# THE SEARCH FOR NOVEL INHIBITORS OF COACTIVATOR-ASSOCIATED ARGININE METHYLTRANSFERASE 1

by

Shona Anne Sedlock

B.Com., The University of Victoria, 2007

### A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate and Postdoctoral Studies (Pharmaceutical Sciences)

## THE UNIVERSITY OF BRITISH COLUMBIA (Vancouver)

August 2015

© Shona Anne Sedlock, 2015

#### Abstract

Coactivator-associated arginine methyltransferase 1 (CARM1) is a member of the protein arginine methyltransferase (PRMT) family of enzymes, and is also known as PRMT4. PRMTs catalyse the transfer of methyl groups from *S*-adenosyl-L-methionine (SAM) to the side chain of arginine residues in substrate proteins. The dysregulation of CARM1 contributes to the onset and progression of breast and prostate cancer. For this reason, CARM1 is a target for inhibition, yet CARM1 inhibitors to date either lack selectivity or fail to show anti-proliferative effects in cells. This work aims to identify novel small molecules that can be further developed to inhibit CARM1 activity in cancer cells.

Using the crystal structure (2Y1X) of the CARM1 catalytic domain in complex with CMPD-2 and *S*-adenosyl-L-homocysteine as a model for downstream screening, a computer-aided drug discovery and design (CADD) pipeline was developed, enabling docking to identify novel inhibitors. Over 76,600 compounds alongside known CARM1 inhibitors were screened using the industry standard proprietary software suite Accelrys Discovery Suite 4.5®. LibDock and CDOCKER algorithms were deployed independently and in series. Subsequently, a P81 filterbinding assay assessed the top hits from the *in silico* screening for CARM1 inhibition *in vitro* to generate IC<sub>50</sub> values. A lead CARM1 inhibitor, Diamine 12 was used as a positive control, which scored highly with LibDock and CDOCKER and had an IC<sub>50</sub> value of  $1.3\pm1 \mu$ M with the P81 filter-binding assay. Top-ranked hits identified using Accelrys® showed some binding interactions in the arginine-binding cavity yet little to no activity *in vitro*. The work here fundamentally addresses the development of a workflow that provides a platform for discovery consisting of *in vitro* and *in silico* screening methods. Future work will involve expanding on the

ii

findings here to identify novel CARM1 inhibitors to be developed into therapeutic agents for the treatment of breast and prostate cancers.

### Preface

This thesis contains work completed wholly in Drs. Adam Frankel, Tara Leah Klassen's and David Grierson's laboratories. The work would not have been possible without the generous collaboration of Dr. Tara Leah Klassen, Dr. David Grierson, Dr. Mark Bedford, Dr. Tom Pfeifer and Dr. Eric Therrien.

In Chapter 2, I expressed and purified CARM1 and GST-PABP1(437-488). I separated CARM1 and GST-PABP1(437-488) samples by gel electrophoresis and performed the Bradford assay and densitometry on the PABP1 fragment as well as the Edelhoch method on CARM1 in order to quantify the proteins. I performed all methylation activity assays and designed and conducted all P81 filter-binding assays. I performed the  $IC_{50}$  experiment for Diamine 12, plotted and calculated the  $IC_{50}$  value using SigmaPlot. I also completed all mass spectrometry assays with the assistance of Andras Szeitz. In Chapter 3, I performed all in silico docking using LibDock and CDOCKER with the Accelrys Discovery Suite 4.5<sup>®</sup>. In Chapter 4, I conducted all analysis on both the in vitro screening and in silico docking results.

In Chapter 2, the plasmid DNA of GST-PABP1(437-488) was a generous gift from Dr. Mark Bedford. The library of GPS compounds for *in silico* docking was received from Dr. Grierson in an excel spreadsheet. With the help of Ying Gong, we were able to draw all 384 compounds in Accelrys Draw 4.2 to be converted into .mol files. This was important as only .sd and .mol files are compatible with Accelrys<sup>®</sup>. The GPS compounds were received as 20 mM stocks in DMSO from Dr. Grierson. The library of CCBN and Diver Set compounds were provided generously by Dr. Tom Pfeifer at the CDRD as .sd files and compounds for *in vitro* screening were retrieved by Jack Yang at the CDRD. Diamine 12 and a peak purity analysis were received from Dr. Eric Therrien. In Chapter 3, the in silico docking using RosettaLigand<sup>®</sup> was performed by Dr. Jennifer Bui at UC San Diego. Dr. Tara Leah Klassen was an integral in providing insight, method development and analysis for the in silico docking with the Accelrys<sup>®</sup>.

I contributed to the ChemBioChem publication (Thomas, D. et al. Protein arginine Nmethyltransferase substrate preferences for different N $\eta$ -substituted arginyl peptides. ChemBioChem 15, 1607–1613 (2014)). I also had an accepted abstract for the 2015 AAPS Biotechnology conference in San Francisco and was asked to contribute to the AAPS blog where my article 'Computer Modeling that Expedites Drug Discovery' was posted on June 9, 2015 on the AAPS blog website.

### **Table of Contents**

| Abstractii  |
|---|
| Prefaceiv   |
| Table of Contentsv  |
| List of Tables vi   |
| List of Figures vii   |
| List of Abbreviationsix   |
| Acknowledgementsx   |
| 1. Introduction   |
| 1.1 Protein arginine <i>N</i> -methyltransferase structure and activity |
| 1.2 CARM1   |
| 1.3 Hypothesis and research objectives                                  |
| 2. Methods  |
| 2.1 Reagents and sources for the <i>in vitro</i> assays                 |
| 2.2 In vitro assays   |
| 2.3 Computer-aided drug discovery and design (CADD)                     |
| 3. Results  |
| 3.1 Rosetta ligand results  |
| 3.2 LibDock and CDOCKER results   |
| 3.3 CDOCKER only results  |
| 3.4 LibDock scores of 2Y1X vs. 2Y1W                                     |
| 4. Conclusions  |
| References  |
| Appendix  |

### List of Tables

| Table 1.1. CARM1 inhibitors   | . 26 |
|---|------|
| Table 2.1. A comparison of assays used to screen CMPD-1, CMPD-2, Diamine 12, GPS and        |      |
| CDRD compounds  | . 42 |
| Table 2.2. A Comparison of RosettaLigand, LibDock and CDOCKER                               | . 46 |
| Table 3.1. Predicted K <sub>I</sub> values for CARM1-binding compounds from the GPS library | . 59 |
| Table 3.2 Interactions of false positives and top CARM1 inhibitors                          | . 73 |

## List of Figures

| Figure 1.1. Activity classifications of PRMTs   | 4       |
|---|---------|
| Figure 1.2. Conserved structure of human PRMT catalytic core                                | 6       |
| Figure 1.3. Dimeric arrangement of Type I PRMTs   |         |
| Figure 1.4. A schematic illustration illustrating the structural mechanism of Type I PRMT   | Гs 8    |
| Figure 1.5. Active site of CARM1 with key glutamate residues involved in methyl transfe     | er 11   |
| Figure 1.6. The catalytic domain of CARM1 in complex with inhibitor CMPD-2 and SAI          | Η       |
| (2Y1X)  | 16      |
| Figure 1.7 Surface representation and binding pocket of CMPD-2 (2Y1X)                       | 17      |
| Figure 1.8. Surface representation of CARM1 <sub>apo</sub> and CARM1 <sub>holo</sub> (2Y1X) | 19      |
| Figure 1.9. Residues within the CARM1 active site stabilize SAH and cause conformatio       | nal     |
| changes   | 19      |
| Figure 1.10. SAH binding to CARM1 causes large structural changes                           | 20      |
| Figure 2.1. Purification of recombinant CARM1   | 30      |
| Figure 2.2. Purification of GST-PABP1(437-488)  | 30      |
| Figure 2.3. Time-dependent methylation of PABP1 by CARM1                                    | 33      |
| Figure 2.4. Effect of DMSO on the standard calibration curve using the UPLC-MS/MS as        | ssay to |
| quantify SAH  | 36      |
| Figure 2.5. The linear range of the P81 filter-binding assay                                | 39      |
| Figure 2.6. The K <sub>m</sub> of the P81 filter-binding assay                              | 39      |
| Figure 2.7. Effect of DMSO on the P81 filter-binding assay                                  | 40      |
| Figure 2.8. Schematic representation of RosettaLigand, LibDock and CDOCKER algorith         | 1ms.48  |
| Figure 2.9. Decision tree outlining the selection and rationale for developing the CARM1    | 2Y1X    |
| protein model used for CADD studies   | 50      |
| Figure 2.10 Difference between LibDock score with water minus without water vs. LibDoc      | ock     |
| score without water   | 53      |
| Figure 2.11 Ligand preparation for in silico docking using the Accelrys Biovia Discovery    | Studio  |
| for CADD studies  | 54      |
| Figure 2.12 Tautomer generated of CMPD-2.   | 55      |
| Figure 3.1. '40 set' screened using the P81 filter-binding assay                            | 61      |
| Figure 3.2 CMPD-2 docked using LibDock shows same interactions as crystal structure         | 63      |
|   | vii     |

| Figure 3.3. Activators of CARM1 used as negative controls for docking                   | 4 |
|---|---|
| Figure 3.4. LibDock vs. biased CDOCKER scores 40 set and controls                       | 6 |
| Figure 3.5. Leading CARM1 inhibitors occupy the arginine-binding cavity                 | 7 |
| Figure 3.6 Clusters centers of CARM1 inhibitors orientation in binding pocket           | 7 |
| Figure 3.7. IC <sub>50</sub> curve for Diamine 12 using the P81 filter-binding assay    | 8 |
| Figure 3.8. LibDock vs. biased CDOCKER scores for GPS and CDRD compounds7               | 0 |
| Figure 3.9. In vitro results of 5 top-ranked GPS compounds                              | 2 |
| Figure 3.10. In vitro results of top-ranked CDRD compounds, GPS383 and GPS347 with 4%   |   |
| DMSO  | 2 |
| Figure 3.11 Hydrogen-bond interaction profiles for false positive and actual inhibitors | 4 |
| Figure 3.12. Biased vs. Unbiased CDOCKER scores for controls, GPS and CDRD compounds    |   |
|   | 6 |
| Figure 3.13. LibDock scores for 2Y1W vs. 2Y1X for GPS and CDRD compounds                | 8 |

### List of Abbreviations

| aDMA             | $\omega - N^G, N^G$ -dimethylarginine                     |
|------------------|---|
| BSA              | Bovine serum albumin                                      |
| CADD             | Computer-aided drug discovery and design                  |
| CARM1            | Coactivator-associated arginine methyltransferase 1       |
| CHARMm           | Chemistry at HARvard Molecular Mechanics                  |
| CMPD1-2          | Compound 1 and 2  |
| СРМ              | Counts per minute   |
| CTD              | C-terminal domain   |
| DMSO             | Dimethyl sulfoxide  |
| DPM              | Disintegrations per minute                                |
| ERα              | Estrogen-receptor α                                       |
| GOLD             | Genetic optimization for Ligand Docking                   |
| GRIP1            | Glutamate receptor-interacting protein 1                  |
| GST-GAR          | Glutathione-S-Transferase - Glycine and Arginine rich     |
| HKMT             | Histone lysine methyltransferase                          |
| HuD              | Hu antigen D  |
| HuR              | Hu antigen R  |
| IC <sub>50</sub> | Half maximal inhibitory concentration                     |
| K <sub>d</sub>   | Dissociation constant                                     |
| LC-MS/MS         | Liquid chromatography tandem mass spectrometry            |
| MCF7             | Michigan Cancer Foundation-7                              |
| MMA              | $\omega - N^G$ -monomethylarginine                        |
| PABP1            | Poly(A) binding protein 1                                 |
| PDB              | Protein data bank   |
| PMT              | Protein methyltransferase                                 |
| PRMT1-9          | Protein arginine N-methyltransferase 1-9                  |
| RMSD             | Root-mean-square deviation                                |
| SAH              | S-adenosyl-L-homocysteine                                 |
| SAM              | S-adenosyl-L-methionine                                   |
| SAP49            | Spliceosome-associated protein 49                         |
| SAR              | Structure-activity relationship                           |
| sDMA             | $\omega - N^{\eta^1}$ , $N^{\eta^2}$ -dimethylarginine    |
| SDS-PAGE         | Sodium dodecyl sulfate polyacrulamide gel electrophoresis |
| SH3              | SRC homology domain 3                                     |
| SiRNA            | Small interfering RNA                                     |
| SRC-1            | Steroid receptor coactivator-1                            |
| TCA              | Trichloroacetic acid                                      |
| vHTS             | Very high-throughput screening                            |

#### Acknowledgements

First, I would like to acknowledge and thank my supervisors, Drs. Adam Frankel and David Grierson, who both continuously exceeded my expectations of a supervisor. Without their support, insightful contributions and ability to assist while still allowing me to think and make decisions on my own, this work would not have been possible. I also wanted to personally thank Dr. Frankel for allowing me to conduct the majority of this work in his laboratory and for always having a positive outlook.

I would also like to express my sincere gratitude to Dr. Tara Leah Klassen, who not only provided me access to her laboratory and computers to conduct all of the *in silico* docking but spent countless hours providing in-depth knowledge and insight from her many years of experience with computer modelling. I also extend thanks to my former colleague, Dr. Dylan Thomas, who helped initially train me in the laboratory and provided answers to questions after he had left the Faculty and to my current colleague Mynol I. Vhuiyan, who helped me overcome technical hurdles and trained me in new techniques.

Special thanks are owed to my mom (Shevaun Sedlock), dad (Robert Sedlock), sisters (Jessica and Sarah Sedlock), as well as my fiancé (Fraser Taylor-Mitchell) for their unending love and support. I am so blessed to have the family that I do, and I could not have gotten through this work without them.

#### 1. Introduction

#### 1.1 Protein arginine *N*-methyltransferase structure and activity

#### Post-translational modification

The primary sequence of a protein is dictated by the genetic code <sup>1</sup>. After translation, the chemical complexity of the protein can be altered through posttranslational modifications <sup>2</sup>. When post-translational modifications occur on basic histone tails, they affect the binding of transcriptional factors and alter chromatin structure which can impact gene expression without altering the original polypeptide sequence <sup>3</sup>. These modifications are dynamic and have been shown to influence transcription, affect DNA processes such as repair, replication and recombination and effectively regulate the function of eukaryotic cells <sup>2</sup>. An ever-growing list of posttranslational modifications and the complexity of their action is only beginning to be understood.

