Exploring the Mechanism and Product Specificity of Protein Arginine Methyltransferase 1 and Drug Discovery for Diabetes and Malaria

by

Symon Mukigi Gathiaka

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
May 9, 2015

Keywords: chemistry, computational, molecular dynamics, drug discovery, PRMT, PPAR, *pf*TrxR

Approved by

Orlando Acevedo, Chair, Associate Professor of Chemistry
Konrad Patkowski, Assistant Professor of Chemistry
Holly Ellis, Associate Professor of Chemistry
Stewart Schneller, Professor of Chemistry

Abstract

In this work, three main projects are studied: (1) the mechanism and product specificity of protein arginine methyltransferase 1 (PRMT1) (2) the design and development of agonists that target peroxisome proliferator activated receptors (PPAR) as potential drugs to manage diabetes and (3) the characterization of *plasmodium falciparum* thioredoxin reductase (*pf*TrxR) inhibitors, as antimalarial drugs, using *in silico* techniques and biological assays.

Chapter 1 is concerned with the introduction of the three proteins that were investigated in chapter 3 (PRMT1), chapter 4 (PPAR) and chapter 5 (*pf*TrxR). Chapter 2 is mainly concerned with the overview and theoretical background of the computational methods used in this research. Chapter 3 reports the results of the simulations carried out on PRMT1 in order to elucidate how its active site is engineered for specific product formation. The role of specific residues (Glu144, Glu153 His293 and Met48) in catalysis and controlling product specificity are discussed. Chapter 4 presents the results of a joint computational and experimental study that lead to the discovery of compound **9**, a dual PPARδ/γ agonist that doesn't elicit undesirable effects associated with full PPARγ agonists. Chapter 5 presents the results of the determination of antiplasmodial activity and binding affinity of certain compounds towards *pf*TrxR using combined computational and experimental approaches.

Acknowledgments

I would like to thank my advisor for believing in me. His mentoring through the years has been invaluable and I would not have come this far without him. To all his research collaborators especially Drs. Amin, Calderon and Hevel, thank you for the science “fun” you have enabled me to experience. I would also like to thank my committee members Drs. Patkowski, Ellis and Schneller, for your availability and support throughout the later years of my graduate career. To my past and present lab mates: Samba, Kira, Billy, Caley, Bin, Nicole, Robel, Boykin and Brian for being wonderful colleagues. To my family: Dorcas, Claire and Angela, I will always be indebted for making tough times tolerable.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

PRMT	Protein arginine methyltransferase
SAM	S- adenosyl methionine
MMA	Mono-methylated arginine
ADMA	Asymmetric di-methylated arginine
SDMA	Symmetric di-methylated arginine
SAH	S-adenosyl-L-homocystein
PPAR	Peroxisome proliferator-activated receptor
LBD	Ligand binding domain
DBD	DNA binding domain
RXR	Retinoid X receptor
TZD	Thiazalodinedione
AF-2	Transactivation factor 2
*Pf*TrxR	*Plasmodium falciparum* Thioredoxin reductase
*h*TrxR	*human* Thioredoxin reductase
MD	Molecular dynamics
aMD	Accelerated molecular Dynamics
WT	Wild type

Chapter 1: Introduction

1.1 Protein arginine methyltransferase 1 (PRMT1)

1.1.1 Biological methylation

When genes coding for specific proteins are expressed, a specific part of the DNA is copied into messenger RNA (mRNA) through transcription. During translation, the mRNA then serves as a template for the protein's synthesis with the decoding to amino acids done by ribosomes. The amino acids are then folded into a three-dimensional protein. To extend their polypeptide diversity beyond the limits of the encoded amino acids, eukaryotic cells use posttranslational modifications which expand the information encoded by the human genome, from about 25,000 genes to over 1 million proteins that compose the human proteome.[1] Proteins can be modified by modifications such as phosphorylation, methylation, acetylation and ubiquitylation, creating binding sites for specific protein domains where the modifications modulate molecular interactions, protein localization and stability.[2,3]

Methyltransferases (MTases) catalyze the transfer of methyl groups from a methyl donor such as S-adenosyl methionine (SAM) to the methyl acceptor such as a peptide substrate. The substrate molecule enters the active site of the enzyme and binds by a number of non-bonded interactions to different protein residues. The positively charged sulfur atom of SAM attracts electron density from the methyl group while the substrate, usually the nucleophilic entity with lone pair of electrons or a partial negative charge, abstracts the methyl group. Fig. 1.1 shows a general reaction mechanism where a base is required to abstract a proton from the substrate.[4,5]

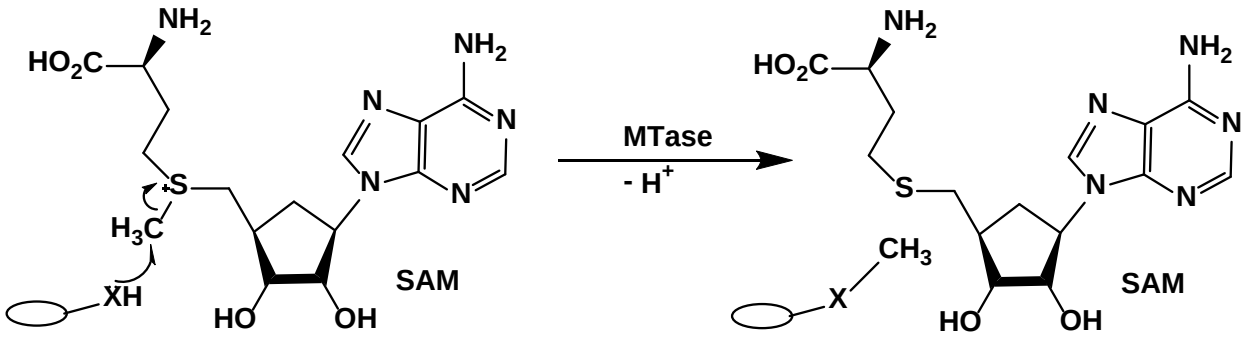


Figure 1.1 The general S_N2 mechanism for SAM-dependent enzymes.

1.1.2 Protein arginine methyltransferases

Protein arginine methylation is a common modification involved in cellular processes like signaling, RNA processing, gene transcription, and cellular transport function.[6] Arginine methylation is catalyzed by the protein arginine methyltransferases (PRMTs) and their significance in human diseases has been increasingly studied, especially in cardiovascular diseases[1,7] and cancer[8-10].

So far, 11 PRMT isoforms have been identified with the length of the proteins varying between 316 and 956 amino acid residues (Fig. 1.2).[11] Depending on the type of methylarginine product generated, PRMTs are divided into four major classes (Fig. 1.3). Type I and II PRMTs both catalyze mono-methylation of the guanidinium nitrogen of specific arginine residues in substrates. The mono-methylation product (MMA) can either be further methylated to an asymmetric (ADMA) by Type I or a symmetric di-methylated arginine (SDMA) by Type II methyltransferases. Type III enzymes catalyze only the mono-methylation of the ω nitrogen in arginine residues in proteins while type IV generates a mono-methylated arginine residue on the δ nitrogen of the substrate.[12]

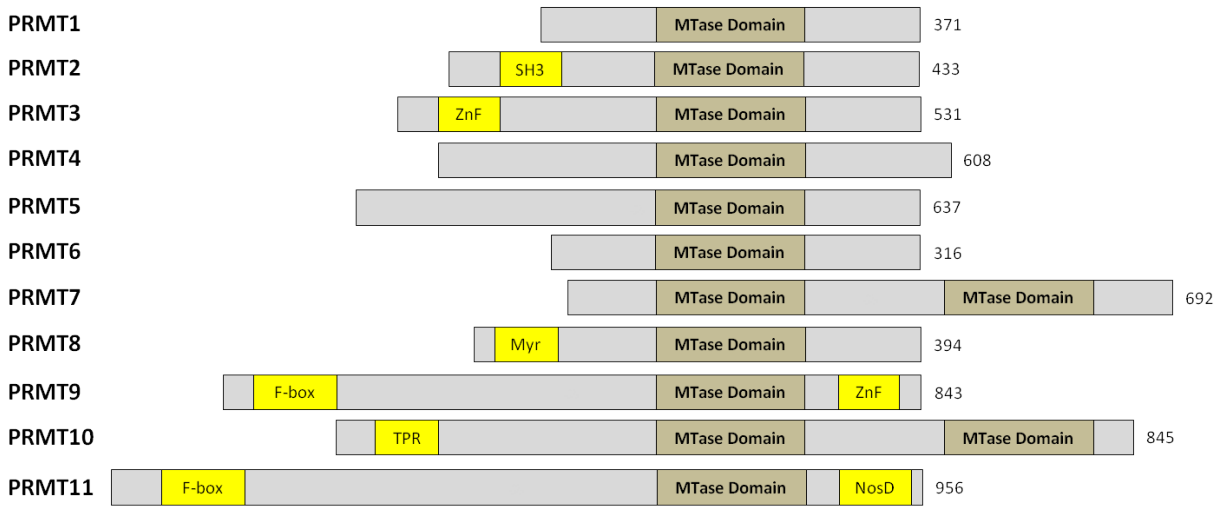


Figure 1.2 Representation of the eleven human PRMTs, all containing a conserved methyltransferase (MTase) domain (Adapted from wolf[11]).

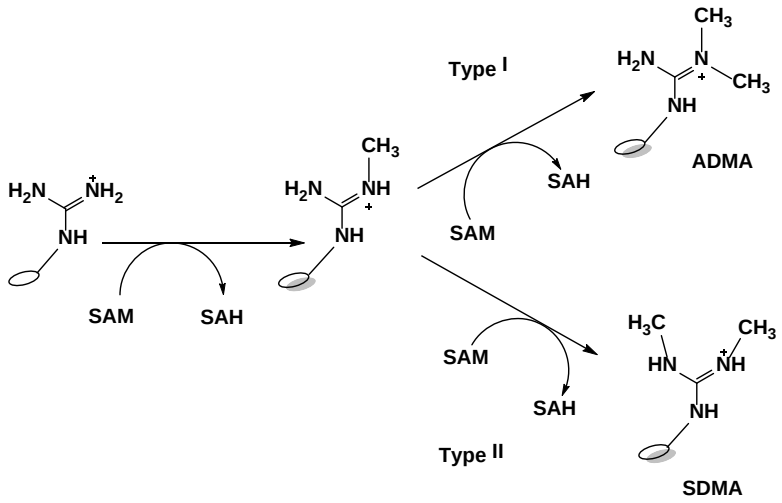


Figure 1.3 Methylation of the arginine side chain by PRMTs.

With the biological significance of these proteins gaining appreciation over the last couple of years, the current understanding of their biochemistry is limited. Adding to the complexity of the PRMT field, different methylation statuses (MMA, ADMA, or SDMA state) of the same substrate can lead to distinct biological outputs.[13] For example, symmetric

dimethylation of arginine 3 of histone H4 (H4R3) leads to repression of gene expression, while asymmetric dimethylation of H4R3 is associated with gene activation.[14]

1.1.3 Protein arginine methyltransferase 1 (PRMT1)

PRMT1 is the predominant type I PRMT in mammalian cells, accounting for 85% of cellular PRMT activity and is involved in the transfer of the methyl group from SAM to the guanidino group of arginines in protein substrates, resulting in exclusively the MMA and ADMA products.[11,15]

1.1.4 PRMT1 structure

The crystal structure of rat PRMT1 was first determined by Zhang et al. in complex with the reaction product S-adenosyl-L-homocysteine (SAH) and a 19 residue R3 peptide substrate containing three arginines.[15] However, the reported structure contained only a single arginine residue as the substrate as the rest of the residues did not have clear electron densities. The monomeric protein is a two domain structure: the SAM binding domain and the β-barrel domain. The structure contains four major parts: N-terminal, SAM binding site, the β-barrel and the dimerization arm (Fig. 1.4). As seen in the structure, the N-terminal constrains the bound SAH.

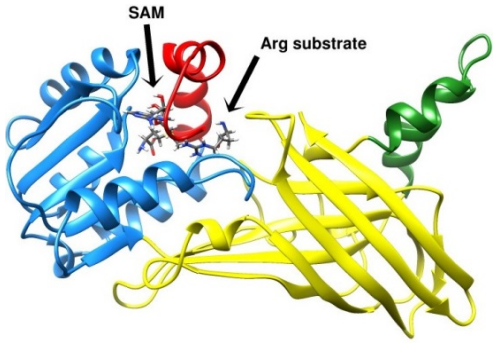


Figure 1.4 The monomer structure. N-terminal (red), SAM binding (blue), β-barrel (yellow) and dimerization arm (green).

The whole protein is an oligomer with a hydrophobic dimer interface between the dimerization arm and the SAM binding site. Beyond showing that that the dimerization of the protein is essential for SAM binding, Zhang et al. reasoned that the interface in the ring-like dimer would allow the processive production of the methylation product since isolated products of the enzyme reaction are usually completely or almost completely dimethylated (Fig. 1.5). This is conceivable by the cooperation of multiple active sites where the product of the first methylation reaction enters the other part of the dimer without releasing the mono-methylated substrate or re-binding SAM.

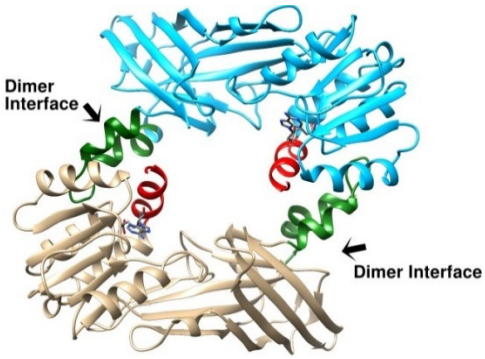


Figure 1.5 The ring-like dimer of rat PRMT1 with the N-terminal shown in red and the dimerization arm shown in green.

1.1.5 Mechanism of PRMT1

Rust et al. probed the mechanism of this enzyme.[16] In their study, they evaluated the role played by some highly conserved active site residues (Tyr39, Arg54, Glu100, Glu144, Glu153, Met155 and His293) in SAM recognition, substrate binding and catalysis. Their hypothesis was that catalysis involves three conserved active site residues: Arg 54, Glu144 and Glu153 (Fig. 1.6A). Arg54 and Glu144 help position the substrate for the attack while Glu153 plays a role in increasing the nucleophilicity of the guanidinium moiety of the substrate arginine through

electronic effects. They proposed a mechanism that involves the formation of a dication intermediate (Fig. 1.6B).

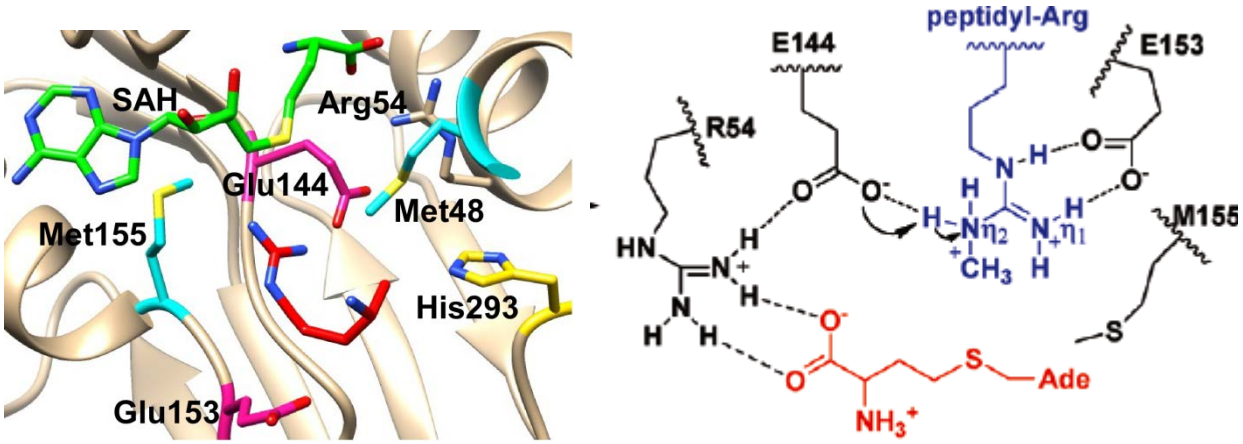


Figure 1.6 (A) The conserved active site residues Arg54, Glu144, Glu153. (B)1.6b: Rust et al.'s proposed dication intermediate (Adapted from Rust et al.[16]).

The proposed mechanism was supported by their findings that indicated the reaction being mainly driven by proximity effects. The prior deprotonation of the substrate guanidinium may not be necessary for the rate-limiting step of the mechanism.

1.1.6 Regulation of PRMT1

To understand the biological function of PRMTs, it is imperative to understand how the respective states of methylation are achieved as this is the key to product specificity. The conserved Met48 in PRMT1 has previously been shown to be involved in substrate recognition and activity which impacts on the degree of methylation in the substrate.[13] In addition, the M48F mutant (Phe is the equivalent residue in PRMT5 – a type II PRMT) generated both ADMA and SDMA.[5] However, the mutant enzyme was highly impaired and the production of SDMA was low. Complicating the role of M48 in specificity is that M48A and M48L exclusively produce ADMA as the dimethylation product.

Interestingly, the mutation of the conserved phenylalanine (F379) to methionine at the corresponding location in PRMT5 converts it to a more active enzyme that catalyzes both symmetric and asymmetric dimethylation of arginine in seemingly equal amounts.[14] Furthermore, no such activity was detected with the F379A mutation. This necessitates a deeper evaluation of the elements of PRMTs that dictate how its active site is engineered to control product specificity.

1.2 Peroxisome Proliferator Activated receptors (PPAR)

1.2.1 Nuclear receptors

The nuclear receptor superfamily is a class of proteins that work with other proteins to regulate transcription, the first step in gene expression where a particular segment of DNA is copied into RNA. They are one of the most abundant classes of transcriptional regulators in biology and they control diverse functions from development, metabolism to homeostasis.[17,18] Nuclear receptors have a common structural organization. This includes a DNA binding domain (DBD) which is a highly conserved region found approximately along the center of the polypeptides and targets the receptor to highly specific DNA sequences. The second region is the ligand binding domain (LBD) and is located in the C-terminal half of the receptor and recognizes specific endogenous and exogenous ligands directing specificity to biological response. Other regions are the variable N- and C- terminal domains that are highly variable both in sequence and size as well as a variable length hinge region that links the DBD and LBD.[17,19]

The superfamily can be divided into two groups: type I are the steroid receptors while type II contains the hitherto orphan receptors, such as, peroxisome proliferator activated receptors (PPARs), and the retinoid X receptors (RXRs). In order to bind their suitable response elements, almost all of the type II's, in particular PPARs, form heterodimers with RXR.[20,21]

1.2.2 RXR:PPAR Heterodimer and gene expression – PPAR mechanism of action

RXRs are composed of three subtypes (α, β, and γ) and their ligands are retinoids, a set of structural and functional analogues of vitamin A. PPARs are also ligand-activated transcription factors constituted by three subtypes α, β, and γ and regulate the expression of numerous genes involved in a range of important biological processes.[22,23] For their function, PPARs require RXR as a heterodimeric companion. They undergo a conformational change upon agonist binding which results in the release of histone deacetylase (HDAC) co-repressors, enabling the heterodimerization with RXR. To this complex, the recruitment of RNA polymerase II and co-activators with histone acetyl transferase (HAT) activity then takes place and the heterodimer (PPAR and RXR) then binds to the peroxisome proliferator responsive element (PPRE) in target genes leading to chromatin remodeling and eventual gene expression (Fig. 1.7).[22,23]

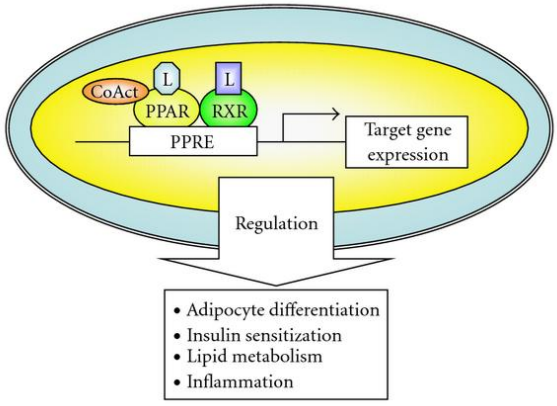


Figure 1.7: The RXR:PPAR Heterodimer activation of gene transcription (Adapted from Shimizu et al.[22]).

1.2.3 PPAR subtypes

Peroxisomes are organelles found in eukaryotic cells and contain enzymes involved in metabolic reactions, including energy metabolism processes.[24] In the early 1990's, an orphan receptor that induced peroxisome proliferation in rodent hepatocytes in response to exogenous

chemicals (peroxisome proliferators) was cloned and characterized and the isoform later labelled PPARα.[25] Since the discovery of PPARα, two other PPAR subtypes have been identified and labelled PPARβ/δ and PPARγ, with the individual subtypes encoded by different genes and having a slightly different tissue distribution with disparate physiological roles and endogenous agonist preference. Unlike in PPARα, neither PPARδ nor PPARγ is activated by peroxisome proliferators in rodents and none of the PPARs cause peroxisome proliferation in human.[26,27] Each of the three isotypes either activates or suppresses different genes with only partial overlap in activity. All subtypes are involved in lipid homeostasis and energy balance through glucose homeostasis.[28]

PPARα is found primarily in metabolically active tissues such as the heart, liver, kidney, skeletal muscle, macrophages and brown adipose tissue. In the liver, PPARα is centrally involved in lipid metabolism and plays an important role in fatty acid uptake, transport and catabolism and its activation lowers lipid levels.[29,30] The activation of PPARα prevents high-fat diet-induced obesity, increase insulin sensitivity and mitigate metabolic syndrome symptoms. The endogenous activators of PPARα include fatty acids and eicosanoids and the activation of the pharmaceutically relevant hypolipidemic drugs (fibrates e.g. clofibrate, fenofibrate, and bezafibrate) causes a decrease in circulating lipid levels in humans.[23,31]

The gene for PPARγ codes for four distinct variant of the isoform, PPARγ1-4.[32] These variants are found in several tissues including liver, muscle and heart and adipose tissue. PPARγ is involved in the differentiation of adipocytes (fat cells), regulation of genes involved in energy homeostasis, lipoprotein metabolism, and insulin sensitivity.[28,33] The endogenous ligands for PPARγ include fatty acids and eicosanoid derivatives which activate PPARγ at μm concentrations. Its activation is of clinical importance since it's involved in improving insulin

sensitivity. The thiazolidinedione (TZD) class of antidiabetic drugs (e.g., rosiglitazone and pioglitazone, troglitazone and ciglitazone) lower blood glucose levels in humans at the nm ranges *in vitro*.[30,33]

Unlike PPARα and γ, the PPARδ subtype has been studied less extensively. It is expressed ubiquitously in human tissues though some tissues like the brain, liver, adipose tissue, and skeletal muscles exhibit higher concentrations. It is involved in fatty acid oxidation, and in blood cholesterol and glucose homeostasis. Its activation in adipocytes results in fatty acid oxidation and utilization and has therefore become of interest in the management of obesity and obesity related disorders. Ligand activation has also been shown to decrease serum triglycerides and increase insulin sensitivity. Endogenous agonists of this subtype include fatty acids and prostacyclin (PGI_2).[23,28]

1.2.4 PPAR structure

The PPAR proteins have domains that are very common in virtually all nuclear receptor superfamily (Fig. 1.8). The A/B domain (N-terminal) contains the ligand-independent AF-1 (transactivation function) site and is least conserved. The DNA binding domain (two zinc finger motifs) is a highly conserved region found approximately along the center.

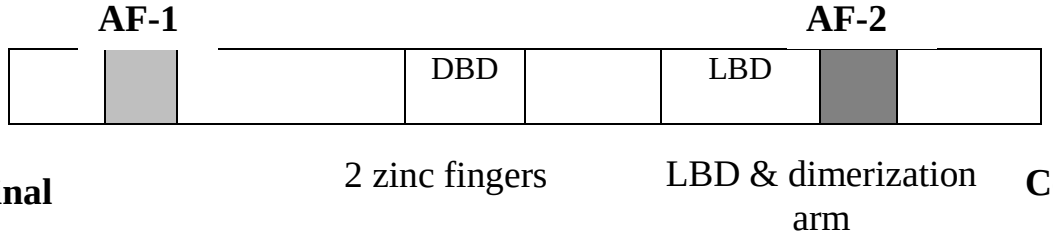


Figure 1.8: The domain structure of the PPARs.[27]

The zinc fingers bind PPREs in the regulatory region of PPAR-responsive genes. The extreme C-terminal section of the receptor is also conserved and contains the ligand binding domain (LBD)

which encompasses a ligand-dependent transactivation function (AF-2). Additionally, the LBD is required for receptor dimerization and interaction with auxiliary proteins.[26,27,33]

PPARs LBD structures have been studied extensively (due to their importance as antidiabetic [34,35] among other drug targets[23]). X-ray apo and agonist co-crystal structures of PPARα[36,37], PPARγ[35,38] and PPARδ[39,40] are available with good resolution.

1.2.5 PPAR's LBDs as drug targets

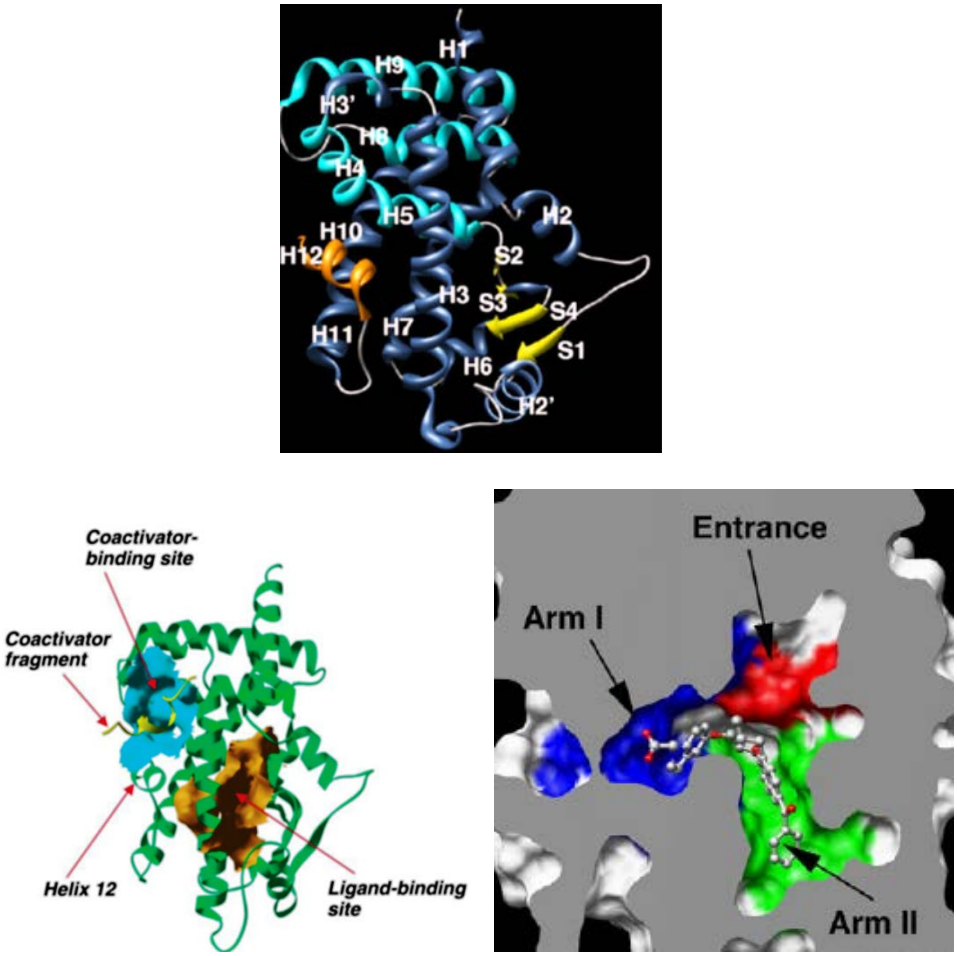


Figure 1.9: (A) The secondary structure elements H1-H12 and β-strands S1-S4 with the AF-2 helix H12 shown in orange (B) and (C) PPARγ LBD. (Adapted from Berger at al.[33] and Zoete et al.[41]).

The overall LBD structure is common in all 3 subtypes and is composed of 13 α-helices and a small 4-stranded β-sheet (Fig. 1.9A). In comparison with other nuclear receptors, it has a large cavity (Fig. 1.9B) which allows the receptors to interact with a wide variety of structurally distinct natural and synthetic ligands. It is Y- shaped and extends from the C-terminal AF-2 helix to the β-sheet between helices 3 and 6 (Fig. 1.9C).[33,41,42]

The cavity starts from the entrance and extends into arm I and arm II pockets. Arm I is substantially polar while the entrance and arm II are mainly hydrophobic. The binding cavity is composed of 34 residues with 80% conservation across the subtypes. Natural and synthetic ligands usually conform to a standard pharmacophoric model in which they are composed of a polar head that binds in arm I and a hydrophobic tail that binds in arm II and the entrance. Most agonists interact with the conserved polar residues of Arm I (together with a residue of the AF2 helix) which include Thr289, His323, His449 and Tyr473 (AF2) of PPARδ, Ser280, Tyr314, His440 and Tyr464 of PPARα, and Ser289, His323, His449 and Tyr473 (AF2) of PPARγ, respectively. These residues are part of a hydrogen-bonds network involving the carboxylate group of fatty acids and eicosanoic acids, the natural agonists.[41,43]

Due to their various metabolic and therapeutic actions, PPARs have become major drug targets. Synthetic ligands that interact with PPAR-γ can either be full agonists, partial agonists, or antagonists. The structural basis of full and partial agonists has been extensively studied whereas antagonists have received less attention.[41,44] The full or partial character of synthetic ligands seems to be provided by specific H-bonds with the conserved polar residues of arm I. For most full agonists, H-bonding typically occurs with Ser289, His323, Tyr473 and His449. H-bonding with Tyr473 is crucial to the stabilization of the AF2 helix H12 which allows the binding of coactivators that lead to the activation of the genes responsible for adipogenesis

(formation of fat cells) and insulin sensitivity.[41,44] Partial agonists lead to a lesser degree of AF-2 stabilization and allow for binding of coactivators responsible for insulin sensitivity without affecting adipogenesis. Antagonists on the other hand show high affinity, but do not activate PPAR-γ. [41,44] Recent efforts have seen research into the new generation of drugs, i.e., PPARα/γ dual agonists that have the capacity of insulin sensitization without the associated side effects.[28]

1.3 *Plasmodium falciparum* Thioredoxin reductase (*Pf*TrxR)

1.3.1 Malaria

Malaria remains a huge public health and economic development threat in the tropical and sub-tropical regions despite century long efforts to eradicate the disease. Approximately 40% of the world population lives in high risk areas and 97 countries and territories had ongoing malaria transmission in 2014.[45,46] In 2013, According to the World Health Organization's World Malaria Report estimates, 198 million cases of malaria occurred worldwide resulting in ~ 500,000 deaths. Globally, majority of the deaths occurred to under-five year olds.[45,46]

The disease is caused by infection with the *plasmodium* parasite transmitted by the female *Anopheles* mosquitoes. There are five major plasmodium species that infect humans namely; *P. falciparum, P. vivax, P. ovale P. malariae P. and knowlesi*. Of this species, *P. falciparum* is the most deadly while the rest cause a comparably milder form of the disease that is rarely fatal. The control strategies currently being employed include early diagnosis and treatment in combination with preventive measures targeting mosquito control. Unfortunately, the malaria parasite is rapidly acquiring and spreading resistance to anti-malarial drugs and so far, there are no new drugs in advanced stages of development or effective vaccine in the pipeline.[45,47-49] This could be due to the fact that the parasites have a complicated life cycle.

