main/Supplementary-RK>