#### Arginine methylation

An important post-translational modification that was discovered in the 1960s is arginine methylation <sup>4</sup>. Arginine methylation occurs when the nitrogens of arginine residues within polypeptides are post-translationally modified to contain methyl groups <sup>1</sup>. Arginine is unique among the amino acids as its guanidine group contains five potential hydrogen bond donors that can make favourable interactions with biological hydrogen bond acceptors <sup>5</sup>. Methylation can affect hydrogen-bonding patterns by creating added hydrophobicity and only slightly altering the pK<sub>a</sub> of the guanidino group <sup>3</sup>. Arginine methylation is catalyzed by protein arginine *N*-methyltranferases (PRMTs) both on nuclear and cytoplasmic proteins <sup>6</sup>. PRMTs are a family of

nine proteins, some of which catalyze the posttranslational methylation from *S*-adenosyl-Lmethionine (SAM) to nitrogen atoms on the guanidino group of arginine residues within substrate proteins <sup>7</sup>. The reaction results in the formation of methylated arginine residues and *S*-adenosyl-Lhomocysteine (SAH). PRMTs can be classified into three main types: type I, type II and type III. Both type I and type II PRMTs transfer two methyl groups, however, type I enzymes produce asymmetric  $N_{9^1}$ ,  $N_{9^1}$ -dimethylarginine (aDMA) and type II enzymes produce symmetric  $N_{9^1}$ ,  $N_{9^2}$ dimethylarginine (sDMA) (Figure 1.1). Type III PRMTs transfer only one methyl group to form  $N_9$ -monomethylarginine (MMA).

Type I, type II and type III activities of PRMTs that generate aDMA, sDMA and MMA respectively are shown in Figure 1.1 . Type I enzymes include PRMT1, 2, 3, 4, 6 and 8<sup>8</sup>. PRMT5 and PRMT9 are the only type II enzymes, <sup>8,9</sup> and PRMT7 is the only type III enzyme <sup>10</sup>. PRMT1 is the predominant type I arginine methyltransferase in mammals accounting for approximately 85% of cellular PRMT activity <sup>11</sup>.

How does arginine methylation affect protein function? Arginine is a positively charged amino acid with a basic pK<sub>a</sub> of approximately 12.48 <sup>3,12</sup>. Studies have shown that methylation does not change the positive charge of the arginines, but increases hydrophobicity and may prevent hydrogen bonding and/or introduce steric constraints <sup>13</sup>. This influences protein-protein interactions that can occur. For example, the WD40-repeat protein 5 (WDR5) has been reported to selectively bind to histone H3 (a PRMT substrate) and induce conformational changes in the protein that accompany histone binding. However, when arginine methylation occurs, WDR5 does not bind to histone H3 <sup>14</sup>. Arginine methylation is a key process that affects protein-protein

interactions and cellular processes including intracellular localization, transcriptional regulation, RNA metabolism and DNA damage repair <sup>13,15</sup>. PRMTs have been shown to affect cell growth, proliferation and differentiation, can be deregulated in cancer and other disease conditions and are emerging as promising therapeutic targets <sup>6</sup>.



#### Figure 1.1. Activity classifications of PRMTs

Type I, type II and type III activities of PRMTs that generate aDMA, sDMA and MMA, respectively, as shown. Figure is adapted from reference <sup>16</sup>.

#### Conserved structural features of PRMTs

PRMT structure is highly conserved among all eukaryotes <sup>11,15</sup>. Crystal structures have been determined for PRMT1, PRMT3, PRMT4 (also known as co-activator-associated arginine methyltransferase 1 or CARM1), PRMT5 and PRMT6 <sup>17–19</sup>. PRMTs are organized around a conserved catalytic core composed of an N-terminal Rossmann fold where SAM binds, and a C-terminal β-barrel domain where the substrate binds (Figure 1.2) <sup>6</sup>. Some PRMTs also contain protein-protein interaction modules that extend beyond the catalytic core. Examples include an SH3 domain in PRMT2 <sup>20</sup>, a zinc finger in PRMT3 <sup>21</sup>, a pleckstrin homology domain in CARM1<sup>18</sup>, and a triosephosphate isomerase (TIM) barrel in PRMT5 (Figure 1.2) <sup>6</sup>. It is suggested that these structural domains recruit other proteins necessary to form a catalytically active complex, facilitate substrate binding, and/or contribute to oligomerization of PRMT monomers <sup>6</sup>. Lastly, PRMTs contain a dimerization arm that extends from the C-terminal β-barrel domain to make reciprocal interactions with a second PRMT subunit, forming an active dimer (Figure 1.2) <sup>11,17,18,22</sup>.



#### Figure 1.2. Conserved structure of human PRMT catalytic core

The catalytic domain of PRMTs contains a Rossman fold (green) and the C-terminal domain showing a  $\beta$ -barrel domain (purple). The  $\alpha$ -helix (blue) is found in the N-terminal domain and the dimerization arm is found in the C-terminal domain (cyan). In PRMT2, PRMT3, CARM1, PRMT5, and PRMT9 there are also structural domains involved in different protein-protein interactions (brown). Abbreviations are: SRC Homology 3 Domain (SH3), zinc finger (ZnF), pleckstrin homology (PH), triosephosphate isomerase (TIM) barrel, and transient receptor potential (TRP). Figure adapted from <sup>6</sup>.

#### Type I and III PRMTs

The solved structures of type I PRMTs, to date, have been shown to dimerize in a head-to-tail arrangement where the dimerization arm extending from the  $\beta$ -barrel makes hydrophobic interactions with the Rossman fold of another subunit (Figure 1.3)<sup>6,11,17,22,23</sup>. This, however, is not the case with PRMT7, the only type III enzyme. Although no crystal structure of mammalian PRMT7 has been reported, CePRMT7 (C. elegans) has been shown to have a slight variation to this head-to-tail arrangement <sup>6</sup>. The N-terminal domain adopts a canonical PRMT structure while the C-terminus, unique to PRMT7, mimics a second type I PRMT subunit <sup>6</sup>. Interestingly, an important pair of glutamic acid residues conserved in active PRMTs thought to signify an active PRMT monomer/catalytic monomer, are absent from the C-terminal CePRMT7 core <sup>24</sup>. Also of importance in type I and III PRMTs is an  $\alpha$ -helix, composed of  $\alpha$ -X and  $\alpha$ -Y, at the N-terminal domain (Figure 1.2)<sup>6</sup>. The folding of  $\alpha$ -X and  $\alpha$ -Y upon SAM binding has been shown to complete the formation of the substrate binding site and facilitate recruiting and positioning the substrate <sup>6</sup>. SAM becomes buried in the core of the Rossmann fold and in turn, the  $\beta$ -barrel and  $\alpha$ -helix at the substrate-binding site become clearly defined (Figure 1.4)<sup>6</sup>. The binding of SAM alone, however, may not be a sufficient explanation of the folding of this  $\alpha$ -helix as it has been shown to fold in a variety of conformational states in the holo structures of PRMT1, PRMT3, CARM1, and PRMT6<sup>11,18,25</sup>. CARM1 also demonstrates a profound structural rearrangement of the  $\alpha$ -helix upon SAM binding, which will be discussed in detail in section 1.2<sup>18</sup>.



#### Figure 1.3. Dimeric arrangement of Type I PRMTs

Type I PRMTs have been shown to dimerize in a head-to-tail arrangement where the  $\beta$ -barrel (purple) interacts with the Rossman fold (green) of another monomer as shown here with CARM1. The dimerization arm is shown in blue. Structure was compiled in PyMol version 1.7.4 (PDB file 2Y1X).



#### Figure 1.4. A schematic illustration illustrating the structural mechanism of Type I PRMTs

The folding of  $\alpha$ -X and  $\alpha$ -Y upon SAM binding has been shown to complete the formation of the substrate-binding site and facilitates in recruiting and positioning the substrate. Figure was adapted from reference <sup>6</sup>.

#### Type II PRMTs

PRMT5 and PRMT9 are the only type II enzymes <sup>69</sup>. A recent crystal structure of PRMT5 (4GQB) shows that the N-terminal Rossman fold and C-terminal  $\beta$ -barrel are conserved in this type II PRMT, however, the  $\alpha$ -X and  $\alpha$ -Y helices are not <sup>6</sup>. Instead, PRMT5 has a short  $\alpha$ -Y helix preceded by a loop that links the Rossman fold to a PRMT5-specific TIM-barrel at the N-terminal domain (Figure 1.2) <sup>6</sup>. The purpose of the TIM-barrel is to recruit MEP50, a protein which act as a structural scaffold for the recruitment of substrates <sup>19</sup>. Another structure unique to PRMT5 is its tetrameric form <sup>6</sup>. Unlike type I and III PRMTs which have a dimerization arm, PRMT5 makes extensive interactions between the TIM barrel of one subunit and the Rossman fold of another <sup>6</sup>. Two dimers are arranged to form a tetramer at which the center lies the substrate binding site <sup>6</sup>. PRMT9 was also only recently discovered as a type II enzyme <sup>9</sup>. PRMT9 was shown to methylate the spliceosome-associated protein SAP145 at arginine 508, which takes on the form of monomethylated arginine (MMA) and symmetrically dimethylated arginine (sDMA). To date, PRMT5 and PRMT9 are the only mammalian enzymes capable of depositing the sDMA mark <sup>9</sup>.

#### PRMT active sites and key residues involved in methyl transfer

There is a rich literature base of crystallographic and enzyme kinetic studies of PRMTs that have helped to define their mechanism of catalysis <sup>26</sup>. PRMTs use a bimolecular nucleophilic substitution ( $S_N 2$ ) methyl transfer mechanism in which the lone pair of electrons of the arginine nitrogen atom attacks the electrophilic methylsulphonium cation of SAM at a 180° angle to the leaving group, to form a penta-coordinate carbon transition state <sup>26,27</sup>. The transition state

collapses, with methyl group relocation to the nitrogen atom of the arginine side chain to form methylarginine and the co-product SAH.

A common structural feature that distinguishes protein methyltransferases (PMTs) from other proteins that use SAM as a cofactor is the overall structure of their catalytic sites <sup>18</sup>. This consists of a SAM-binding pocket that is accessed from one face of the protein, and a narrow, hydrophobic, acceptor (arginine) channel that extends to the opposite face of the protein surface <sup>18</sup>. Crystal structures have shown that there are two distinct binding modes for SAM or SAH in the cofactor-binding pockets. When SAM binds in the active site, there is an alignment of the SAM methylsulphonium cation at the base of the acceptor-binding channel <sup>18</sup>. Because PRMTs are capable of performing multiple rounds of arginine methylation to produce either MMA, sDMA or aDMA, it is has been shown that PRMTs follow and ordered, sequential mechanism in which SAM binds before the arginine-containing substrate <sup>26,28</sup>.

There are key residues involved in methyl transfer that are highly conserved across the PRMT family. The "double-E" loop found in the active site of all PRMTs consists of two invariant glutamate residues that form a stable salt bridge with the N<sub> $\varepsilon$ </sub> and N<sub> $\eta$ 2</sub> of the substrate <sup>29</sup>. These two glutamate residues are required for enzymatic activities, and the mutation of either of them has been shown to greatly decrease enzymatic activity <sup>26</sup>. For example, the catalytic core of CARM1 contains Glu258 and Glu267 located in the arginine-binding cavity that extends towards the site of methyl transfer (Figure 1.5) <sup>17</sup>. The two glutamate residues have also been shown to be key binding residues for top CARM1 inhibitors <sup>17</sup>.



#### Figure 1.5. Active site of CARM1 with key glutamate residues involved in methyl transfer

The "double-E" residues (Glu258 and Glu267) of CARM1's active site are shown in complex with CMPD-2, a top CARM1 inhibitor <sup>17</sup>. The "double-E" loop is conserved across PRMTs and key for methyl transfer. Figure was made in Accelrys Discovery Suite 4.5 ® using 2Y1X.

#### 1.2 CARM1

#### Discovery of CARM1

CARM1 was first identified in 1999 by Chen et al. (1999) when it was discovered to associate with glucocorticoid receptor-interacting protein 1 (GRIP1), which is a member of the p160 family of coactivators <sup>30</sup>. The p160 family of coactivators (GRIP1, SRC-1, AIB) mediate transcriptional activation by nuclear hormone receptors <sup>30</sup>. Using a yeast two-hybrid system to screen a mouse 17-day embryo complementary DNA (cDNA library), Chen et al. discovered a cDNA clone encoding a 608-residue protein that was found to bind to the COOH-terminal amino acids 111 to 1462 of GRIP1. When this protein bound, it was determined to enhance transcriptional activation by nuclear receptors, but only when GRIP1 or steroid receptor coactivator 1a (SRC-1a) was coexpressed. The central portion of this coding region was recognized as having extensive homology to the PRMT family. The protein was tested for methyltransferase activity where it was determined that the protein preferentially methylated histone H3 and was identified as CARM1. The direct consequence of this methylation event is the enhancement of gene transcription <sup>17</sup>. CARM1 was found to specifically enhance the function of AD2 activation domain and this correlated with its ability to bind GRIP1. CARM1 was later found to be a positive co-regulator for a number of other transcriptional modulators as well <sup>31</sup>, and additional roles for CARM1 have been suggested in muscle differentiation as well as protein repair, chromatin regulation, mRNA stabilization and gene splicing<sup>17</sup>. Mutations of critical residues in the catalytic domain of CARM1 have been shown to compromise transcriptional activation, suggesting the integrity of the methyltransferase domain of CARM1 is important for its coactivator function <sup>31</sup>.

#### CARM1 substrates

CARM1 has been shown to methylate a wide-array of substrate proteins <sup>32</sup>. CARM1 has been shown to methylate arginine residues harbouring a PGM motif in proline-, glycine- and methionine-rich regions <sup>5</sup>. This sequence specificity is due to a unique C-terminal extension, which provides a lower ridge of the peptide-binding groove that connects the cofactor and the active site <sup>32</sup>. Furthermore, this C-terminal extension generates a neutral molecular surface charge <sup>32</sup>. CARM1 substrates can be classified into three categories. In the first category, CARM1 methylates histone H3 and p300/CBP, which alters chromatin structure and in turn impacts transcriptional initiation <sup>33</sup>. In the second category, methylation by CARM1 of proteins involved in splicing such as SmB, U1C, SAP49 and CA150 implicates CARM1 in regulating alternative splicing <sup>33</sup>. In the third category, methylation of RNA-binding proteins such as HuR, HuD and PABP indicates a role for CARM1 in mRNA processing and stability <sup>33</sup>. The primary substrates of CARM1 include histone H3 (at H3R17), AIB1 (a steroid receptor coactivator amplified in breast cancer 1), p300 (a coactivator also known as EP300 or E1A binding protein that activates transcription), CBP (CREB-binding protein that also activates transcription) and RNA pol II CTD (C-terminal domain of RNA polymerase II that activates polymerase activity)<sup>5</sup>. The variety of substrates CARM1 methylates is suggestive of the enzyme's importance.

#### CARM1 and cancer

Several studies have shown that CARM1, in particular, is overexpressed in aggressive prostate and breast cancer cells <sup>34–37</sup>. CARM1 has been shown to be up-regulated in the nucleus of prostate cancer cells during the progression to castration resistance <sup>33,35</sup>. Over-expression of CARM1 is seen in both androgen-stimulated and castration-resistant prostate cancer tumours <sup>33,35</sup>. SiRNA mediated knockdown of CARM1 expression has been shown to inhibit cell proliferation, induce apoptosis and inhibit androgen-induced gene expression <sup>35</sup>. Similarly, CARM1 is a co-activator of estrogen receptor  $\alpha$  (ER $\alpha$ ), an important therapeutic target in breast cancer <sup>38</sup>. It has been shown that CARM1 is essential for estrogen-induced cell cycle progression in the MCF-7 breast cancer cell line <sup>39</sup>. The knockout or silencing of CARM1 impedes estrogen-stimulated gene growth of breast cancer cells and evidence also links CARM1, cyclin E, and steroid coactivator overexpression to high-grade breast cancer tumours <sup>31</sup>. Aggressive breast tumours that overexpress CARM1 have also been shown to have high levels of the oncogenic co-activator AIB1<sup>5</sup>. The overexpression of AIB1 is reported in breast cancer and is associated with a poor prognosis <sup>5</sup>. CARM1 is of particular interest as a cancer target because of its function as a nuclear receptor coactivator. Although the extent of the role of CARM1 in cellular regulation is not yet known, its many functions in transcription activation by nuclear hormone receptors <sup>40</sup> and other transcription factors such as p53<sup>41</sup>, make it a probable target for the treatment of breast and prostate cancers <sup>17</sup>. In 2014, prostate and breast cancer alone accounted for 52% of all new cancer cases in Canada<sup>42</sup>. For this reason, CARM1 has been focused on as a plausible target for breast and prostate cancer drug development <sup>31,36,37</sup>.