1.3.2 The life cycle of the *Plasmodium* parasite

Their life cycle (Fig. 1.10) involves the sexual part (in the mosquito) and the asexual part (in the host). A mosquito ingests the sexual stages of the malaria parasite (gametocytes) during blood feeding from a malaria infected host. These gametes then fertilize in the mosquito mid-gut resulting into zygotes that then develop into oocysts. The oocysts then grow and divide producing sporozoites that invade the salivary glands. Infection of a new host starts when a mosquito bites for a blood meal. The mosquito's saliva containing sporozoites enters the bloodstream and reaches the liver infecting hepatocytes. This stage shows no symptoms. The sporozoites then proliferate in the liver cells producing a very invasive form referred to as the merozoites, which are then released into the bloodstream invading the host's red blood cells where the asexual blood cycle begins. In the red blood cells, the cytoplasm gets consumed followed by binary divisions that produce more merozoites that burst the host cells to infect a new round of red blood cells. This is when illness and complications of malaria begins and that's why most antimalarial drugs target the blood stage.[50-52]

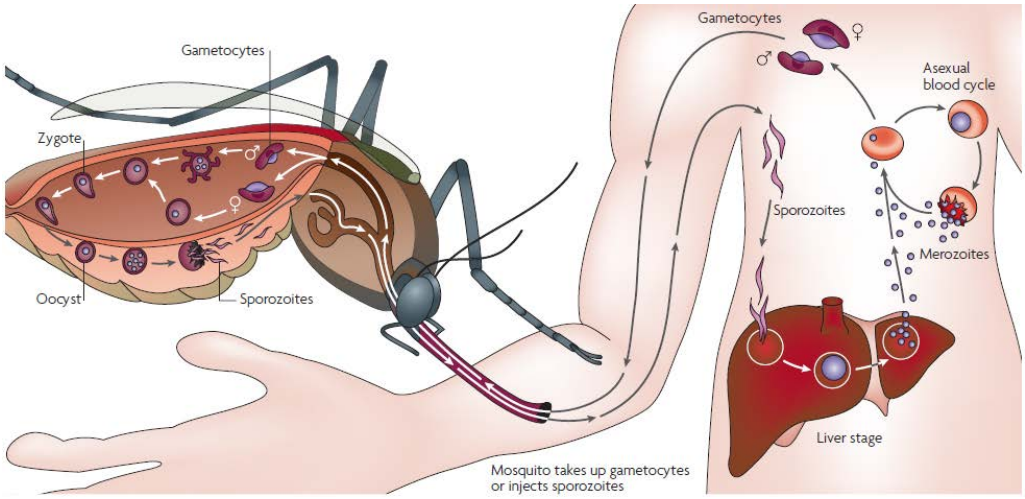


Figure 1.10: The life cycle of the malaria parasite (Adapted from Su et al.[52]).

1.3.2 Oxidative reductases as drug targets

The extremely complex life cycle of *Plasmodium* require specialized protein expression for intracellular and extracellular survival in host environments.[53] During the asexual blood cycle, as the parasites attack the red blood cells, they decompose hemoglobin as a source of proteins and this releases toxic heme. Heme together with activated macrophages of the host exposes the parasite to oxidative stress.[54] To cope with the high fluxes of reactive oxygen species (ROS), the parasite has its own battery of defense mechanisms and produces its own antioxidant enzymes that involve a network of NADPH-dependent redox enzymes.[55] The two major redox systems that regulates cellular antioxidative capabilities are the glutathione/glutathione disulfide (GSH/GSSG) that governs the antioxidative capacity of cells and the dithiol/disulfide of thioredoxins (Trx_{red}/Trx_{ox}) systems that form a distinct but complementary redox system.[56] Disrupting the antioxidative systems in the malaria parasite holds promise and these two systems have become targets for drug development in order to inhibit the survival of the parasite in host cells.[57,58]

1.3.3 *Pf*TrxR structure

P. falciparum Thioredoxin reductase (*Pf*TrxR) is a homodimeric flavoenzyme that reduces thioredoxin (Trx) to an active dithiol containing form acting as an important antioxidant defense by donating electrons to a range of target molecules.[59]

$$NADPH + H^+ \; Trx\text{-}S_2 \longrightarrow NADP^+ + Trx\text{-}(SH)_2$$

The only *Pf*TrxR crystal structure available (PDB ID: 4B1B) shows that the monomeric unit contains three domains: a NADPH binding domain, a FAD binding domain, and a monomer–monomer interface.[55] Comparison between the human and parasite TrxRs shows that they have

46% sequence identity and overlay with an RMSD of 0.91 Å between 374 monomer atom pairs (Fig. 1.11).

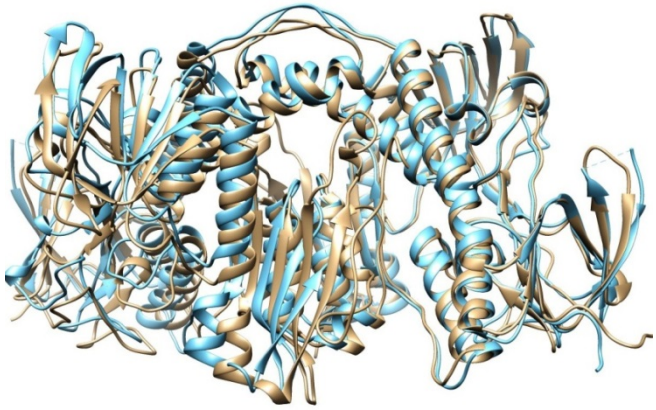


Figure 1.11: Superposition of *h*TrxR[60] (brown) with *pf*TrxR[55] (blue) crystal structures.

The structural differences between the *Pf*TrxR and its corresponding human protein (*h*TrxR) can be exploited for the development of selective inhibitors. First, in the *h*TrxR, the active site contains a Cys-selenocysteine (Sec) pair which is absent in *Pf*TrxR, an attractive target for selectivity.[61] Second, the dimer interface can be exploited for selective noncompetitive inhibitors between the two enzymes. *Pf*TrxR is narrower than in *h*TrxR due to the presence of Tyr101 and His104 (Fig. 1.12) and can therefore host smaller molecules. Their counterparts in the human isoform are Gln and Leu and this difference can determine the chemical nature of suitable inhibitors.[55]

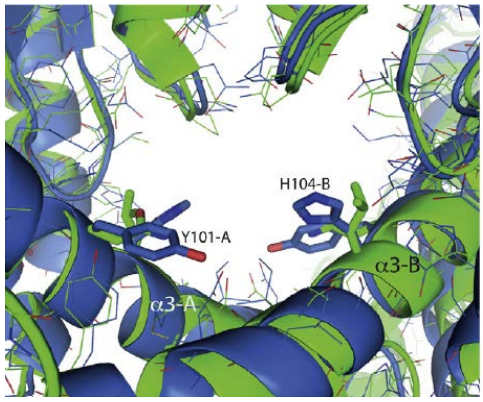


Figure 1.12: *Pf*TrxR (blue) in comparison with *H*TrxR (green) (Adapted from Boumis et al.[55]).

Chapter 2: Methods

2.1 Molecular Dynamic Simulations and Theory

A major tool that can be used in the theoretical study of biological systems is molecular dynamics (MD) simulations. Based upon the principles of physics, MD simulations can readily provide information on the internal motions of a macromolecule as a function of time at the atomistic level.[62-64] Alder and Wainwright were the first to introduce molecular dynamics and they applied the theory to calculate equilibrium properties on hard spheres[65]. The first simulations were however done by Rahman[66] in the 1960's for liquid argon while the first protein simulation was carried out by McCammon in the 1970's[67]. The techniques have greatly improved over the years. At present, MD is routinely applied to complex systems[68], including in complimenting experimental procedures such as the refinement of NMR structures[69].

In this work, molecular dynamics simulations using AMBER have been performed on Protein Arginine Methyltransferase 1 (PRMT1) in chapter 3, in order to investigate how its active site is engineered to control product specificity. This section contains the underlying theory of molecular dynamics implementation in the AMBER biomolecular package[70] as well as a basic overview of the organizational structure of the suite of programs.

2.1.1 Ensemble Averages

Simulations of molecular systems compute macroscopic properties from microscopic interactions and the connection between the two is made through statistical averages. The properties of the system depend on the microscopic state (positions and momenta) of the number of particles (N) that contain the system. An ensemble is a collection of all possible systems which have different microscopic states but have an identical macroscopic or thermodynamic state. The ensemble average is given by equation 2.1 below:

$$< A >_{ensemble} = \iint dp^N dr^N A(p^N, r^N)\, \rho(p^N, r^N) \quad (2.1)$$

Where:

$A(p^N, r^N)$ - is the property of interest, p is the momenta and r the position.

(< >) - indicates an ensemble average or expectation value.

$\rho(p^N, r^N)$ – is the probability density of the ensemble, i.e., the probability of finding a configuration with momenta p^N and positions r^N.

Molecular dynamics simulations thus aims at generating enough representative conformations that samples ideally all the possible states. With enough sampling, information on the structural, dynamic and thermodynamic properties of the system of interest can be computed. Different macroscopic variable constraints can result in different types of ensembles.[71,72] The constraints usually involves the number of particles (N), volume (V), energy (E), temperature (T), chemical potential (μ), and pressure (P). Examples include,

1) The microcanonical ensemble (NVE) – is the state where the number of particles (N), the volume (V) and the energy (E) of the system are fixed. In this state, the system is referred to as isolated.
2) The canonical ensemble (NVT) - the number of particles (N), the volume (V) and the temperature (T) of the system are fixed. Helmholtz free energy is derived from this ensemble.
3) Grand canonical Ensemble (μVT) - is the state where the chemical potential (μ), the volume (V) and the temperature (T) of the system are fixed.
4) Isobaric-Isothermal Ensemble (NPT) - is the state where the number of particles (N), the pressure (P) and the temperature (T) of the system are fixed.

2.1.2 The Equation of Motion and Time-integration Algorithms

Molecular dynamics simulation is based on Newton's equation of motion where the solution of the equation 2.2 yields the time evolution of a collection of interacting particles.

$$\mathbf{F}_i = m_i \mathbf{a}_i = m_i \frac{d\boldsymbol{v}_i}{dt} = m_i \frac{d^2\mathbf{r}_i(t)}{dt^2} \tag{2.2}$$

Where:

a, v and r – are the acceleration, velocity and position of the particles respectively. A trajectory that encompasses the positions, velocities and acceleration of particles as they vary with time can then be achieved by integrating the above second order differential equations. The trajectory allows the state of the system to be predictable at any time while its analysis provides values of averaged properties. The position and velocity vectors describe the time evolution of the system in phase space and are usually propagated with a finite time interval using numerical integrators. The integration is broken down into many small steps. Each step is separated in time by a fixed time interval Δt and the aim is to express the positions $r_i(t + \Delta t)$ at time $t + \Delta t$ in terms of the already known positions at time t.[73]

There are many algorithms for integrating the equations of motion but one of the most accepted numerical integrators is the Verlet algorithm[74] due to its simplicity and stability. Using Taylor expansion of the coordinate of a particle, around time t, the algorithm's expression can be described as equation 2.3 below:

$$\mathbf{r}_i(t + \Delta t) \approx 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t) + \frac{\mathbf{F}_i(t)}{m_i} \Delta t^2 \tag{2.3}$$

One disadvatage of this algorithm is that it lacks an explicit velocity term, however, the velocities are calculated from the knowledge of the trajectory by dividing the difference in positions at time $t + \Delta t$ and $t - \Delta t$ by $2\Delta t$. The leap frog algorithm [75] was introduced in 1970 by Hockney to include the velocity explicitly into the Verlet algorithm.

The most resource-consuming part of molecualar dynamics simulation is calculating the force acting on each atom with the fastest motions in the system determining the intergration step (Δt). Bonds involving hydrogen atoms are the fastest, but not crucial most of the time. They are regulary eliminated to fasten the calculation by imposing constraints on the lengths. In AMBER, this is usually done using the SHAKE algorithm[76] which reduces and averages out the highest frequency vibrations by limiting the degrees of freedom.

2.1.3 The Force Field and the Interatomic Potentials

Using a classic model to describe the molecular dynamics of a system, the basic consideration is the fact that interatomic forces are responsible for the holding together the collection of atoms. The interaction law describes the potential energy of the atoms as a function of their positions and the force acting on each atom is then related to the gradient of the potential energy as shown by equation 2.4 below:

$$\mathbf{F}_i = -\nabla \boldsymbol{r}_i V(\boldsymbol{r}_1 \cdots \mathbf{r}_N) \tag{2.4}$$

Models of atomic force fields are usually based on empirical potentials which contain a function that represents the potential energy of the system together with the related parameters that are required. The total energy of the system is usually computed from the sum of the bonded and non-bonded energies as shown by equation 2.5 below:

$$\boldsymbol{E}_{total} = \boldsymbol{E}_{bond} + \boldsymbol{E}_{angle} + \boldsymbol{E}_{dihedral} + \boldsymbol{E}_{non\text{-}bond} \tag{2.5}$$

The basic force field used in AMBER takes the functional form shown in equation 2.6 below:

$$V\,(r^n) = \sum_{bonds} K_r\ (r - r_{eq})^2\ +\ \sum_{angles} K_\theta\ (\theta - \theta_{eq})^2$$

$$+ \sum_{dihedrals} \frac{V_n}{2}\,[1\ +\ cos(n\varphi - \gamma)] + \sum_{i<j}\ \left[\frac{A_{ij}}{R_{ij}^{12}} + \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon_r R_{ij}}\right] \tag{2.6}$$

Where the potential energy V is expressed as a function of the positions r of n atoms and K_r, r_{eq}, K_θ, θ_{eq}, V_n, n, γ, A_{ij}, B_{ij}, ε_r, q_i and q_j are all empirically defined parameters. In this

function, r_{eq} and θ_{eq} are equilibrium structural parameters whereas K_r, k_θ and V_n are force constants.

The first term is used to represent the stretching of a bond between two atoms and is modeled as a simple harmonic oscillator. The second term describes the angle bending of three covalently bound atoms, also modeled as a harmonic oscillator. The third term is for the torsional energetics and utilizes a Fourier series expansion. n is the multiplicity (number of minima in a 360° rotation), φ is the dihedral angle and γ is the phase angle that determines the location of the minima. The non-bonded terms are a combination of Lennard-Jones and Coulomb terms. The Lennard-Jones 12-6 function models the *Van der Waals* interactions while the Coulomb potential represents electrostatic interactions.[77,78] The electrostatic and van der Waals interactions described above assume that the total non-bonded interaction energies are the sum of all atom pairs. However, three-body interactions aren't additive. To account for these effects, the AMBER force field therefore incorporates an efficient way to deal with this caveat by using "effective" pair potentials when calculating the non-bonding interactions.

Computing the non-bonded interactions is very computationally demanding. The van der Waals interaction between atoms decreases rapidly with distance and therefore, a common practice is to truncate the potential. However, electrostatic interactions are longer ranged and truncating the potential may introduce unacceptable error in the energy.

2.1.4 Long-range Integrations

The Ewald summation technique[79] , is a method for computing long-range interactions in periodic systems. It was first developed as a method for calculating electrostatic energies of ionic crystals, and is now commonly used for calculating long-range interactions in molecular

mechanics simulations. In this method, each particle interacts with all the other particles in the simulation box and with all of their images in an infinite periodic array. The position of each of these boxes, with reference to the central box, assumed to be a cube for simplicity, of side L containing N charges, can be described by a translation vector ($\pm iL,\pm jL,\pm kL$), where i, j, k are positive integers. For the interaction of atoms within the central box, the electrostatics are calculated using the Coulomb potential whereas those between the central box and in the image boxes, the distance is modified by the translational vector n (= n_xL, n_yL, n_zL). The total electrostatic energy is given by equation 2.7 below:

$$V = \frac{1}{2} \sum_{|n|}^{'} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j}{4\pi\varepsilon_0 |r_{ij} + n|}$$

$$|\boldsymbol{n}| = (n_x l, n_y L, n_z L) \tag{2.7}$$

This potential now includes the interactions between all pairs of charges, q_i and q_j in the central box and all image boxes positioned at a cubic lattice point $|\boldsymbol{n}|$. The prime on the first summation indicates that the series does not include interaction $i = j$ for $n = 0$.

The problem with the Ewald sum (2.7) is that it converges extremely slowly. Amber uses the Particle Mesh Ewald (PME)[80] summation method approach which uses a grid based system to replace the point charges with their continuous coordinates. The atom centered point charges are then distributed on the grid taking care to reproduce the original location of the charges potential. The PME algorithm rapidly converges to zero at long distances, hence the direct sum converges quickly.

2.1.5 Periodic Boundary Conditions

Modeling a solution requires a very large number of solvent molecules which complicates the molecular dynamics simulation. Again, having a box of solvent molecules

creates challenges due to the solvent molecules at the edge of the solvent being surrounded by a vacuum. Periodic boundary conditions are therefore employed in order to have a relatively small number of solvent molecules and reproduce the properties of the bulk.[81] Using this technique, the particles in the simulation are enclosed in a box which is then replicated in all directions to give a periodic array (Fig. 2.1).

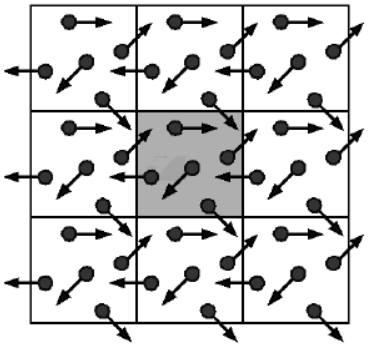


Figure 2.1 Periodic boundary conditions in two dimensions.

The coordinates of each particle in the image boxes can be computed by adding or subtracting integral multiples of the box sides. The number of particles within the central box remains constant since if during the simulation a particle leaves the box, it is replaced by an image particle that enters from the opposite side at the same time (Fig. 2.1).[81]

2.1.6 Accelerated Molecular Dynamics

Due to the computational resources required by biological systems, the simulation is normally restricted to the nanosecond time scale. Adequate consideration of the energy landscape for these systems by molecular dynamics simulation is challenging due to the number of metastable states present separated by large barriers. The presence of multiple potential energy wells gets the system stuck in some local minima for extended periods during simulation and it is paramount for proper sampling to have a way of simulating transitions between potential energy wells in order to perform thorough simulations of systems. There are a number of approaches

that have been used to tackle this challenge.[82] Examples include replica exchange[83], self-guided molecular dynamics[84] and conformational flooding[85].

In this study, accelerated molecular dynamics (aMD) was used. This is a bias potential implemented in AMBER which was introduced by McCammon[82]. The technique adds a bias potential that enables the surmounting of high energy barriers and only requires the evolution of a single copy of the system, i.e., it doesn't require advance information of the location of either the potential energy wells or saddle points.[86]

The bias boost potential function $\Delta V(\mathbf{r})$, which is a continuous positive value, is introduced whenever the true potential value $V(\mathbf{r})$ gets below a certain chosen value E (Fig 2.2) and the simulation is performed using the modified potential $V^*(\mathbf{r}) = V(\mathbf{r}) + \Delta V(\mathbf{r})$ showed by the dashed lines.

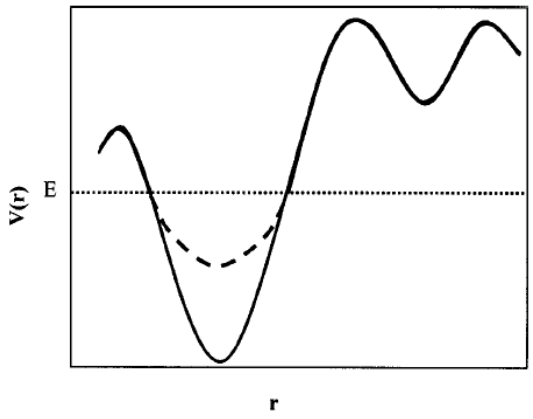


Figure 2.2 A representation of the normal potential, the biased potential, and the threshold boost energy, E (Adapted from Hamelberg et al.[82]).

When the true potential function $V(\mathbf{r})$ is greater than the chosen value E, the simulation is performed on the true potential, $V^*(\mathbf{r}) = V(\mathbf{r})$. The minima and the barriers are still maintained as

such thereby conserving the shape of the real potential which is recovered by a reweighting procedure.

The modification of the potential is described by equation 2.8 below:

$$\Delta V\ (r) = \frac{(Ep - V\ (r))^2}{(\alpha P + Ep - V(r))} + \frac{(Ed - Vd\ (r))^2}{(\alpha D + Ed - V(r))} \quad (2.8)$$

Where:

V(r) and *Vd(r)* – the normal potential and the normal torsion potential respectively.

Ep and *Ed* - average potential and dihedral energies. They are the reference energy for comparing the present position and therefore the relationship to the boosting factor to be applied.

αP and αD - factors that determine the strength of the boost to be applied.

AMBER supports acceleration based on the entire potential or only the torsional terms of the potential. It also allows the possibility to boost the whole potential with an extra boost to the torsions[86].

2.1.7 Overview of the Amber Simulation Package

Amber[87] was originally developed in 1978 at Peter Kollman's laboratory[88]. The term refers to a suite of programs that enables the carrying out of molecular dynamics simulations for proteins, nucleic acids and carbohydrates. The programs work in synergy in the input preparation, carrying out simulations and analyzing the outputs.[70] The current main scientific groups involved in the development of MD methods particular in Amber are over 40 researchers, external collaborators and contributors.[86]

The information flow is depicted in Fig. 2.3 where the three main steps involved are: input preparation, simulations and analyzing the outputs.

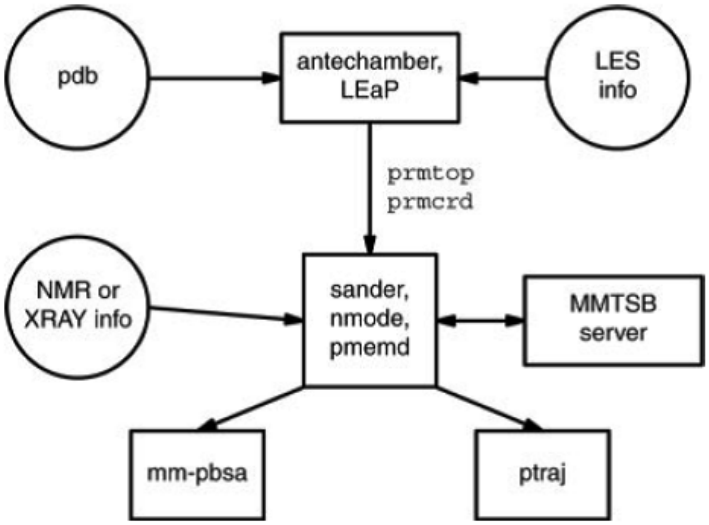


Figure 2.3 The basic information flow in the AMBER program suite (Adapted from Case et al.[70])

Preparation programs: The preparation programs are mainly *antechamber* and *LEaP*. *Leap* is the primary program to create a new system in Amber, or to modify existing systems. It creates inputs, solvates them and prepares the force field parameters that are required. *Antechamber* prepares the residues whose parameters are not part of the standard libraries using the general AMBER force field (GAFF).

Simulation programs: *Sander* is the main energy minimizer and molecular dynamics program. *Pmemd*, whose name stands for "Particle Mesh Ewald Molecular Dynamics", is a sander version optimized for speed and for parallel scaling. The *Pmemd.cuda* is the code that runs on GPUs.

Analysis programs: The *Ptraj* analysis program analyzes MD trajectories for clustering of trajectories, hydrogen bonding occupancies, entropy and distance calculations among other variables. The *Mm-pbsa.py* program carries out energy analysis of snapshots from a molecular dynamics simulation.

2.2 Protein modeling

X-ray crystallography is the most widely used technique for determining high resolution 3D protein structures. However, crystallization requires several years to elucidate a new crystal structure. NMR is the second widely used and accepted method but is not always suitable for biological systems. In the absence of an experimental crystal structure, protein modeling can be used to predict the 3D structure with comparable accuracy to experimental results.[89] Protein structure prediction using computational methods can be divided into two general approaches. The first one is comparative (homology) modeling which relies on similarity of the amino acid sequence with at least one known structure. The second approach is the *de novo* or *ab initio* method which depends on amino acid sequence only to predict the structure, without relying on known related structures.[90]

In this work, homology modeling technique has been used utilizing the comparative modeling program MODELLER 9.10.[91] This was done to model the missing portion of the N-terminal helix and re-orient Glu153 in PRMT1 (chapter 3).

2.2.1 Homology modeling

Homology modeling is based on 2 main ideas[89]:

1) If the amino acid sequence of a protein is known, it is feasible to obtain the structure since the unique 3D structure of a protein is determined by the sequence.

2) Rost's rule[92] that states if the length of two protein sequences and percent of identical residues lies in the "safe" region in the Fig. 2.4 below, the two proteins are certain to adopt a similar structure. This is based on the observation that, evolutionally changes associated with the structure are much slower than the related sequence and even distant relatives fold into similar structure.[93]

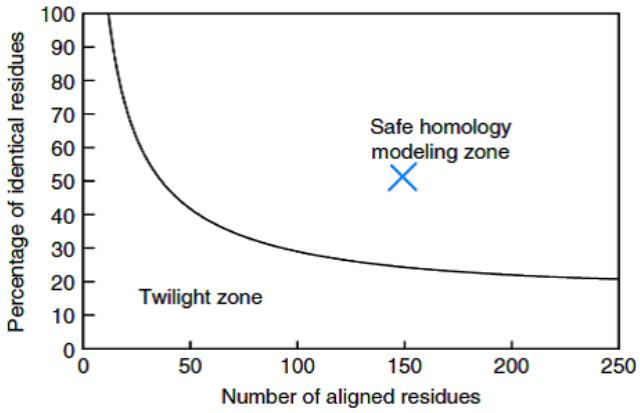


Figure 2.4 The two zones that demarcate the percentage of sequence identity that's safe to guarantee similarity in structure given a certain aligned residue length (Adapted from Bourne et al.[89]).

The method usually involves three steps[94]:

Step 1: Template recognition and initial alignment. Template(s) recognition involves sequence searches from the PDB database[95] to identify similar sequences by percentage identity. This is usually done using the Basic Local Alignment Search Tool (BLAST)[96], a computer algorithm available for use online at the National Center for Biotechnology Information (NCBI) website. BLAST align and compare a query protein sequence with the database of sequences containing available crystal structures (the Protein Data Bank), which makes it an important tool for template recognition and initial alignment between the target protein and the suitable template(s).[97]

Step 2: Model building. To construct the 3D of the entire protein or part of the protein, the software package MODELLER can be used to generate the model. The user provides the sequence alignment of a sequence to be modeled with known related structures and the program calculates a model with all non-hydrogen atoms. MODELLER builds a structure by satisfaction of spatial restraints that include distances, angles, dihedral angles, pairs of dihedral angles and

other spatial features which are either homology-derived, obtained from the CHARMM-22 force field and/or from a representative set of known protein structures. The spatial restraints are then expressed as probability density functions (pdf) that are summed into an objective function. The optimization of the created model is then done by a combination of conjugate gradients and molecular dynamics with simulated annealing.[98]

Step 3: Model Validation. Errors in homology models mainly depend on the percentage identity of the amino acid sequence between the target and the template as well as the imperfections of the template. It is therefore imperative in homology modeling to evaluate the quality of the created model.[89] There are various quality validation tools that are available for accessing correctness of the overall fold, errors over localized regions and stereochemical properties of the model. Examples of these tools are PROCHECK[99], WHATCHECK[100], WHATIF[101], PROSA[102], VERIFY 3D[103] and ERRAT program[104]. In summary, the entire process of protein structure prediction may be represented by Fig. 2.5 below.

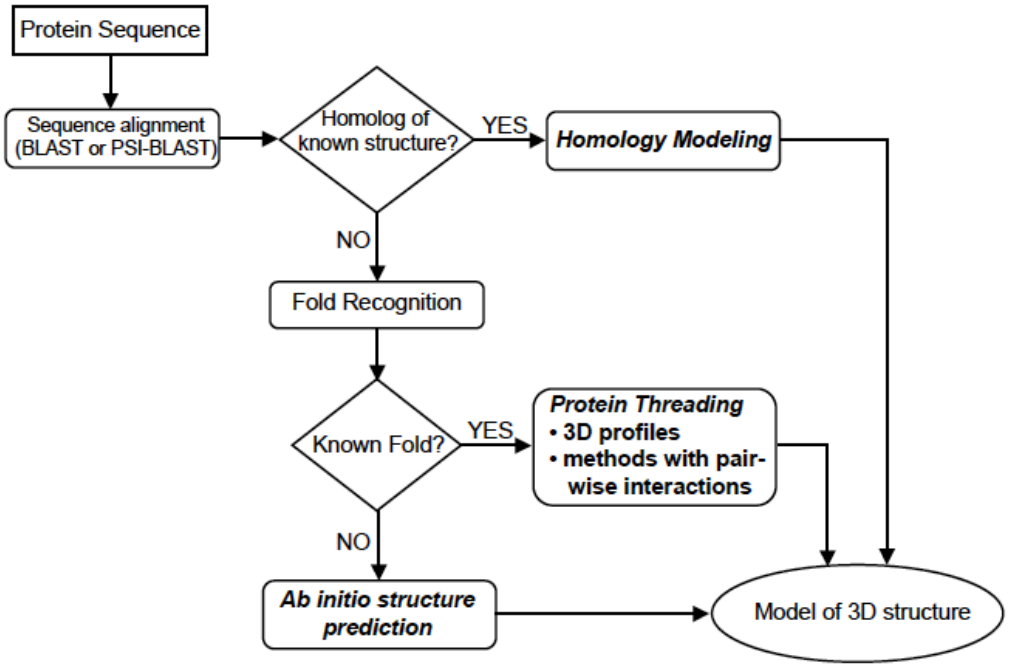


Figure 2.5 The protein structure prediction flow chart.

2.3 Molecular Docking

Molecular docking of small-molecule ligands on two targets, the peroxisome proliferator activated receptors (PPAR) and *plasmodium falciparum* thioredoxin reductase (*pf*TrxR), have been carried out in this work. The target proteins are discussed in chapter 4 and 5. Docking tools were used in order to establish high affinity agonists for the PPAR and inhibitors for *pf*TrxR. This section is an overview on molecular docking and the programs used in the study.

2.3.1 Ligand-Protein Binding Affinities

Modern drug discovery programs are increasingly incorporating computer modeling of ligand-protein interactions. Drug discovery researchers who are most adept with computational tools have the potential of delivering new drug candidates quicker and at a highly reduced cost.[105] The different classes of computer methods can be grouped based on the fastest (robust but least rigorous physically) to slowest (most rigorous physically). The classes can be listed as follows[106]:

1) The molecular docking methods. Examples include AutoDock[107], Surflex[108], Glide[109] and GOLD[110]. These methods make the most approximations, especially on solvation, but are reasonably fast and generally reliable which is desirable for screening large libraries.
2) Approximate free energy methods. Fewer approximations are made and also account for solvent and protein motions.
3) Relative binding free energy methods. They take into account solvent and protein motions fully, but require some prior knowledge on how the ligand binds to the protein.
4) Absolute binding free energy methods. Are the most physically rigorous methods and therefore requires a lot of computer resources. No prior knowledge of the complex formed between the protein and the ligand is required to predict affinities.