#### Structure of CARM1

The crystal structure of the catalytic domain of CARM1 has been determined, which shows structural and active site residue conservation common to members of the PRMT family <sup>18</sup>. The two most recent crystal structures of CARM1 (Protein Data Bank codes 2Y1X and 2Y1W) were co-crystallized respectively with SAH and a pyrazole inhibitor (CMPD-2) (Figure 1.6), and with sinefungin and an indole inhibitor (CMPD-1) <sup>17</sup>.

Full-length CARM1 is 608 residues, however, the catalytic core region is 350 residues in length (residues 134-483) and has been shown to methylate substrate peptides *in vitro*<sup>17</sup>. The N-terminal domain of the catalytic core is formed by three  $\alpha$ -helices ( $\alpha$ X- $\alpha$ Z) followed by a mixed  $\alpha/\beta$  Rossman fold (residues 166-287), which is a defining structural feature for the most common SAM-dependent methyltransferases <sup>32</sup>. The C-terminal domain contains a  $\beta$ -barrel-like substrate-binding domain (Figure 1.6) (residues 290-299 and 378-478)<sup>17</sup>. The N-terminal domain is responsible for binding SAM and catalysis<sup>17</sup>, and the C-terminal domain contains several divergent loops that may be involved in substrate specificity, as well as the conserved 'THWxQ' loop that forms part of the substrate-binding groove<sup>17</sup>. At the C-terminal domain the  $\beta$ -barrel is interrupted by a 39-residue two-helix bundle (residues 300-338) containing a hydrophobic tip (residues 316-322), which acts as a dimerization arm to make extensive interactions with a second molecule to form a CARM1 homodimer<sup>17</sup>. The purpose of this interaction is to stabilize the enzyme-substrate complex <sup>32</sup>. The catalytic domain of CARM1 was co-crystallized with a top CARM1 inhibitor, CMPD-2, and with SAH (Figure 1.7) <sup>17</sup>.





The crystal structure of CARM1's catalytic domain is 608 residues long and consists of a two-domain structure: an N-terminal domain (green) containing a mixed  $\alpha/\beta$  Rossman fold, which is the SAM binding domain, and a C-terminal  $\beta$ -barrel-like substrate-binding domain (orange). The C-terminal domain is interrupted by a two-helix bundle capped by a hydrophobic tip, which acts as a dimerization arm (blue) to make interactions with a second CARM1 monomer to form a homodimer. Structure was compiled in Accelrys Discovery Studio 4.5®.



Figure 1.7 Surface representation and binding pocket of CMPD-2 (2Y1X)

Panel A shows a surface representation of the binding positions of CMPD-2 and SAH in the catalytic domain of CARM1. Panel B shows the binding interactions of CMPD-2 within the substrate-binding groove and arginine-binding pocket.

#### Key residues regulate cofactor stabilization in CARM1

Conserved residues have been implicated in the stabilization of tertiary and quaternary structure of CARM1 and of the apo- and holoenzyme states. CARM1 can be in either an inactive *apo* state (lacking cofactor) or in an active *holo* state (cofactor bound). The crystal structure of the CARM1 catalytic core in the *apo* (2V7E) and *holo* (2V74) states reveal cofactor-dependent formation of a substrate binding groove providing a specific access channel for arginine to the active site <sup>32</sup>. When SAH binds, the ordered  $\alpha$ X helix forms a lid covering the cofactor pocket, while the  $\alpha$ Y helix forms the upper ridge of the substrate-binding groove (Figure 1.8) <sup>32</sup>. This complex results in the cofactor pocket being completely buried from the exterior, making the methylsulfonium in SAM only accessible to the target substrate arginine in the arginine-binding pocket via a narrow tunnel.

There are three interactions with SAH that have an impact on the conformational structure of CARM1 (Figure 1.9 and Figure 1.10) <sup>18</sup>. Arg169, present on the  $\alpha$ Z-helix, interacts with the carboxyl group of SAH, Glu215 forms a hydrogen bond with the two ribose hydroxyl groups, and Glu244, Ser272, and Val243 interact with N6, N1, and N7 respectively on the adenine moiety of SAH <sup>18</sup>. Furthermore, the hydroxyl group of Ser217 forms a strong hydrogen bond with the carbonyl oxygen of Tyr154 to lock SAM inside the binding cavity <sup>17</sup>. All of these residues are highly conserved among the PRMT family <sup>32</sup>. It has also been shown that CARM1 inhibitors bind only in the presence of SAH <sup>17</sup>. In addition, several residues are repositioned to properly form the arginine-binding cavity and the surrounding pocket <sup>17</sup>.





A surface representation of the active site in the *apo* (orange) and *holo* (blue) states of CARM1. As shown in the *apo* form, the  $\alpha$ X helix is disordered and the cofactor packet is accessible. In the *holo* form, the  $\alpha$ X helix (cyan) forms a lid covering the cofactor pocket and at the same time creating an arginine-binding pocket <sup>32</sup>. Figures were created using the Accelrys Discovery Suite 4.5®.



Figure 1.9. Residues within the CARM1 active site stabilize SAH and cause conformational changes

Residues Arg169, Glu215, Glu244, Ser272 (not shown in figure for clarity) and Val243 are involved in stabilizing SAH. Figure was generated using Accelrys discovery suite 4.5® from the PDB 2Y1X <sup>17 18</sup>.



#### Figure 1.10. SAH binding to CARM1 causes large structural changes

Sequence alignment of the backbone primary structure of unliganded CARM1 (2V7E), shown in yellow, and liganded (with SAH) CARM1 (2Y1X) shown in blue. The reference protein used was 2Y1X. Large structural changes occur in the presence of SAH, particularly near the catalytic site, that lead to proper formation of the arginine-binding cavity and the surrounding pocket allowing inhibitors to bind. The calculated RMSD for this overlay was 24.5 Å. Figure was constructed using Accelrys Discovery Suite 4.5®.

#### Ordered bisubstrate mechanism

CARM1 uses a sequential ordered mechanism. This means that SAM binds first, which promotes formation of the peptide-substrate binding groove to then allow the substrate to bind <sup>28</sup>. Following methyl transfer, the methylated substrate dissociates first, and the generation of SAH promotes dissociation of the dimeric complex, with SAH dissociating last <sup>29,43</sup>. This mechanism is supported by structural investigations of CARM1, which show that SAM binding occurs first, resulting in a conformational change that forms the binding site for protein substrates as discussed above <sup>33,43</sup>. This observation suggests that arginine methylation by CARM1 proceeds in an ordered mechanism in which the cofactor binds first and the intermediate monomethylated arginine substrate must be released from the active site to allow replenishment of the cofactor in order for the second methylation reaction to occur <sup>32</sup>. It is also supported by studies showing that the K<sub>m</sub> for MMA was lower than that of unmethylated arginine containing peptide for both CARM1 and PRMT6 <sup>44</sup>. This means the monomethylated peptide will preferentially be methylated in a homogenous, free diffusion system<sup>28</sup>.

#### Efforts to target CARM1

To date, there have been a limited number of publications describing novel inhibitors of CARM1 originating either from molecular modelling studies or high-throughput screenings <sup>31</sup>. The majority of these hits lacked selectivity and showed no anti-proliferative effects <sup>31</sup>. The exact mechanism of inhibition is usually unknown and remains unclear as some inhibitors identified can inhibit the methylation of some substrates, but not others <sup>33</sup>. The natural product sinefungin (Table 1.1, panel A), determined as an inhibitor of methyltransferases in 1978 <sup>45</sup>, acts by competing for the SAM-binding site and has been confirmed to be a potent yet promiscuous inhibitor of CARM1 (IC<sub>50</sub> ~

1-3 $\mu$ M) <sup>46</sup>. It is important to note that IC<sub>50</sub> values derived under different conditions using different assays are not comparable <sup>43</sup>.

In 2004, AMI-1 (Table 1.1, panel B), a symmetrical sulfonated urea, was described as the first specific CARM1 inhibitor, and AMI-5 (Table 1.1, panel C) was identified as the most potent, but less selective compound <sup>47</sup>. AMI-1 specifically inhibits arginine and not lysine methylation, does not compete for the SAM binding site, and is cell-permeable and non-toxic to cells <sup>48</sup>. The IC<sub>50</sub> value of AMI-1 was not determined, but the IC<sub>50</sub> of AMI-5 was later determined to be 53.7  $\mu$ M <sup>47</sup>. In 2007, the AMI-5 chemical structure was chosen as a template and from it a series of simplified analogues were designed <sup>47</sup>. From these new compounds, the 1,5-bis(3,5-dibromo-4-methoxy) analogue 12 (Table 1.1, panel D) behaved as a potent and selective CARM1 inhibitor with an IC<sub>50</sub> value of 27  $\mu$ M, yet was inactive against PRMT1 (the most predominant PRMT in cells), showing the selectivity of the inhibitor for CARM1 <sup>47</sup>.

In 2008, a paper was published that also developed simplified analogues of AMI-5<sup>49</sup>. A number of new compounds bearing two *ortho*-bromo- and *ortho*,*ortho*-dibromophenol moieties were tested against a panel of PRMTs including PRMT1 and CARM1, as well as human SET7 (a Histone Lysine Methyltransferase, HKMT) in order to determine selectivity <sup>49</sup>. Compound 1a was particularly active against CARM1 and to a lesser extent PRMT1, suggesting the compound showed some selectivity.

Another group, also in 2008, reported the identification and hit-to-lead optimization of inhibitors of CARM1 <sup>33</sup>. From a high-throughput screening effort, a pyrazole amide derivative was identified

and closely related analogues were established as 'hits'. From this, a pyrazole amide 7b compound (Table 1.1, panel E) was identified as a potent and selective inhibitor of CARM1 with an IC<sub>50</sub> value of 0.08 $\mu$ M <sup>33</sup>. In 2009, the pyrazole amide, CMPD-2, was optimized. A selective pyrazole based inhibitor 7f (Table 1.1, panel F) was identified as a potent and selective inhibitor of CARM1 with an IC<sub>50</sub> value of 0.04 $\mu$ M <sup>50</sup>. In comparison to compound 7b, 7f was found to be significantly less active against PRMT1 and PRMT3 with IC<sub>50</sub> > 25  $\mu$ M, suggesting selectivity in binding and inhibition. By exchanging the amide unit with a 1,3,4-oxadiazole moiety, the authors were able to improve the permeability of the compound <sup>31</sup>.

Further pyrazole-type CARM1 inhibitors were developed in 2009. A series of *N*-benzyl-1heteroaryl-3-(trifluoromethyl)-1*H*-pyrazole-5-carboxamides targeting CARM1 were designed and synthesized <sup>51</sup>. These compounds showed high potency due to the nature of the heteroaryl fragment with thiophene analogues <sup>51</sup>. In particular, compound 7a (Table 1.1, panel G) was a potent inhibitor of CARM1 with an IC<sub>50</sub> of 0.06  $\mu$ M <sup>51</sup>. The problem with these compounds was that both the phenyl and thiophene-based inhibitors had poor bioavailability when dosed via intraperitoneal injection in mice <sup>31</sup>. To improve the pharmacokinetic profile, a new series of thiophene-based inhibitors of CARM1 were synthesized by another laboratory that showed improved clearance, volumes of distribution, and half-lives <sup>31</sup>. These compounds contained an *N*<sup>1</sup>-benzyl-*N*<sup>2</sup>methylethane-1,2-diamine unit <sup>31</sup>. Of these compounds, Diamine 12 (Table 1.1, panel H) showed the highest potency and had an IC<sub>50</sub> of 0.20 $\mu$ M <sup>31</sup>. The flexible ethylene fragment is shown to be productive in binding <sup>31</sup>. Overall, the inhibitory activities of the diamines were of the same order of magnitude as their predecessors <sup>31</sup>. Benzo[*d*]imidazole inhibitors were also identified by hit-tolead optimization and SAR development in 2009 <sup>52</sup>. These inhibitors showed both potency and selectivity with a lead compound being 17b (Table 1.1, panel I) <sup>52</sup>. Compound 17b had an IC<sub>50</sub> of 0.84  $\mu$ M, but was significantly less active again PRMT1 and 3 with IC<sub>50</sub> values > 25 $\mu$ M <sup>52</sup>.

In 2010, peptides containing a single arginine residue substituted at the guanidine nitrogen ( $N^{\eta}$ ) with an ethyl group bearing zero to three fluorine atoms (R1-1, -2, -3 and -4) were synthesized <sup>43</sup>. These  $N^{\eta}$ -substituted arginyl peptide inhibitors showed strong inhibition of PRMT1 and 6, but weak inhibition of CARM1 (minimum IC<sub>50</sub> of 168 ± 19 µM obtained with R1-4) <sup>43</sup>. Peptidic partial bisubstrate hybrids composed of a minimal SAM fragments (lacking adenosine) covalently linked to a model substrate peptide designed for CARM1 inhibitors were also developed in 2011 <sup>16</sup>. Selective inhibition of CARM1 was seen with compounds containing a general SAM mimic with a variable peptide scaffold <sup>16</sup>.

Also in 2010, a plant-derived inhibitor called TBBD (ellagic acid) (Table 1.1, panel J) selectively inhibited methylation at arginine 17 of histone H3 (H3R17), but not at arginine 26<sup>53</sup>. This molecule was shown to bind to the substrate, histone H3 (specifically at the proline residue at position 16), inhibit methylation of H3R17, and in turn repress the expression of *p21*, a p53-responsive gene, thus implicating a possible role for H3R17 methylation in tumour suppressor function <sup>53</sup>. The IC<sub>50</sub> reported for this inhibitor was 25  $\mu$ M <sup>53</sup>. TBBD could be used to investigate the roles of methylation in both apoptosis and tumour suppressor pathways <sup>53</sup>.

In 2011, a panel of xenoestrogens that shared features with previously discovered AMI-1 were screened <sup>54</sup>. The authors showed that tamoxifen, a xenoestrogen used as a breast cancer treatment for patients who are estrogen receptor (ER) positive, also was a weak CARM1 inhibitor. An IC<sub>50</sub>

of 43.31  $\mu$ M was determined and tamoxifen was shown to inhibit CARM1 in a cellular setting as well <sup>54</sup>.