The binding free energy has two components, $\Delta G = \Delta H - T\Delta S$, where H and S are enthalpy and entropy, respectively. The binding affinity (dissociation constant – $1/K_a$) is related to the binding free energy by equation 2.9 below:

$$\Delta G^{\circ} = -RT \ln \left(C^{\circ} \frac{1}{K_a}\right) \tag{2.9}$$

Where:

R – gas constant

T – absolute temperature

C° - the standard concentration (1M)

Docking techniques usually start with a validated structure of the protein and a known ligand structure and then aims to generate structural information from the bound conformation. Many methodologies typically treat the protein as a semi-rigid structure, allowing for some motions of the active site residues in the protein (requires more resources - time and computer power), while the ligands are typically allowed to sample different conformations. Docking is usually achieved through two interrelated steps: 1) by exploring multiple ligand conformations and orientations, and in some cases even different potential binding sites, 2) by ranking these conformations based on their free energy of binding via a scoring function that is normally a physical or knowledge-based potential. The scoring functions are usually simplified with a simple solvent model such as a distance-dependent dielectric for the sake of speed. These approximations affect the predictions since, for instance, they do not capture the loss of interactions between ligand polar groups with the solvent upon ligand binding.[111] Some energy models in docking techniques neglect configuration entropy whereas others account for it based on the number of rotatable bonds in a ligand. In addition, any restriction of the protein residues and ligand's motion when bound will lead to a penalty in the binding affinity due to neglect of

configuration entropy.[112] In the best case scenario, the sampling algorithm should reproduce experimental binding mode of known ligand-protein complexes and among the generated conformations, the scoring function should rank it highest.[106,113]

An example of a scoring function is AutoDock's semiempirical free energy force field.[114] It estimates the free energy of binding using six pairwise terms (*V*) (equation 3.0) to approximate the ligand-protein interaction and an empirical method for solvent contribution.

$$\Delta G = [V_{\text{bound}} - V_{\text{unbound}}]^{L-L} + [V_{\text{bound}} - V_{\text{unbound}}]^{P-P} + [(V_{\text{bound}} - V_{\text{unbound}})^{P-L} + \Delta S_{\text{conf}}] \quad (3.0)$$

Where *P* and *L* are the protein and ligand respectively.

The aim of the empirical technique is to incorporate the enthalpic and entropic contributions (ΔS_{conf}) in easily evaluated terms. The entropic contribution includes all the torsional degrees of freedom and is taken to be directly related to the number of rotatable bonds in the ligand. Each of the pair-wise energetic terms includes evaluations for dispersion/repulsion (first term), H-bonding (second term), electrostatics (third term) and desolvation potential (fourth term) (equation 3.1).[114]

$$V = W_{vdw} \sum_{i,j} \left[\frac{\mathrm{A_{ij}}}{\mathrm{r_{ij}}^{12}} - \frac{\mathrm{B_{ij}}}{\mathrm{r_{ij}}^{6}} \right] + W_{hbond} \sum_{i,j} E(t) \left[\frac{C_{ij}}{{r_{ij}}^{12}} - \frac{D_{ij}}{{r_{ij}}^{10}} \right] + W_{elec} \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} + W_{sol} \sum_{i,j} (S_i V_j + S_j V_i) e^{(-{r_{ij}}^2/2\sigma^2)} \quad (3.1)$$

To calibrate the empirical free energy to match experimental values, the weighing constants *W* are usually optimized. The parameters *A* and *B* for the dispersion/repulsion term are adopted from the AMBER force field whereas the parameters *C* and *D* are allocated to give a potential well depth of 5 kcal/mol at 1.9 Å for O-H and N-H, and 1 kcal/ mol at 2.5 Å for S-H. Directionality of the H-bond is heightened by limiting the number of bonds available to each

point in the grid to the definite number of hydrogen bonds. The desolvation potential is based on the volume of atoms (*V*) that shelter a given atom from solvent weighed by a solvation parameter (*S*) and an exponential term that is based on the distance with a weighting factor σ that is set to 3.5 Å.[114]

2.3.2 Overview of AutoDock, AutoDock Vina and Surflex Dock

AutoDock: AutoDock calculations are performed in several steps: 1) preparation of coordinate files using AutoDockTools which include defining the required charges and atom types in the input PDB structure. 2) Pre-calculation of atomic affinities using AutoGrid. Autodock uses a grid-based method that allows rapid evaluation of the binding energy of trial conformations. 3) Docking of ligands using AutoDock, and 4) analysis of results using AutoDockTools.

The main method for conformational searching is the Lamarckian genetic algorithm whereas a semi-empirical free energy force field is used to predict the free energies of binding of small-molecule ligands to target protein structures.[107]

AutoDock Vina: AutoDock Vina was developed in the same laboratory as AutoDock. It is faster than AutoDock by two orders of magnitude and has parallelism implemented. Vina considerably improves the accuracy of the binding mode predictions relative to AutoDock.[115] AutoDockTools is used in the preparation of input files and analysis of results. The Iterated Local Search global optimizer is used for conformational searching whereas a knowledge based potential with a conformation-dependent part of the scoring function referred to as “Vina” is used for ranking the orientations.

Surflex-Dock: Surflex utilizes an idealized active site ligand (referred to as a protomol) as a target to produce conformations of molecules or molecular fragments. It combines the

Hammerhead scoring function with a molecular similarity method (morphological similarity) in generating and ranking the conformations.[108,116]

Chapter 3: Mechanism and Product Specificity of PRMT1: A Combined Simulation and Experimental Study

3.1 Introduction

When the structure of PRMT5 (forms SDMA exclusively) was solved, a conserved phenylalanine, Phe-379, was found at the corresponding location in the active site as Met-48 in rat PRMT1 (Fig. 3.1).[14]

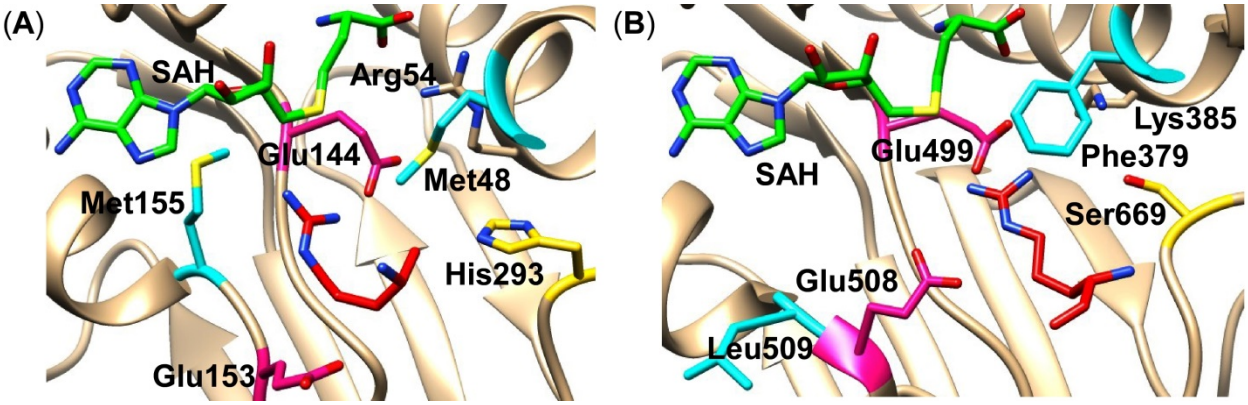


Figure 3.1 (A) The PRMT1 active site bound with S-adenosyl-L-homocysteine (SAH) and the substrate Arg as observed in the crystal structure (Protein Data Bank (PDB) ID 1OR8).[15] (B) The PRMT5 active site bound with S-adenosyl-L-homocysteine (PDB ID 3UA3)[14] with the substrate Arg modeled into the active site. The position of the reacting Arg was approximated by overlaying with the 4GQB PRMT5[117] crystal structure that contained the Arg substrate and an analog of S-adenosyl-L-homocysteine. The phenylalanine is critical for PRMT5 to specify symmetric addition of the second methyl group as changing it to a methionine converts PRMT5 to an enzyme that catalyzes both symmetric and asymmetric dimethylation of arginine.

Mutation of Met-48 to phenylalanine in rat PRMT1 enables the remodeled PRMT1 to generate both ADMA and SDMA, although the amount of SDMA was relatively limited[5]. The low amount of SDMA formed in PRMT1-M48F prompted us to evaluate the free energies of activation, $\Delta G^{\ddagger}$, of each methyl group transfer step in MMA, ADMA, and SDMA formation using gas-phase quantum mechanical (QM) calculations in order to establish the energetic barriers of the enzyme-free reactions. Our study revealed unique energetic challenges for SDMA-forming methyltransferases and highlights the exquisite control of product formation by active site residues in the PRMTs.

Additionally, we investigated the mechanism and product specificity of type 1 methyl transferases by employing accelerated-Molecular Dynamics (aMD)[10] simulations of rat PRMT1 protein. The following are the simulated complexes:

Wild type protein

1. with SAM and the arginine substrate (WT-Arg) (Fig. 3.2).
2. with SAM and the monomethylated arginine substrate methylated on the Nη1 nitrogen (WT-MMANη1).
3. with SAM and the monomethylated arginine substrate methylated on the Nη2 nitrogen (WT-MMANη2).

Mutants

4. M48F with SAM and the arginine substrate (M48F-Arg).
5. M48F with SAM and the monomethylated arginine substrate methylated on the Nη1 nitrogen (M48F-MMANη1).

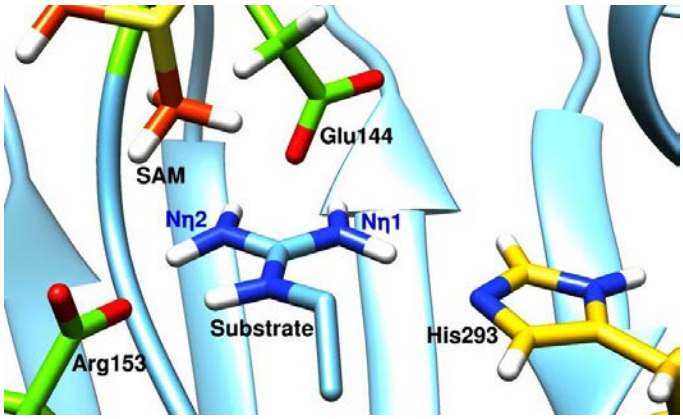


Figure 3.2: The positioning of the substrate's guanidino group indicating the labeling of the two nitrogens, Nη1 & Nη2.

6. M48F with SAM and the monomethylated arginine substrate methylated on the Nη2 nitrogen (M48F-MMANη2).

7. H293A with SAM and the arginine substrate (H293A-Arg).

8. H293S-M48F with SAM and the un-methylated arginine substrate (H293S-M48F-Arg).

9. H293S-M48F with SAM and the monomethylated arginine substrate methylated on the Nη1 nitrogen (H293S-M48F-MMANη1).

10. H293S-M48F with SAM and the monomethylated arginine substrate methylated on the Nη2 nitrogen (H293S-M48F-MMANη2).

Molecular dynamics enabled us to shed more light on the mechanism and product specificity from a dynamical point of view.

3.2 Methods

3.2.1 Gas-phase Calculation of the Activation Energy for Methylarginine Formation

All geometries and energies in the present study were computed using density functional theory[118] and the high-level composite method CBS-QB3[119] as implemented in the Gaussian 09

program[120]. For the density functional theory calculations, the B3LYP [121,122] and M06-2X[123] methods with the 6-311++G(2d,p) basis set have been employed. Frequency calculations were used to characterize all structures as minima or first-order saddle points and to provide thermodynamic corrections.

3.2.2 Enzyme Preparation for Molecular dynamics Simulations

The published PRMT1 crystal structure (PDB ID: 1OR8) has 3 limitations:

(1) No electron density was given for the first N-terminal helix that supplies residues to the active site.

(2) Glu153, one of the active site residues involved directly in catalysis[16], is oriented away from the active site, which may be due to the crystallization conditions (pH 4.7)[15]. Under these conditions, Glu153 and the substrate would both be protonated (Glu153's pKa is ~6.2)[16] and lead to poor electrostatic interactions.

(3) Apart from the arginine that undergoes methylation, the rest of the substrate's side chain densities were not sufficiently resolved to allow identification of the amino acids (the R3 peptide - RGG repeat sequence).

A model of PRMT1 that includes the missing portion of the N-terminal helix and a re-modeled Glu153 region was therefore created through homology modeling. A standard Protein BLAST was performed using Position specific iteration-basic local alignment search tool (PSI-BLAST)[124] against the PDB[95] database using the default general algorithm parameters. Only the PRMT3 structure (PDB ID: 1F3L), with a sequence identity of 51%, has the extra N-terminal residues and was thus used as the template to append 15 residues (Ala26 – Ala40) to the PRMT1 structure. The comparative modeling program MODELLER 9.10[91] was used.

A sequence alignment of the original structure (1OR8) and the target sequence was performed and an initial model of PRMT1 with the extra residues constructed. This resulted in a model with a disordered region at the added residues. The rest of the protein (His41 – Arg353) in this initial model was then restrained during optimization where default protocols were adjusted for a thorough optimization and refinement by a combination of conjugate gradients and molecular dynamics using 1F3L as the template to re-model the secondary structure of the region.

In order to re-orient Glu153, residues Cys149 to Leu160 were manually sliced from the 1OR8 model and re-build using 1F3L as the template since the conserved corresponding glutamate in 1F3L was considered appropriately oriented. The same protocol used for constructing the missing N-terminal helix was utilized.

3.2.3 Molecular Dynamics Simulations

All MD simulations were performed using AMBER[70,125] with minor modifications from protocols described in our previous study.[126] The re-modeled structures were prepared using the leap module where the appropriate hydrogen atoms were added. The ternary systems were solvated in an octahedral box of TIP3P[127] water molecules with solvent layers 10 Å from the edge of the protein, and the system neutralized with Na^+ counterions. The AMBER GAFF force field[128] was used to generate the parameters for Adomet while ff99SB force field[129] was used to construct the topology files. Parameters derived by Papamokos et al.[130] were used for the modified amino acid residues, i.e., mono-methylated or di-methylated arginines of the peptide substrate.

Minimization was started with 200 steps of conjugate gradient for the water molecules only, followed by 10,000 steps for the entire system. The system was then gradually heated from

0 to 300 K at NVT for 50 ps using the weak-coupling algorithm with a temperature coupling value of 2.8 ps, then, equilibrated at NPT for 500 ps at a constant pressure of 1 atm using a coupling value of 2.0 ps. Following additional 500 ps of NVT equilibration, 300 ns periodic boundary dynamics production simulations were carried out for the whole system at NVT with a time step of 1fs for each protein complex using the GPU-accelerated version of AMBER 12.[131] The coordinates were saved every 5,000 steps (5 ps) while the energy data was dumped every 10,000 steps (10 ps). The SHAKE[132] algorithm was used to restrict all covalent bonds involving hydrogen atoms with the Particle Mesh Ewald (PME) method[133] employed to treat long-range electrostatic interactions with a non-bonded cutoff of 10 Å.

MD Analysis: Clustering, hydrogen bond analysis, distance calculations, root-mean square fluctuation (rmsf) and root-mean square deviation (rmsd) calculations were computed using the ptraj and cpptraj programs within AmberTools.[134]

3.3 Results and Discussion

3.3.1 Computational Simulation to Understand the Energetics for Different Methylarginine Formation

To better understand why the formation of SDMA is limited, it was questioned whether the formation of SDMA in the absence of protein was energetically more costly as compared with MMA and ADMA. Calculations have been applied for studies with protein lysine methyltransferases (PKMTs) suggesting that the energetics of the methyl transfer reactions, at least in part, may determine the product specificity of protein lysine methyltransferases.[135-138] Recent calculations on PRMT1 predicted a more facile ability to form ADMA as compared with MMA, but SDMA formation was not investigated.[139] New gas-phase calculations were performed here to establish the free energies of activation for MMA, ADMA, and SDMA

production in PRMTs. Modeling entire proteins in enzymatic reactions typically requires extensive computational resources and may not always be necessary to predict the energetics or product formation. For example, truncated methyl transfer reaction models involving dimethylammonium, tetramethylammonium, and trimethylsulfonium to dimethylamine have been successful in furthering the understanding of catechol O-methyl-transferase.[140] As such, in our PRMT1 reaction model (Fig. 3.3), AdoMet was truncated at the Cα of the methionine moiety and at the oxolane ring of the adenosyl moiety. This is believed to appropriately represent the electronic structure of the cofactor because the main role of the positively charged sulfur atom of AdoMet is to attract electron density from the methyl group as the arginine abstracts the methyl group.

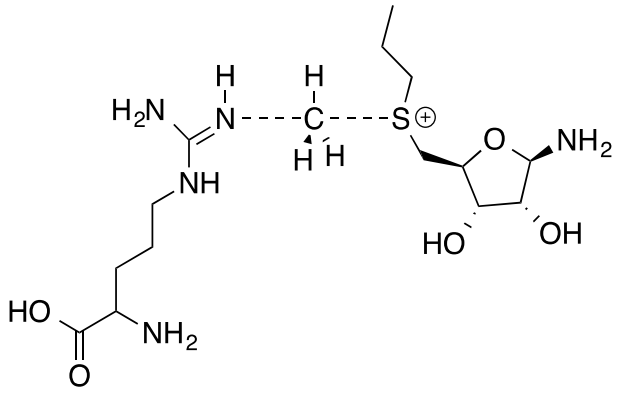


Fig 3.3: MMA-forming transition state model featuring a truncated AdoMet and a nucleophilic arginine.

The calculations were attempted with both protonated and deprotonated arginine in place of the nucleophilic arginine in Fig. 3.3 as experimental solvent isotope effects have indicated that prior deprotonation of the substrate guanidinium may not be required for PRMT1 methyl transfer.[16] However, the energetics of the methyl transfer to the protonated arginine were predicted to be prohibitively large, i.e. $\Delta G^{\ddagger}$ of 110.5, 98.0, and 107.2 for MMA, ADMA, and SDMA production, respectively (Table 3.1).

Table 3.1: Computed gas-phase free energies of activation, $\Delta G^{\ddagger}$, (kcal/mol) for the MMA, ADMA, and SDMA methyl-transfer reactions using a protonated arginine.

	MMA	**ADMA**	**SDMA**
CBS-QB3	110.5	98.0	107.2

The extreme activation barriers reflect the repulsive nature of the positively charged residue coming in close contact with the positively charged sulfur atom on AdoMet. The difference is also reflected in the perturbed transition structure distances (Table 3.2).

Table 3.2: Computed transition structure distances (Å) between (AdoMet)S–C(methyl) and (methyl)C–N(protonated arginine) for the MMA, ADMA, and SDMA methyl-transfer reactions.

	MMA		**ADMA**		**SDMA**	
	S···C	C···N	S···C	C···N	S···C	C···N
CBS-QB3	2.89	1.72	2.72	1.84	2.85	1.72

It is clear that the protein environment is crucial for modulating the charges during the reaction and could be indicative of a concomitant proton transfer during reaction. Thus, we further focused on arginines in a nucleophilic state where the proton had been previously abstracted to have the electronics necessary to perform the nucleophilic attack.

The computed transition structure distances between the S_N2 reacting atoms (AdoMet)S–C(methyl) and (methyl)C–N(arginine) given in Table 3.4 are generally in close agreement between the different methods. CBS-QB3 has been shown to be particularly accurate in the prediction of methyl transfer reactions[140] and is assumed to produce the best gas-phase results for

these reactions among the methods applied. Illustrations of the optimized transition structures for ADMA and SDMA formation are given in Figure 3.4.

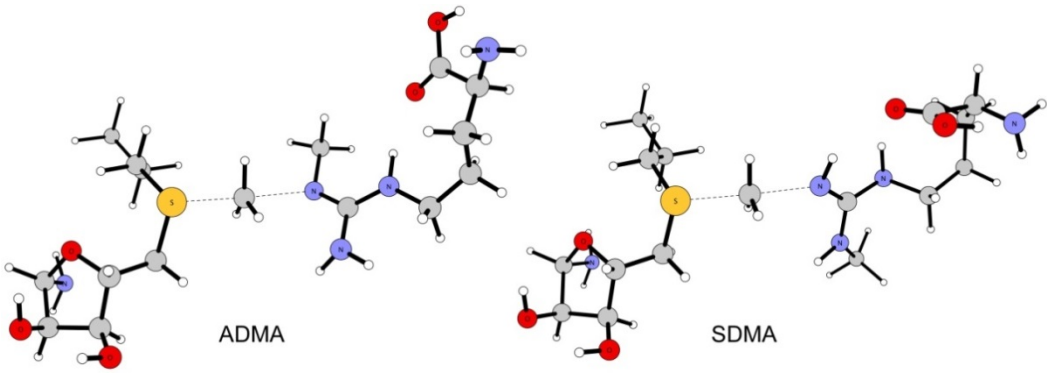


Figure 3.4: Transition structures for the ADMA and SDMA reactions computed using CBS-QB3.

Table 3.3: Computed gas-phase free energies of activation, $\Delta G^{\ddagger}$, (kcal/mol) for the MMA, ADMA, and SDMA methyl-transfer reactions.

	MMA	ADMA	SDMA
CBS-QB3	9.5	10.2	13.4
M06-2X[a]	13.3	10.9	17.0
B3LYP[a]	11.6	13.5	17.2

[a] 6-311++G(2d,p) basis set.

Comparison of the computed $\Delta G^{\ddagger}$ between MMA and ADMA formation finds that the preferred product differs based on the method used, e.g. CBS-QB3 and B3LYP favor MMA over ADMA in contrast to M06-2X (Table 3.3).

Comparison with recent simulations by Zhang et al.[139] that carried out mixed quantum and molecular mechanical (QM/ MM) ONIOM calculations on the entire protein utilizing B3LYP/6-31G(d) on a protonated arginine, part of AdoMet, and residues Arg-54, Glu-144, and Glu-153 found nearly identical energies and geometries as the current gas-phase work. For example, the QM/MM B3LYP/6-31G(d) method and the current gas-phase B3LYP/6-

311G++(2d,p) method yielded energy barriers of 11.76 and 11.6 kcal/mol, respectively. The QM/MM calculation of ADMA production gave a nearly identical activation energy (11.63 kcal/mol) to MMA production; however, an MP2/6-31G(d) single point energy calculation significantly increased the difference in the energy barriers of MMA and ADMA, i.e. 19.08 and 14.94 kcal/mol, respectively. This strongly suggests a more facile formation of ADMA. The large energetic preference for MMA of ~4 kcal/mol is unrealistic as this value would suggest an increase of several orders of magnitude in the rate constant for ADMA, whereas experimental k_{cat} measurements yielded more muted values of 0.39 min^{-1} as compared with 0.79 min^{-1} for the unmodified and monomethylated substrate, respectively.[141] The current M06-2X and B3LYP methods also overestimate the energy differences. This highlights that higher theory levels such as CBS-QB3, despite their large computational costs, may be required to properly reproduce experimental rates. The computed free energies of activation, $\Delta G^{\ddagger}$, predicted a lower barrier for the ADMA reaction relative to SDMA at all theory levels (Table 3.3).

Table 3.4: Computed transition structure distances (Å) between (AdoMet)S–C(methyl) and (methyl)C–N(arginine) for the MMA, ADMA, and SDMA methyl-transfer reactions.

	MMA		**ADMA**		**SDMA**	
	S···C	C···N	S···C	C···N	S···C	C···N
CBS-QB3	2.29	2.14	2.26	2.17	2.27	2.15
M06-2X[a]	2.24	2.07	2.22	2.08	2.23	2.08
B3LYP[a]	2.30	2.13	2.27	2.16	2.29	2.14

The CBS-QB3 $\Delta\Delta G^{\ddagger}$ of 3.2 kcal/mol favoring ADMA formation is a significant value and would represent several orders of magnitude difference in the enzymatic rate constant. The energy difference is in line with the exclusive formation of ADMA observed experimentally for WT-PRMT1. The higher $\Delta\Delta G^{\ddagger}$ for SDMA can be rationalized when considering that methyl

groups are electron-donating, making the unmethylated guanidino nitrogen on the SDMA-forming arginine a weaker nucleophile relative to the adjacent ADMA forming methylated nitrogen atom. However, given the limitations of a gas-phase environment, it is important to expand the model to include the full protein complex.

3.3.2 Analysis of the Modeled Molecular Dynamics Starting Structure

The quality of the obtained structure (Fig. 3.5) was extensively evaluated. For example, the stereochemistry of the model was analyzed with Procheck[142] whereas other qualities of the modeled structure were evaluated using ProSA (Protein Structure Analysis)[102]. A Ramachandran's plot (backbone dihedral angles ψ vs φ of amino acid residues in a protein structure) showed that of the 15 inserted PRMT1 residues, 13 of them fell within the most favored region, 2 in the additionally allowed and generously allowed regions, whereas none is in the disallowed area (Fig. 3.6). The PROSA curves represent the residue interaction energies with negative values corresponding to stable parts of the molecules. The inserted region (26-40) has negative values suggesting reliability of the model (Fig. 3.7). Computational modeling of missing residues is a common practice and generally provides accurate results if care is taken.[143]

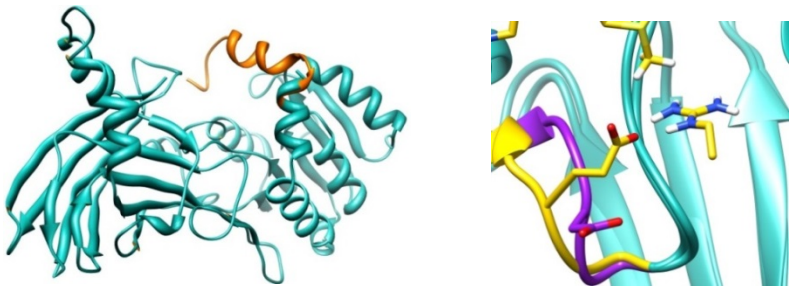


Figure 3.5 (A) PRMT1 structure with the appended N-terminal shown in orange. (B) The original Glu153 orientation is shown in purple while the re-modeled Glu153 region is shown in yellow.

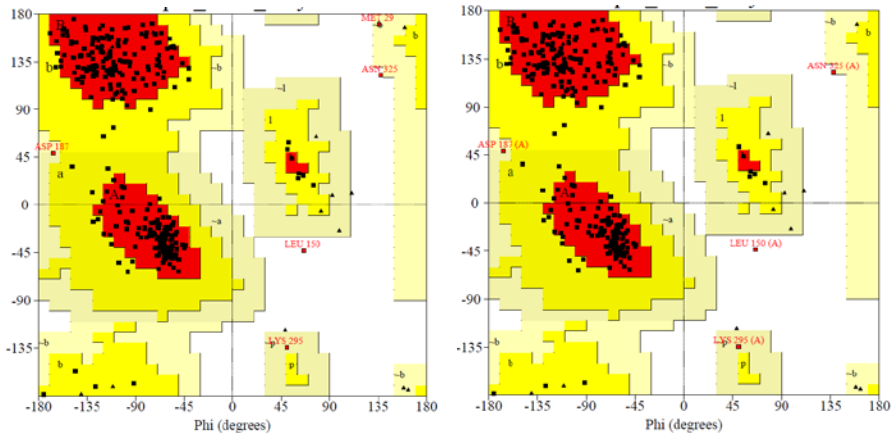


Figure 3.6 (A) Ramachandran Plot of the model (B) Ramachandran Plot for 1RO8

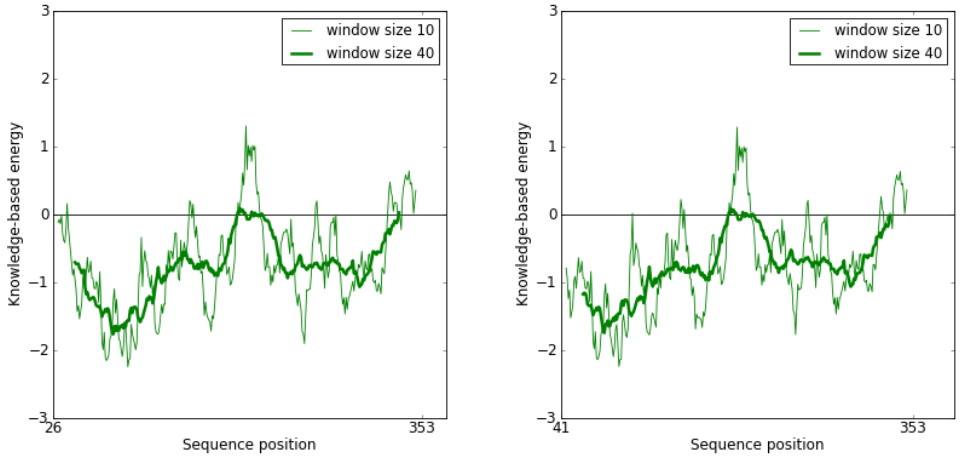


Figure 3.7 (A) ProSA energy plot for the model (B) ProSA energy plot for 1OR8

In order to extend the substrate in the active site, the software program BOMB[144] was used to add Gly residues on either side of the crystallographically resolved Arg. The peptide substrate Gly-Gly-Arg-Gly was grown into the active site region of the Adomet bound PRMT1.

3.3.3 Overall Molecular Dynamics Analysis

To evaluate the connection between the structural changes that occur within the active site of PRMT1 bound to the cofactor and the arginine substrate in the respective methylated states, a detailed analysis of the MD simulation was performed. The root-mean-square deviation

(RMSD) and the root-mean-square fluctuations (RMSF) of the trajectories were calculated in order to monitor the structural stability and global conformational changes of the production runs. Fig. 3.8 shows the RMSD of the backbone protein atoms with average values of ~3.5 Å over 300ns for all complexes confirming good stability.

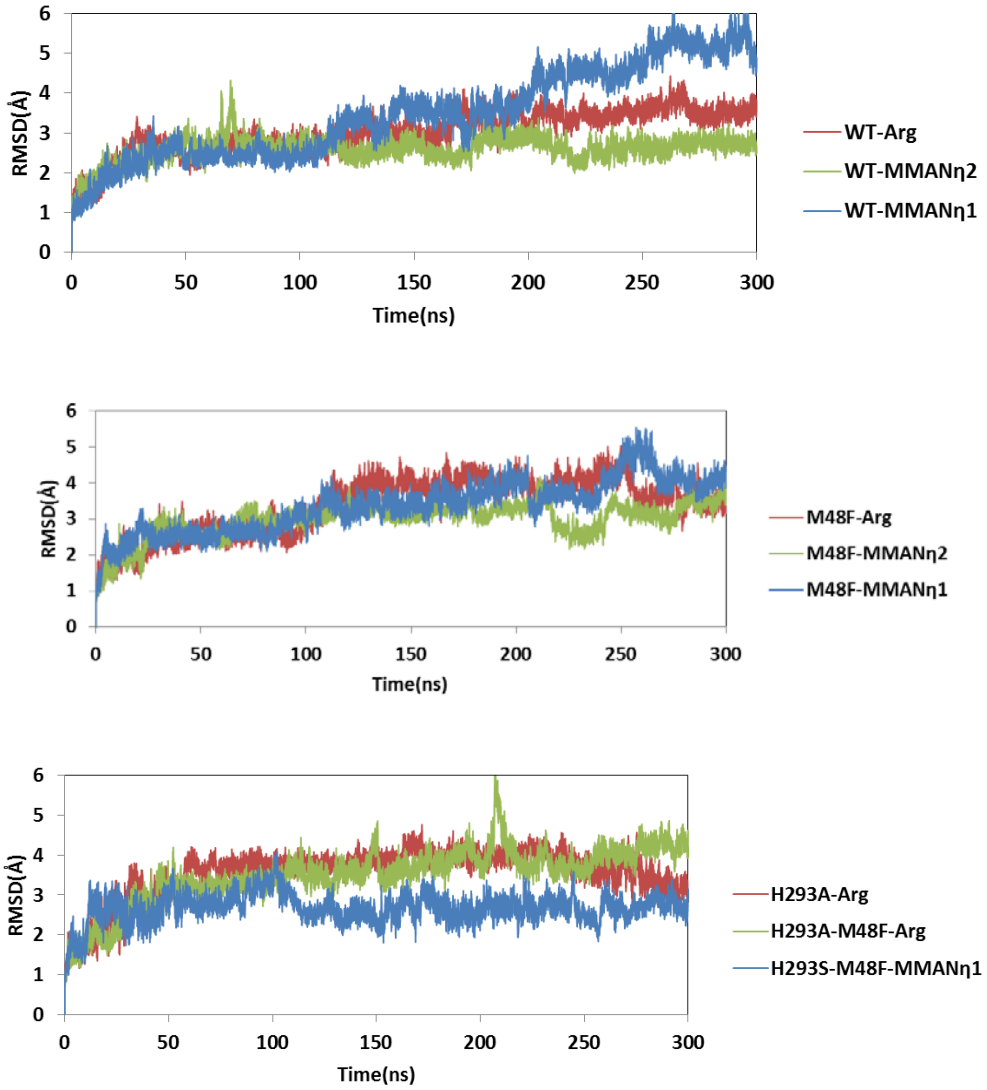


Figure 3.8: RMSD of the protein backbone atoms (N, Cα and C) for the respective simulations relative to first frame.