Indole and pyrazole inhibitors, CMPD-1 and CMPD-2 (Table 1.1, panel K and L) were developed in 2011 by Bristol-Myers Squibb Research and Development <sup>17</sup>. The major findings of the paper were that the inhibitors bind in the arginine-binding cavity and the surrounding pocket that exists at the interface between the N- and C-terminal domains <sup>17</sup>. It was also shown that the inhibitors bind to the CARM1 catalytic domain only in the presence of the cofactor SAH <sup>17</sup>. The authors reported IC<sub>50</sub> values of 0.030  $\pm$  0.015 and 0.027  $\pm$  0.009  $\mu$ M for CMPD-1 and CMPD-2, respectively <sup>17</sup>. Since the indole and pyrazole inhibitors bind in the arginine-binding cavity, they provided valuable information concerning the types of interactions that could exist with the natural substrate. These inhibitors showed CARM1 selectivity over PRMT1 and PRMT3, but were not tested against any other PRMTs.

To date, the  $N^1$ -benzyl- $N^2$ -methylethane-1,2-diamines (Diamine 12) <sup>31</sup> and the indole (CMPD-1) and pyrazole (CMPD-2) <sup>17</sup> derivatives represent the most potent and selective inhibitors of CARM1 <sup>46</sup>. A brief overview of the assays used to confirm these compounds will be compared to the assay used in this thesis in Section 2.2.

#### Table 1.1. CARM1 inhibitors

| Name                        | IC <sub>50</sub> | Selective<br>Over | Structure  | Ref.                  | Name                       | IC <sub>50</sub> | Selective<br>Over         | Structure   | Ref.                  |
|-----------------------------|------------------|-------------------|--|-----------------------|----------------------------|------------------|---------------------------|-------------|-----------------------|
| A)<br>Sinefungin            | ~1-3 µM          | None              | 148  | 45<br>1978            | <b>G)</b> 7a               | 0.06 µM          | N/A                       | and         | <sup>51</sup><br>2009 |
| <b>B)</b> AMI-1             | N/A              | PKMTs             | Janay  | <sup>55</sup><br>2004 | <b>H)</b><br>Diamine<br>12 | 0.2 µM           | PRMT1,<br>Set 7/9<br>PKMT | Pan         | <sup>31</sup><br>2009 |
| <b>C)</b> AMI-5             | 53.7 μM          | None              | And the second s | 55<br>2004            | <b>I)</b> 17b              | 0.84 µM          | PRMT1,<br>PRMT3           | -otop       | <sup>52</sup><br>2009 |
| <b>D)</b><br>Analogue<br>12 | 27 μM            | PRMT1             | And  | 47<br>2007            | J) TBBD                    | 25 µM            | At H3R17<br>only          |             | <sup>53</sup><br>2010 |
| <b>E)</b> 7b                | 0.08 µM          | PRMT1,<br>PRMT3   | jon<br>ari   | <sup>33</sup><br>2008 | <b>K)</b><br>CMPD-1        | 0.03 µM          | PRMT1,<br>PRMT3           | door        | <sup>17</sup><br>2011 |
| <b>F)</b> 7f                | 0.04 µM          | PRMT1,<br>PRMT3   | Job<br>an  | <sup>52</sup><br>2009 | L)<br>CMPD-2               | 0.027<br>µМ      | PRMT1,<br>PRMT3           | tools<br>ar | <sup>17</sup><br>2011 |
#### 1.3 Hypothesis and research objectives

The objectives of this research were aimed at identifying novel small molecule inhibitors of CARM1. Specifically, this involved conducting a high-throughput screening of over 76,600 novel small molecules using the Accelrys Discovery Suite 4.5<sup>®</sup>. Potential top hits from the *in silico* docking were then verified using an *in vitro* assay to generate  $IC_{50}$  values. The purpose was to determine if computer modelling could act as an accurate predictor of novel CARM1 inhibitors. The specific research hypothesis is as stated, *"in silico docking can act as a better predictor of novel CARM1 inhibitors than structural similarity to known inhibitors."* The studies found herein describe decisions made in developing an *in silico* model template used for computer modelling, as well as the P81 filter-binding assay developed to conduct an *in vitro* screening to generate  $IC_{50}$  values and compare the biological data to the *in silico* representation.

#### 2. Methods

#### 2.1 Reagents and sources for the *in vitro* assays

#### Expression and purification of CARM1 and Poly(A) binding protein 1 (PABP1)(437-488)

Recombinant His<sub>6</sub>-tagged human CARM1 (H<sub>6</sub>-CARM1) and GST-PABP1(437-488) were expressed using methods routinely utilized in Dr. Frankel's laboratory <sup>28</sup>. The expression vectors pET28a-hCARM1 and pGEX-2T-PABP1(437-488) (a generous gift from Dr. Mark Bedford at the MD Anderson Cancer Center) were first transformed into BL21 pLysS cells (Stratagene). CARM1 and the PABP1 fragment were induced with 1.0 mM isopropyl 1-thio- $\beta$ -D-galactoside (IPTG) at 25°C overnight (16 h) in LB medium (Fisher Scientific) containing an additional 1.0% glucose, 50 µg/mL kanamycin and 35 µg/mL chloramphenicol. The cells were harvested by centrifugation in a Beckman model J2-21 centrifuge at 10,000 g for 15 minutes. The cell pellets were stored at - 80°C until protein purification.

CARM1 and PABP1(437-488) were purified using fast protein liquid chromatography (FPLC) (Figure 2.1 and Figure 2.2). After removal from the -80°C, the pellets were resuspended in lysis buffer (50 mM HEPES-KOH, pH 7.6, 1.0 M NH<sub>4</sub>Cl, 10mM MgCl<sub>2</sub>, 0.1% lysozyme, 25 U/mL DNase I. 0.2 Triton X-100, 7.0 mM  $\beta$ -mercaptoethanol, 1.0 μM mΜ phenylmethanesulphonylfluoride and complete EDTA-free protease inhibitor cocktail tablets) (Roche product code 04693132001) at 2 mL per gram of cells (wet weight). The cells were sonicated using a Branson Sonifier 450 on ice for eight 30 s pulses at 50% duty cycle with 30 s pauses in between. Recombinant protein CARM1 was first purified via a 5.0 mL HisTrap FF affinity column (GE Healthcare) per 2.0 L bacterial culture using an established method and

according to the manufacturer's directions <sup>56</sup> The eluent from the first step was purified using a HiLoad 26/60 Superdex 200 pg column (GE Healthcare). The CARM1 samples were collected and exchanged into a storage buffer (100 mM HEPES-KOH, pH 8.0, 200mM NaCl, 1mM DTT, 10% glycerol and 2mM EDTA). PABP1(437-488) was purified via 3 x 1 mL GST affinity columns (GE Healthcare), the samples collected and exchanged into 50 mM Tris pH 8.5, 200 mM NaCl, 0.5 mM dithiothreitol (DTT), 50% (v/v) glycerol. Both CARM1 and the PABP1 fragment were concentrated using Amicon Ultra ultracentrifugal filters with a 10-kDa molecular weight cut-off (Millipore). The concentrated samples were frozen in liquid nitrogen and stored at -80°C.



#### Figure 2.1. Purification of recombinant CARM1

The SDS-page gel shows lanes 1 and 12 as molecular weight ladders. Lane 2 is the cell lysate, lane 3 is the flow through from loading the nickel column, lane 4 is the wash fraction, lane 5 and 6 are the elution fractions collected from the His column containing CARM1. Lane 7 is the flow through from the SEC column, lanes 8, 9, and 10 are fractions from the SEC column that appeared to contain CARM1. Lane 11 shows the concentrated CARM1 sample, indicating that it was successfully isolated. Lanes 2-11 each contain  $0.2 \,\mu$ L of sample.



#### Figure 2.2. Purification of GST-PABP1(437-488)

Each lane contains  $0.5 \mu$ L of sample loaded into an SDS-page gel. Lanes 1 and 9 are molecular weight ladders. Lane 2 is the cell lysate, lane 3, the flow through from loading the glutathione (GST) column, lane 4 is flow-through from loading the binding buffer into the GST column and lane 5-8 were fractions collected from the GST column containing the PABP1 fragment (lane 6 is the concentrated sample of lane 5 and lane 8 is the concentrated sample of lane 7).

#### Protein quantification and spectral characteristics

CARM1 was quantified spectrophotometrically using the NanoDrop 2000 UV-Vis spectrophotometer. The Edelhoch method <sup>57</sup> was utilized where the sample was dissolved in a 6 M guanidine hydrochloride buffer and the absorbance of the protein was measured at 280 nm ( $\varepsilon_{280}$  = 63,020 cm<sup>-1</sup> M<sup>-1</sup> as calculated using methods from <sup>58</sup>). The final concentration was calculated using the Beer-Lambert law to be 17.5  $\mu$ M. The Edelhoch method provides a more accurate way to measure the concentration of a pure protein solution by measuring the absorbance arising from the intrinsic chromophores tryptophan, tyrosine and cystine (oxidised cysteine in a disulphide bond) <sup>59</sup>.

PABP1(437-488) was quantified using the Bradford assay <sup>58,60</sup> with bovine serum albumin (BSA) used as the protein standard <sup>61</sup>. Since the sample of PABP1 was not pure, further steps were taken in order to generate a more accurate concentration calculation. The purified proteins were separated on SDS-PAGE (Figure 2.2) and subsequent densitometry of Coomassie blue-stained bands was conducted using ImageJ 1.47 (<u>http://rsb.info.nih.gov/ij</u>). The PABP1 band accounted for ~ 34% of all bands shown in lane 8 of the SDS-PAGE gel, which was used to calculate a more accurate concentration of the PABP1 fragment. PABP1(437-288) was determined to have a final concentration of 253  $\mu$ M.

#### Methylation activity assays

In order to determine that the substrate PABP1(437-488) was being methylated by CARM1, the following experiment was conducted: CARM1 (1  $\mu$ M) was incubated (37°C) from 0 – 960 min

with PABP1 (200  $\mu$ M) and <sup>14</sup>C-SAM (50  $\mu$ M,; PerkinElmer NEC363050UC, 1.81 GBq mmol<sup>-1</sup>). Reactions mixtures were separated electrophoretically on a tricine gel (16.5%). The gel was fixed using glutaraldehyde (5%; Sigma-Aldrich) according to previous methods to prevent peptide diffusion from the gels <sup>62</sup>. The gels were then stained with Coomassie Blue, dried, and exposed to a storage phosphor screen (GE healthcare) for 96 h before development on a Typhoon 9400 imager (GE Healthcare) at 50- $\mu$ m resolution (Figure 2.3).



### Incubation time (min)



Methylation of GST-PABP1(437-488) was assessed by radioactive labeling of the CARM1 substrate over the course of 0 to 960 min. Reactions were separated on a 16.5% Tricine gel and subsequently stained with Coomassie blue (bottom). The phosphor image (top) shows a time-dependent increase in substrate methylation.

#### Novel small molecule libraries and Diamine 12

Three novel compound libraries were screened via *in silico* docking. The top hits were further assessed for inhibitory activity *in vitro*.

*GPS library:* Dr. Grierson's medicinal chemistry team synthesized a 384-compound library comprised of new drug-like structures. All compounds were stored in 96-well plates as 20 mM stocks in DMSO. The structures of all 384 compounds were drawn in Accelrys Draw 4.2 and saved as .mol files to be imported into the Accelrys Discovery Suite 4.5®.

*CCBN library:* The CCBN library consisted of 25,990 novel compounds provided generously by Dr. Tom Pfeifer at the Centre for Drug Research and Development (CDRD). All compounds were in .sd format, compatible with Accelrys® and available as 5 mM stocks in DMSO.

*Diver Set library:* The Diver Set library consisted of 50,000 novel compounds provided generously by Dr. Pfeifer at the Centre for Drug Research and Development (CDRD). All compounds were in .sd format, compatible with Accelrys® and available as 5 mM stocks in DMSO.

*Diamine 12:* Diamine 12 (Table 1.1, panel H) was received as a generous gift from Dr. Eric Therrien at Molecular Forecaster Inc. in Montreal, Canada. Diamine 12 was reported as a CARM1 inhibitor and patented in 2009 <sup>31</sup>. The reported IC<sub>50</sub> value was 0.2  $\mu$ M <sup>31</sup> and the compound was used as a positive control both for the *in silico* and *in vitro* screenings. Approximately 3.4 mg of Diamine 12 was received as a solid powder, which was dissolved in DMSO to make a 20 mM

stock. The purity of the Diamine 12 sample was determined by Dr. Eric Therrien using a diode array UV detector (Appendix, Figure A.1.).

#### 2.2 In vitro assays

The studies described herein utilize two *in vitro* assays including a liquid chromatography coupled mass spectrometry assay and a radioactivity-based P81 filter-binding assay.

*Mass spectrometry assay:* Reaction mixtures were incubated at 37 °C. After 2 h (within the linear range of the assay)  $^{63}$ , the reaction was stopped via flash freezing. The samples were thawed on ice and the SAM and peptide substrates were spin filtered through a 10-kDa molecular weight cut-off flat bottom filter (VW 82031-354) at 4 °C for 10 min in order to separate CARM1 from the reaction volume. The filtrate was loaded into a 96 micro-well plate (VWR CA62408-948) and samples were analyzed via LC-MS/MS on a ABSciex QT5500 mass spectrometer. Separation and quantification of SAH was achieved utilizing a Waters Acquity UPLC BEH C<sub>18</sub> reverse phase column (2.1 x 100 mm) with the mobile phases 0.1% aqueous formic acid (buffer A) and 0.1% formic acid in acetonitrile (buffer B). A standard calibration curve of 0 to 5000 nM (the linear dynamic range of the assay  $^{63}$ ) was used in order to quantify SAH.

A major drawback of using the mass spectrometry assay is that DMSO appeared to directly affect the standard calibration curve (Figure 2.4). Because all GPS compounds were already dissolved in DMSO as 20 mM stocks, avoiding DMSO was impossible. The DMSO effect would obscure any inhibitory effects that may be evident, therefore, the mass spectrometry could not be used.



# Figure 2.4. Effect of DMSO on the standard calibration curve using the UPLC-MS/MS assay to quantify SAH

The y-axis is the measured concentration of SAH (ng/mL) and the x-axis is the actual concentration of SAH (nM). At only 5% DMSO, there was an increase in the amount of SAH measured.

#### P81 filter-binding assay

CARM1 activity was also measured in methylation assays using *S*-adenosyl-L-[*methyl*-<sup>14</sup>C]methionine (<sup>14</sup>C-SAM; PerkinElmer NEC363050UC, 1.81 GBq mmol<sup>-1</sup>) as the radioactive methyl donor. Reactions consisting of CARM1, PABP1(437-488) and <sup>14</sup>C-SAM in methylation buffer were spotted on to P81 phosphocellulose filter paper (Whatman P-81 phosphocellulose filter paper, 46 x 57 cm, No.: 3698-915 Fisher Scientific) and left to air-dry at room temperature (25 °C) for 20 minutes. The paper was then washed 5 times with 20 mL of buffer (Na<sub>2</sub>CO<sub>3</sub>/NaHCO<sub>3</sub>, pH 9.0). The P81 filter binding assay immobilizes <sup>14</sup>C-methylated peptides by binding positively charged amino acids <sup>64</sup>. The filter paper was transferred to a scintillation vial containing 2.5 mL of Ultima Gold scintillation cocktail (PerkinElmer).  $\beta$ -emission of the immobilized peptides was quantified by a Tri-Carb 3110TR liquid scintillation counter in disintegrations per minute (DPM) <sup>64</sup>.