Fig. 3.9 shows the positional fluctuations of the backbone atoms for all the residues with the dimerization arm being very flexible. This encompasses the residues Arg188 to Lys215 and

is made of three helices linked by a loop: αhelix 188-194, 3/10 helix 196-199 and α-helix 210-214. The region undergoes drastic conformational changes during the simulations of between 10-12Å and it appears Asp187 and Glu216 acts as hinges. The enzyme is active as a dimer and based on a careful examination of the dimeric unit, the dimerization arm wouldn't be expected to fluctuate this drastically.

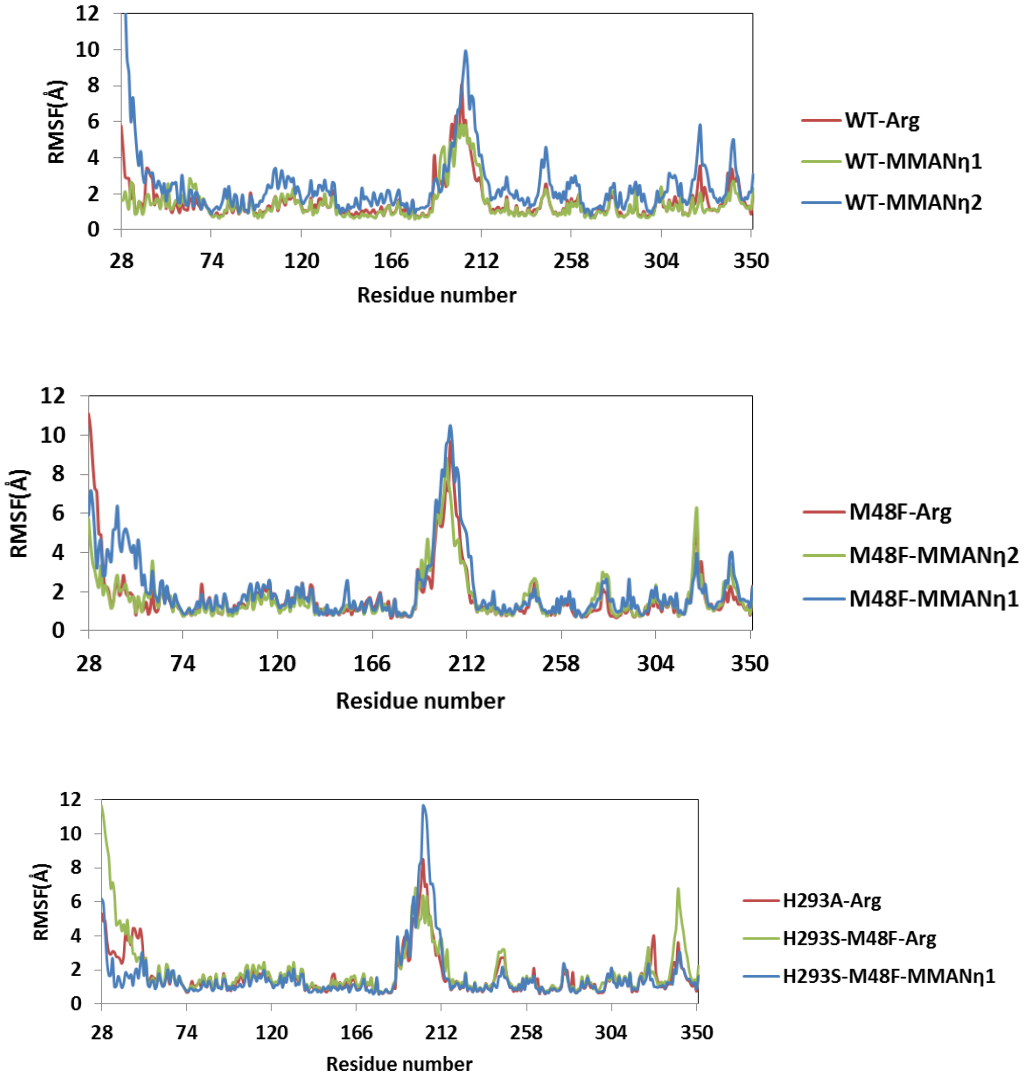


Figure 3.9: RMSF by residue of the C, Cα and N atoms relative to the average structure.

On the other hand, considering the substrate undergoes a two-step reaction, these fluctuations might suggest that the motion of this region has significant impact on substrate binding and/or

product release. The hydrophobic interactions between the dimerization arm of one subunit and the SAM binding site of the other subunit may play a large role.

The calculated fluctuations of the SAM binding site residues (28-145) show that these residues are modestly flexible (Fig. 3.9). Among these residues, 26-50 is the N-terminus region whereas the others are part of the core binding site. The first two helices (43-63) of the N-terminus lie at the entrance of the binding site and they fluctuate appreciably as expected since they should open and close for substrate binding and product release. In fact, this region's flexibility is evidenced by the lack of clear electron density in the co-crystal structure by Zhang et al.[15]

3.3.4 Intermolecular Interactions and the Chemical Step

SAM interactions: Key interactions between SAM and active site residues involve long lasting H-bonding, e.g., with Arg54, Thr81, and Glu100 have been monitored over the course of the MD simulation.

Table 3.5: H-bonds formed between SAM and active site residues.

		Complexes						
		WT-Arg	WT-MMANη2	M48F-Arg	M48F-MMANη2	H293A-Arg	H293S-M48F	H293S-M48F-MMANη1
H-acceptor	H-donor	%occupancy						
GLU100@OE1+OE2	SAM@H14	99.4	99.5	99.5	99.0	99.4	98.9	99.1
GLU100@OE1+OE2	SAM@H15	99.9	99.9	99.9	99.9	99.9	99.9	99.9
SAM@OXT+O	THR81@HG1	79.1	93.2	90.3	71.4	67.3	87.1	97.0
SAM@O+OXT	ARG54@HH22	76	49.8	87.5	45.4	60.2	18.4	32.1
SAM@O+OXT	ARG54@HH12	36.2	65.6	33.3	39.4	47.1	45.5	81.1
GLU129@OE1	SAM@H16	4.9	0	3.9	13.2	14.8	1.3	13.8
GLU129@OE2	SAM@H17	8.2	0	2.8	14.6	10.0	1.4	17.4

The carboxylate group of Glu100 forms H-bonds with the hydroxyl groups of the ribose moiety of SAM and these constitutes very high occupancy of above 98% for all the simulations (Table 3.5). Since the ribose ring is at the middle of SAM, the residue's role is seemingly to anchor the cofactor in position (Fig. 3.10). SAM's methionine moiety is held in position by H-bonds with Thr81 and Arg54 for significant part of the simulation time for all the systems whereas the adenosyl part appears to be positioned by non-electrostatic interactions as only Glu129 forms any appreciable H-bonds and for very short amount of time, ~10% occupancy across board.

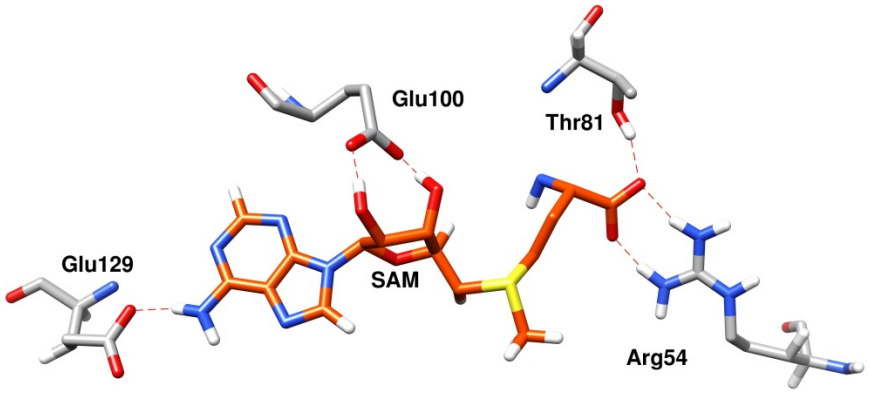
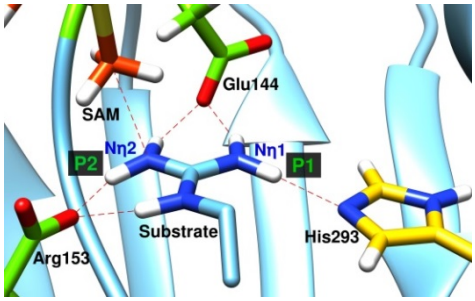


Figure 3.10: Key residues that H-bond with SAM.

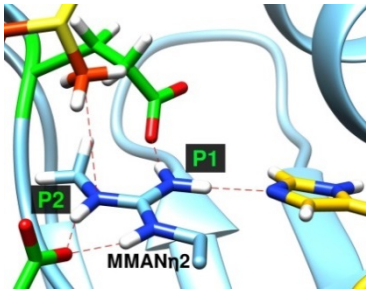
Other important interactions are mainly hydrophobic with residues Met155, Met48 and the side chain rings of Tyr39 and Phe36. However, residues 155 and 48 have been shown to be particularly important for SAM binding[15], catalysis[5] and product specificity[5,13]. The simulations find that the side chain of Met155 forms a key electrostatic interaction with the purine ring that helps in positioning of the cofactor.

Substrate interactions: It is well documented that enzymes can enhance the rate of reactions by lowering the activation barrier through transition state stabilization.[145] This stabilization is achieved by enzyme-substrate intermolecular interactions to yield high affinity complexes. For the WT complex, the percentage occupancy of the H-bonds formed by both

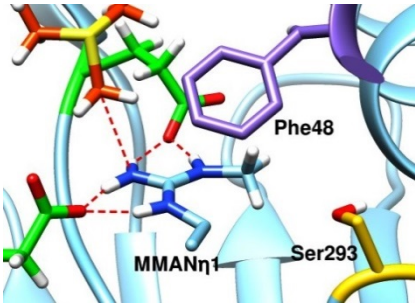
oxygens of the carboxylates of Glu144 (94.5 and 53.2%), and Glu153 (92.2 and 62.6%), the imidazole of His293 (60.4%) and the substrate are consistently high (Table 3.6).



(A) WT-Arg



(B) WT-MMANη2



(C) H293S-M48F-MMANη1

Figure 3.11: Key residues that H-bond with the substrate. P1 and P2 are positions 1 (Nη1) and 2 (Nη2) respectively. Only the reacting part of the substrate is shown. The rest is omitted for clarity.

Table 3.6: H-bonds formed between substrate and active site residues.

		Complexes						
		WT-Arg	**WT-MMANη2**	**M48F-Arg**	**M48F-MMANη2**	**H293A-Arg**	**H293S-M48F-Arg**	**H293S-M48F-MMANη1**
H-acceptor	**H-donor**	**%occupancy**						
GLU144@OE1+OE2	ARG@HH12	94.5	55.1	94.3	17.1	74.5	16.3	85.62
GLU144@OE1+OE2	ARG@HH22	53.2	N/A	74.1	N/A	70.8	22.1	68.1
GLU153@OE1+OE2	ARG@HE	92.2	83.8	65.4	68.3	41.7	20.3	27.9
GLU153@OE1+OE2	ARG@HH21	62.6	81.6	25.1	85.5	47.0	20.4	21.2
HIS293@NE2	ARG@HH11	60.4	49.0	61.0	25.2	N/A	N/A	N/A

These key electrostatic interactions (Fig. 3.11A) are critical in regard to the positioning of the substrate's guanidinium and in increasing the nucleophilicity of the guanidino moiety for the methylation reaction, an otherwise poor nucleophile.

M48F and H293A single mutations in M48F-Arg or H293A-Arg systems respectively don't have a dramatic impact on the H-bonds occupancies. However, double mutation in the H293S-M48F-Arg system significantly reduces the substrate's ability to effectively interact with both Glu144 and Glu153. This implies Met48 and His293 play a synergistic role in substrate positioning for PRMT1 and therefore the double mutant could be severely impaired.

The introduction of a methyl group on position P2 (Nη2) (Fig. 3.11B) reduces the interaction with Glu144 but doesn't severe it (WT-MMANη2 and M48F-MMANη2). This coupled with the presence of a hydrophobic methyl group in a polar active site seems to be the reason why MMA containing peptides are poor substrates for the enzyme.[146,147]

Experimental evidence indicated His293 is important for catalysis with its main role being unclear. This was concluded from H293Q and H293A mutants having effects on both the kinetic parameters for SAM and the peptide substrate with evidence refuting its acting as a general base.[16]

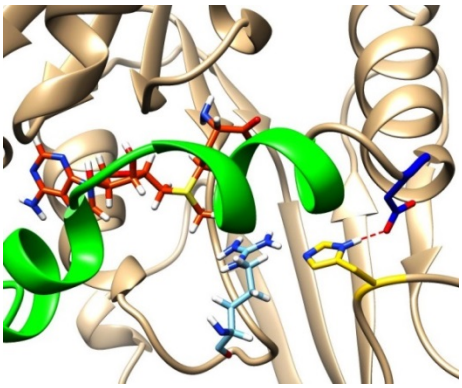


Figure 3.12: The salt bridge between His293 (yellow) and Asp51 (blue) believed to be structurally important. The N-terminal helices shown in green.

His293's main role was therefore hypothesized to be the formation of a salt bridge with a conserved residue Asp51, thereby playing a critical part in forming the two-helix boundary that separates the SAM and peptide binding portions of the active site (Fig. 3.12).

In all the simulations, this salt bridge is ubiquitous implying justification of this hypothesis (Table 3.7). However, the integrity of the active site in the H293A system remains intact during the 300 ns simulation, perhaps suggesting H293's involvement in modulating the nucleophilicity of the substrate, and therefore rate enhancement, cannot be undervalued.

H-bond network between Arg54-SAM-Glu144, Tyr39-Glu153, Asp51-His293 and the Substrate: The simulations have confirmed the intermolecular interactions between Arg54, SAM, Glu144 and the substrate are very important in stabilizing the ternary complex for the methylation. Arg54 makes key H-bonds with both SAM and Glu144, one of the crucial residues in catalysis (Fig. 3.13). The interaction fixes SAM and the glutamate in position for the reaction (Table 3.5 & 3.7).

A similar interaction involving Tyr39 and Glu153 is ever-present with occupancies above 90% especially in the WT systems (Fig. 3.13 & Table 3.7). The importance of this interaction was experimentally highlighted by the enzyme's pH profile where the protonation and deprotonation of Tyr39 and Glu153 must be achieved for optimal PRMT1 activity.[16] The simulations show that Arg54, Tyr39 and Asp51 play a significant role in forming electrostatic interactions with Glu144, Glu153 and His293 respectively, orienting them and influencing their electronic arrangement for effective modulation of the substrate's nucleophilicity for catalysis.

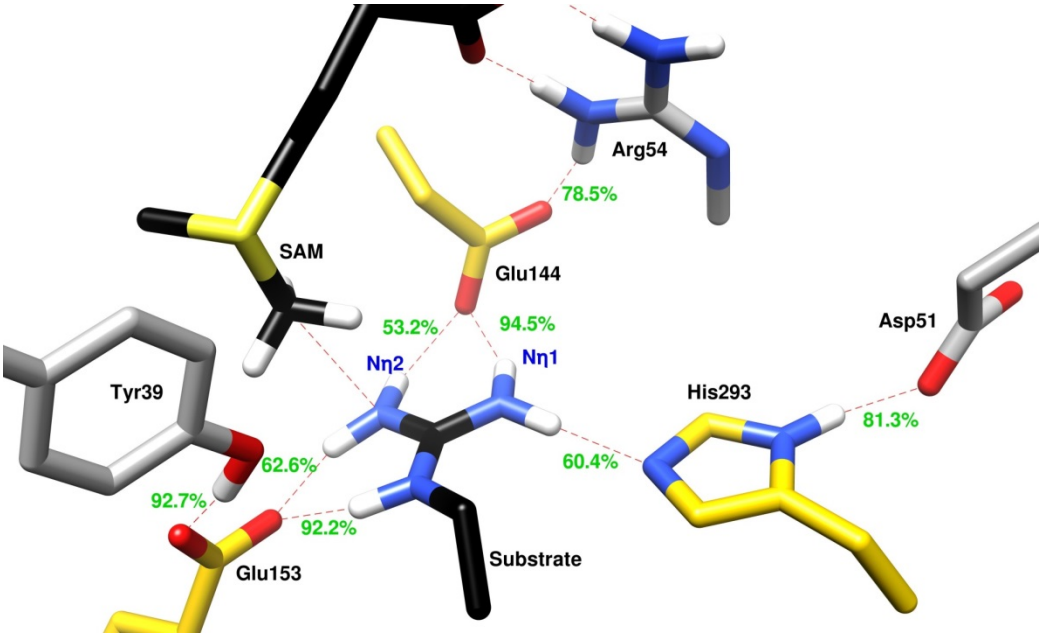


Figure 3.13: H-bond network involving Arg54-SAM-Glu144, Tyr39-Glu153, Asp51-His293 and the Substrate.

Table 3.7: H-bond network in the active site.

		Complexes						
		WT-Arg	**WT-MMA-Nη2**	**M48F-Arg**	**M48F-MMA-Nη2**	**H293A-Arg**	**H293S-M48F-Arg**	**H293S-M48F-MMA-Nη1**
H-acceptor	**H-donor**	**%occupancy**						
Glu144@OE1+OE2	Arg54@HH11	78.5	75.5	91.7	83.8	65.7	68.0	95.0
Glu153@OE1+OE2	Tyr39@HH	92.7	97.8	90.8	76.7	70.8	22.0	95.2
ASP51@OD1+OD2	HIS293@HD1	81.3	87.1	84.3	83.5	N/A	N/A	N/A

3.3.5 Proposed Mechanism

Different methyltransferases have diverse structural requirements for the way they bind the substrate and cofactor but they all share the same basic mechanism. The positively charged sulfur atom on SAM causes the S-CH_3 bond to be polarized. A base in the enzyme's active site

abstracts a proton from the arginine substrate making it more nucleophilic. SAM then attracts the electron density from the methyl group as the nucleophile attacks the methyl group.[148]

His293: Based on evaluation of the crystal structure[15], Rust et al.[16] concluded that His293 is >6 Å from the approximate position of the substrate guanidinium, a distance too far to abstract the proton or directly influence catalysis. However, in the PDB structure the cofactor is in the product state. The manual addition of the methyl group to convert SAH to SAM together with the re-orientation of Glu153 required the substrate to re-position within the active site. Minimization of all solvated ternary systems before MD simulations reduce the distance between the guanidinium nitrogen and His293, enabling the realization of the H-bond (Fig. 3.11A, 3.11B). This is persistent in the WT simulations (Table 3.6) suggesting its absence in H293A should have a significant effect on the nucleophilicity of the substrate, potentially explaining the experimental decrease on the catalytic potency of this mutant.[16]

Glu153: The previously reported mechanism by Rust et al.[16] involve the formation of a dication intermediate with Glu144 responsible for the abstraction of the proton. They found Glu144's H-bonding ability to both the substrate and Arg54 is far more important than its charge whereas for Glu153, both the charge and position are essential for catalysis. This fact downplays Glu144 as the general base and points to Glu153. Additionally, the pH profile of the enzyme was attributed to the deprotonation and protonation of Glu153 and Tyr39[16] further supporting this view.

From the current simulations, Glu144 only abstracts the proton during the second methylation step if MMA binds with the methyl group on Nη1. This would result to the loss of the interaction between the substrate and His293. Alternatively, with MMA's methyl group on Nη2, the hydrogen that ought to be abstracted in ADMA formation is H-bonded to Glu153 and

not Glu144 (Fig. 3.11B). MMA bound this way does not disrupt interactions between the substrate and active site residues. Based on this and the persistent Tyr39-Glu153-substrate H-bond network, it is possible that Glu153 behaves as the general base. The dication formed during both MMA and ADMA formation has three active site residues (Glu144, Glu153 and His293) to stabilize the charges (Fig. 3.14).

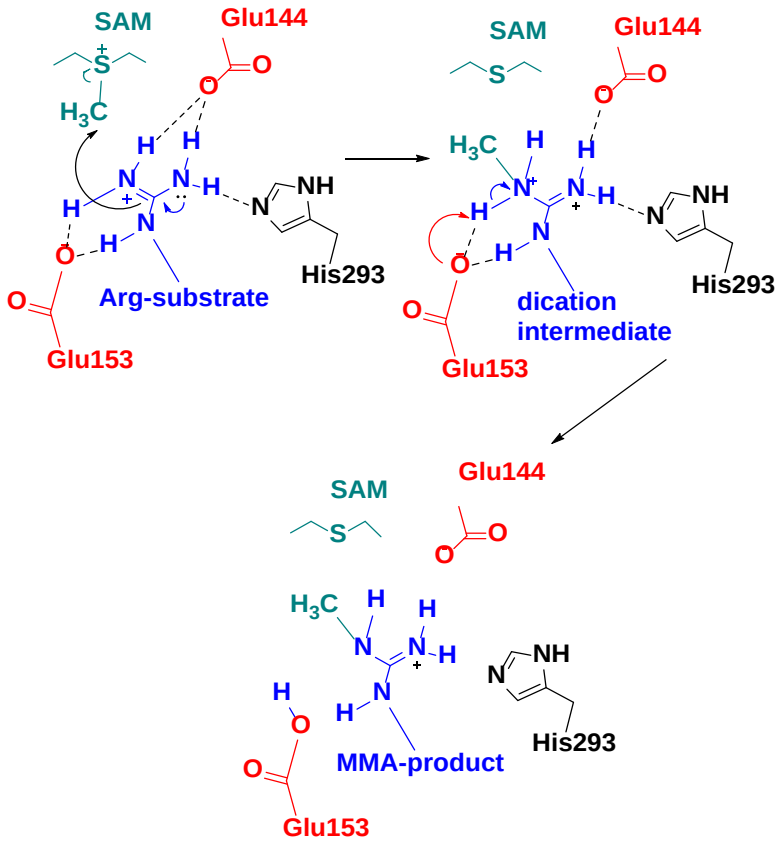


Figure 3.14: Alternative catalytic mechanism of PRMT1. His293, Glu144 and Glu153 H-bond with the guanidinium nitrogen to activate arginine for the nucleophilic S_N2 reaction.

3.3.6 Substrate Binding and Origin of Product Specificity

A potential way for controlling product specificity in PRMTs is to control the binding of MMA by active site residues for the methylation that leads to either ADMA or SDMA. To probe why PRMT1 exclusively makes ADMA, the preferred orientations by the substrate in the various

ternary systems were evaluated by comparing the distributions of d + 0.5(cosθ) over time for all the trajectories (Fig. 3.15, 3.18, 3.19 and 3.20). d is the attack distance between arginine's nitrogens (Nη1 and Nη2) and SAM's methyl carbon while θ is the respective attack angle N···C···S.. The angle (θ) is in radians and therefore, the most ideal attack angle (180˚ = π radians) should have a value of -1 (cos π) whereas the least ideal (90˚) have a value of 0. The d value is thus scaled downwards by a favorable angle (90-180) and upwards if the angle is unfavorable (between 0 - 90).

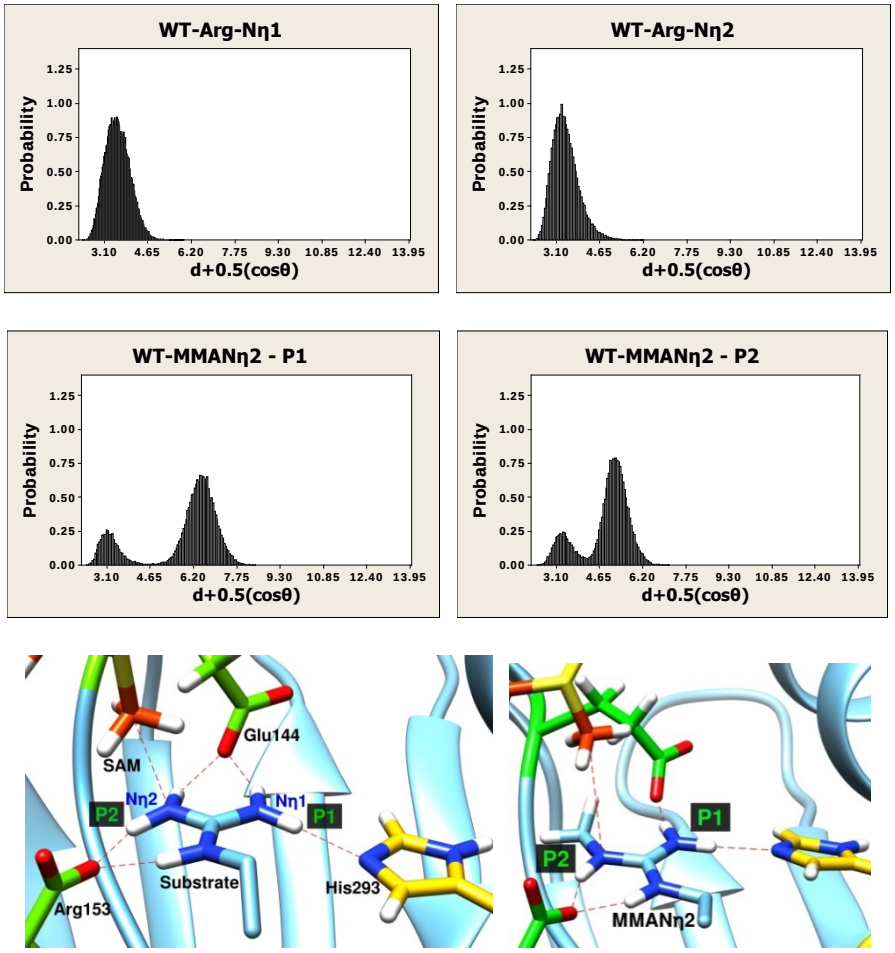


Figure 3.15 Probability distributions of d + 0.5(cosθ) calculated from the MD trajectories. The WT-MMANη2 - P1 distribution has MMA methylated on Nη2 and the second methylation

occurring on position 1 (P1) leading to SDMA. WT-MMANη2 - P2 has MMA methylated on Nη2 and the second methylation occurs on P2 leading to ADMA.

The average distances were evaluated to quantify the proximity of the arginine substrate's guanidino moiety (Nη1 and Nη2) to the methyl group on SAM (Fig. 3.16).

WT complexes: The WT protein bound with MMA substrate (WT-MMANη2) has a wider spread and shifts to larger values in distribution compared to the WT-Arg (Fig. 3.15). This would be expected since MMA substrates bind poorer to PRMT1, impeding a consistent formation of near-attack conformers. The distribution in WT-MMANη2 - P1 has shifts to slightly larger values implying the expected fact that WT-MMANη2's second methylation should be on position 2 (P2) leading to ADMA.

In the WT-Arg complex, the nitrogen – carbon attack distance between the substrate's nitrogen and SAM's carbon averaged 3.8 and 4.0 Å for P2 and P1 (Fig. 3.16) with the reactants forming productive near-attack conformations.

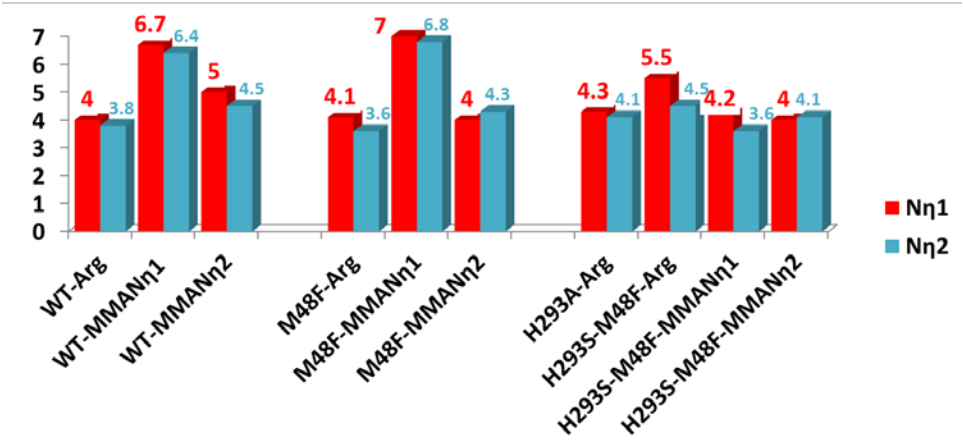


Figure 3.16: The average near-attack distance between the nucleophiles nitrogens (Nη1 or Nη2) and the carbon of SAM's methyl group.

For WT-MMANη2, the average was 4.5 and 5.0 respectively and the substrate is in an orientation suitable for methylation. Through electrostatics, Glu153 and His293 confine the substrate in position for the reaction (Fig. 3.17).

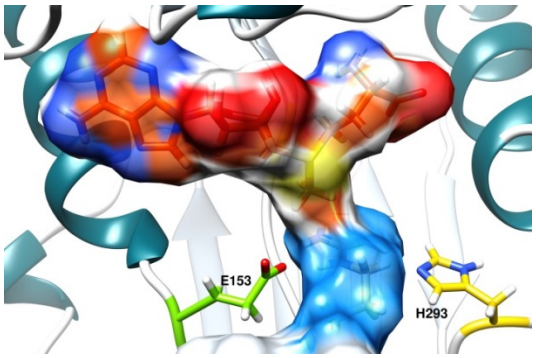


Figure 3.17: Glu153 and His293 help localize the substrate for methylation using H-bonds.

However, for WT-MMANη1, the distributions for either the P1 or P2 show widespread tabulated frequencies, highlighting poor substrate binding when MMA is methylated on the Nη1 nitrogen (Fig. 3.18). Additionally, the average distances are 6.4 and 6.7 for Nη2 and Nη1 respectively (Fig. 3.16).

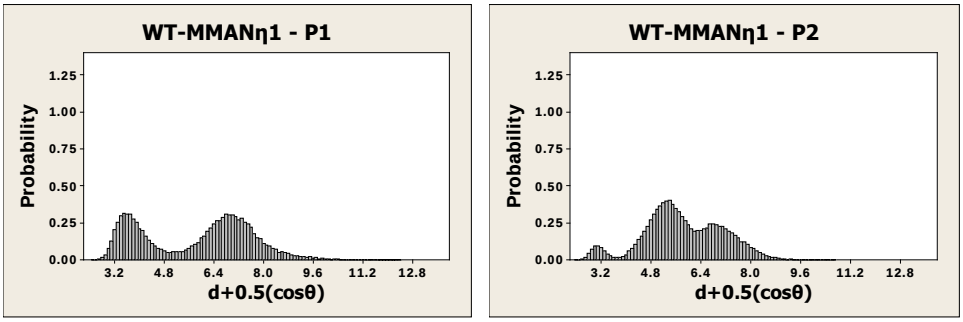


Figure 3.18 WT-complex (WT-MMANη1).

A plot of the attack distances of the 60,000 frames over the 300ns trajectory (not shown) indicate that the guanidino moiety veer away to much longer values (10-15Å) from SAM's CH_3 group with the first 2 helices of the N-terminal eventually opening up the active site. Over time, the substrate gets improperly placed, a confirmation that this orientation is very unfavorable.

This can be explained since MMA methylated on Nη1 precludes the formation of the substrate-His293 interaction while causing some repulsion in the active site. Overall, the consistently lower distances of Nη2 over Nη1 in all trajectories demonstrate preference for the first and second methylation taking place on Nη2.

In addition to other factors[5], this simulation results find the product specificity of PRMT1 is partially achieved by controlling the binding of the substrate for the formation of conformers that lead to ADMA.