There are many advantages to using the P81 filter-binding assay. First, it requires a much smaller reaction volume than the previously described LC-MS/MS assay. At only 35  $\mu$ L, the amount of CARM1, <sup>14</sup>C-SAM and PABP1(437-488) required will be approximately four-fold less. Second, no separation step is needed, which simplifies the assay procedure, decreases the cost, and greatly enhances the assay speed <sup>64</sup>. Last, the 'DMSO effect' was not evident with the P81 filter-binding assay. Three preliminary tests were conducted in order to determine: a) the linear range of the assay (Figure 2.5) b) the apparent K<sub>m</sub> determination of PABP1(437-488)(Figure 2.6), and c) the DMSO effect (Figure 2.7). The linear range was shown to be within a 2 h timeframe, the determined K<sub>m</sub> was 1.1  $\mu$ M, and there was little to no effect of DMSO up to 5%. The importance of determining the linear range was to ensure the incubation time fell within the working range of the assay. Because the linear range was within 2 h, all reactions were incubated for a 2 h timeframe.

The  $K_m$  was determined to be 1.092  $\mu$ M. To ensure the signal-to-noise ratio was high enough to detect any inhibitory effect, the concentration of the substrate used (PABP1) was 5 times greater than the  $K_m$ . In Figure 2.7, it is shown that up to 5% DMSO, there is little to no effect on the assay. Above 10% DMSO, however, there is a decrease in activity. In order to minimize the DMSO effect, samples and controls were consistently incubated with 4% DMSO.



#### Figure 2.5. The linear range of the P81 filter-binding assay

The reaction consisted of 1  $\mu$ M of CARM1 incubated with 20  $\mu$ M PABP1(437-488) and 50  $\mu$ M <sup>14</sup>C-methylated SAM for 0 – 960 min in duplicate. The linear range was shown to be within a 2-h timeframe.



#### Figure 2.6. The K<sub>m</sub> of the P81 filter-binding assay

CARM1 was incubated with increasing concentrations of PABP1 (0 – 20  $\mu$ M) and 50  $\mu$ M <sup>14</sup>C-methylated SAM for 120 min in duplicate. The K<sub>m</sub> as determined from the nonlinear fit of the baseline-corrected data was determined to be 1.1  $\mu$ M. The data was plotted and the calculations conducted using GraphPad Prism 6.



В.



#### Figure 2.7. Effect of DMSO on the P81 filter-binding assay

CARM1 methylation of PABP1(437-488) at (A) DMSO concentrations (0-20%) and (B) DMSO concentrations (0-5%) which reveals that DMSO has little effect on the P81 filter-binding assay at low concentrations.

Α.

# Comparison of the assays used to confirm CMPD-1, CMPD-2 and Diamine 12 to the P81 filter-binding assay

 $IC_{50}$  values derived under different conditions using different assays are not comparable <sup>43</sup>. It is therefore important to understand the differences between the P81 filter-binding assay and the assays used to identify top CARM1 inhibitors. CMPD-1 and CMPD-2 were screened using a methylation-based filter assay and Diamine 12 was screened using a P30 filtermat assay (Table 2.1). Both of these assays used a radiolabeled filter-binding assay to bind positively charged peptides, similar to the P81 filter-binding assay.

| Table 2.1. A comparison of assays used to screen CMPD-1, CMPD-2, Diamine 12, GPS and CDRD compounds |
|---|
|---|

|                    | Assay Name                   | Reaction  | Buffer  | Incubation  | Stop                                    | Quantifying the samples  |
|--------------------|------------------------------|---|---|-------------|---|--|
| CMPD-1 &<br>CMPD-2 | Methylation-<br>based filter | <ul> <li>6 nM GST-<br/>CARM1<br/>(human)</li> <li>1 μM histone H3</li> <li>0.05 μM <sup>3</sup>H-SAM</li> </ul> | 20 mM<br>Tris/HCl (pH<br>8), 200 mM<br>NaCl, 0.4 mM<br>EDTA | 60 – 90 min | Add TCA                                 | <ul> <li>Precipitate reaction with BSA overnight</li> <li>Filter and wash</li> <li>Read in a Top Count after addition of MicroScint-20</li> </ul>  |
| Diamine 12         | P30 Filtermat                | <ul> <li>10 μM GST-<br/>CARM1<br/>(mouse)</li> <li>6.5 μM histone H3</li> <li>2 μM <sup>3</sup>H-SAM</li> </ul> | 50 mM Tris-<br>HCl (pH 9),<br>2.5 mM DTT                    | 30 – 35 min | Add 60 μM<br>SAH                        | <ul> <li>Spot onto P30 Filtermat paper</li> <li>Wash paper 2 X (15 min)</li> <li>Read counts with a Wallac<br/>Microbeta counter (CPM) after<br/>addition of wax scintillant</li> </ul>  |
| GPS &<br>CDRD      | P81 filter-<br>binding       | <ul> <li>1 μM His-CARM1<br/>(human)</li> <li>4.5 μM PABP1</li> <li>50 μM <sup>14</sup>C-SAM</li> </ul>          | 50 mM<br>HEPES (pH<br>8), 10 mM<br>NaCl, 1 mM<br>DTT        | 120 min     | Add<br>samples to<br>liquid<br>nitrogen | <ul> <li>Spot onto P81 filter paper</li> <li>Air-dry 20 min at room temp.</li> <li>Wash 5 X and air-dry overnight</li> <li>Read counts with a Tri-Carb 3110<br/>TR counter (DPM) after addition<br/>of Ultima Gold scintillation<br/>cocktail</li> </ul> |

The reported IC<sub>50</sub> values using each of the assays outlined in Table **2.1**, were  $0.030 \pm 0.015$  and  $0.027 \pm 0.009 \,\mu\text{M}$  for CMPD-1 and CMPD-2 respectively <sup>17</sup>, and  $0.2 \,\mu\text{M}$  for Diamine  $12^{31}$ . When the P81 filter-binding assay was used to confirm Diamine 12, the determined IC<sub>50</sub> value was  $1.3 \pm 1 \,\mu\text{M}$  (Figure 3.7). The differences in IC<sub>50</sub> values can be attributed to differences in substrate, enzyme and cofactor concentrations, assay conditions and protocols, and human error. Different scintillation counters were used to quantify the samples for all compounds which also can affect the results.

#### 2.3 Computer-aided drug discovery and design (CADD)

#### Software and associated docking algorithms

Computer-aided drug discovery and design (CADD) has played a major role in the development of therapeutically important small molecules for over three decades <sup>65</sup>. Docking is a computational technique which acts to sample the conformations of small molecules constrained within a designated protein binding sites. These conformational poses are then computationally evaluated generating a ranking score to identify which molecular conformations best complements the protein binding site with the most favorable molecular interactions. It is important to note that the scoring function can yield a number of scored parameters reflecting the 1) number 2) nature and 3) type of molecule: target interaction which differ by algorithm (e.g. RosettaLigand K<sub>d</sub> vs CDOCKER Interaction Energy vs. LibDock Score). These scored parameters are used to assign the highest score to the 'correct pose' which is the computer derived pose that recapitulates the native pose observed in crystal structure in the presence of the same pharmacophore. Ultimately those molecules with the highest ranked scores are prioritized for additional analysis including functional assays and lead optimization.

CADD studies described below were performed in two stages using distinct software each with distinct algorithms and scoring functions. The initial molecular screening using SAR principles was performed using RosettaLigand <sup>66</sup> (Dr. Jennifer Bui from UC San Diego) a publically available standalone algorithm yielding ranked molecules with computational K<sub>d</sub> values. All subsequent modeling was performed using the industry standard proprietary software suite Accelrys Discovery Suite 4.5® (*Ref. Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, Release 4.5, San Diego: Dassault Systèmes, 2015.).* This CADD pipeline enables the high-throughput molecular screening of compound libraries through a coordinated but customizable workflow of life science modeling and simulation tools allowing target identification followed by *in silico* lead optimization.

The selected algorithms all employ: (1) a method to explore the conformational space of the ligand and protein target and (2) a scoring function to evaluate the proposed binding modes referred to as 'poses' <sup>67</sup>. In addition to RosettaLigand, the Accelrys Biovia Discovery Studio (DS4) docking and scoring algorithms LibDock <sup>68</sup>, and CDOCKER <sup>69</sup> were deployed independently and in series for lead identification. For ease of comparison, Table 2.2 summarizes the key components of RosettaLigand, LibDock and CDOCKER algorithms.

Evaluation of the standalone RosettaLigand algorithm and the DS4 LibDock and CDOCKER algorithms reveals limitations of RosettaLigand for this study. Specifically, RosettaLigand frequently fails to 1) recapitulate the native binding mode as the lowest energy binding mode,

44

2) lacks ligand internal energy for ligand refinement in the scoring function and 3) is incapable of docking multiple ligands simultaneously. <sup>70</sup> This is in direct contrast to both CDOCKER and LibDock which independently meet these analytical requirements. For molecules such as CARM1, the ability to co-dock ligands is vital for accurate drug docking as it has been shown that the binding of SAH causes considerable conformational changes to the protein template, allowing top CARM1 inhibitors to also bind, and forming inherent interactions within the binding cleft <sup>17</sup>. Therefore DS4 was selected for all molecular modeling of the compound libraries.

| Table 2.2. | A Comparison | n of RosettaLigand, | LibDock and CDOCKER |
|------------|--------------|---------------------|---------------------|
|------------|--------------|---------------------|---------------------|

| <u>Algorithm</u> | <u>Type of</u> | <u>Usage</u>    | Docking            | Ligand           | Target Protein  | Interaction  | Advantages                  | <u>Disadvantages</u>      |
|------------------|----------------|-----------------|--------------------|------------------|-----------------|--------------|-----------------------------|---------------------------|
|                  | <u>docking</u> |                 | <u>Flexibility</u> | <u>Placement</u> |                 | <u>Sites</u> |                             |                           |
| RosettaLigand    | Bond           | Small to        | Flexible           | Monte Carlo      | Flexible Side   | Used in      | Ligand and backbone         | Fails to recapitulate the |
| (freeware)       | rotation,      | medium sized    | (protein           |                  | Chains          | Scoring      | flexibility explored        | native binding mode as    |
|                  | translation    | libraries       | and                |                  |                 | -            | simultaneously              | the lowest energy and     |
|                  | and/or rigid   | (10-100's of    | molecule           |                  |                 |              | -                           | cannot dock multiple      |
|                  | body           | molecules)      | move)              |                  |                 |              |                             | ligands simultaneously    |
|                  | rotation       | ,               | ,                  |                  |                 |              |                             | Slow in identifying an    |
|                  |                |                 |                    |                  |                 |              |                             | initial starting pose     |
| Libdock          | Hotspots-      | Screening of    | Rigid              | Matching         | Rigid Structure | Used in      | Very fast, multiple pose    | Predictive power reduced  |
| (Biovia DS4)     | based          | large libraries | (molecule          | C C              | C               | Docking and  | generation, docking score   | because ligand poses are  |
| · · · · · ·      | docking        | (vHTS)          | only               |                  |                 | Scoring      | includes interactions       | not fully refined         |
|                  | U              | (>1 million     | moves)             |                  |                 | e            |                             | (less accurate scoring)   |
|                  |                | molecules)      | ,                  |                  |                 |              |                             | х с <i>у</i>              |
| CDOCKER          | Grid based     | Screening of    | Flexible           | Molecular        | Flexible        | Used in      | More accurate in            | Very slow                 |
| (Biovia DS4)     | which uses     | small           | (protein           | Dynamics –       | Repacking       | Docking and  | predicting relative binding | <100,000 size library     |
|                  | CHARMm,        | libraries and   | and                | simulated        |                 | Scoring      | affinity of congeneric      | only                      |
|                  | forcefield     | refinement of   | molecule           | annealing        |                 | e            | compounds, molecular        | 5                         |
|                  | generation.    | docking         | move)              |                  |                 |              | repacking reflects biology. |                           |
|                  | and energy     | poses           |                    |                  |                 |              | protein and ligand energy   |                           |
|                  | minimization   | (10 - 100000)   |                    |                  |                 |              | used in scoring             |                           |
|                  | minimization   | molecules)      |                    |                  |                 |              | used in secting             |                           |
|                  |                | molecules)      |                    |                  |                 |              |                             |                           |
|                  |                |                 |                    |                  |                 |              |                             |                           |
|                  |                |                 |                    |                  |                 |              |                             |                           |

#### A. RosettaLigand algorithm (Monte Carlo)



```
X-ray complexes
```

#### Figure 2.8. Schematic representation of RosettaLigand, LibDock and CDOCKER algorithms

(A) RosettaLigand uses a Monte Carlo algorithm where the pose of the ligand is sequentially modified through bond rotation, translation and/or rigid body rotation in a predefined and static 3D binding pocket. A new conformation is kept if it has a higher score, lower energy or through using a selection criterion <sup>67</sup> Note in RosettaLigand the protein:ligand interaction energy is used only in the scoring phase of the CADD process. This is in contrast to both Libdock and CDOCKER where interaction energy is used in both the docking and ranking process. (B) LibDock, which uses a matching algorithm, first generates a pharmacophore representing the protein and ligand. Distance matrices for both are examined for a match and if there is a match, rotational and translational vectors position the ligand in the same frame of reference as the protein <sup>67</sup>. (C) With CDOCKER random orientations of the conformations are produced by translating the center of the ligand to a specified location within the receptor active site, and performing a series of random rotations. A softened energy is calculated and the orientation is kept if the energy is less than a specified threshold. This process continues until either the desired number of low-energy orientations is found, or the maximum number of low scoring orientations have been tried. This step can be skipped to use the input orientation only such as docking the highest scored pose from LibDock. Each orientation is subjected to simulated annealing molecular dynamics. The temperature is heated up to a high temperature then cooled to the target temperature. A final minimization of the ligand in the rigid receptor using non-softened potential is performed. For each final pose, the CHARMm energy (interaction energy plus ligand strain) and the interaction energy alone are calculated. The poses are sorted by CHARMm energy and the top scoring (most negative, thus favorable to binding) poses are retained <sup>71</sup>.

#### CARM1 X-ray crystal structure template

The crystal structures of the human CARM1 catalytic domain (residues 135-482) (2011) were obtained from RCSB (http://www.rcsb.org/): 1) 2Y1X.pdb, in complex with SAH and an indole inhibitor, CMPD-2 (Table 1.1, panel L) <sup>17</sup> and 2) 2Y1W.pdb, in complex with sinefungin (Table 1.1, panel A) and an indole inhibitor, CMPD-1 (Table 1.1, panel K) <sup>17</sup>. Crystallized under the identical condition of pH 8.5 both 2Y1X and 2Y1W contain are 348 residues of the catalytic domain with resolutions of 2.4 and 2.1 Å, respectively.

#### CARM1 protein target preparation

Proteins downloaded from the PDB must be correctly prepared for CADD docking studies to reflect the biochemical and biological structure (Figure 2.9). Note, both 2Y1X and 2Y1W .pdb files were corrected for a missing Thr478 atom by reordering the protein using DS4. This terminal residue does not interact with inhibitor binding in the arginine-binding cavity (Met260, Glu258, His415, Glu267), substrate binding groove (Asn266, Tyr262, Gln159) or with the binding of SAM (Ser217, Tyr154)<sup>17</sup>.



# Figure 2.9. Decision tree outlining the selection and rationale for developing the CARM1 2Y1X protein model used for CADD studies

The process for protein target selection and model generation involved key decision making steps which must be evaluated based on the selected target (CARM1), the *in vitro* assay and the molecular environment where the inhibitor will be used.

Through iterative protein preparation and docking experiments using recapitulation of CMPD-2 crystal structure molecular interactions and docking pose RSMD as the accuracy cutoff, a series of key protein preparation steps were taken as outlined in Figure 2.9:

- A) 2Y1X was chosen over 2Y1W as it contains both CMPD-2 and SAH co-crystallized during structure generation. Top CARM1 inhibitors do not bind in the absence of SAH <sup>17</sup>.
- B) 2Y1X contains a tetramer of chains A, B, C and D. Chains A and C and chains B and D are mirror images of the active dimer each containing an active binding site. The dimeric structure, A and C contain two arginine binding clefts of identical architecture. For ease of amino acid numbering chain A was selected for further analysis. A random Cl<sup>-</sup> heteroatom X-ray artifact was deleted from within the binding pocket.
- C) The 2Y1X molecule was then subjected to cleaning to 1) Remove alternate protein conformations contained within the .pdb crystal structure which conflict with CHARMm forcefield generation by introducing disordered structural details, 2) Protonate the protein to add hydrogens at pH 8.5, which was the pH at which the protein was crystallized and enables hydrogen bond formation in the docked structure. 3) Corrects missing or incorrectly specified residues to identify those residues lacking structural resolution. The 2Y1X structure did not require any corrections.
- D) Both at physiological pH (7.0) or at the pH of the *in vitro* assay (8.0) yielded identical results to proteins prepared at pH 8.5.

- E) The protein target was prepared in the presence of the co-crystalized SAH molecule. SAH binding causes a significant conformational change in protein structure allowing access to the arginine binding cleft and enabling inhibitors to bind <sup>17</sup>. CMPD-2 was removed from the binding cleft to prepare the arginine binding pocket atoms in the naïve unbound state.
- F) All of the water molecules were removed in the protein preparation step to allow the molecule target interactions to be evaluated directly. Notably, the docking studies in the presence of water negatively impacted the docking scores of potent CARM1 inhibitors. Specifically, these molecules scored lower when 3 selected water molecules, which form the binding pocket, were included (Figure 2.10) <sup>17</sup>. These selected water molecules were water 1 (W1) positioned above the mouth of the arginine-binding cavity, water 2 (W2) acting as a bridging water molecule interconnecting the oxadiazole nitrogen in position 3 with Nε2 of Gln159, and water 3 (W3) making the only polar interaction between the protein and inhibitor linking the indole nitrogen and Nδ2 of Asn266 <sup>17</sup>. Finally, when waters were removed, the docked pose of CMPD-2 recapitulated the crystal structure exactly as described<sup>17</sup>.
- G) The CHARMm (Chemistry at HARvard Molecular Mechanics) forcefield was used in preparation of the protein which is in the Accelrys DS4 CADD suite. CHARMm performs a simulated annealing optimization followed by minimization <sup>17</sup>. This results in distance restraints being applied between the mapped atoms and the location features during the CHARMm simulation which provide more accurate downstream docking in flexible application such as CDOCKER <sup>72</sup>.



# Figure 2.10 Difference between LibDock score with water minus without water vs. LibDock score without water

CARM1 inhibitor controls, Diamine 12 and 7b (red) scored worse when three selected water molecules were included in the CARM1 PDB template <sup>31,33</sup>, whereas false positive compounds displayed a higher LibDock score when water molecules were included.



Figure 2.11 Ligand preparation for *in silico* docking using the Accelrys Biovia Discovery Studio for CADD studies

The process for ligand preparation involved key decision making steps which must be evaluated based on the library to be screened and the chemistry reflected therein.

#### In silico ligand preparation

The small molecule libraries are prepared for CADD in the DS4 software using a predefined protocol which removes duplicates, enumerates isomers and tautomers and generates 3D conformations. As in the protein preparation, here a series of decisions were required to prepare the library for docking (Figure 2.11):

- A) For both the internal library and the commercially available molecules the same process used for ligand preparation were undertaken. The parameters chosen were selected using those where the prepared version of a hand-drawn CMPD-2 molecule recapitulated the crystal structure pose when subjected to CADD docking.
- B) A canonical tautomer was generated for each molecule rather than all tautomers. Because tautomers are constitutional isomers of organic compounds that rapidly interconvert via the relocation of a proton the generation of all tautomers generates compounds that cannot exist in a biological setting (Figure 2.12)<sup>73</sup>. Canonicalization converts any of the tautomeric forms of a given molecule into a single unique representation that is the physiologically preferred form<sup>73</sup>.



#### Figure 2.12 Tautomer generated of CMPD-2.

The tautomer generated of CMPD-2 is not a physiologically preferred structure containing an unrealistic double bond. Thus, only only canonical tautomers were generated for CADD studies. The figure was made using the Accelrys Discovery Suite 4.5<sup>®</sup>.

- C) No isomers were generated during ligand preparation because isomers are compounds that have the same molecular formula but different chemical structures which confound CADD studies from an established well characterized library <sup>51</sup>.
- D) Ligand valence were corrected during ligand preparation to ensure that incorrect valences were removed.

#### Drug docking studies

Initial SAR based docking studies using RosettaLigand were performed by Dr. Jennifer Bui (UCSD). The resulting  $K_d$  values *in silico* for the top 20 ranked poses were not obtained *in vitro*. Analysis and comparison of RosettaLigand and the Biovia DS4 Suite revealed key differences in underlying CADD algorithms suggesting that LibDock and CDOCKER would be more accurate for the proposed CARM1 inhibitor screening studies.

The LibDock and CDOCKER algorithms were deployed independently and in series. First, the binding site was defined based on the CMPD-2 location within the original PDB file. The XYZ coordinates were 28.72, -27.14 and -76.83 respectively with a radius of 9.36 Å and overlaid both the arginine binding pocket and a portion of the SAH co-factor. This was designated the binding site and defined the area which the ligands were freely able to dock. LibDock, a high-throughput rigid docking algorithm, was deployed to generate initial poses and independently scored conformations. LibDock docks the molecules using shape complementarity or interaction site matching algorithms (Figure 2.8, panel B) <sup>68</sup>. This algorithm is very fast, but has a lower predictive power because the ligand poses are not fully refined, thus improved accuracy can be obtained from post-processing selected poses using a secondary search algorithm in series. After initial LibDock

scoring, all CARM1 inhibitors had a LibDock score greater than 120. This was defined as the CADD score for inhibitor development post-processing. All top-ranked poses for each compound with a LibDock Score of greater than 120 were then post-processed by docking in CDOCKER (Figure 2.8, panel C). CDOCKER generates several initial ligand orientations in the CARM1 active site to better refine docking poses while simultaneously allowing protein target flexibility. This allows accurate prediction of relative binding affinity of congeneric compounds. All compounds were independently docked in CDOCKER to compare the docking results between standalone and serial docking.

#### 3. Results

#### 3.1 Rosetta ligand results

#### In silico docking and scoring of GPS compounds

All GPS compounds were initially screened using RosettaLigand by Dr. Jennifer Bui at the University of California, San Diego in order to identify compounds that bind to and potentially inhibit CARM1. Dr. Bui used known CARM1 structures (2Y1X and 2Y1W) for molecular dynamic simulations. She computationally ranked all 384 GPS compounds according to virtual K<sub>I</sub> values. The top 20 ranked compounds are listed in Table 3.1.

| GPS number | Structure                                 | Κ <sub>ι</sub> (μΜ) | GPS number | Structure    | Κ <sub>ι</sub> (μΜ) |
|------------|---|---------------------|------------|--------------|---------------------|
| 332        | HN H ONH                                  | 8.7                 | 343        |              | 20.2                |
| N/A        |   | 11.1                | 296        |              | 20.6                |
| N/A        |   | 13.1                | 270        |              | 20.8                |
| 338        | NH<br>NH<br>NH                            | 13.3                | 114        | HN NO HN     | 21.5                |
| N/A        | ( ( ( ( ( ( ( ( ( ( ( ( ( ( ( ( ( ( (     | 16.0                | 331        |              | 21.8                |
| N/A        |   | 16.2                | 124        |              | 24.6                |
| N/A        |   | 17.1                | N/A        |              | 28.5                |
| N/A        | HAN O O O O O O O O O O O O O O O O O O O | 18.1                | N/A        |              | 28.6                |
| 334        |   | 18.9                | 336        |              | 29.6                |
| 344        |   | 19.7                | N/A        | HN C HN C NH | 29.8                |

Table 3.1. Predicted K<sub>I</sub> values for CARM1-binding compounds from the GPS library

#### In vitro results of the "40 set"

The top 20 ranked compounds as predicted by RosettaLigand docking (Table 3.1) combined with 20 other GPS compounds that had promising structural features similar to Diamine 12, CMPD-1 and CMPD-2 (listed in Table 1.1) primarily an extended amine or amide group that could be directed towards the bottom of the arginine-binding cavity, were screened using the P81 filter-binding assay. The results showed that RosettaLigand docking or selecting compounds based on structural similarities to known inhibitors was not an accurate predictor of what was occurring biologically. All compounds showed little or no inhibitory activity even at 200  $\mu$ M. The positive control in this study was Diamine 12, which did show inhibitory activity at all concentrations tested (Figure 3.1).





The top 20 compounds ranked using RosettaLigand combined with 20 other compounds with structural similarity to CARM1 inhibitors were screened using the P81 filter-binding assay at 20  $\mu$ M (blue) and 200  $\mu$ M (yellow). Diamine 12 was used as a positive control (left side) at 1, 20 and 200  $\mu$ M (green, blue and yellow).

#### 3.2 LibDock and CDOCKER results

#### CMPD-2 docked using LibDock recapitulates the crystal structure (2Y1X)

Docking experiments were performed to assess the accuracy in reproducing the binding mode of the ligand. In particular, CMPD-2 was first docked using LibDock within a root-mean-square difference (RMSD) of 1.2 Å of the crystallographic pose (2Y1X) as shown in Figure 3.2<sup>17</sup>. The binding interactions were also shown to match those described in the crystal structure of CARM1 co-crystallized with CMPD-2 in Figure 3.2<sup>17</sup>. The terminal L-alaninamide is directed towards the bottom of the arginine-binding cavity, where the terminal amino nitrogen interacts with a side chain oxygen and carbonyl oxygens of Glu258 and the carbonyl oxygen of Met260. The carbonyl oxygen of the L-alaninamide hydrogen bonds with NE2 of His415 and the adjacent nitrogen hydrogen bonds with one of the side chain oxygen of Glu267. Outside of the arginine-binding cavity, the benzyl ring is positioned at the mouth of the cavity, the pyrazole moiety sits above the imidazole ring of His415 and the trifluoromethyl group passes between the side chains of Asn162 and Tyr417. The 1,3,4-oxadiazole crosses back over the arginine-binding cavity where the indole group is positioned above the Glu267 side chains and the oxygen of the oxadiazole hydrogen bonds with the hydroxyl oxygen of Tyr262<sup>17</sup>. There are also three water molecules observed. The first interconnects the oxadiazole nitrogen with Nɛ2 of Gln159, the second fills the space under the pyrazole ring and is positioned above the mouth of the arginine-binding cavity and the third links the indole nitrogen and Nô2 of Asn266. The result for LibDock's accuracy and ligand retrieval provided utility in conducting further screening.


C. Overlay of CMPD-2 crystal structure and docked pose



## Figure 3.2 CMPD-2 docked using LibDock shows same interactions as crystal structure

A) Interactions of CMPD-2 in the substrate-binding site as cited directly from the crystal structure  $(2Y1X)^{17}$ . B) Interactions of CMPD-2 in the substrate-binding site as docked using computerized model. (C) The calculated RMSD between the crystal structure and docked pose was 1.2 Å. Figures were generated in Accelrys Discovery Suite 4.5®.

## LibDock and CDOCKER scores of the 40 set and of CARM1 controls

Top CARM1 inhibitors as described in section 1.2 and listed in Table 1.1, along with three CARM1 activators (Figure 3.3) used as negative controls <sup>74</sup>, and the '40 set' were first docked using LibDock. LibDock scores can range from 0 to 200 and CDOCKER interaction energies scores are based on the kcal•mol<sup>-1</sup> where the number generated is actually a –CDOCKER interaction energy score, meaning the higher the score, the "better" the pose <sup>75</sup>.



Figure 3.3. Activators of CARM1 used as negative controls for docking

Compounds 3g, 4a and 4c were discovered as enhancers of CARM1 activity <sup>74</sup> and used as negative controls for docking with LibDock. Structures were generated using Accelrys Discovery Suite 4.5®.

CDOCKER was subsequently used in order to refine each of the top poses and provide a more accurate prediction of relative binding affinities. CDOCKER scores generated subsequently to LibDock scoring are referred to hereon as 'biased CDOCKER' scores. LibDock score vs. biased CDOCKER score of the 40 set and CARM1 inhibitors and activators are plotted in Figure 3.4. The activators, 3g, 4a and 4c, scored 122.82, 101.632 and 103.718, respectively, which was comparable to scores of the 40 set. Top inhibitors had LibDock scores greater than 120 and were found to bind in the arginine-binding cavity, with the exception of AMI-1 (Figure 3.5) which bound across the substrate-binding groove (Figure 3.5, panel A). This may have contributed to AMI-1's lower score. A cluster analysis of CARM1 inhibitors' orientation within the binding pocket was performed (Figure 3.6). In subsequent docking of the GPS and CDRD libraries only a top pose with a LibDock score greater than 120 was docked using CDOCKER. The purpose of this was to focus on compounds that might actually have potential in acting as CARM1 inhibitors. Since no currently known inhibitors had LibDock scores less than 120, this seemed like a logical approach. Some top-ranked GPS and CDRD compounds were also found to bind in the arginine-binding cavity, although not as extensively as known CARM1 inhibitors (Table 3.2).



#### Figure 3.4. LibDock vs. biased CDOCKER scores 40 set and controls

LibDock versus biased CDOCKER energy scores of the 40 set (blue), CARM1 inhibitors (red) and activators (green). The red line indicates that only compounds with LibDock scores greater than 120 were subsequently docked using CDOCKER.



#### Figure 3.5. Leading CARM1 inhibitors occupy the arginine-binding cavity

(A) AMI-1, a weak CARM1 inhibitor, binds in the substrate-binding groove but not in the arginine-binding cavity (B) Diamine 12 is a more recent and potent inhibitor of CARM1, discovered in 2009, that makes interactions both in the substrate binding groove and the arginine- binding pocket. The reported  $IC_{50}$  of Diamine 12 is 0.2  $\mu$ M. Figures were created using the Accelrys Dicovery Suite 4.5®.