M48F complexes: The introduction of a hydrophobic residue in a polar active site shows unexpected distributions since experimental studies found this mutant impaired (Fig. 3.19 vs 3.15). One would expect the M48F systems to cause a dramatic structural change in the active site with near-attack conformations reduced in frequency.

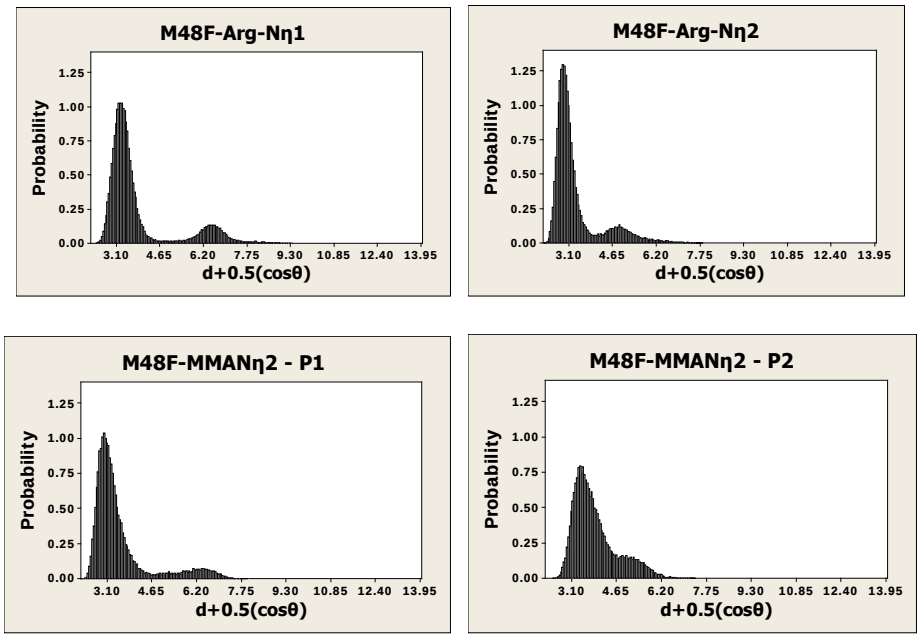


Figure 3.19 M48F-complexes. M48F-MMANη2 - P1 would lead to SDMA whereas M48F-MMANη2 - P2 would lead to ADMA.

The M48F-Arg and M48F-MMANη2 complexes have the guanidino moiety closer to SAM's methyl group than in the WT-Arg and WT-MMANη2, respectively. Additionally, the distances show the substrate's guanidino nitrogens on M48F-MMANη2 have a slight preference for Nη1 over Nη2 attacking SAM, implying a higher propensity for SDMA formation, in line with experimental observations[5]. Perhaps a favorable cation-π interaction between the positively charged guanidino moiety and the phenylalanine's aromatic ring results from an altered positioning. Energetically significant cation-π interactions are ubiquitous in proteins and are regarded as an important noncovalent binding interaction relevant to structural biology.[149]

Distances for the M48F complexes imply this mutation should be comparable to the WT which contradicts its poor experimental activity. It should however be noted that positions alone cannot capture all factors that determine reactivity. The presence of the hydrophobic and inflexible Phe48 in a polar active site for instance halves the substrate-His293 H-bond occupancy from 49.0 (WT-MMANη2) to 25.2 % (M48F-MMANη2) (Table 3.6). This should affect electron redistribution in the nucleophile conceivably depressing the nucleophilicity.

An important attribute of M48 maybe its flexibility coined with non-hydrophobicity. Therefore, M48's lack of negative repulsions in the active site. M48A and M48P formed ADMA exclusively, but reduced catalytic efficiency by 338 and 70 fold respectively (Ala and Pro are hydrophobic but small) whereas M48L formed ADMA reducing efficiency by a factor of only 3 (hydrophobic but flexible).[13]

The simulations for the M48F mutant do not show a clear structural preference for methylation on either nitrogen for the exclusive formation of ADMA or SDMA. This underscores that product specificity in PRMTs is not as straightforward as the Phe/Tyr switch that controls specificity in lysine methyltransferases.[150]

H293A/S complexes: For the H293A-Arg complex, the conformations of the reactants sample sub-optimal orientations with the substrate's reacting moiety being relatively dynamic. The distributions for H293S-M48F-Arg complex (Fig 3.20) are similar to the H293A-Arg system and the substrate is not well localized within the active site. The average distances (Fig. 3.16) are however longer and these observations emphasize the additive efforts of Met48 and His293 are important for its positioning. Clearly, this double mutant should lose considerable activity considering M48F or H293A individually are experimentally impaired.

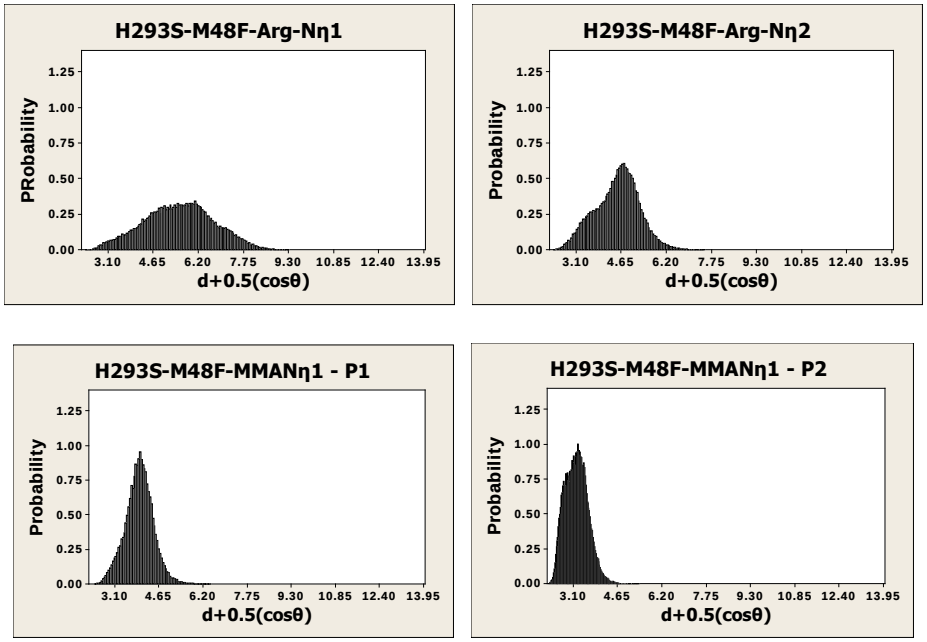


Figure 3.20 H293S-M48F-complexes. H293S-M48F-MMANη1 - P1 would lead to ADMA whereas H293S-M48F-MMANη1 - P2 would lead to SDMA.

The introduction of the methyl group on the Nη1 position of MMA in the H293S-M48F-MMANη1 system helps restrain the substrate (Fig. 3.20). Near-attack conformers are formed more often than the system with the arginine substrate. Perhaps this is due to the hydrophobic interaction between MMA's methyl group and Phe48, combined with the absence of His293

since the repulsion between the histidine and the methyl group is eliminated. It therefore becomes feasible that this mutant should tolerate MMA methylated on the Nη1 nitrogen (MMANη1) and possibly form some SDMA if it retains any activity.

3.4 Conclusions

From the Quantum Chemical Simulations

Gas-phase QM to compute the free energies of activation for MMA, ADMA, and SDMA formation revealed that SDMA formation is more energetically costly than ADMA, which would lead to a significant difference in methylation rate constants for asymmetric and symmetric dimethylation. An energetic argument is consistent with the activity displayed by the M48F-PRMT1 mutant as well as the activity displayed by recombinant PRMT5. This indicates that SDMA-forming methyltransferases overcome the energetic challenges *in vivo*, using mechanisms that may represent another exciting layer of enzymatic regulation.

From the Molecular Dynamics Simulations

1. Since there are no water molecules near the methylation site, Arg54-Glu144, Tyr14-Glu153 and Ser51-His293 hydrogen bond network is important in increasing the nucleophilicity of the otherwise poor guanidino group. Glu153 may be responsible for the proton abstraction in addition to stabilizing the transition state during the methyl transfer reaction.
2. The electrostatic/H-bonding interaction between His293 and the substrate indicates that it could be involved in determining substrate orientation. Its presence enables PRMT1 to yield ADMA exclusively by precluding MMA from binding with the methyl group on the Nη1 position which is more favorable in SDMA formation.

3. Based on the molecular dynamics simulations, it was hypothesized that the PRMT1 mutant H293S and/or H293S/M48F might form SDMA.

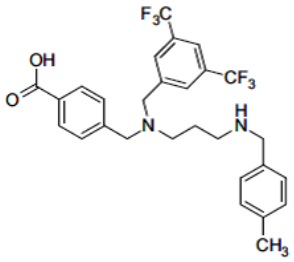
Chapter 4: Design, development and evaluation of novel dual PPARδ/PPARγ Agonists

4.1 Introduction

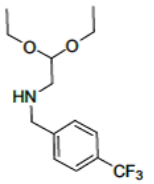
Type 2 diabetes mellitus is at epidemic proportions and is associated with excessive cardiovascular morbidity and mortality. Both impaired insulin secretion and resistance contribute to the development of the disease. The thiazolidinediones (TZDs) represent a class of drugs which improve whole body insulin sensitivity.[151] The beneficial effects of TZDs are attributed to the activation of the nuclear receptor class of transcription factors; PPARγ (Peroxisome Proliferator Activating Receptor). However, TZDs are associated with adverse cardiovascular events mostly due to development of peripheral edema and weight gain.[152] The PPAR family of nuclear receptors constitute three members; PPAR-α, β/δ, and γ that are highly conserved in mammals, each forming a functional heterodimeric complex with 9-cis retinoic acid receptor (RXR). PPAR β/δ activation is associated with improving overall circulating cholesterol levels (HDL, LDL and triglycerides) and is ubiquitously expressed throughout the body. Recently, it has been found that overexpression of PPAR β/δ in skeletal muscle improves the glycolytic muscle fiber type in animal models and thus improves circulating glucose and fatty acid levels.[153] However, agonists for this class of receptors do not have a significant impact upon improving insulin sensitivity.[154] Consequently, because diabetes and metabolic syndrome are associated with defects in glucose oxidation and lipid metabolism, the concept of discovering dual agonists which can activate both PPARβ/δ and PPARγ simultaneously, has emerged as a potential therapeutic target for improving outcomes of diabetes.

Thus, the goal of the current study was to rationally discover new dual PPARγ and PPARδ agonists without the deleterious side effects of full PPARγ agonists. To accomplish this task a joint computational, synthesis and gene regulation study of novel dual PPARγ and PPARδ agonists were carried out. Twenty-three nontraditional compounds have been designed and synthesized as potential agonists. Docking simulations of the proposed compounds within the ligand-binding domain of PPARγ and PPARδ have been performed. Compounds found to be having high and low binding activity on PPARγ/δ as determined by the docking studies were evaluated for biological significance. In vitro cellular assays were utilized to test transcriptional regulation of both PPARγ and PPARδ targets as well as for adiposity and mitochondrial biogenesis.

We studied 23 compounds by docking calculations (as described further below) and predicted that compound **9** (Fig. 4.1) to possess the highest binding affinity in both PPARγ and PPARδ binding pockets. Compound 3-121 (Fig. 4.1) was found to have low binding affinity for PPARγ and average affinity for PPARδ.



Compound 9



Compound 3-121

Figure 4.1: Compound **9** (top) and **3-121** (bottom) structures.

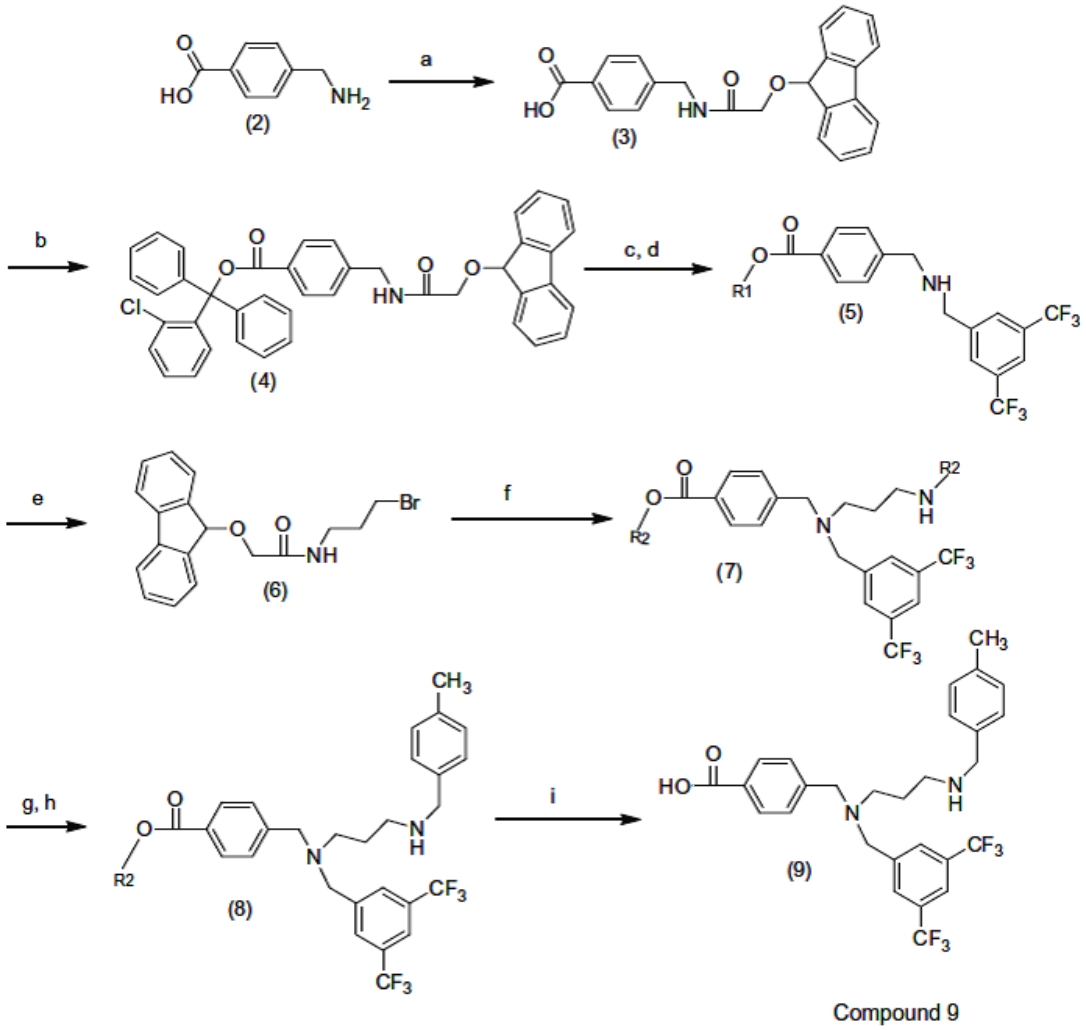
4.2 Methods

4.2.1 Chemical Synthesis

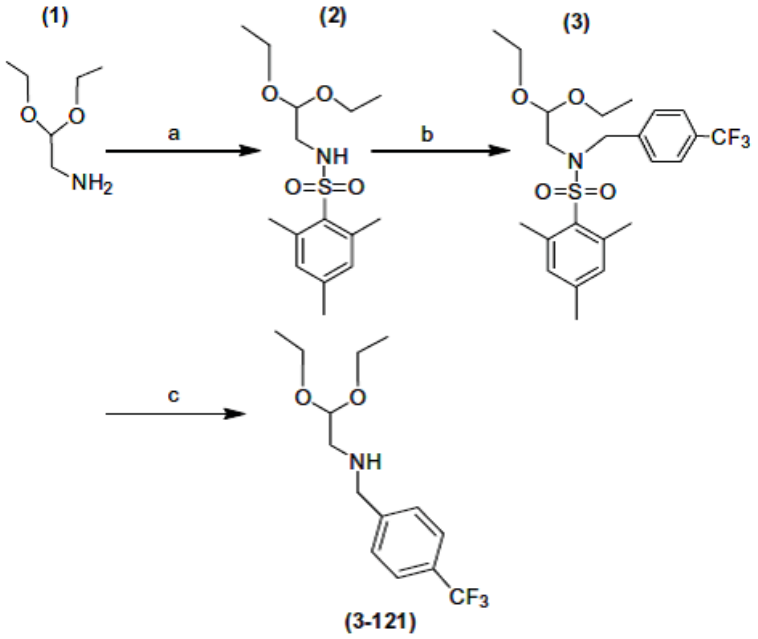
The general synthesis of compounds **9** and **3-121** are depicted in Schemes 1 and 2.

Scheme 1: Treatment of 4-(aminomethyl)benzoic acid **2** with Fmoc chloride, Na2CO3, dioxane, water, at 0 ˚C provided **3**. Nucleophilic substitution of 3 with 2-chlorotrityl chloride resin, D-IPEA, COMU, methanol, DCM, and DMF provided intermediate **4**. Subsequent treatment of **4** with 3,5-bis(trifluoromethyl) benzyl bromide, DIPEA, sodium hydride and DCM, followed by deprotection of Fmoc using DCM, DMF, Piperidine in a 1:1:2 ratio, yielded **5**. The Fmoc protected amino propyl bromide intermediate 6 was created by reacting 3-bromopropylamine hydrobromide with Fmoc chloride, sodium carbonate, dioxane, water, at 0 ˚C. The intermediate compound **6** was reacted with **5**, DIPEA, sodium hydride, DCM, DMF to generate **7**. 4-Methylbenzyl bromide, DIPEA, sodium hydride, DCM, DMF were reacted with **7** for 12 h before subsequent treatment with DCM, DMF, Piperidine in a 1:1:2 ratio to remove the Fmoc and generate **8**. The last step involved to generate the final compound **9** was to clip it from the resin using 90% TFA and DCM for 1.5 h to dissociate resin from final compound **9**.

Scheme 2: Treatment of aminoacetaldehyde diethyl acetal **1** with 2-mesitylenesulfonyl chloride, 10% NaOH, CH2Cl2 for 24 h, provided **2**. Nucleophilic substitution of 4-(triflouromethyl)benzyl bromide, NaH, DMF, under nitrogen, with **2** yielded **3**. The last step required to generate **3-121** was reacting **3** with HBr solution (33% acetic acid), phenol, ethyl acetate, for 24 h, to create **3-121**.



Scheme 1: Reagents and conditions: (a) FMOC, Na2CO3, Dioxane, Water, at 0 ˚C; (b) 2-Chlorotrityl chloride resin, DIPEA, COMU, Methanol, DCM, DMF; (c) 3,5-Bis (trifluoromethyl) benzyl bromide, DIPEA, NaH, DCM; (d) DCM, DMF, Piperidine, 1:1:2; (e) 3-bromopropylamine hydrobromide, FMOC Chloride, Na2CO3, Dioxane, Water at 0 ˚C; (f) (8), DIPEA, NaH, DCM, DMF; (g) 4-methylbenzyl bromide, DIPEA, NaH, DCM, DMF; (h) DCM, DMF, Piperidine, 1:1:2; (I) 90% TFA, DCM.



Scheme 2: Reagents and conditions: (a) 2 Mesitylenesulfonyl chloride, 10% NaOH, CH2Cl2; (b) 4-(Trifluoro-methyl) benzyl bromide, NaH, DMF; (c) HBr, Phenol, EtoAC.

4.2.2 Computational Methods

AutoDock4[155] AutoDock Vina[156] and Surflex-Dock from Sybyl-X[157] were used to dock known and proposed agonists of PPARγ and PPARδ. Initial Cartesian coordinates for the protein–ligand structures were derived from a reported 1.95 Å crystal structure of PPARγ (PDB ID: 3ET3)[34] and a 2.00 Å crystal structure of PPARδ (PDB ID: 3GZ9).[158] In the case of PPARγ, the present model included the active site and all residues within 15 Å of it. Clipped residues were capped with acetyl or N-methylamine. The reduced PPARγ model consisted of ~2800 atoms and 162 residues out of 275 residues. The PPARδ model used the protein in its entirety. The protein targets were prepared for molecular docking simulation by removing water molecules and bound ligands. Auto-Dock Tools (ADT)[155] was used to prepare and analyze the docking simulations for the AutoDock and AutoDock Vina programs, whereas Sybyl-X was

used to prepare, carry out runs, and analyze the Surflex-Dock simulations. All ligands were constructed using PyMOL[6] with subsequent geometry optimizations carried out using the semi empirical method PDDG/PM3[159-161] and the BOSS program[162]. Polar hydrogens were added, and in the case of AutoDock, Gasteiger charges were assigned. Nonpolar hydrogen was subsequently merged. The protonation state of the ligands was adjusted to the species assumed predominant at physiological pH, specifically, carboxylic acid moieties were deprotonated. Conjugate gradient minimizations of the systems were performed using MCPRO[162] and GROMACS[163]. A grid was centered on the catalytic active site region and included all amino acid residues within a box size set at x = y = z = 70 grid points and 26 Å for AutoDock and AutoDock Vina, respectively. AutoGrid 4 was used to produce grid maps for AutoDock calculations where the search space size utilized grid points of 0.375 Å.

AutoDock Vina details: Standard flexible protocols of AutoDock Vina using the iterated local search global optimizer[155] algorithm were employed to evaluate the binding affinities of the molecules and interactions with the receptors. All ligands and active site residues, as defined by the box size used for the receptors, were set to be rotatable. Calculations were carried out with the exhaustiveness of the global search set to 100, number of generated binding modes set to 20, and maximum energy difference between the best and the worst binding modes set to 5. Following completion of the docking search, the final compound pose was located by evaluation of AutoDock Vina's empirical scoring function where the conformation with the lowest docked energy value was chosen as the best.

AutoDock details: The Lamarckian genetic algorithm[164] was chosen to search for the best conformers. The rigid roots of each ligand were defined automatically and the amide bonds were made non-rotatable. The docking process used for AutoDock mirrored that of a recent

study by Perryman and McCammon.[165] For example, 100 conformers were considered for each compound. The population size was set to 150, and the individuals were initialized randomly. The maximum number of energy evaluation was set to $9X10^{10}$. The docking parameters used were: maximum number of generations was $9X10^{4}$, maximum number of top individuals that automatically survived set was 1, mutation rate of 0.02, crossover rate of 0.8, step sizes were 2 Å for translations, 50˚ for quaternions, and 50˚ for torsions, a cluster tolerance of 2 Å was employed, and the maximum number of iterations in the pseudo-Solis-and-Wets minimization/local search was increased to 3000.

Surflex-Dock details: Following the receptor's preparation, all hydrogen bonds were added and the side chain amides in all Asn and Gln oriented to maximize hydrogen bonding. Minimization was performed for 100 iterations using the AMBER FF99 force field.[70] Prior to docking runs, the Surflex docking algorithm[157] required the generation of a protomol,[166] an idealized representation of the binding site that defines the search area. For this purpose, the prepared receptor files were loaded and the protomol generation constructed based on protein residues that constitute the active site, decreasing the threshold and increasing bloat parameters values to 0.25 and 2 Å, respectively. The threshold has a default value of 0.50 and decreasing this number increases the volume while the bloat on the other hand has a default of 0 and increasing this value inflates the protomol. During the docking runs, all other adjustable parameters were left at their default values.

4.2.3 Biological Assays

To validate the specificity of the compounds towards activation of PPARγ/δ targets, the compounds were tested for their capacity to bind to select PPARγ/δ Peroxisome Proliferator Response Elements (PPRE). The PPRE are unique sites located in the promoter region where

PPARs bind and transcriptionally activate the target genes. AP2-PPRE is the PPARγ target involved in adipocyte growth and differentiation. PDK4–PPRE is a PPARδ target involved in regulating energy metabolism in the cell.

The physiological effects of our compounds upon lipid accumulation were evaluated; 3T3L1 adipocytes were cultured and treated with compounds for a period of 6 days. Lipid accumulation was determined by oil red-o staining, and measured the absorbance spectrophotometrically at a wavelength of 510 nm and standardized to total protein concentrations. Data from lipid accumulation studies were compared to cells treated with rosiglitazone (full PPARγ agonist).

Recently it has been found that PPARδ agonists (GW0742) induce mitochondrial biogenesis.[167] Therefore to better understand the level of PPARδ agonism of our compounds, C_2C_{12} skeletal muscle cells were cultured and were treated with the compounds for 4 days. Changes associated with markers for mitochondrial biogenesis were determined by quantitative real time PCR.

4.3 Results and Discussion

4.3.1 *In silico* Validation

Design of small molecules that bind to a biological target have made great advancements in technique recently.[168] Of particular interest, virtual screening methods that dock ligands into a receptor, allow for a large number of compounds to be compared quickly, including PPARγ.[169] However, virtual screening methods often neglect important statistical and chemical contributions in favor of computational efficiency.[170] To examine the accuracy of the present docking methods, rosiglitazone (PDB ID: 2PRG),[171] indeglitazar (PDB ID: 3ET3),[34] and D321 (PDB ID: 3GZ9)[158] in PPARγ and PPARδ have been calculated using AutoDock, AutoDock Vina, and Surflex-Dock and the resultant poses compared to the crystal structures. The flexible

protocols of AutoDock Vina yielded computed RMSD values 0.88, 0.83, and 0.07 Å for rosigilitazone, indeglitazar, and D321, respectively, for all atoms between the controls and the docked conformations. AutoDock's computed RMSD values were 1.87, 1.40 and 1.07, while Surflex-Dock's were 1.16, 1.81 and 0.70 Å for the three agonists. Illustrations of the calculated versus crystal structure binding poses are given in the Supplementary data (Figs. S1–S9) and show that AutoDock Vina reproduced the controls significantly better than the other docking programs.

Consequently, AutoDock Vina was used to dock and interpret data for the current compounds. Twenty-three proposed synthetic agonists have been docked with both PPARγ and PPARδ using AutoDock Vina to elucidate the interactions with active site residues.

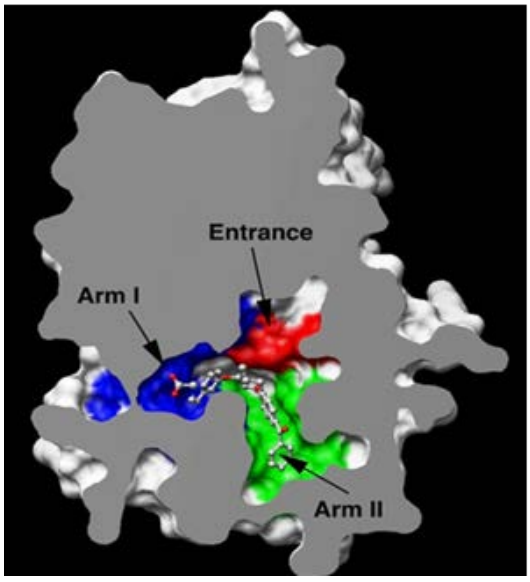


Figure 4.2: The ligand binding domain of PPARγ bound with an agonist. (Adapted from Zoete et al.[172])

The PPAR ligand binding domain resembles a large Y-shaped cavity starting from the entrance and extending into Arm I and Arm II pockets (Fig. 4.2).[172] Arm I is substantially polar while the entrance and Arm II are primarily hydrophobic. Out of the 23 compounds studied, the

calculations predicted compound **9** to possess the highest binding affinity for both PPARγ and PPARδ (Table 4.1).

Table 4.1: Predicted binding affinities (kcal/mol) for proposed agonists in PPARγ and PPARδ

PPARγ		PPARδ	
Molecule	Affinity	Molecule	Affinity
9	−12.0	**9**	−10.8
8	−11.8	**4-20**	−10.6
7	−11.1	**8**	−10.6
3-91	−11.0	**3-75**	−10.4
4-20	−10.2	**3-71**	−10.2
4	−10.0	**3-91**	−10.2
3-95	−9.9	**5**	−10.1
5	−9.7	**7**	−9.8
6	−9.6	**4**	−9.6
3-71	−9.4	**3-95**	−9.5
3-65	−9.4	**2**	−9.3
3-75	−9.3	**6**	−9.2
4-21	−9.2	**3**	−9.2
4-23	−9.2	**3-65**	−9.0
3-137	−8.8	**4-21**	−8.9
2	−8.8	**3-121**	−8.6
3	−8.4	**4-23**	−8.6
3-115	−8.3	**3-115**	−8.3
1	−7.9	**1**	−8.2
3-121	−7.7	**3-137**	−7.7
3-133	−7.5	**3-133**	−7.4
4-19	−7.1	**4-19**	−7.1
3-81	−7.0	**3-81**	−6.9
Rosi. (PPARγ)	−9.2	D321 (PPARδ)	−12.2

4.3.2 Compound 9 Binding to PPARγ:

Inspection of the in silico PPARγ/**9** complex suggests that the carboxylate group of compound **9** forms a hydrogen bond at a distance of 2.02 Å with the entrance residue Glu343 while the 2˚ amine donates a tighter hydrogen bond of 1.85 Å with the polar Ser289 residue of Arm I (Fig. 4.3). The noncanonical X···H···π hydrogen bond has been shown to be of great importance in proteins and **9**'s trifluoromethyl disubstituted phenyl ring is revealed to accept a N--H···π hydrogen bond of ca. 2.69 Å from His449. It is interesting to note that this interaction

holds the ring moiety in position making **9** avoid contact with Tyr473 which is ca. 4.8 Å away. This residue is crucial to the stabilization of the AF2 helix H12 which allows the binding of co-activators that lead to the activation of the genes responsible for adipogenesis.[169]

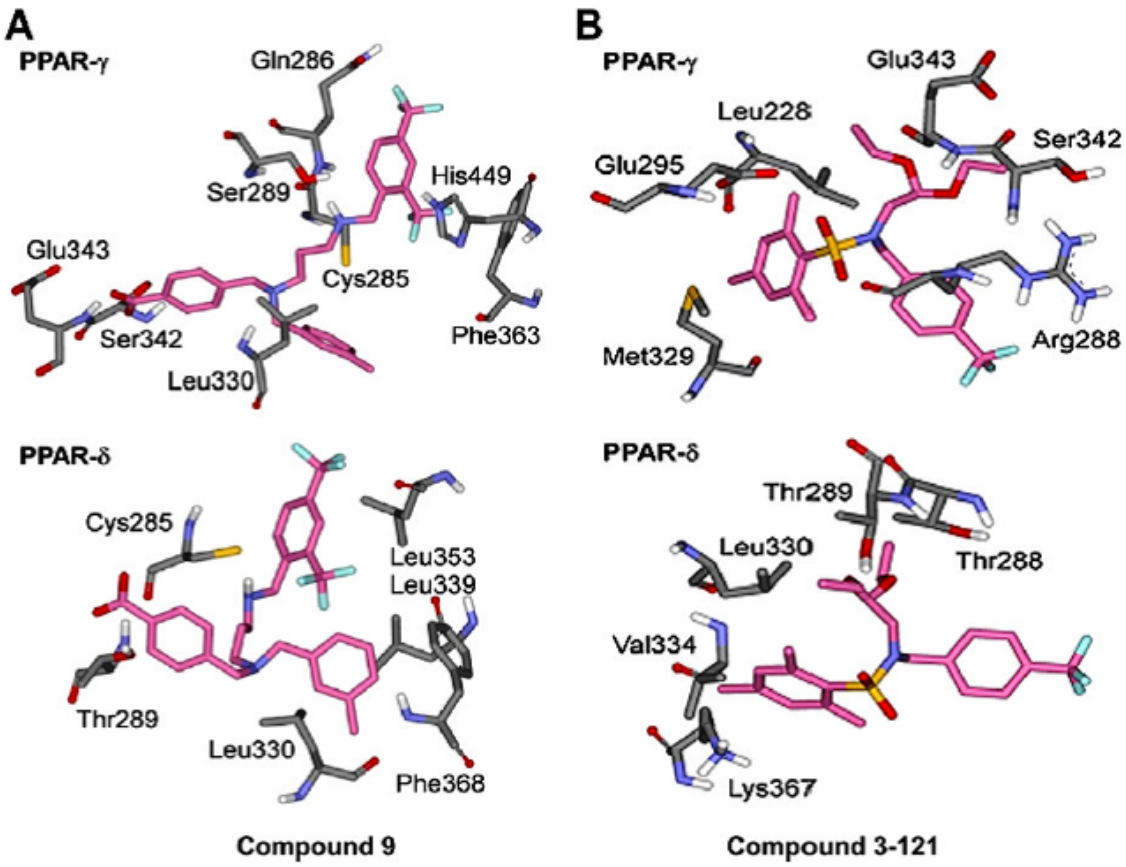


Figure 4.3: Predicted images of compounds **9** and **3-121** bond patterns. (A) Compound **9** bound to the active site of PPARγ (top) and PPARδ (bottom) with key residues shown. (B) Compound **3-121** bound to the active site of PPARγ (top) and PPARδ (bottom) with key residues shown.