#### Figure 3.6 Clusters centers of CARM1 inhibitors orientation in binding pocket

A cluster analysis was performed of the top pose for the known CARM1 inhibitors in order to compare binding orientations.

## In vitro results for Diamine 12

Diamine 12 scored the highest by both the LibDock and CDOCKER algorithms (Figure 3.4). The compound was a generous gift from Dr. Eric Therrien at Methylgene in Montreal <sup>31</sup>. It acted as a positive control with the P81 filter-binding assay to determine if the *in silico* and *in vitro* data aligned. An IC<sub>50</sub> curve was generated as shown in Figure 3.7. The IC<sub>50</sub> value was calculated to be  $1.3\pm1 \mu$ M.



Figure 3.7. IC<sub>50</sub> curve for Diamine 12 using the P81 filter-binding assay

The IC<sub>50</sub> curve of Diamine 12 screened with the P81 filter-binding assay shows an IC<sub>50</sub> value of  $1.3 \pm 1 \mu$ M. The curve and calculation were made using SigmaPlot.

## LibDock and CDOCKER biased scores for all GPS and CDRD compounds

Once the *in silico* model was shown to predict higher LibDock and CDOCKER scores for known CARM1 inhibitors and lower comparative scores for the 40 set, all GPS and CDRD compounds were docked using LibDock. Only poses with a LibDock score greater than 120 were subsequently docked using CDOCKER to generate a CDOCKER biased score. Using either 2Y1X or 2Y1W as a docking template was shown to produce close to the same docking and scoring results as shown in Figure 3.13 which explains most of the variability of the response data around its mean. The top GPS and CDRD compounds shown in yellow and green in Figure 3.8 were screened using the P81 filter-binding assay. The red box in Figure 3.8 highlights the compounds who ranked highly both with LibDock and CDOCKER algorithms. Their interactions were compared in Table 3.2. In Figure 3.11 the hydrogen-bond interaction profile of these compounds is compared. Hydrogen-bond made to Glu258 are in dark blue, to Met260 in brown, to Glu267 in grey and to His415 in yellow. All other hydrogen-bonds are in blue. Actual inhibitors were shown to have more points of contact with the molecules which are likely influencing the on/off kinetics of the false positives. Furthermore, false positives lacked many of the hydrogen-bonds to the four key residues that the inhibitors exhibited. For this reason, hydrogen-bonding interactions should be evaluated alongside LibDock and CDOCKER scores.



#### Figure 3.8. LibDock vs. biased CDOCKER scores for GPS and CDRD compounds

The landscape shows the LibDock versus biased CDOCKER energy scores of all GPS compounds (blue), the top 5 GPS compounds (yellow), all CDRD compounds (grey) and the top CDRD compounds (green). CARM1 inhibitors are shown in red. The red box indicates the compounds selected to evaluate interactions in more detail (Table 3.2) and their hydrogen-bond interactions are evaluated in Figure 3.11.

### In vitro results for top ranked GPS and CDRD compounds

The top 5 GPS compounds were screened using a P81 filter-binding assay at 1, 20 and 200  $\mu$ M. The results are shown in Figure 3.9. It appears that GPS383 and GPS347 could be potential inhibitors. It should be noted that there were solubility issues with both of these compounds and for this reason additional DMSO was added for a total assay concentration of 21%. Looking back at Figure 2.7, at 21%, the decrease in activity could be contributed to the high DMSO content. For this reason, GPS383 and GPS347 were screened alongside the top CDRD compounds and GPS33, GPS346 and GPS373 (which also scored highly in the *in silico* docking) again at 200  $\mu$ M. This time, the amount of DMSO in every sample was kept consistent at 4%. Diamine 12 was again used as a positive control. The results are shown in Figure 3.10. The top-ranked GPS and CDRD compounds using LibDock and CDOCKER subsequently showed negligible inhibitory activity of CARM1 (Figure 3.10) even though they scored highly with LibDock and CDOCKER. These compounds were referred to as 'false positives.'



#### Figure 3.9. In vitro results of 5 top-ranked GPS compounds

The top 5 compounds ranked using LibDock and CDOCKER were screened using the P81 filter-binding assay at 1  $\mu$ M, 20  $\mu$ M (blue) and 200  $\mu$ M (yellow). Diamine 12 was used as a positive control (left side) at 1, 20 and 200  $\mu$ M (green, blue and yellow). (Data for GPS384 was not obtained at 200  $\mu$ M data due to sample loss during the assay.)



#### Figure 3.10. In vitro results of top-ranked CDRD compounds, GPS383 and GPS347 with 4% DMSO

The top 12 CDRD compounds, GPS383, GPS347, GPS33, GPS346 and GPS373 ranked using LibDock and CDOCKER were screened using the P81 filter-binding assay at 200  $\mu$ M. Diamine 12 was used as a positive control.



# Table 3.2 Interactions of false positives and top CARM1 inhibitors



## Figure 3.11 Hydrogen-bond interaction profiles for false positive and actual inhibitors

Hydrogen-bond interactions were evaluated for the top-ranked false positives and actual inhibitors as highlighted in the red box in Figure 3.8.

## 3.3 CDOCKER only results

CDOCKER was run independently of LibDock in order to compare the scoring. The results are shown in Figure 3.12. Biased CDOCKER scoring was shown to be a better scoring method than unbiased CDOCKER scoring for two reasons. First, the biased CDOCKER scoring is better at distinguishing between positive controls and false positives. As shown in Figure 3.12, there is a linear relationship between unbiased and biased CDOCKER interaction energy scores for CARM1 inhibitors only. The fitted line for the CARM1 inhibitors shows an R<sup>2</sup> value of 0.8255, whereas the GPS and CDRD values were only 0.0616 and 0.2836, respectively. This means that 82.55% of the variance in the unbiased CDOCKER scores is predictable from the biased CDOCKER scores for the inhibitors but none is for the GPS and CDRD compounds. In fact, some of the compounds with generally low biased CDOCKER scores had much higher unbiased scores. Biased CDOCKER scores were first screened using LibDock. LibDock uses a matching algorithm of rotational and translational vectors to position the ligand first in the same frame as reference as the protein. With unbiased CDOCKER scoring this initial matching step is not conducted meaning the ligands are subjected directly to high temperature molecular dynamics to generate ligand conformations without being matched first. CDOCKER is a better algorithm for the refinement of docking poses once a match has been established, rather than an initial docking algorithm <sup>69</sup>. Because we already know that CARM1 inhibitors have a high binding affinity, using CDOCKER initially does not have as significant impact on the scoring as the LibDock matching phase is not as crucial.



## Figure 3.12. Biased vs. Unbiased CDOCKER scores for controls, GPS and CDRD compounds

A comparison of the biased and unbiased CDOCKER interaction energies for controls (red), GPS compounds with LibDock Scores >120 (blue) and CDRD compounds with LibDock Scores >120 (grey).

The second reason why the biased CDOCKER scoring is better than unbiased CDOCKER scoring, is that there is a significant decrease in time required for the screening with biased CDOCKER scoring. Generating unbiased CDOCKER scores for 76,600 compounds took approximately 211 hours, whereas it took only 29 hours to generate biased CDOCKER scores for this same set of compounds. Even accounting for the time to conduct the initial LibDock scoring, which was approximately 16 hours, this is a huge time difference. This is not only because the compounds couldn't initially be filtered to only screen poses with LibDock scores greater than 120, but also because under the docking parameters, the 'random conformations' must be set above zero. The number of starting random conformations generated from equilibration and minimization of the starting ligand structure must be defined. For compounds already screened with LibDock, the random conformations can be set to zero because the original input structure is used. Furthermore, also under the docking parameters, the 'orientations to refine' must also be set above zero. It is important to define the number of rotated ligand orientations to refine for each of the conformations from dynamics, whereas compounds previously screened with LibDock can be set to zero, as rotations were also already performed.

## 3.4 LibDock scores of 2Y1X vs. 2Y1W

A comparison of the LibDock scores for the GPS and CDRD compounds was made using the PDB files 2Y1X and 2Y1W. The 2Y1X structure is the catalytic domain of CARM1 cocrystallized with SAH and CMPD-2, and the 2Y1W structure is the catalytic domain of CARM1 coco-crystallized with sinefungin and CMPD-1. The exact protocol used to prepare the protein and ligands and dock the compounds using LibDock was conducted as outlined in section 2.3. The results are shown in Figure 3.13.



## Figure 3.13. LibDock scores for 2Y1W vs. 2Y1X for GPS and CDRD compounds

A comparison of the LibDock scores using 2Y1W and 2Y1X as a docking template for known CARM1 inhibitors (red), all GPS compounds (orange) and all CDRD compounds (green) is shown. The R<sup>2</sup> value for all CDRD compound is 0.8464, indicating that most of the data points sits closely to the fitted regression line.

# 4. Conclusions

The long-term goal of this research is to identify novel inhibitors of CARM1 with attributes that previously identified inhibitors lack. These attributes include having anti-proliferative effects and showing selectivity towards CARM1. The work here fundamentally addresses the development of a workflow that provides a platform for discovery consisting of *in vitro* and *in silico* screening methods, new to the Frankel laboratory. The *in vitro* P81 filter-binding assay provides a straightforward method to screen novel compounds and measure PRMT activity. The in silico docking protocol provides a baseline for further docking where the *in vitro* data can be used as a control in evaluating new docking algorithms. Using the industry standard proprietary software suite, Accelrys Discovery Suite 4.5<sup>®</sup>, a system has been established to screen further compounds. By providing a clearer picture of the landscape of the LibDock and CDOCKER scores for both the GPS and CDRD compounds and comparing their interactions to known CARM1 inhibitors, it provides a baseline for subsequent docking which may help lead to the discovery of novel inhibitors of CARM1. Furthermore, second generation compounds can be designed directly within the program that enhance binding affinity prior to the compounds ever having to be synthesized in the laboratory.

There were two major findings from this work. First, top inhibitors of CARM1 were shown to bind in the substrate-binding groove and make intensive interactions (primarily hydrogen bonds) in the arginine binding cavity. The interactions of CMPD-1 and CMPD-2 have been described previously <sup>17</sup>, however, the binding interactions of Diamine 12 were unknown. The findings from this work show that Diamine 12 also made polar interactions with CARM1 residues in the argininebinding cavity common to CMPD-1, CMPD-2 and 7f (optimized CMPD-2) (Table 1.1). By evaluating and quantifying the hydrogen-bond interaction profiles of the false positives and top positive controls (Figure 3.11), it was shown that actual inhibitors made more points of contact with the molecules which likely influenced the on/off kinetics of the false positives. Furthermore, false positives lacked many of the hydrogen-bonding interactions to the four key residues (Glu258, Met260, Glu267 and His415) that the inhibitors showed. For this reason, hydrogen-bonding interactions should be evaluated alongside LibDock and CDOCKER scores.

Second, novel compounds that had LibDock and CDOCKER scores relatively close to known CARM1 inhibitor scores, showed no inhibitory effect in vitro (Figure 3.8). The top-ranked compound, GPS383, had a LibDock score of 156.384 and a CDOCKER score of 60.187 whereas the top CARM1 inhibitor, Diamine 12, had a LibDock score of 152.483 and a CDOCKER score of 71.1185. Since Diamine 12 is a known potent inhibitor of CARM1, it would be expected that GPS383, having comparable LibDock and CDOCKER interaction scores, would have an inhibitory effect in vitro. This was not the case. GPS383 showed little to no inhibitory activity in vitro (only a 7% decrease in enzyme activity at 200 µM), whereas Diamine 12 showed an 83% drop in enzyme activity at 200  $\mu$ M (Figure 3.10). The scoring between the two compounds were expected to have been more pronounced. There are many factors that could have affected the scoring. First, there may be on/off kinetics of the compounds when binding to CARM1. At a snapshot in time, the binding affinity may score high, but in a biological setting the compound may bind for a brief amount of time, not allowing much time to inhibit CARM1 activity. Second, the compounds could be binding in areas other than the arginine-binding pocket. The binding site was defined as the site where CMPD-2 was found to bind when co-crystallized with CARM1  $(2Y1X)^{17}$ , but it is unknown whether the compounds screened would necessarily prefer the same binding site. Potential inhibitors from the GPS and CDRD libraries may have been overlooked strictly based on where the binding site was predefined. Third, crystal structures used as models

for *in silico* docking are merely snapshots in the timescale of protein dynamics. A compound scoring high at one point in time using computer modelling may have no effect biologically as all steps of the methylation processes cannot be accounted for.

There are two directions this project could go. First, determining a better way to filter the data and second, conducting structure activity relationships in silico. Better filtering the data means a different algorithm could be used or the pipeline could be altered. For example, GOLD (Genetic Optimization for Ligand Docking; Cambridge Crystallographic Data Centre) might be a good choice for the next docking algorithm to try as it utilizes a flexible ligand and target, and it also has the ability to include different solvation parameters including specific water molecules <sup>76</sup>. There are dozens of different receptor-ligand interaction products that sample conformations of small molecules in protein binding sites differently. Furthermore, different scoring functions of each of these algorithms assess which of the conformations best complements the protein binding site. Also, conducting structure activity relationships in silico to predict better inhibitory activity is an attractive option with the data already obtained. Once a lead compound is identified, modifications can be made to the compound using in silico design to see if binding affinity and scoring is improved. Using the 'sketch molecules tools' in Accelrys<sup>®</sup>, new small molecules can be created and edits to structures can be made. If a lead compound is discovered, it is a matter of making changes directly in the program to alter any functional groups, change bond orders or elements of selected atoms. Newly created structures can be screened immediately and compared to parent compounds for improved binding. This will greatly reduce the time required to synthesize second generation compounds in the laboratory.

Second generation compounds synthesized will be tested for selectivity against other PRMTs and protein lysine methyltransferases (PKMTs). There are a number of assays that can be used, however, because of PMT's slow enzymatic turnover, most PMT-activity assays detect reaction products over measuring the depletion of starting materials <sup>7</sup>. The methylated substrate and SAH can be quantified directly (*e.g.*, autoradiography, mass spectrometry or *anti*-methylarginine/lysine antibodies) or quantified indirectly after processing them into various derivatives (*e.g.*, enzyme-coupled calorimetric assay and shot-gun mass spectrometry) <sup>7</sup>. The best assays to determine selectivity against other PRMTs and PKMTs are the P81 filter-binding assay, an antibody-based PMT-activity assay or a SAH-based chemogenic assay <sup>7</sup>. Mass spectrometry assay should be avoided as even a small amount of DMSO was shown to impact the standard calibration curve of the assay (Figure 2.4).

Last, top-ranked compounds will be screened using cultured cancer cells for anti-proliferative or cytotoxicity effects. The assay best used to measure these effects is a live content imaging system to find compounds that show dose-dependent cell-based activity. Using time-lapse microscopy, cell-growth and apoptosis can be monitored over many hours, days or weeks. By quantifying cellular dynamics in real time, toxicity and efficacy profiles of lead compounds can be determined.

CARM1 is a prominent and epigenetic target in the treatment of prostate and breast cancer, yet progress in identifying therapeutics remains idle. The work provided herein addresses the development of a workflow providing a platform for future discovery. Both the *in silico* and *in vitro* screening methods provide useful resources for the identification and improvement of novel compounds in inhibiting CARM1 and potentially leading to the discovery of therapeutic agents for cancer treatment.