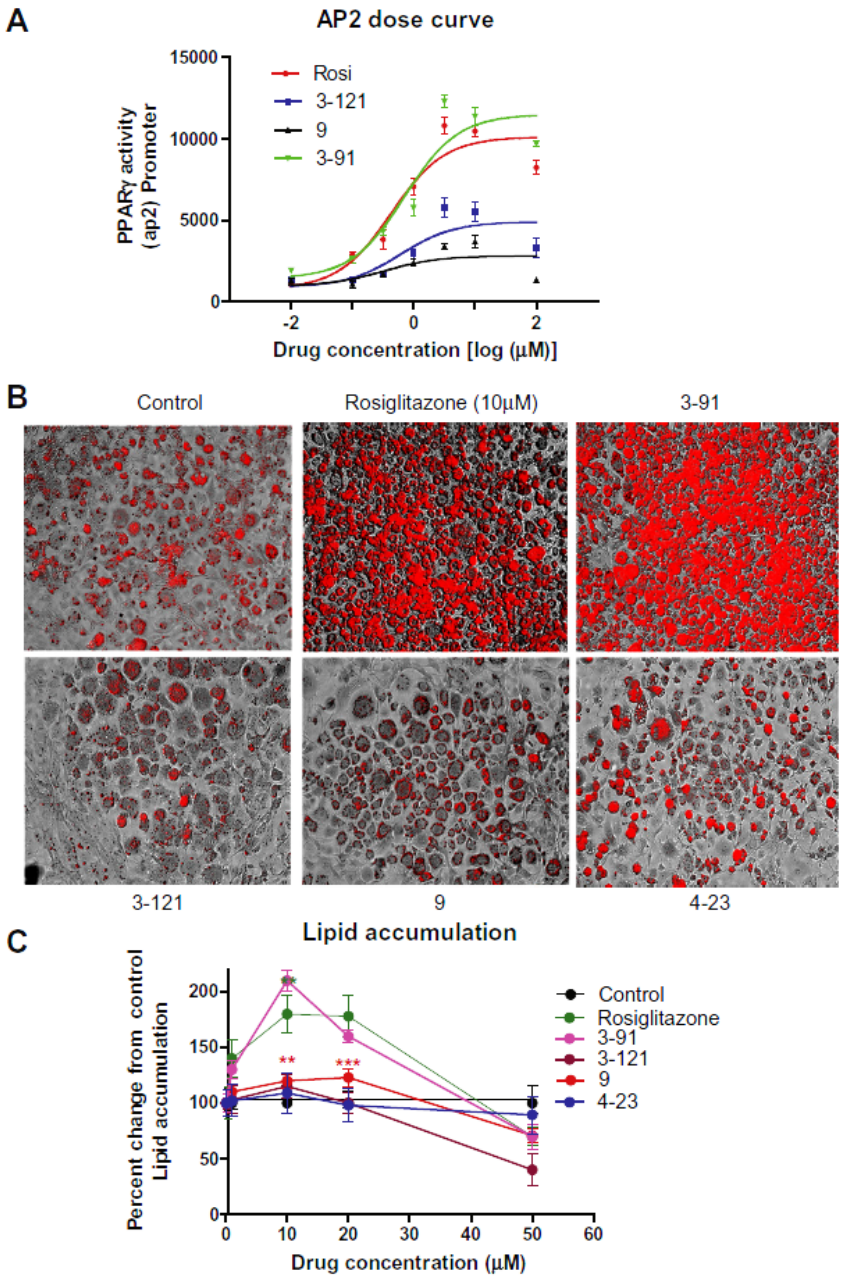


Figure 4.4: PPARγ biological studies. (A) AP2-PPRE dose–response curve. Data from luciferase assays shows that compound **3-91** and rosiglitazone (10 µM) have strong binding affinity for the AP2-PPRE. However compounds **9** and **3-121** show less binding affinity for the AP2-PPRE. (B) Compound 3-91 induces lipid accumulation comparable to rosiglitazone in 3T3-L1 adipocytes, as demonstrated by the oil red-o stain (red color).

Compounds **9**, **3-121** and **4-23** induce less lipid accumulation. (C) Lipid accumulation levels were quantitated by measuring the absorbance at a wavelength of 510 nm and graphically represented as a percent change from vehicle treated cells (control). *P <0.05,**P <0.005, ***P <0.0001.

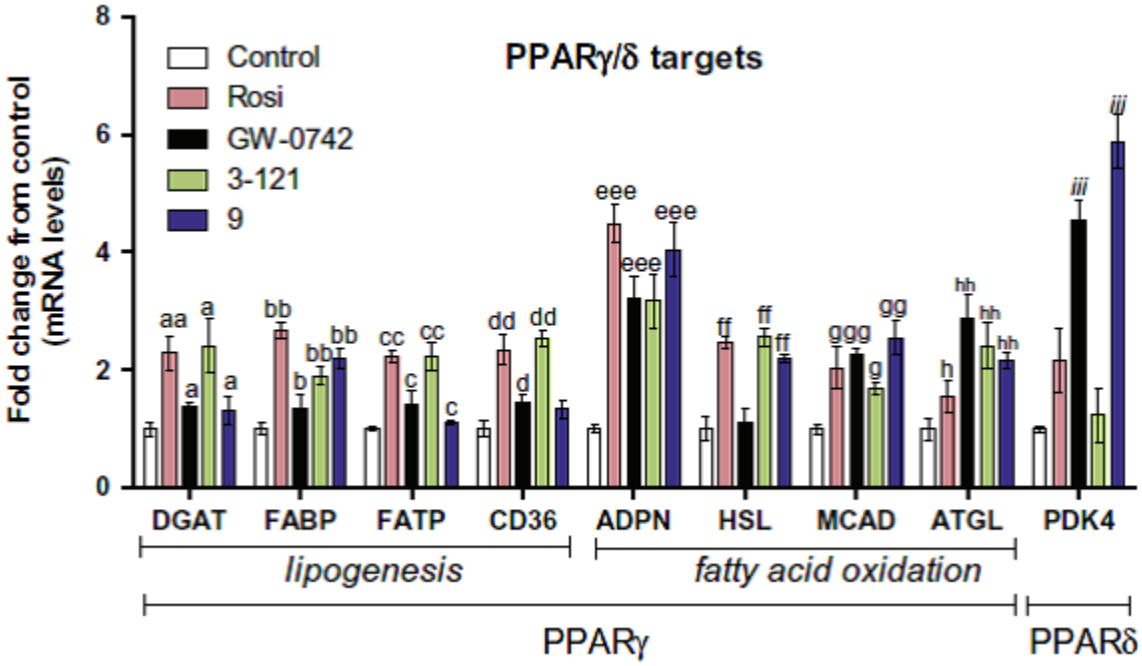


Figure 4.5: MRNA analyses of PPARγ/δ targets. Rosiglitazone, GW0742, 3-121 and 9 (10 lM dose for all compounds), were applied to 3T3-L1 adipocytes to determine effects on PPARγ/δ target mRNA expression levels. Compound **9** increased the gene expression of PPARγ targets MCAD, HSL, ATGL and adiponectin. Furthermore compound **9** also increased the gene expression of PPARδ target, PDK4. Values represent the fold change from control mRNA expression levels, where a, b, d, g, h; P <0.05, aa, bb, cc, dd, ee, ff, gg, hh; P <0.005, eee, ggg, iii; P <0.0001.

Our *in silico* data were confirmed by transcriptional assays which demonstrate that compound **9** minimally activates the AP2-PPRE (Fig. 4.4A). The results appear consistent with the present lipid accumulation assays, where compound 9 has negligible effects upon lipid accumulation in adipocytes (Fig. 4.4B). Furthermore compound **9** marginally induced the expression of genes associated with lipid accumulation and synthesis such as FATP, CD36 and DGAT (Fig. 4.5). However, compound 9 elevated the expression of genes associated with fatty acid oxidation (MCAD, HSL and ATGL) and insulin sensitivity (adiponectin) (Fig. 4.5).

4.3.3 Compound 9 Binding to PPARδ:

Docking studies of our compounds on PPARδ shows that the carboxylate group of the receptor forms a hydrogen bond of 3.03 Å with Thr289 in Arm I which is part of the His323, His449 and Tyr473 hydrogen bond network. This site is specifically involved in the carboxylate group of fatty acids and eicosanoic acids (the endogenous ligands) binding, which are natural ligands for PPARδ in circulation.[172] The 2˚ amine also donates a H-bond of 2.87 Å to the Sulfur atom of the polar residue Cys285 in Arm I. Furthermore, the methyl substituted phenyl ring in compound **9** located between the entrance and Arm II and surrounded by the hydrophobic residues Leu330, Leu339 and Phe368 stabilize the ligand binding domain.

The biological significance of these modifications were determined by luciferase assays (Fig. 4.6A) which demonstrate that compound **9** binds to PDK4–PPRE, a direct PPARδ target. We also found an increase of PDK4 gene expression in adipocytes (Fig. 4.5). Furthermore, an increased gene expression of mitochondrial markers was observed in Figure 4.6B. Interestingly, compound **9** significantly increased cytochrome oxidase c 4-2 subunit which is involved in inducing mitochondrial respiration and biogenesis in response to ischemia in the heart.[173] This

suggests a possible mechanism by which compound **9** can offer cardio protection against ischemic injury.

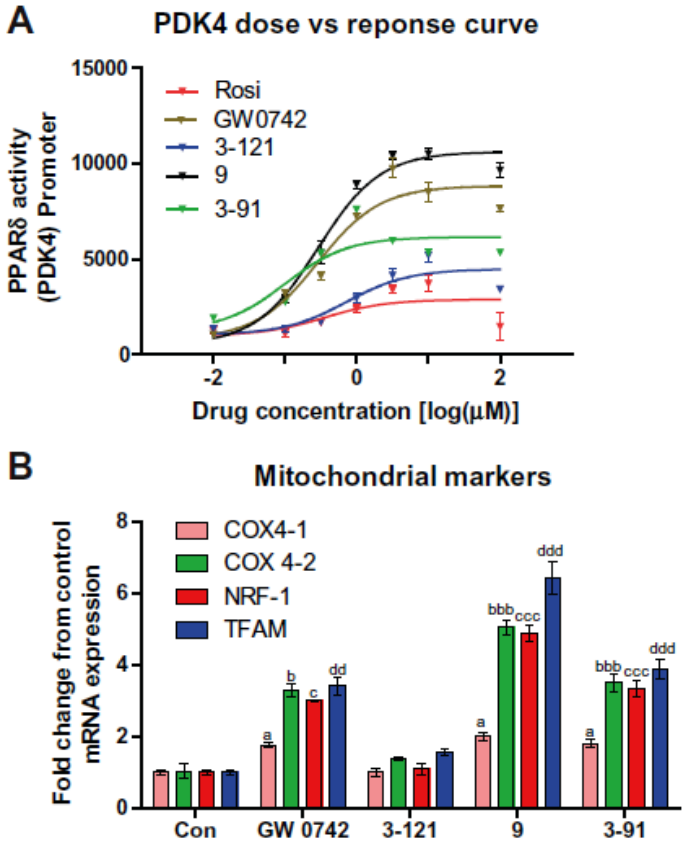


Figure 4.6: PPARδ biological studies. (A) PDK4–PPRE dose–response curve. Data from luciferase assays shows that compound **9** and GW0742 have strong binding affinity for the PDK4–PPRE. However compounds **3-91** and **3-121** show less binding affinity for the PDK4–PPRE. (B) Compound **9** and GW0742 increases the gene expression of mitochondrial markers in C_2C_{12} skeletal muscle cells as demonstrated by quantitative RT-PCR. Data are means ± SEM from 3 independent experiments. Values were set fold change from control. *P <0.05, **P <0.005,***P <0.0001.

4.3.4 Binding of the other Compounds:

To further validate the significance of the AutoDock Vina modeling of PPARγ/δ agonism, we tested another compound, **3-121**. Out of the 23 compounds tested, compound **3-121** was ranked in the lower end of the relative binding affinities for both receptors (Table 1.1). Our analysis predicted that the ether oxygen atoms of **3-121** accept two hydrogen bonds of ca. 2.17 and 2.86 Å from the backbone NH group of the polar residue Glu343 in the entrance region of PPARγ (Fig. 4.3B). The calculations do not predict binding with any of the polar residues located in Arm I; this appears consistent with the lipid accumulation studies which demonstrate an insignificant induction of adiposity with compound 3-121 compared to full PPARγ agonists (Fig. 4.4). The docking studies also predicted **3-121** to bind weakly to PPARδ due to the absence of the conserved hydrogen bonding. These binding predictions correlate well with results from the luciferase assays which indicates that **3-121** did not activate the PPARγ target AP2 or the PPARδ target PDK4.

Compound **3-91** was predicted to display high PPARγ activation capacity because it possesses the same binding mode as the full PPARγ agonist rosiglitazone. This was demonstrated in our docking analysis which predicted that the PPARγ receptor forms two hydrogen bonds with the isoindole-dione substituent group of **3-91**, the hydrogen on His323 and a carbonyl oxygen, and the hydrogen on His449 and the nitrogen at distances of 2.38 and 2.66 Å, respectively (Supplementary data S10). In addition, the OH group of Ser289 also forms a hydrogen bond of 2.57 Å with the N of the isoindole-dione substituent group. This binding mode is consistent with known full agonists, for example, rosiglitazone.[174] Our cellular assays which tested the biological effects of compound **3-91** on lipid accumulation verified it to be a strong PPARγ activator and thus induced comparable levels of lipid accumulation to rosiglitazone.

However **3-91** does not form substantial hydrogen bonds with PPARδ despite having a larger binding affinity relative to 3-121.

4.4 Conclusion

We have developed a series of novel dual PPARδ/γ agonists which spans from full agonists to partial agonists for PPARγ as well as PPARδ. In regards to PPARγ activation, our goals were to develop compounds which will lack the deleterious side effects found with current TZDs, including weight gain and the eventual heart failure.[175] We have found that compound **9** does not interact with Tyr473 of the PPARγ binding pocket, a residue crucial to the stabilization of the AF2 helix H12 and has been shown to lead to adipogenesis.[176] Compound **9** selectively activates PPARγ target genes to have negligible effects on lipid accumulation while activating other genes involved in regulating fatty acid oxidation(HSL, ATGL and MCAD). We have also found that compound 9 is a potent PPARδ activator and the immediate future interest is to evaluate its ability to promote insulin sensitization in animal models.

Chapter 5: Characterization of *Pf*TrxR inhibitors using *in silico* techniques and antimalarial assays

5.1 Introduction

The current study is aimed to employ the combined approach of *in silico* studies and antimalarial assays for identification of specific *P. falciparum* thioredoxin reductase *(Pf*TrxR) inhibitors as scaffolds for lead optimization. The compounds 1,4-napthoquinone (1,4-NQ), bis-(2,4-dinitrophenyl)sulfide (2,4-DNPS), 4-nitrobenzothiadiazole (4-NBT), 3-dimethylaminopropiophenone (3-DAP) menadione (MD) and curcuminoids were tested for antimalarial activity against both chloroquine (CQ)-sensitive (D6) and chloroquine (CQ)-resistant (W2) strains of *plasmodium falciparum* through *in vitro* and LC-MS-based functional assays. Analysis of non-covalent interactions with *Pf*TrxR was carried out through molecular docking.

5.2 Methods

5.2.1 Computational Methods

AutoDock Vina[177] was used to dock inhibitors to the respective targets. Initial Cartesian coordinates for the protein-ligand structures were derived from reported crystal structures of *h*TrxR (PDB ID: 3QFA)[60] and *Pf*TrxR (PDB ID: 4B1B)[55].

The protein targets were prepared for molecular docking simulation by removing water molecules and bound ligands. AutoDockTools (ADT)[155] was used to prepare the docking simulations whereas Chimera was used to analyze the docking poses.

All ligands were constructed using PyMol[178] with subsequent geometry optimizations carried out using the semi-empirical method PDDG/PM3[159,179,180]. Polar hydrogens were added. ADME properties logP, logS, polar surface area, and apparent Caco-2 permeability for each ligand were computed using QikProp[181,182]. Conjugate gradient minimizations of the systems were performed using GROMACS[183]. A grid was centered on the catalytic active site region and included all amino acid residues within a box size set at x = y = z = 20 Å.

AutoDock Vina details: Standard flexible protocols of AutoDock Vina using the Iterated Local Search global optimizer[177] algorithm were used to evaluate the binding affinities of the molecules and interactions with the receptors. All ligands and docking site residues, as defined by the box size used for the receptors, were set to be rotatable. Calculations were carried out with the exhaustiveness of the global search set to 100, number of generated binding modes set to 20 and maximum energy difference between the best and the worst binding modes set to 5. Following completion of the docking search, the final compound pose was located by evaluation of AutoDock Vina's empirical scoring function where the conformation with the lowest docked energy value was chosen as the best.

5.2.2 Biological Assays

Chemicals and enzymes: Solvents used for LC-MS analysis were purchased from Fischer Scientific International (Atlanta, GA). Buffer salts, bis-2, 4-dinitrophenyl sulfides and demethoxycurcumin were purchased from Sigma-Aldrich (Allentown, PA). Curcumin was purchased from ChromaDex, Irvine, CA. All other reagents were purchased from Sigma–

Aldrich. Deionized water generated by a Milli-Q water system (Millipore, MA) was used in the experiments. PfTrxR (M_r 59 kDa) enzyme was provided as a gift by Prof. Katja Becker, Justus-Liebig University, Giessen, Germany The recombinant PfTrxR was prepared and purified using silver-stained SDS page according to the procedure published by Kanzok et al.[184] The specific activity of PfTrxR (1.9 U/mg) was determined by DTNB [5, 5'-dithiobis (2-nitrobenzoic acid)]. Protein concentration of enzymes was determined by Bradford method[185].

Antimalarial assay: Briefly, antimalarial activity of the compounds were determined *in vitro* on chloroquine sensitive (D6, Sierra Leone) and resistant (W2, IndoChina) strains of *p. falciparum*. The 96-well microplate assay is based on the effect of the compounds on growth of asynchronous cultures of *p. falciparum*, as determined by the fluorometric SYBR green assay[186].

Cytotoxicity assay: Cytotoxicity in terms of cell viability was evaluated using 3T3 cells by AlamarBlue assay[187] and VERO cells by Neutral Red assay for the curcuminoids[188]. This assay was conducted on compounds designated as active in the *Pf*TrxR functional assay and the antiplasmodial phenotypic screening. Chloroquine was used as a reference compound for cytotoxicity study.

Reactive oxygen species (ROS) in red blood cells assay: Accelerated generation and accumulation of reactive oxygen intermediates (superoxide radical, hydroxyl radical and hydrogen peroxide) are mainly responsible for oxidative stress.[189] The intraerythrocytic formation of ROS was monitored in real-time with 2'7'-dichlorofluorescein diacetate (DCFDA), a fluorescent ROS probe[190]. Human erythrocytes collected in citrate phosphate anticoagulant were used. The erythrocytes were washed twice with 0.9 % saline and suspended in PBSG at a hematocrit of 10%. A 60 mM stock of DCFDA was prepared in DMSO and added to the erythrocyte suspension in PBSG (10% hematocrit) to obtain the final concentration of 600 µM.

Erythrocyte suspension containing 600 µM of DCFDA was incubated at 37°C for 20 min and centrifuged at 1000 g for 5 min. The pellet of DCFDA loaded erythrocytes was suspended in PBSG to 50% hematocrit and used for kinetic ROS formation assay. The assay was directly set up in a clear flat-bottom 96 well microplate. The reaction mixture contained 40 µL of DCFDA loaded erythrocytes, the test compounds (concentration as mentioned) and potassium phosphate buffer (100 mM, pH 7.4), to make up the final volume to 200 µL. The controls without drug were also set up simultaneously. Each assay was set up at least in duplicate. The plate was immediately placed in a microplate reader programmed to kinetic measurement of fluorescence (excitation 488 nm and emission 535 nm) for 2 hours with 5 min time intervals.

5.3 Results and Discussion

5.3.1 Biological studies

The in vitro antimalarial activity of the five known inhibitors of *Pf*TrxR (1,4-NQ, 2,4-DNPS, 4-NBT, 3-DAP, MD)[58,191] (Fig. 5.1 and Table 5.1) was evaluated against both CQ-sensitive (D6 clone) and CQ resistant (W2 clone) strains of *P. falciparum*, while cell cytotoxicity was determined against 3T3 cells (Table 5.1) using the procedure described earlier.

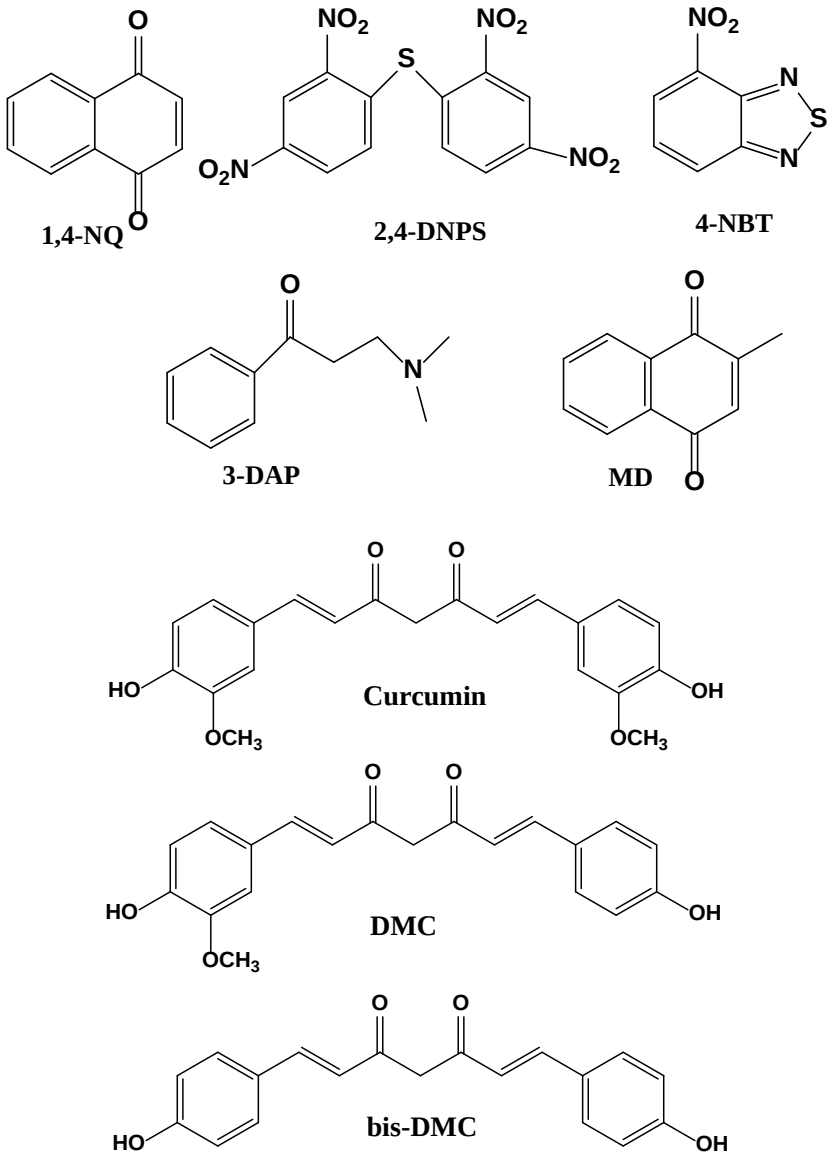


Figure 5.1 The *Pf*TrxR inhibitors.

The compounds 1,4-NQ and 4-NBT were found to be the most active against the two strains of P. falciparum, MD and 2,4-DNPS were moderately active, and 3-DAP was inactive. In terms of antiplasmodial activity against the W2 strain, 1,4-NQ and 4-NBT showed IC50 value of < 20 μM.

Table 5.1 *Pf*TrxR inhibitory and antiplasmodial activities of tested compounds

Test compounds	***Pf*TrxR IC_{50} (µM)**	***Pf* (D6) CQ sensitive IC_{50}(µM)**	**SI D6**	***Pf* (W2) CQ resistance IC_{50}(µM)**	**SI W2**	**3T3 IC_{50}(µM)**
1,4-NQ	0.75	8.9±2.3	4.6	16.7±3.7	2.4	38.5±0.76
2,4-DNPS	0.5	91.2±11.3	0.8	72.3±11.3	1.0	79±3.51
4-NBT	2	8.3±2.1	10	9.8±1.9	8	80±1.15
3-DAP	15.4	>100	>1	>100	>1	>100
MD	1.6	18.5±1.9	3.8	28.3±5.6	2.5	70.5±3.69
CQ		0.055±0.006		0.440±0.045		NC

*IC_{50} values, preparation of *Pf*TrxR and optimal experimental conditions for the *Pf*TrxR functional assay were reported in Andricopulo et al.[58,191,192].

**Values are mean ± S.D. of triplicate observations; SI: Selectivity index. NC: no cytotoxicity up to concentration much higher than the concentration responsible for its antiplasmodial activity.

The low correlation between the high *Pf*TrxR inhibitory activity and moderate antiplasmodial activity of 2,4-DNPS could be explained by an inability to penetrate the cell membranes. Accordingly, 2,4-DNPS is predicted to have poor Caco-2 and MDCK cell line permeability. Table 5.2 gives the computed octanol/water partition coefficient (logP), solubility in water (logS), polar surface area, and apparent Caco-2 and MDCK permeability for all compounds given in Figure 5.1, the enol form of 3-DAP, and CQ.

Table 5.2 Predicted physico-chemical properties of the compounds

Molecule	log P octanol/water (M)	log S aqueous solubility (M)	Polar Surface Area (PSA)	Apparent Caco-2 Permeability (nm/sec)	Apparent MDCK Permeability (nm/sec)
1,4 NQ	0.486	-0.764	57.199	954	470
2,4 DNPS	1.406	-3.444	177.135	5	2
4-NBT	0.746	-1.414	75.626	382	385
3-DAP	1.249	-0.313	32.353	819	441
3-DAP-enol	1.800	-1.170	25.296	842	454
MD	0.880	-1.312	55.456	1225	616
Chloroquine	4.276	-3.585	24	1364	1862
Range 95% of Drugs	**(-2.0 / 6.5)**	**(-6.5 / 0.5)**	**(7.0 / 330.0)**	**(<25 poor, >500 great)**	**(<25 poor, >500 great)**

The lack of antiplasmodial activity of 3-DAP, a Mannich base, may be due to (i) non-specific alkylation of cellular thiol groups, and also (ii) due to the absence of active transport to red blood cells and parasites. The correlation between inhibition of *Pf*TrxR in the enzyme inhibition assays and antiplasmodial activity in cell culture allows for a better evaluation of biological activities of inhibitor compounds. The active compounds namely, 1,4-NQ, 2,4-DNPS, 4-NBT and MD showed more toxicity than 3-DAP against the 3T3 cell line. The 3T3 cells are epithelial cells that reflect toxicity against proliferating mammalian cells.

In order to test the five *Pf*TrxR inhibitors for their ability to induce signs of oxidative stress by accelerated generation and accumulation of reactive oxygen intermediates (superoxide radical, hydroxyl radical and hydrogen peroxide)[189] the intraerythrocytic formation of ROS was monitored in real-time for 120 min with 2′7′-dichlorofluorescein diacetate (DCFDA), a fluorescent ROS probe[190]. Among the compounds tested 4-NBT, MD and 1,4-NQ caused a significant increase in oxidative stress (Fig. 5.2).

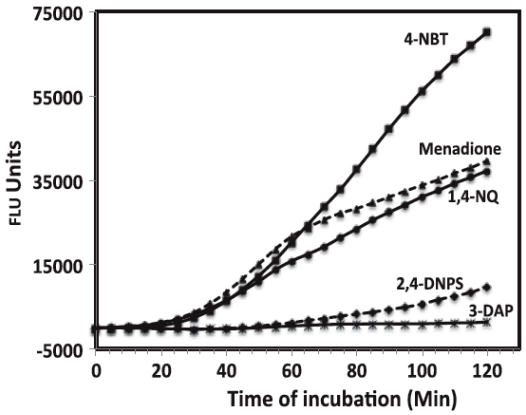


Figure 5.2 Formation of reactive oxygen species (ROS), as indicated by increase in fluorescence. DCFDA loaded human erythrocytes by five *Pf*TrxR inhibitors.

Whereas, 3-DAP and 2,4-DPNS did not cause the production of ROS. These results suggest that 4-NBT, MD and 1,4-NQ compromises the capability of erythrocytes to scavenge reactive oxygen intermediates. The accumulated intra-erythrocytic oxidative stress by these compounds may be responsible for the inhibition of *h*TrxR enzyme. The erythrocytes, our target cells, have higher capacity to produce oxidative stress than 3T3 cell line used for cytotoxicity assessment.

The 1,4-NQ chemical features and the ability to generate ·OH suggest the proficiency in altering intracellular redox status.[189] The antimalarial naphthoquinones (1,4-NQ and MD) are believed to perturb the major redox equilibria of the targeted *P. falciparum* infected red blood cells, which might be removed by macrophages. This perturbation results in development arrest and death of the malaria parasite at the trophozoite stage.[190]

Based on the report of Mulabagal and Caldero´n[193] regarding curcumin and DMC as *Pf*TrxR ligands, we tested these compounds for their ability to inhibit *Pf*TrxR. Curcumin and DMC displayed good binding affinity for the *Pf*TrxR target enzyme but curcumin displayed less than 50% inhibition of *Pf*TrxR at 10 μM when tested in the functional assay compared to DMC

which displayed more than 50% *Pf*TrxR inhibition at 10μM with an IC_{50} value of 2.0 μM. The fact that curcumin has been shown to inhibit rat TrxR by alkylating the cysteine/selenocysteine catalytic active site residues[194] supports our data and prediction that curcumin selectively inhibits mammalian TrxR but not the *Pf*TrxR. The in vitro antiplasmodial activity of curcumin and DMC was evaluated against both CQ-sensitive (D6 clone) and CQ-resistant (W2 clone) strains of *P. falciparum*, while cell cytotoxicity was determined against Vero cells (Table 5.3).

Table 5.3 Inhibition of *Pf*TrxR, predicted binding affinities and antiplasmodial activity of curcuminoids.

Test compounds	***Pf*TrxR IC_{50} (μM)**	**Computed dimer-interface affinity (kcal/mol)**	***Pf* (D6) CQ sensitive IC_{50} (μM)**	**SI D6**	***Pf* (W2) CQ resistance IC_{50} (μM)**	**SI W2**	**VERO IC_{50} (μM)**
Curcumin	NA	-9.4	15.9 ± 2.1	4.0	41.2 ± 6.2	1.6	64.6 ±1.73
DMC	2.03 ± 1.05	-9.6	17.7 ± 3.1	5.0	82.7 ± 10.3	1.1	89.7 ± 4.51
Bis-DMC	NA	-9.8	ND	ND	ND	ND	ND
CQ			0.055 ± 0.006		0.440 ± 0.045		NC

Notes: NA, not active i.e, *Pf*TrxR inhibition was < 50% at 10μM; ND, not determined; NC, no cytotoxicity up to concentration much higher than the concentration responsible for its antiplasmodial activity.

The two compounds were active against the D6 strain of *P. falciparum* and moderately active against the resistant W2 strain. The antiplasmodial activity of curcumin (IC_{50}: 4.21 μM) against CQ-resistant strain MP-14 of *P. falciparum* has been reported by Mishra et al.[195]. The difference in IC_{50} values between our results and others could be due to the use of different strains of the parasite, protocol to measure the parasite growth inhibition and the detector used for the measurement. Furthermore, the curcuminoids showed cytotoxicity against Vero cells

from 64 μM and selectivity index of four-six folds and onefold against D6 and W2, respectively. Curcumin and DMC were tested for their ability to induce signs of oxidative stress by accelerated generation and accumulation of reactive oxygen intermediates (superoxide radical, hydroxyl radical and hydrogen peroxide). The intraerythrocytic formation of ROS was monitored in real time for 120 min with 2070-dichlorofluorescein diacetate, a fluorescent ROS probe. Neither of these two compounds was able to increase the ROS in healthy erythrocytes by a potential inhibition of mammalian TrxR.

5.3.2 Binding studies

Since these compounds were active against *Pf*TrxR as well, molecular docking was used to study their interactions with *Pf*TrxR to gain further insight into the mode of interaction for these molecules. MD[196], DNPS and 4-NBT have been proposed to bind at the intersubunit region in *Pf*TrxR's. However, 1,4-NQ and 3-DAP bind to the reductase covalently precluding the use of docking calculations. For example, 1,4-NQ is an inhibitor of TrxR that behaves as a subversive substrate.[196] The compound 3-DAP inactivates TrxR by alkylating the C-terminal redox active catalytic Cys-Cys pair.

Table 5.4 Comparison between computed binding affinities at the dimer interface in *Pf*TrxR and experimental IC_{50} values

Molecule	**Computed Binding Affinity (kcal/mol)**	**Exptl. *Pf*TrxR IC_{50}(μM)**
2,4-DNPS	-8.4	0.5
MD	-7.9	1.6
4-NBT	-6.0	2

This is achieved by the formation of a reactive α, β-unsaturated ketone intermediate after it undergoes deamination in solution.[191] Therefore 3-DAP acts an alkylator. The calculations predict the same activity trend observed in the experimental IC50 values for the non-covalent inhibitors, 2,4-DNPS, MD, and 4-NBT in PfTrxR (Table 5.4).

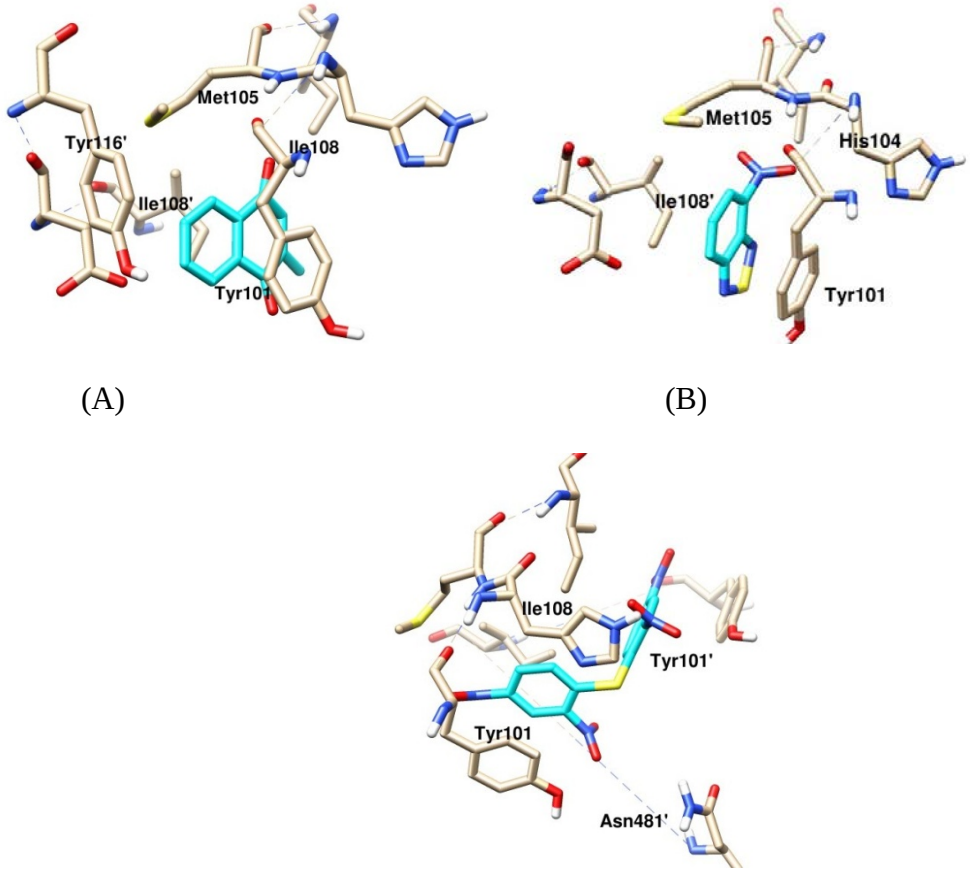


Figure 5.3 The predicted binding poses for the inhibitors showing the main interactions with the dimer interface residues. (A) *Pf*TrxR/MD, (B) *Pf*TrxR/4-NBT, and (C) *Pf*TrxR/2, 4-DNPS complexes.

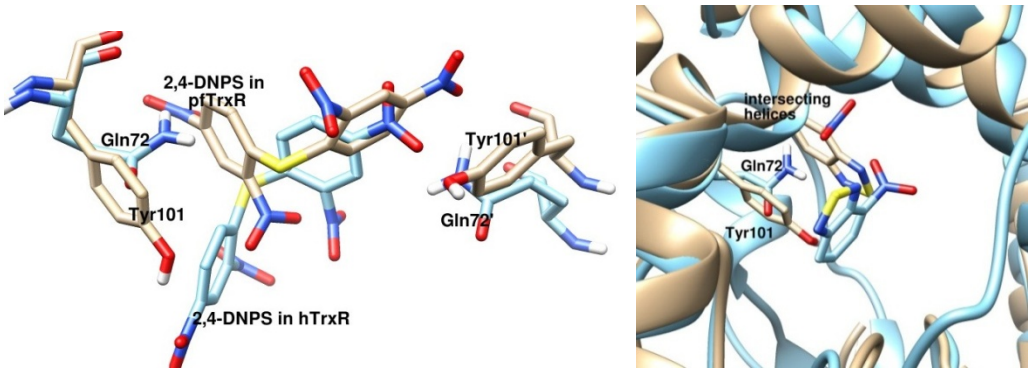
For the *Pf*TrxR/MD complex, pi stacking interactions are predicted to form between the inhibitor's phenyl ring and Tyr101 side chain ring. The backbone nitrogen of Met105 is in close proximity to the carbonyl group of MD; however, the predicted angle between N-H and O of 85°

impedes hydrogen bonding. The molecule further forms hydrophobic interactions with the phenyl ring of Tyr116′ and the side chains of Ile108 from both subunits (Fig. 5.3). Similar to MD, 4-NBT's phenyl ring also has pi-pi stacking with Tyr101's phenyl ring, but forms hydrophobic interactions only with Ile108 from subunit B in the large cavity. The nitro group causes the molecule to twist subtly compared to MD in order to better interact with the electrostatic surface created by the peptide bond between His104 and Met105's and sulfur (Figure 3). Compared to the size of the cavity, MD and 4-NBT are small molecules and do not fully interact with most of the residues lining the wall of the dimer interface. 2,4-DNPS forms the only electrostatic interaction at a distance of 3.9 Å with Asn481′. Pi stacking interactions are formed between one of the inhibitor's phenyl rings and Tyr101 side chain ring with the other phenyl ring of the molecule forming a parallel displaced pi stacking interaction with Tyr101′ (subunit B). As with MD and 4-NBT, the side chains of Ile108 from both subunits form hydrophobic interactions with 2,4-DNPS. Most of the interactions the three molecules are forming with the proteins are with the intersecting helices between the two subunits of the enzymes.

Table 5.5 Comparison between computed binding affinities at the dimer interface and experimental IC_{50} values in *Pf*TrxR and *h*TrxR.

Molecule		***Pf*TrxR**	***h*TrxR**
2,4-DNPS	Exptl. IC_{50}(μM)	0.5	4
	Calc. Binding Affinity (kcal/mol)	-8.4	-8.1
4-NBT	Exptl. IC_{50}(μM)	2	50
	Calc. Binding Affinity (kcal/mol)	-6.0	-5.7

The experimental activities for 2,4-DNPS and 4-NBT show selectivity between the parasite and human isoform of thioredoxin reductase (Table 5.5). The experimental values for 2,4-DNPS show an 8-fold selectivity for *Pf*TrxR, whereas 4-NBT has a 25-fold selectivity. While the docking simulations correctly predicted binding trends, limitations in the method, including potentially inaccurate scoring functions, the use of rigid proteins, and a lack of solvation could have contributed to its inability to reproduce the large differences observed in the IC_{50} values. The docked poses, however, have considerable differences within the cavity, which could point to the observed selectivity (Fig. 5.4).



(A) 2,4-DNPS (B) 4-NBT

Figure 5.4 The docking pose differences of (A) 2,4-DNPS and (B) 4-NBT between the *Pf*-(brown) and *h*-TrxR (blue). The docked structures point to the difference in conformation between Try101 (*pf*TrxR) and Glu72 (*h*TrxR) which is proposed to have significant contribution to the observed experimental selectivity.

The presence of Tyr101 in *Pf*TrxR enables 2,4-DNPS to form a favorable pi stacking interaction with the phenyl ring of the molecule, whereas its counterpart in *h*TrxR is a Gln72 that orients the molecule to avoid steric clashes. This results in the second ring of the molecule forming a parallel displaced pi stacking interaction with Tyr101′, whereas a hydrogen bond between the

nitro group and Gln72' in the *h*TrxR is realized. The effect of this substitution on 4-NBT seems to be the fact that the presence of Gln72 pushes the molecule deep into the large cavity precluding the interaction with the residues of the intersecting helices between the subunits. 1,4-NQ and 4-NBT can be considered to be attractive leads for further optimization as these compounds display good *Pf*TrxR inhibitory and antiplasmodial activity. A thorough examination of the residues making any form of interaction with the small molecules showed that no other, including His104 (*Pf*TrxR) and its counterpart in *h*TrxR (Leu75), influences the differences in binding between the parasite and human isoform. Figure 5.5 shows 2,4-DNPS docked in both proteins especially showing the positions of the His104 and Leu75 as an example.

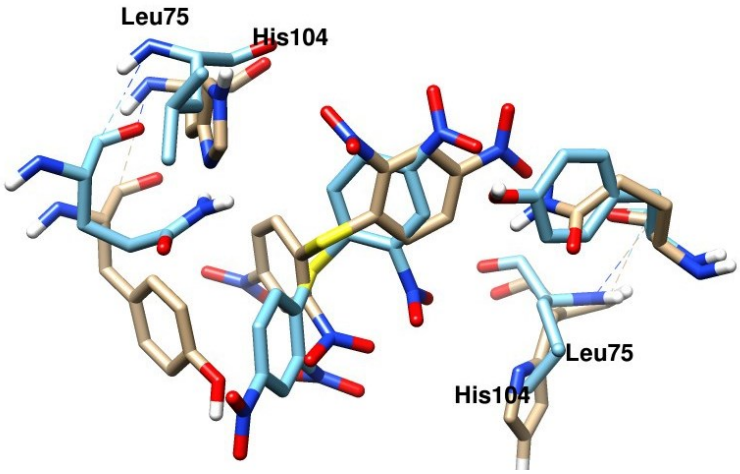


Figure 5.5 The predicted binding mode of 2,4-DNPS in the *Pf*TrxR (beige) and *h*TrxR (blue) showing the position of the residues Try101 (PfTrxR) and Glu72 (hTrxR).

To validate the docking protocols, the bound cofactor, flavin adenine dinucleotide (FAD), was re-docked as the control (with crystallographically bound water molecules included) as the structure used does not have a bound ligand and there are no alternative crystal structures of the target protein. By visual inspection, the flexible docking protocols of AutoDock Vina reasonably

reproduced the experimental binding pose of FAD indicating a good accuracy for the parameters used in the present docking methodology (Figure S12).

Curcumin and DMC were predicted to bind to the dimer interface of *Pf*TrxR at the intersecting helices between the subunits. An alternative curcuminoid, bis-DMC, that lacks the methoxy substituents on the phenyl moieties (Fig. 5.1) was also found to favour the *Pf*TrxR dimer interface. Intriguingly, all three curcuminoids were predicted to bind in a nearly identical fashion, regardless of the presence or lack of methoxy substituents (Fig. 5.6).

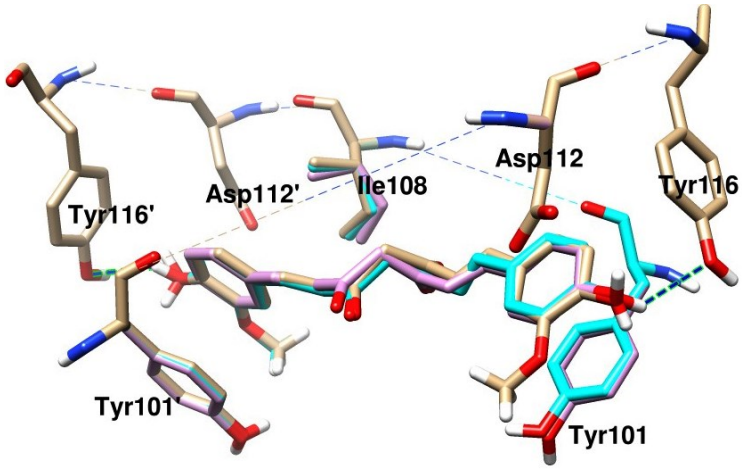


Figure 5.6 The interactions between DMC (cyan), curcumin (tan) and bis-DMC (purple) with *Pf*TrxR.

Examination of the crystal structure[55] reveals that the residues from the intersecting helices that interact with the ligands are Leu98, Tyr101, Ala102, His104, Met105, Ile108, Asp112, Tyr116 and Pro480 from both subunits A and B. It is evident that the first phenol moiety (demethoxylated in DMC) interacts with the residues on subunit A, whereas the other moiety extends towards subunit B (Figure S13). The results suggest that the methoxy groups may contribute to the difference in inhibition towards *Pf*TrxR. In curcumin, the ortho-methoxy group can form an intramolecular hydrogen bond with the phenolic hydrogen, making the H-

atom abstraction from the ortho-methoxy phenols surprisingly easy. However, elucidating how phenyl methoxy groups mediate inhibition of PfTrxR is more difficult.

Earlier theoretical studies have shown that hydrogen bonding between ortho-methoxy oxygen and phenolic hydrogen in curcumin influences the planarity, conformation and ability to undergo oxidation.[197] From the current docking simulations, the phenyl moiety formed a pi–pi stacking interaction with the Tyr101 from each respective subunit. The hydroxyl group in DMC formed a hydrogen bond with a distance of 2.5Å to the oxygen of Tyr116 in subunit A, albeit at an angle of 110˚. Curcumin's OH group is bent in the opposite direction because of the conformational change, and thus, the H bond is not realized. For subunit B, the opposite trend occurs. Curcumin forms a 2.4Å H bond with the Tyr 101' at an angle of 138˚ and DMC's OH group bends to avoid the interaction (Fig. 5.6). The hydrophobic a, b-unsaturated chain interacts with the side chain of Ile 108 from subunit A. The experimental data show a marked difference between the two compounds with respect to *Pf*TrxR inhibitory activity, whereas docking analysis predicted indistinguishable binding affinities despite the subtle differences in their binding poses (Table 5.3).

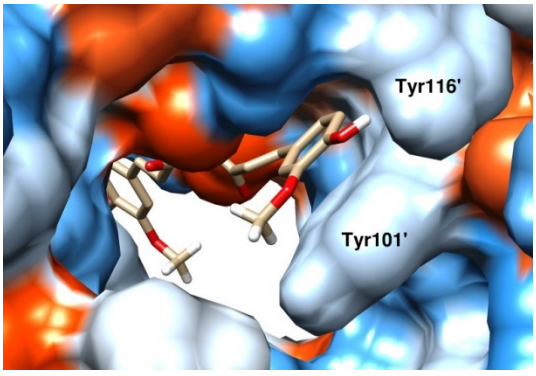


Figure 5.7 Hydrophobicity surface for curcumin-docked *Pf*TrxR. Color scale: white to blue for the most hydrophilic residues and orange to red for the most hydrophobic residues.

The methoxyphenyl moiety of the curcuminoids is shown to lay in a narrow groove in the dimer inter-subunit interface, sandwiched by Tyr101 and Tyr116, thereby locking the rings into position (Fig. 5.7).

The structural differences in the curcuminoids could point to the differences in experimental values. For example, the pi–pi interactions between the phenyl moieties of the curcuminoids and the Tyr residues (Fig. 5.6) suggest that the methoxy groups on curcumin and DMC are parallel to the hydroxyl group of Tyr 101 and may be less favorable due to poorer sterics. The lack of a methoxy group in bis-DMC would not incur the same energetic penalty, and therefore, the binding affinity is slightly enhanced according to the calculations. Comparisons to the mammalian TrxR are difficult to make as curcumin binds covalently[194], precluding the use of docking calculations. It may be reasonable to assume that curcuminoid derivatives DMC and bis-DMC would also bind covalently to TrxR. Based on the little correlation between the results from phenotypic screening and docking studies of curcumin and DMC to support the activity through *Pf*TrxR inhibition, the antiplasmodial activity might be due to other mechanism of actions such as inhibition of $Ca2^{+}$ - ATPase (*Pf*ATP6)[198], S-adenosyl-L-homocysteine hydrolase (*Pf*SAHH)[199] or any other target.

5.4 Conclusions

In this study, tools for the identification of *Pf*TrxR inhibitors using phenotypic screening and docking studies have been validated for their potential use for antimalarial drug discovery project.

References

(1) Bachand, F. *Eukaryotic cell* **2007**, *6*, 889.

(2) Esse, R.; Leandro, P.; Rivera, I.; Almeida, I. T. d.; Blom, H. J.; Castro, R. *Deciphering Protein Arginine Methylation in Mammals*, 2012.

(3) Chen, C.; Nott, T. J.; Jin, J.; Pawson, T. *Nature reviews. Molecular cell biology* **2011**, *12*, 629.

(4) Struck, A. W.; Thompson, M. L.; Wong, L. S.; Micklefield, J. *Chembiochem : a European journal of chemical biology* **2012**, *13*, 2642.

(5) Gui, S.; Gathiaka, S.; Li, J.; Qu, J.; Acevedo, O.; Hevel, J. M. *The Journal of biological chemistry* **2014**, *289*, 9320.

(6) Schubert, H. L.; Blumenthal, R. M.; Cheng, X. *Trends in biochemical sciences* **2003**, *28*, 329.

(7) Leiper, J.; Nandi, M.; Torondel, B.; Murray-Rust, J.; Malaki, M.; O'Hara, B.; Rossiter, S.; Anthony, S.; Madhani, M.; Selwood, D.; Smith, C.; Wojciak-Stothard, B.; Rudiger, A.; Stidwill, R.; McDonald, N. Q.; Vallance, P. *Nature medicine* **2007**, *13*, 198.

(8) Chen, D.; Ma, H.; Hong, H.; Koh, S. S.; Huang, S. M.; Schurter, B. T.; Aswad, D. W.; Stallcup, M. R. *Science* **1999**, *284*, 2174.

(9) Hong, H.; Kao, C.; Jeng, M. H.; Eble, J. N.; Koch, M. O.; Gardner, T. A.; Zhang, S.; Li, L.; Pan, C. X.; Hu, Z.; MacLennan, G. T.; Cheng, L. *Cancer* **2004**, *101*, 83.

(10) Majumder, S.; Liu, Y.; Ford, O. H., 3rd; Mohler, J. L.; Whang, Y. E. *The Prostate* **2006**, *66*, 1292.

(11) Wolf, S. S. *Cellular and molecular life sciences : CMLS* **2009**, *66*, 2109.

(12) Miranda, T. B.; Sayegh, J.; Frankel, A.; Katz, J. E.; Miranda, M.; Clarke, S. *The Biochemical journal* **2006**, *395*, 563.

(13) Gui, S.; Wooderchak, W. L.; Daly, M. P.; Porter, P. J.; Johnson, S. J.; Hevel, J. M. *The Journal of biological chemistry* **2011**, *286*, 29118.

(14) Sun, L.; Wang, M.; Lv, Z.; Yang, N.; Liu, Y.; Bao, S.; Gong, W.; Xu, R. M. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 20538.

(15) Zhang, X.; Cheng, X. *Structure* **2003**, *11*, 509.

(16) Rust, H. L.; Zurita-Lopez, C. I.; Clarke, S.; Thompson, P. R. *Biochemistry* **2011**, *50*, 3332.

(17) Olefsky, J. M. *The Journal of biological chemistry* **2001**, *276*, 36863.

(18) Robinson-Rechavi, M.; Escriva Garcia, H.; Laudet, V. *Journal of cell science* **2003**, *116*, 585.

(19) Huang, P.; Chandra, V.; Rastinejad, F. *Annual review of physiology* **2010**, *72*, 247.

(20) Kastner, P.; Mark, M.; Chambon, P. *Cell* **1995**, *83*, 859.

(21) Mangelsdorf, D. J.; Evans, R. M. *Cell* **1995**, *83*, 841.

(22) Shimizu, M.; Moriwaki, H. *PPAR research* **2008**, *2008*, 181047.

(23) Peters, J. M.; Shah, Y. M.; Gonzalez, F. J. *Nature reviews. Cancer* **2012**, *12*, 181.

(24) Gabaldon, T. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **2010**, *365*, 765.

(25) Issemann, I.; Green, S. *Nature* **1990**, *347*, 645.

(26) van Bilsen, M.; van der Vusse, G.; Gilde, A.; Lindhout, M.; van der Lee, K. J. M. *Mol Cell Biochem* **2002**, *239*, 131.

(27) Ferre, P. *Diabetes* **2004**, *53 Suppl 1*, S43.
(28) Grygiel-Gorniak, B. *Nutrition journal* **2014**, *13*, 17.
(29) Kersten, S. *European journal of pharmacology* **2002**, *440*, 223.
(30) Bishop-Bailey, D.; Wray, J. *Prostaglandins & other lipid mediators* **2003**, *71*, 1.
(31) Gervois, P.; Torra, I. P.; Fruchart, J. C.; Staels, B. *Clinical chemistry and laboratory medicine : CCLM / FESCC* **2000**, *38*, 3.
(32) Sundvold, H.; Lien, S. *Biochem Biophys Res Commun* **2001**, *287*, 383.
(33) Berger, J.; Moller, D. E. *Annual review of medicine* **2002**, *53*, 409.
(34) Artis, D. R.; Lin, J. J.; Zhang, C.; Wang, W.; Mehra, U.; Perreault, M.; Erbe, D.; Krupka, H. I.; England, B. P.; Arnold, J.; Plotnikov, A. N.; Marimuthu, A.; Nguyen, H.; Will, S.; Signaevsky, M.; Kral, J.; Cantwell, J.; Settachatgull, C.; Yan, D. S.; Fong, D.; Oh, A.; Shi, S.; Womack, P.; Powell, B.; Habets, G.; West, B. L.; Zhang, K. Y.; Milburn, M. V.; Vlasuk, G. P.; Hirth, K. P.; Nolop, K.; Bollag, G.; Ibrahim, P. N.; Tobin, J. F. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 262.
(35) Nolte, R. T.; Wisely, G. B.; Westin, S.; Cobb, J. E.; Lambert, M. H.; Kurokawa, R.; Rosenfeld, M. G.; Willson, T. M.; Glass, C. K.; Milburn, M. V. *Nature* **1998**, *395*, 137.
(36) Cronet, P.; Petersen, J. F.; Folmer, R.; Blomberg, N.; Sjoblom, K.; Karlsson, U.; Lindstedt, E. L.; Bamberg, K. *Structure* **2001**, *9*, 699.
(37) Xu, H. E.; Stanley, T. B.; Montana, V. G.; Lambert, M. H.; Shearer, B. G.; Cobb, J. E.; McKee, D. D.; Galardi, C. M.; Plunket, K. D.; Nolte, R. T.; Parks, D. J.; Moore, J. T.; Kliewer, S. A.; Willson, T. M.; Stimmel, J. B. *Nature* **2002**, *415*, 813.
(38) Gampe, R. T., Jr.; Montana, V. G.; Lambert, M. H.; Miller, A. B.; Bledsoe, R. K.; Milburn, M. V.; Kliewer, S. A.; Willson, T. M.; Xu, H. E. *Molecular cell* **2000**, *5*, 545.
(39) Takada, I.; Yu, R. T.; Xu, H. E.; Lambert, M. H.; Montana, V. G.; Kliewer, S. A.; Evans, R. M.; Umesono, K. *Molecular endocrinology* **2000**, *14*, 733.
(40) Fyffe, S. A.; Alphey, M. S.; Buetow, L.; Smith, T. K.; Ferguson, M. A.; Sorensen, M. D.; Bjorkling, F.; Hunter, W. N. *Molecular cell* **2006**, *21*, 1.
(41) Zoete, V.; Grosdidier, A.; Michielin, O. *Biochimica et biophysica acta* **2007**, *1771*, 915.
(42) Willson, T. M.; Brown, P. J.; Sternbach, D. D.; Henke, B. R. *Journal of medicinal chemistry* **2000**, *43*, 527.
(43) Batista, F. A. H.; Trivella, D. B. B.; Bernardes, A.; Gratieri, J.; Oliveira, P. S. L.; Figueira, A. C. M.; Webb, P.; Polikarpov, I. *PLoS One* **2012**, *7*, e33643.
(44) Lewis, S. N.; Bassaganya-Riera, J.; Bevan, D. R. *PPAR research* **2010**, *2010*, 10.
(45) Gardner, M. J.; Hall, N.; Fung, E.; White, O.; Berriman, M.; Hyman, R. W.; Carlton, J. M.; Pain, A.; Nelson, K. E.; Bowman, S.; Paulsen, I. T.; James, K.; Eisen, J. A.; Rutherford, K.; Salzberg, S. L.; Craig, A.; Kyes, S.; Chan, M. S.; Nene, V.; Shallom, S. J.; Suh, B.; Peterson, J.; Angiuoli, S.; Pertea, M.; Allen, J.; Selengut, J.; Haft, D.; Mather, M. W.; Vaidya, A. B.; Martin, D. M.; Fairlamb, A. H.; Fraunholz, M. J.; Roos, D. S.; Ralph, S. A.; McFadden, G. I.; Cummings, L. M.; Subramanian, G. M.; Mungall, C.; Venter, J. C.; Carucci, D. J.; Hoffman, S. L.; Newbold, C.; Davis, R. W.; Fraser, C. M.; Barrell, B. *Nature* **2002**, *419*, 498.
(46) WHO **2014**.
(47) Crompton, P. D.; Pierce, S. K.; Miller, L. H. *The Journal of clinical investigation* **2010**, *120*, 4168.
(48) Mendis, K.; Rietveld, A.; Warsame, M.; Bosman, A.; Greenwood, B.; Wernsdorfer, W. H. *Tropical medicine & international health : TM & IH* **2009**, *14*, 802.

(49) Alonso, P. L.; Brown, G.; Arevalo-Herrera, M.; Binka, F.; Chitnis, C.; Collins, F.; Doumbo, O. K.; Greenwood, B.; Hall, B. F.; Levine, M. M.; Mendis, K.; Newman, R. D.; Plowe, C. V.; Rodriguez, M. H.; Sinden, R.; Slutsker, L.; Tanner, M. *PLoS medicine* **2011**, *8*, e1000406.

(50) Francis, S. E.; Sullivan, D. J., Jr.; Goldberg, D. E. *Annual review of microbiology* **1997**, *51*, 97.

(51) Sturm, A.; Amino, R.; van de Sand, C.; Regen, T.; Retzlaff, S.; Rennenberg, A.; Krueger, A.; Pollok, J. M.; Menard, R.; Heussler, V. T. *Science* **2006**, *313*, 1287.

(52) Su, X.; Hayton, K.; Wellems, T. E. *Nature reviews. Genetics* **2007**, *8*, 497.

(53) Florens, L.; Washburn, M. P.; Raine, J. D.; Anthony, R. M.; Grainger, M.; Haynes, J. D.; Moch, J. K.; Muster, N.; Sacci, J. B.; Tabb, D. L.; Witney, A. A.; Wolters, D.; Wu, Y.; Gardner, M. J.; Holder, A. A.; Sinden, R. E.; Yates, J. R.; Carucci, D. J. *Nature* **2002**, *419*, 520.

(54) Sarma, G. N.; Savvides, S. N.; Becker, K.; Schirmer, M.; Schirmer, R. H.; Karplus, P. A. *Journal of molecular biology* **2003**, *328*, 893.

(55) Boumis, G.; Giardina, G.; Angelucci, F.; Bellelli, A.; Brunori, M.; Dimastrogiovanni, D.; Saccoccia, F.; Miele, A. E. *Biochem Biophys Res Commun* **2012**, *425*, 806.

(56) Wrenger, C.; Schettert, I.; Liebau, E. *Oxidative Stress in Human Infectious Diseases – Present and Current Knowledge About Its Druggability*, 2013.

(57) Schirmer, R. H.; Müller, J. G.; Krauth-Siegel, R. L. *Angewandte Chemie International Edition in English* **1995**, *34*, 141.

(58) Andricopulo, A. D.; Akoachere, M. B.; Krogh, R.; Nickel, C.; McLeish, M. J.; Kenyon, G. L.; Arscott, L. D.; Williams, C. H., Jr.; Davioud-Charvet, E.; Becker, K. *Bioorganic & medicinal chemistry letters* **2006**, *16*, 2283.

(59) Snider, G. W.; Dustin, C. M.; Ruggles, E. L.; Hondal, R. J. *Biochemistry* **2014**, *53*, 601.

(60) Fritz-Wolf, K.; Kehr, S.; Stumpf, M.; Rahlfs, S.; Becker, K. *Nature communications* **2011**, *2*, 383.

(61) Theobald, A. J.; Caballero, I.; Coma, I.; Colmenarejo, G.; Cid, C.; Gamo, F. J.; Hibbs, M. J.; Bass, A. L.; Thomas, D. A. *Biochemistry* **2012**, *51*, 4764.

(62) Karplus, M.; McCammon, J. A. *Nature structural biology* **2002**, *9*, 646.

(63) van Gunsteren, W. F.; Berendsen, H. J. C. *Angewandte Chemie International Edition in English* **1990**, *29*, 992.

(64) Karplus, M.; Kuriyan, J. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 6679.

(65) Alder, B. J.; Wainwright, T. E. *The Journal of Chemical Physics* **1957**, *27*, 1208.

(66) Rahman, A. *Physical Review* **1964**, *136*, A405.

(67) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585.

(68) Khalili-Araghi, F.; Ziervogel, B.; Gumbart, J. C.; Roux, B. *The Journal of general physiology* **2013**, *142*, 465.

(69) Dolenc, J.; Missimer, J. H.; Steinmetz, M. O.; van Gunsteren, W. F. *Journal of biomolecular NMR* **2010**, *47*, 221.

(70) Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *Journal of computational chemistry* **2005**, *26*, 1668.

(71) Leach, A. R. *2nd ed.; Pearson Prentice Hall: Essex* **2001**.

(72) Cramer, C. *John Wiley & Sons: West Sussex* **2004**.

(73) Meller, J. *ENCYCLOPEDIA OF LIFE SCIENCES Nature Publishing Group* **2001**.

(74) Verlet, L. *Physical Review* **1967**, *159*, 98.