# References

- 1. Bedford, M. T. & Richard, S. Arginine methylation an emerging regulator of protein function. *Mol. Cell* 18, 263–72 (2005).
- 2. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* 21, 381–395 (2011).
- 3. Zhang, R. *et al.* Theoretical Insights into Catalytic Mechanism of Protein Arginine Methyltransferase 1. *PLoS One* 8, 1–11 (2013).
- 4. Allfrey, V., Faulkner, R. & Mirsky, A. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Sci. United States* 315, 786–794 (1964).
- 5. Yang, Y. & Bedford, M. T. Protein arginine methyltransferases and cancer. *Nat. Rev. Cancer* 13, 37–50 (2013).
- 6. Schapira, M. & Ferreira de Freitas, R. Structural biology and chemistry of protein arginine methyltransferases. *Med. Chem. Commun.* 5, 1779–1788 (2014).
- Luo, M. Current chemical biology approaches to interrogate protein methyltransferases. ACS Chem. Biol. 7, 443–63 (2012).
- 8. Cheng, D., Vemulapalli, V. & Bedford, M. T. *Methods applied to the study of protein arginine methylation. Methods in enzymology* 512, (Elsevier Inc., 2012).
- 9. Yang, Y. *et al.* PRMT9 is a type II methyltransferase that methylates the splicing factor SAP145. *Nat. Commun.* 6, 6428 (2015).
- Feng, Y., Hadjikyriacou, A. & Clarke, S. G. Substrate specificity of human protein arginine methyltransferase 7 (PRMT7): the importance of acidic residues in the double E loop. *J. Biol. Chem.* 289, 32604–16 (2014).
- 11. Zhang, X. & Cheng, X. Structure of the predominant protein arginine methyltransferase PRMT1 and analysis of its binding to substrate peptides. *Structure* 11, 509–520 (2003).
- Schmidt, C. L. A., Kirk, P. L. & Appleman, W. K. The Apparent Dissociation Constants Of Arginine And Of Lysine And The Apparent Heats Of Ionization Of Certain Amino Acids. J. Biol. Chem. 88, 285–293 (1930).
- 13. Kimata, Y. *et al.* Genetic Evidence for a Role of BiP / Kar2 That Regulates Ire1 in Response to Accumulation of Unfolded Proteins. *Mol. Biol. Cell* 14, 2559–2569 (2003).
- 14. Couture, J.-F., Collazo, E. & Trievel, R. C. Molecular recognition of histone H3 by the WD40 protein WDR5. *Nat. Struct. Mol. Biol.* 13, 698–703 (2006).
- 15. Bedford, M. T. Arginine methylation at a glance. J. Cell Sci. 120, 4243-6 (2007).
- 16. 't Hart, P., Lakowski, T. M., Thomas, D., Frankel, A. & Martin, N. I. Peptidic partial bisubstrates as inhibitors of the protein arginine N-methyltransferases. *Chembiochem* 12, 1427–32 (2011).
- 17. Sack, J. S. *et al.* Structural basis for CARM1 inhibition by indole and pyrazole inhibitors. *Biochem. J.* 436, 331–9 (2011).

- 18. Troffer-Charlier, N., Cura, V., Hassenboehler, P., Moras, D. & Cavarelli, J. Functional insights from structures of coactivator-associated arginine methyltransferase 1 domains. *EMBO J.* 26, 4391–4401 (2007).
- 19. Antonysamy, S. *et al.* Crystal structure of the human PRMT5:MEP50 complex. *Proceedings of the National Academy of Sciences* 109, 17960–17965 (2012).
- 20. Pak, M. L. *et al.* A Protein Arginine N -Methyltransferase 1 (PRMT1) and 2 Heteromeric Interaction Increases PRMT1 Enzymatic Activity. 1, (2011).
- 21. Swiercz, R., Person, M. D. & Bedford, M. T. Ribosomal protein S2 is a substrate for mammalian PRMT3 (protein arginine methyltransferase 3). *Biochem. J.* 386, 85–91 (2005).
- 22. Thomas, D., Lakowski, T. M., Pak, M. L., Kim, J. J. & Frankel, A. Förster resonance energy transfer measurements of cofactor-dependent effects on protein arginine N-methyltransferase homodimerization. *Protein Sci.* 19, 2141–51 (2010).
- 23. Dou, Y. *et al.* Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nat. Struct. Mol. Biol.* 13, 713–9 (2006).
- 24. 't Hart, P. *et al.* Analogues of the HIV-Tat peptide containing Nη-modified arginines as potent inhibitors of protein arginine N-methyltransferases. *Medchemcomm* 3, 1235 (2012).
- 25. Zhang, X., Zhou, L. & Cheng, X. Crystal structure of the conserved core of protein arginine methyltransferase PRMT3. *EMBO J.* 19, 3509–3519 (2000).
- 26. Copeland, R. a, Solomon, M. E. & Richon, V. M. Protein methyltransferases as a target class for drug discovery. *Nat. Rev. Drug Discov.* 8, 724–32 (2009).
- 27. Smith, B. C. & Denu, J. M. Chemical mechanisms of histone lysine and arginine modifications. *Biochim. Biophys. Acta* 1789, 45–57 (2009).
- 28. Lakowski, T. M. & Frankel, A. A kinetic study of human protein arginine N-methyltransferase 6 reveals a distributive mechanism. *J. Biol. Chem.* 283, 10015–25 (2008).
- Yue, Y., Chu, Y. & Guo, H. Computational Study of Symmetric Methylation on Histone Arginine Catalyzed by Protein Arginine Methyltransferase PRMT5 through QM/MM MD and Free Energy Simulations. *Molecules* 20, 10032–10046 (2015).
- 30. Nef, P. *et al.* Regulation of Transcription by a Protein Methyltransferase. 284, 2174–2178 (1999).
- 31. Therrien, E. *et al.* 1,2-Diamines as inhibitors of co-activator associated arginine methyltransferase 1 (CARM1). *Bioorg. Med. Chem. Lett.* 19, 6725–32 (2009).
- 32. Yue, W. W., Hassler, M., Roe, S. M., Thompson-Vale, V. & Pearl, L. H. Insights into histone code syntax from structural and biochemical studies of CARM1 methyltransferase. *EMBO J.* 26, 4402–12 (2007).
- 33. Purandare, A. V *et al.* Pyrazole inhibitors of coactivator associated arginine methyltransferase 1 (CARM1). *Bioorg. Med. Chem. Lett.* 18, 4438–41 (2008).
- 34. Dowden, J. *et al.* Small molecule inhibitors that discriminate between protein arginine N-methyltransferases PRMT1 and CARM1. *Org. Biomol. Chem.* 9, 7814–21 (2011).
- 35. Majumder, S. *et al.* Involvement of Arginine Methyltransferase CARM1in Androgen Receptor Function and Prostate Cancer Cell Viability. 1301, 1292–1301 (2006).

- 36. Castellano, S. *et al.* Identi fi cation of Small-Molecule Enhancers of Arginine Methylation Catalyzed by Coactivator-Associated Arginine Methyltransferase 1. (2012).
- 37. Messaoudi, S. El et al. (CARM1) is a positive regulator of the Cyclin E1 gene. 103, 13351–13356 (2006).
- 38. Zeng, H. *et al.* A TR-FRET-based functional assay for screening activators of CARM1. *Chembiochem* 14, 827–35 (2013).
- 39. Frietze, S., Lupien, M., Silver, P. a & Brown, M. CARM1 regulates estrogen-stimulated breast cancer growth through up-regulation of E2F1. *Cancer Res.* 68, 301–6 (2008).
- 40. Kraus, W. L. & Wong, J. Nuclear receptor-dependent transcription with chromatin. *Eur. J. Biochem.* 269, 2275–2283 (2002).
- 41. An, W., Kim, J. & Roeder, R. G. Ordered cooperative functions of PRMT1, p300, and CARM1 in transcriptional activation by p53. *Cell* 117, 735–48 (2004).
- 42. Canada, S. Canadian Cancer Statistics Special topic : Liver cancer. 1–114 (2013).
- 43. Lakowski, T. M., 't Hart, P., Ahern, C. A., Martin, N. I. & Frankel, A. Nη-substituted arginyl peptide inhibitors of protein arginine N-methyltransferases. *ACS Chem. Biol.* 5, 1053–63 (2010).
- 44. Thomas, D. *et al.* Protein arginine N-methyltransferase substrate preferences for different Nη-substituted arginyl peptides. *ChemBioChem* 15, 1607–1613 (2014).
- 45. Vedel, M., Lawrence, F., Robert-Gero, M. & Lederer, E. The antifungal antibiotic sinefungin as a very active inhibitor of methyltransferases and of the transformation of chick embryo fibroblasts by Rous sarcoma virus. *Biochem. Biophys. Res. Commun.* 85, 371–376 (1978).
- 46. Dowden, J. *et al.* Small molecule inhibitors that discriminate between protein arginine N-methyltransferases PRMT1 and CARM1. *Org. Biomol. Chem.* 9, 7814–21 (2011).
- 47. Mai, A. *et al.* Synthesis and biological validation of novel synthetic histone/protein methyltransferase inhibitors. *ChemMedChem* 2, 987–91 (2007).
- 48. Cheng, D. *et al.* Small molecule regulators of protein arginine methyltransferases. *J. Biol. Chem.* 279, 23892–9 (2004).
- 49. Mai, A. *et al.* epigenetic multiple ligands: mixed histone/protein methyltransferase, acetyltransferase, and class III deacetylase (sirtuin) inhibitors. *J. Med. Chem.* 51, 2279–90 (2008).
- 50. Huynh, T. *et al.* Optimization of pyrazole inhibitors of Coactivator Associated Arginine Methyltransferase 1 (CARM1). *Bioorg. Med. Chem. Lett.* 19, 2924–7 (2009).
- 51. Allan, M. *et al.* N-Benzyl-1-heteroaryl-3-(trifluoromethyl)-1H-pyrazole-5-carboxamides as inhibitors of coactivator associated arginine methyltransferase 1 (CARM1). *Bioorg. Med. Chem. Lett.* 19, 1218–23 (2009).
- 52. Wan, H. *et al.* Benzo[d]imidazole inhibitors of Coactivator Associated Arginine Methyltransferase 1 (CARM1)--Hit to Lead studies. *Bioorg. Med. Chem. Lett.* 19, 5063–6 (2009).
- 53. Selvi, B. R. *et al.* Identification of a novel inhibitor of coactivator-associated arginine methyltransferase 1 (CARM1)-mediated methylation of histone H3 Arg-17. *J. Biol. Chem.* 285, 7143–52 (2010).

- 54. Cheng, D. & Bedford, M. T. Xenoestrogens regulate the activity of arginine methyltransferases. *Chembiochem* 12, 323–9 (2011).
- 55. Cheng, D. *et al.* Small molecule regulators of protein arginine methyltransferases. *J. Biol. Chem.* 279, 23892–9 (2004).
- 56. Lakowski, T. M. & Frankel, A. Kinetic analysis of human protein arginine N-methyltransferase 2: formation of monomethyl- and asymmetric dimethyl-arginine residues on histone H4. *Biochem. J.* 421, 253–61 (2009).
- 57. Edelhoch, H. Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* 6, 1948–1954 (1967).
- 58. Simonian, M. H. Spectrophotometric determination of protein concentration. *Curr. Protoc. Toxicol.* Appendix 3, A.3G.1–7 (2004).
- 59. Facility, B. B. & Sharpe, T. Biozentrum Biophysics Facility Tim Sharpe, 14 May 2013. 1–2 (2013).
- 60. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254 (1976).
- 61. Bio Rad. Quick Start Bradford Instruction Manual. 1–35 (2014).
- 62. 't Hart, P. *et al.* Analogues of the HIV-Tat peptide containing Nη-modified arginines as potent inhibitors of protein arginine N-methyltransferases. *Medchemcomm* 3, 1235 (2012).
- 63. Lakowski, T. M., Zurita-Lopez, C., Clarke, S. G. & Frankel, A. Approaches to measuring the activities of protein arginine N-methyltransferases. *Anal. Biochem.* 397, 1–11 (2010).
- 64. Wu, J., Xie, N., Feng, Y. & Zheng, Y. G. Scintillation proximity assay of arginine methylation. *J. Biomol. Screen.* 17, 237–44 (2012).
- 65. Kortagere, S. & Ekins, S. Troubleshooting computational methods in drug discovery. J. Pharmacol. Toxicol. Methods 61, 67–75 (2014).
- Rezácová, P., Borek, D., Moy, S. F., Joachimiak, A. & Otwinowski, Z. Crystal structure and putative function of small Toprim domain-containing protein from Bacillus stearothermophilus. *Proteins* 70, 311– 319 (2008).
- Moitessier, N., Englebienne, P., Lee, D., Lawandi, J. & Corbeil, C. R. Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *Br. J. Pharmacol.* 153 Suppl , S7–S26 (2008).
- 68. Rao, S. N., Head, M. S., Kulkarni, A. & LaLonde, J. M. Validation studies of the site-directed docking program LibDock. *J. Chem. Inf. Model.* 47, 2159–2171 (2007).
- 69. Wu, G., Robertson, D. H., Brooks III, C. L. & Vieth, M. Detailed analysis of grid-based molecular docking: A case study of CDOCKER - A CHARMm based MD docking program. *J. Comput. Chem.* 24, 1549–1562 (2003).
- 70. Kaufmann, K. W. & Meiler, J. Using RosettaLigand for Small Molecule Docking into Comparative Models. *PLoS One* 7, (2012).
- Erickson, J. a., Jalaie, M., Robertson, D. H., Lewis, R. a. & Vieth, M. Lessons in Molecular Recognition: The Effects of Ligand and Protein Flexibility on Molecular Docking Accuracy. J. Med. Chem. 47, 45–55 (2004).

- 72. Steffen, C. *et al.* TmoleX--a graphical user interface for TURBOMOLE. *J. Comput. Chem.* 31, 2967–2970 (2010).
- 73. Sitzmann, M., Ihlenfeldt, W. D. & Nicklaus, M. C. Tautomerism in large databases. J. Comput. Aided. Mol. Des. 24, 521–551 (2010).
- 74. Castellano, S. *et al.* Identification of small-molecule enhancers of arginine methylation catalyzed by coactivator-associated arginine methyltransferase 1. *J. Med. Chem.* 55, 9875–9890 (2012).
- 75. Abdel-Hamid, M. K. & McCluskey, A. In Silico docking, molecular dynamics and binding energy insights into the bolinaquinone-clathrin terminal domain binding site. *Molecules* 19, 6609–6622 (2014).
- 76. Al-Balas, Q. *et al.* Generation of the first structure-based pharmacophore model containing a selective 'zinc binding group' feature to identify potential glyoxalase-1 inhibitors. *Molecules* 17, 13740–13758 (2012).

# Appendix

# Appendix figures



#### Figure A.1. Peak purity analysis of Diamine 12

The two panels show the chromatograms of the peak purity analysis of Diamine 12 using a diode array UV detector. As the spectra are sufficiently alike, the peak is considered pure. Data collected and experiment conducted by Dr. Eric Therrien (Molecular Forecaster Inc.).