(75) Hockney, R. W. *Methods in Computational Physics* **1970**, *9*, 136.

(76) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *Journal of Computational Physics* **1977**, *23*, 327.

(77) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *Journal of computational chemistry* **2004**, *25*, 1157.

(78) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *Journal of the American Chemical Society* **1995**, *117*, 5179.

(79) Ewald, P. P. *Ann. Phys.* **1921**, *369*, 253.

(80) Darden, T.; York, D.; Pedersen, L. *The Journal of Chemical Physics* **1993**, *98*, 10089.

(81) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *The Journal of Chemical Physics* **1953**, *21*, 1087.

(82) Hamelberg, D.; Mongan, J.; McCammon, J. A. *J Chem Phys* **2004**, *120*, 11919.

(83) Sugita, Y.; Okamoto, Y. *Chemical Physics Letters* **1999**, *314*, 141.

(84) Wu, X.; Wang, S. *The Journal of Chemical Physics* **1999**, *110*, 9401.

(85) Grubmuller, H. *Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics* **1995**, *52*, 2893.

(86) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2013**, *3*, 198.

(87) Case, D. A.; et, a. l. *San Francisco, CA: University of California* **2012**.

(88) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham Iii, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Computer Physics Communications* **1995**, *91*, 1.

(89) Krieger, E.; Nabuurs, S. B.; Vriend, G. In *Structural Bioinformatics*; John Wiley & Sons, Inc.: 2005, p 509.

(90) Baker, D.; Sali, A. *Science* **2001**, *294*, 93.

(91) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M. Y.; Pieper, U.; Sali, A. *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]* **2007**, *Chapter 2*, Unit 2 9.

(92) Rost, B. *Protein engineering* **1999**, *12*, 85.

(93) Chothia, C.; Lesk, A. M. *The EMBO journal* **1986**, *5*, 823.

(94) Xu, D.; Xu, Y.; Uberbacher, E. C. *Current protein & peptide science* **2000**, *1*, 1.

(95) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic acids research* **2000**, *28*, 235.

(96) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *Journal of molecular biology* **1990**, *215*, 403.

(97) Lobo, I. *Nature Education* **2008**, *1*, 215.

(98) Sali, A.; Blundell, T. L. *Journal of molecular biology* **1993**, *234*, 779.

(99) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *Journal of Applied Crystallography* **1993**, *26*, 283.

(100) Hooft, R. W.; Vriend, G.; Sander, C.; Abola, E. E. *Nature* **1996**, *381*, 272.

(101) Vriend, G. *Journal of molecular graphics* **1990**, *8*, 52.
(102) Wiederstein, M.; Sippl, M. J. *Nucleic acids research* **2007**, *35*, W407.
(103) Eisenberg, D.; Luthy, R.; Bowie, J. U. *Methods in enzymology* **1997**, *277*, 396.
(104) Colovos, C.; Yeates, T. O. *Protein science : a publication of the Protein Society* **1993**, *2*, 1511.
(105) Jorgensen, W. L. *Science* **2004**, *303*, 1813.
(106) Mobley, D. L.; Dill, K. A. *Structure*, *17*, 489.
(107) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. *Journal of computational chemistry* **2009**, *30*, 2785.
(108) Jain, A. N. *Journal of medicinal chemistry* **2003**, *46*, 499.
(109) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *Journal of medicinal chemistry* **2004**, *47*, 1739.
(110) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *Journal of molecular biology* **1997**, *267*, 727.
(111) Gilson, M. K.; Zhou, H. X. *Annual review of biophysics and biomolecular structure* **2007**, *36*, 21.
(112) Chang, C.-e. A.; Chen, W.; Gilson, M. K. *Proceedings of the National Academy of Sciences* **2007**, *104*, 1534.
(113) Meng, X. Y.; Zhang, H. X.; Mezei, M.; Cui, M. *Current computer-aided drug design* **2011**, *7*, 146.
(114) Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. *Journal of computational chemistry* **2007**, *28*, 1145.
(115) Trott, O.; Olson, A. J. *Journal of computational chemistry* **2010**, *31*, 455.
(116) Jain, A. N. *Journal of computer-aided molecular design* **2007**, *21*, 281.
(117) Antonysamy, S.; Bonday, Z.; Campbell, R. M.; Doyle, B.; Druzina, Z.; Gheyi, T.; Han, B.; Jungheim, L. N.; Qian, Y.; Rauch, C.; Russell, M.; Sauder, J. M.; Wasserman, S. R.; Weichert, K.; Willard, F. S.; Zhang, A.; Emtage, S. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109*, 17960.
(118) Parr, R. G.; Yang, W.
(119) Ochterski, J. W.; Petersson, G. A.; Montgomery, J. A. *The Journal of Chemical Physics* **1996**, *104*, 2598.
(120) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.; Gaussian, Inc.: Wallingford, CT, USA, 2009.
(121) Becke, A. D. *The Journal of Chemical Physics* **1993**, *98*, 5648.

(122) Lee, C.; Yang, W.; Parr, R. G. *Physical review. B, Condensed matter* **1988**, *37*, 785.

(123) Zhao, Y.; Truhlar, D. *Theor Chem Account* **2008**, *120*, 215.

(124) Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. *Nucleic acids research* **1997**, *25*, 3389.

(125) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham Iii, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput Phys Commun* **1995**, *91*, 1.

(126) Armacost, K.; Musila, J.; Gathiaka, S.; Ellis, H. R.; Acevedo, O. *Biochemistry* **2014**, *53*, 3308.

(127) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of Chemical Physics* **1983**, *79*, 926.

(128) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *Journal of computational chemistry* **2004**, *25*, 1157.

(129) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712.

(130) Papamokos, G. V.; Tziatzos, G.; Papageorgiou, D. G.; Georgatos, S. D.; Politou, A. S.; Kaxiras, E. *Biophysical journal* **2012**, *102*, 1926.

(131) Gotz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. *Journal of chemical theory and computation* **2012**, *8*, 1542.

(132) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J Comput Phys* **1977**, *23*, 327.

(133) Toukmaji, A.; Sagui, C.; Board, J.; Darden, T. *The Journal of Chemical Physics* **2000**, *113*, 10913.

(134) Roe, D. R.; Cheatham, T. E. *Journal of chemical theory and computation* **2013**, *9*, 3084.

(135) Chu, Y.; Yao, J.; Guo, H. *PLoS One* **2012**, *7*, e37674.

(136) Hu, P.; Zhang, Y. *J Am Chem Soc* **2006**, *128*, 1272.

(137) Yao, J.; Chu, Y.; An, R.; Guo, H. *Journal of chemical information and modeling* **2012**, *52*, 449.

(138) Xu, Q.; Chu, Y. Z.; Guo, H. B.; Smith, J. C.; Guo, H. *Chemistry* **2009**, *15*, 12596.

(139) Zhang, R.; Li, X.; Liang, Z.; Zhu, K.; Lu, J.; Kong, X.; Ouyang, S.; Li, L.; Zheng, Y. G.; Luo, C. *PLoS One* **2013**, *8*, e72424.

(140) Gunaydin, H.; Acevedo, O.; Jorgensen, W. L.; Houk, K. N. *Journal of chemical theory and computation* **2007**, *3*, 1028.

(141) Osborne, T. C.; Obianyo, O.; Zhang, X.; Cheng, X.; Thompson, P. R. *Biochemistry* **2007**, *46*, 13370.

(142) Laskowski, R. A. M., M. W.; Moss, D. S.; Thornton, J. M. J. In *J. Appl. Cryst.* 1993; Vol. 26, p 283.

(143) Rossi, K. A.; Weigelt, C. A.; Nayeem, A.; Krystek, S. R., Jr. *Protein science : a publication of the Protein Society* **2007**, *16*, 1999.

(144) Jorgensen, W. L. *Yale University: New Haven, CT.* **2004**.

(145) Wolfenden, R.; Snider, M. J. *Accounts of chemical research* **2001**, *34*, 938.

(146) Obianyo, O.; Osborne, T. C.; Thompson, P. R. *Biochemistry* **2008**, *47*, 10420.

(147) Kolbel, K.; Ihling, C.; Bellmann-Sickert, K.; Neundorf, I.; Beck-Sickinger, A. G.; Sinz, A.; Kuhn, U.; Wahle, E. *The Journal of biological chemistry* **2009**, *284*, 8274.

(148) Wooderchak, W. L. *Dissertation Utah State University, Utah* **2009**.

(149) Gallivan, J. P.; Dougherty, D. A. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 9459.

(150) Collins, R. E.; Tachibana, M.; Tamaru, H.; Smith, K. M.; Jia, D.; Zhang, X.; Selker, E. U.; Shinkai, Y.; Cheng, X. *The Journal of biological chemistry* **2005**, *280*, 5563.

(151) Artisa, D. R.; Lina, J. J.; Zhanga, C.; Wanga, W.; Mehraa, U.; Perreaultb, M.; Erbeb, D.; Krupkaa, H. I.; Englanda, B. P.; Arnolda, J.; Plotnikova, A. N.; Marimuthua, A.; Nguyena, H.; Willb, S.; Signaevskyc, M.; Kralc, J.; Cantwella, J.; Settachatgulla, C.; Yana, D. S.; Fonga, D.; Oha, A.; Shia, S.; Womacka, P.; Powella, B.; Habetsa, G.; Westa, B. L.; Zhanga, K. Y. J.; Milburna, M. V.; Vlasukb, G. P.; Hirtha, K. P.; Nolopa, K.; Bollaga, G.; Ibrahima, P. N.; Tobinb, J. F. *Proc. Nat. Acad. Sci. USA* **2009**, *106*, 262.

(152) Lago, R. M.; Singh, P. P.; Nesto, R. W. *Lancet* **2007**, *370*, 1129.

(153) Wang, Y. X.; Zhang, C. L.; Yu, R. T.; Cho, H. K.; Nelson, M. C.; Bayuga-Ocampo, C. R.; Ham, J.; Kang, H.; Evans, R. M. *PLoS Biol* **2004**, *2*, e294.

(154) Dimopoulos, N.; Watson, M.; Green, C.; Hundal, H. S. *FEBS letters* **2007**, *581*, 4743.

(155) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. *Journal of computational chemistry* **2009**, *30*, 2785.

(156) Trott, O.; Olson, A. J. *Journal of computational chemistry*, *31*, 455.

(157) Jain, A. N. *J. Med. Chem.* **2003**, *46*, 499.

(158) Connors, R. V.; Wang, Z.; Harrison, M.; Zhang, A.; Wanska, M.; Hiscock, S.; Fox, B.; Dore, M.; Labelle, M.; Sudom, A.; Johnstone, S.; Liu, J.; Walker, N. P.; Chai, A.; Siegler, K.; Li, Y.; Coward, P. *Bioorg.Med.Chem.Lett.* **2009**, *19*, 3550.

(159) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *Journal of computational chemistry* **2002**, *23*, 1601.

(160) Tubert-Brohman, I.; Guimarães, C. R. W.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2005**, *1*, 817.

(161) Tubert-Brohman, I.; Guimarães, C. R. W.; Repasky, M. P.; Jorgensen, W. L. *J. Comput. Chem.* **2003**, *25*, 138.

(162) Jorgensen, W. L.; Tirado-Rives, J. *J. Comput. Chem.* **2005**, *26*, 1689.

(163) Hess, B.; Kutzner, C.; Spoel, D. v. d.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435.

(164) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* **1998**, *19*, 1639.

(165) Perryman, A. L.; McCammon, J. A. *J. Med. Chem.* **2002**, *45*, 5624.

(166) Ruppert, J.; Welch, W.; Jain, A. N. *Protein Sci.* **1997**, *6*, 524.

(167) Miglio, G.; Rosa, A. C.; Rattazzi, L.; Grange, C.; Camussi, G.; Fantozzi, R. *British journal of pharmacology* **2012**, *167*, 641.

(168) Acevedo, O.; Ambrose, Z.; Flaherty, P. T.; Aamer, H.; Jain, P.; Sambasivarao, S. V. *Curr. Pharm. Des.* **2012**, *18*, 1199.

(169) Lewis, S. N.; Bassaganya-Riera, J.; Bevan, D. R. *PPAR Res.* **2009**, *2010*, 861238.

(170) Klebe, G. *Drug Discovery Today* **2006**, *11*, 580.

(171) Nolte, R. T.; Wisely, G. B.; Westin, S.; Cobb, J. E.; Lambert, M. H.; Kurokawa, R.; Rosenfeld, M. G.; Willson, T. M.; Glass, C. K.; Milburn, M. V. *Nature* **1998**, *395*, 137.

(172) Zoete, V.; Grosdidier, A.; Michielin, O. *Biochim Biophys Acta.* **2007**, *1771*, 915.

(173) Fukuda, R.; Zhang, H.; Kim, J. W.; Shimoda, L.; Dang, C. V.; Semenza, G. L. *Cell* **2007**, *129*, 111.

(174) Gampe Jr., R. T.; Montana, V. G.; Lambert, M. H.; Miller, A. B.; Bledsoe, R. K.; Milburn, M. V.; Kliewer, S. A.; Willson, T. M.; Xu, E. *Mol. Cell* **2000**, *5*, 545.

(175) Nesto, R. W.; Bell, D.; Bonow, R. O.; Fonseca, V.; Grundy, S. M.; Horton, E. S.; Le Winter, M.; Porte, D.; Semenkovich, C. F.; Smith, S.; Young, L. H.; Kahn, R.; American Heart, A.; American Diabetes, A. *Circulation* **2003**, *108*, 2941.

(176) Lewis, S. N.; Bassaganya-Riera, J.; Bevan, D. R. *PPAR Res.* **2010**, *2010*, 861238.

(177) Trott, O.; Olson, A. J. *Journal of computational chemistry* **2010**, *31*, 455.

(178) DeLano, W. L. **2002**.

(179) Tubert-Brohman, I.; Guimarães, C. R. W.; Jorgensen, W. L. *Journal of Chemical Theory and Computation* **2005**, *1*, 817.

(180) Tubert-Brohman, I.; Guimaraes, C. R.; Repasky, M. P.; Jorgensen, W. L. *Journal of computational chemistry* **2004**, *25*, 138.

(181) Duffy, E. M.; Jorgensen, W. L. *Journal of the American Chemical Society* **2000**, *122*, 2878.

(182) QikProp **2006**.

(183) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *Journal of chemical theory and computation* **2008**, *4*, 435.

(184) Kanzok, S. M.; Schirmer, R. H.; Türbachova, I.; Iozef, R.; Becker, K. *Journal of Biological Chemistry* **2000**, *275*, 40180.

(185) Bradford, M. M. *Analytical biochemistry* **1976**, *72*, 248.

(186) Co, E. M.; Dennull, R. A.; Reinbold, D. D.; Waters, N. C.; Johnson, J. D. *Antimicrobial agents and chemotherapy* **2009**, *53*, 2557.

(187) Hamalainen-Laanaya, H. K.; Orloff, M. S. *Analytical biochemistry* **2012**, *429*, 32.

(188) Babich, H.; Borenfreund, E. *Applied and environmental microbiology* **1991**, *57*, 2101.

(189) Sivilotti, M. L. *Toxicological reviews* **2004**, *23*, 169.

(190) Ganesan, S.; Chaurasiya, N. D.; Sahu, R.; Walker, L. A.; Tekwani, B. L. *Toxicology* **2012**, *294*, 54.

(191) Davioud-Charvet, E.; McLeish, M. J.; Veine, D. M.; Giegel, D.; Arscott, L. D.; Andricopulo, A. D.; Becker, K.; Muller, S.; Schirmer, R. H.; Williams, C. H., Jr.; Kenyon, G. L. *Biochemistry* **2003**, *42*, 13319.

(192) Munigunti, R.; Calderon, A. I. *Rapid communications in mass spectrometry : RCM* **2012**, *26*, 2051.

(193) Mulabagal, V.; Calderon, A. I. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **2010**, *878*, 987.

(194) Fang, J.; Lu, J.; Holmgren, A. *The Journal of biological chemistry* **2005**, *280*, 25284.

(195) Mishra, S.; Karmodiya, K.; Surolia, N.; Surolia, A. *Bioorganic & medicinal chemistry* **2008**, *16*, 2894.

(196) Morin, C.; Besset, T.; Moutet, J. C.; Fayolle, M.; Bruckner, M.; Limosin, D.; Becker, K.; Davioud-Charvet, E. *Organic & biomolecular chemistry* **2008**, *6*, 2731.

(197) Sandur, S. K.; Pandey, M. K.; Sung, B.; Ahn, K. S.; Murakami, A.; Sethi, G.; Limtrakul, P.; Badmaev, V.; Aggarwal, B. B. *Carcinogenesis* **2007**, *28*, 1765.

(198) Shukla, A.; Singh, A.; Singh, A.; Pathak, L. P.; Shrivastava, N.; Tripathi, P. K.; Singh, M. P.; Singh, K. *Cellular and molecular biology* **2012**, *58*, 182.

(199) Singh, D. B.; Gupta, M. K.; Singh, D. V.; Singh, S. K.; Misra, K. *Interdisciplinary sciences, computational life sciences* **2013**, *5*, 1.

Supplemental Information

AutoDock Vina:

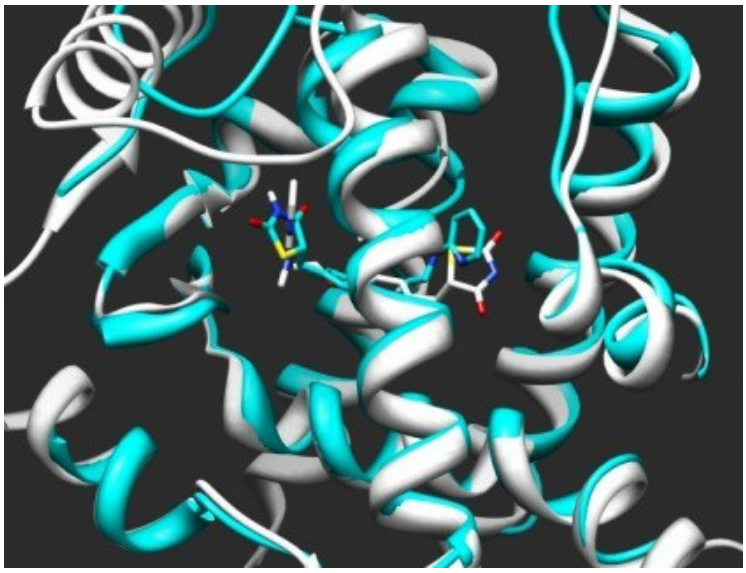


Figure S1. Docked rosigilitazone into PPAR-γ (PDB ID: 2PRG) using AutoDock Vina. RSMD = 0.88 Å.

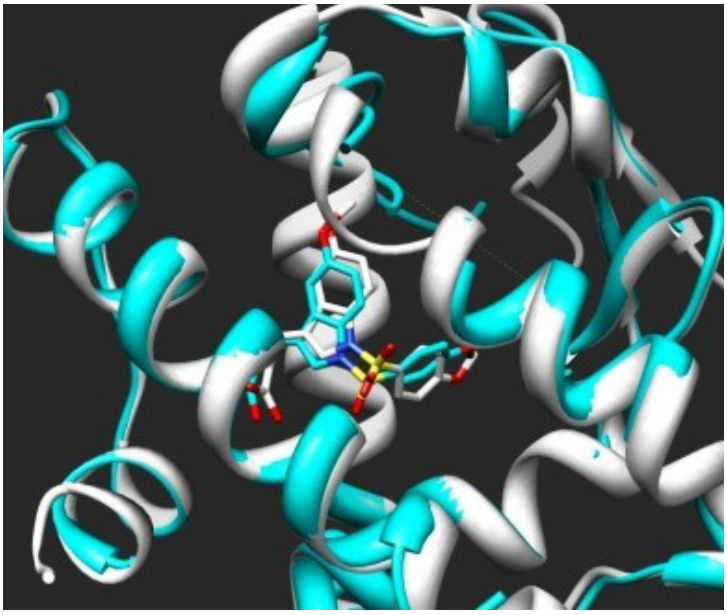


Figure S2. Docked indeglitazar into PPAR-γ (PDB ID: 3ET3) using AutoDock Vina. RSMD = 0.83 Å.

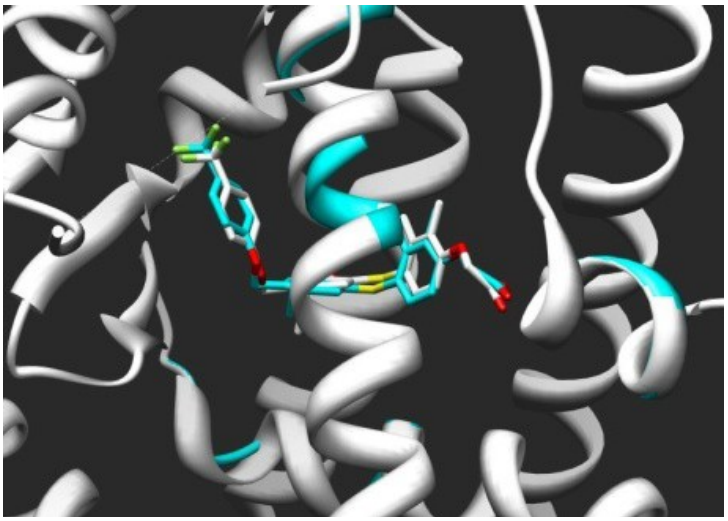


Figure S3. Docked D321 into PPAR-δ (PDB ID: 3GZ9) using AutoDock Vina. RSMD = 0.07 Å.

AutoDock:

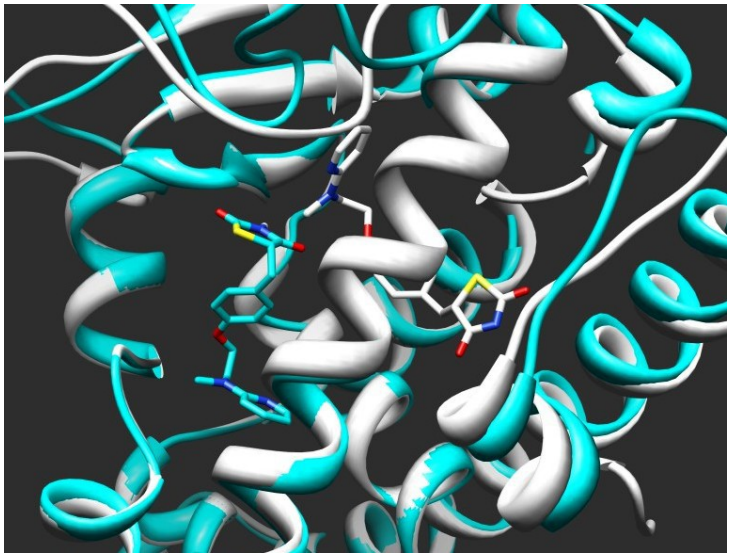


Figure S4. Docked rosigilitazone into PPAR-γ (PDB ID: 2PRG) using AutoDock. RSMD = 1.87 Å.

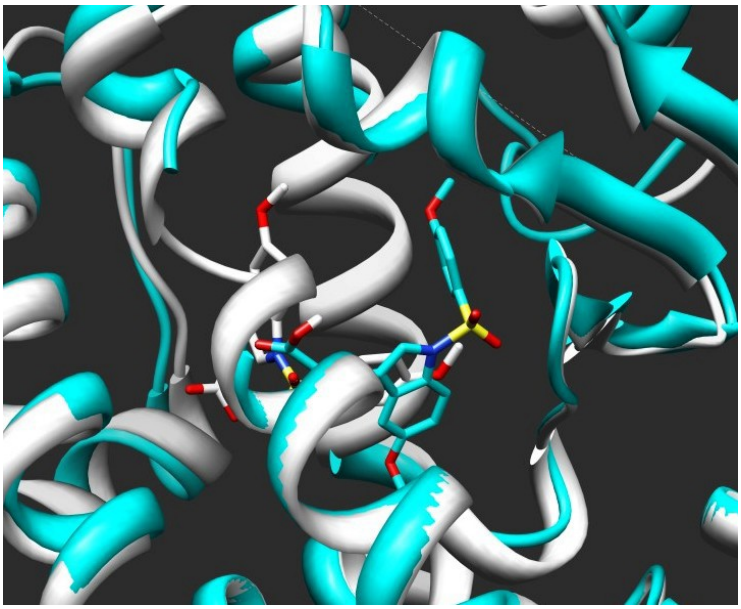


Figure S5. Docked indeglitazar into PPAR-γ (PDB ID: 3ET3) using AutoDock. RSMD = 1.40 Å.

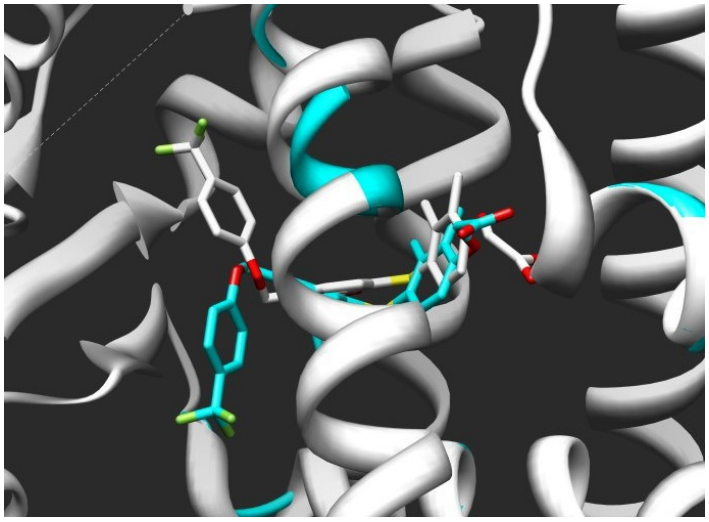


Figure S6. Docked D321 into PPAR-δ (PDB ID: 3GZ9) using AutoDock. RSMD = 1.07 Å.

Surflex-Dock:

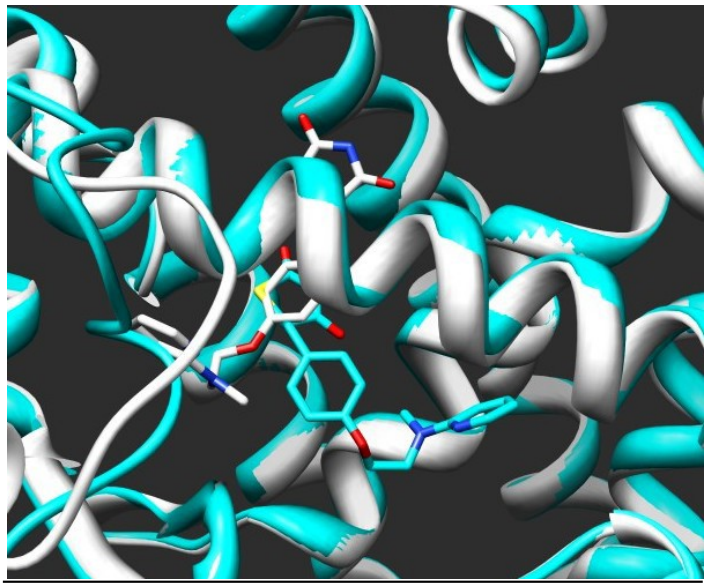


Figure S7. Docked rosigilitazone into PPAR-γ (PDB ID: 2PRG) using Surflex-Dock. RSMD = 1.16 Å.

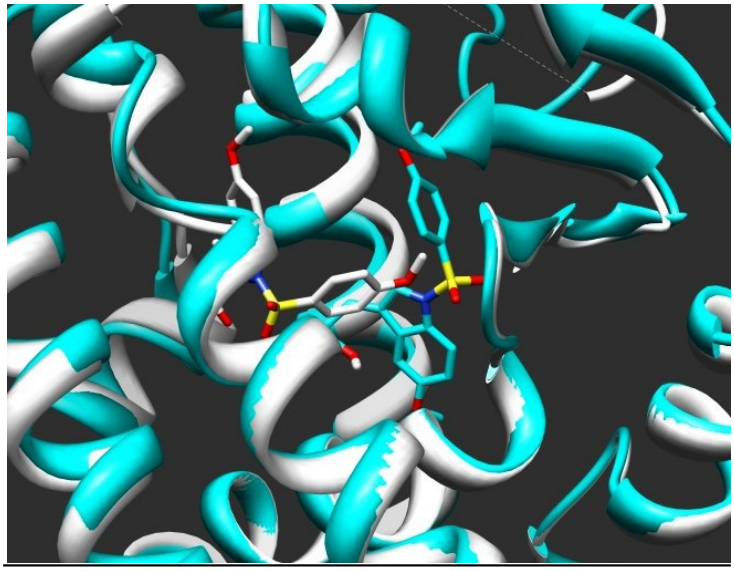


Figure S8. Docked indeglitazar into PPAR-γ (PDB ID: 3ET3) using Surflex-Dock. RSMD = 1.81 Å.

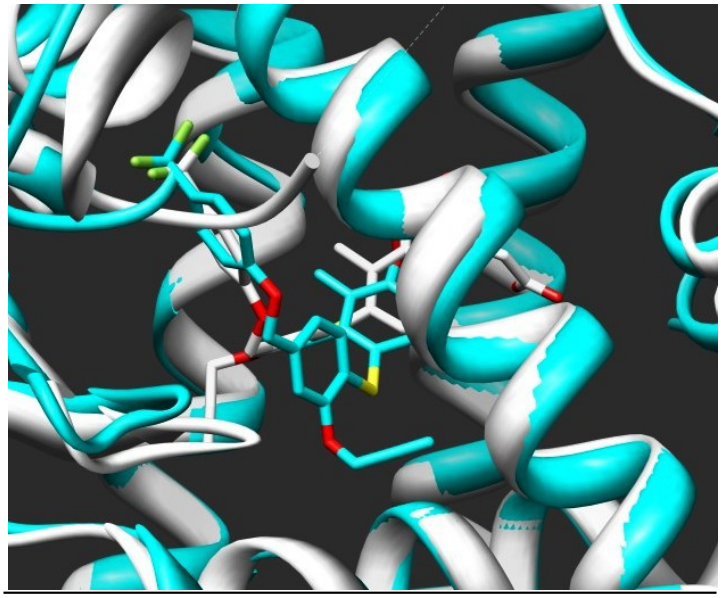


Figure S9. Docked D321 into PPAR-δ (PDB ID: 3GZ9) using Surflex-Dock. RSMD = 0.70 Å.

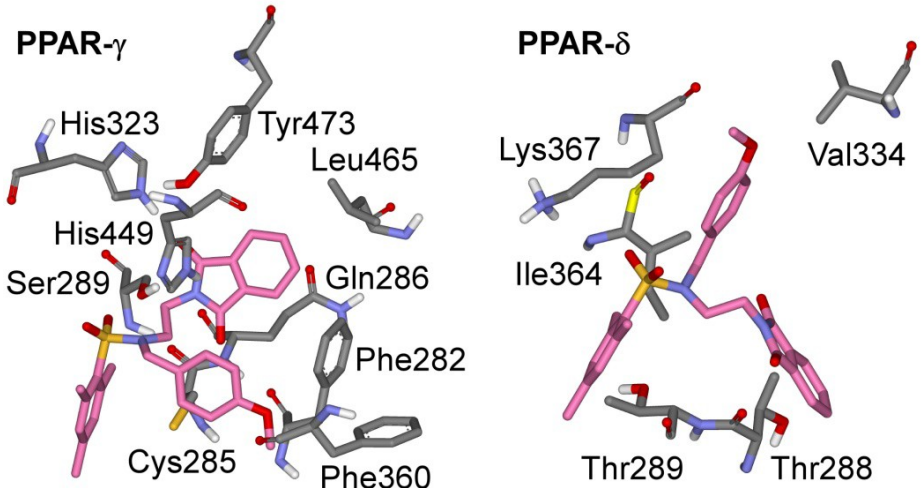


Figure S10. Compound **3-91** bound to the active site of PPAR-γ (left) and PPAR-δ (right) with key

residues shown.

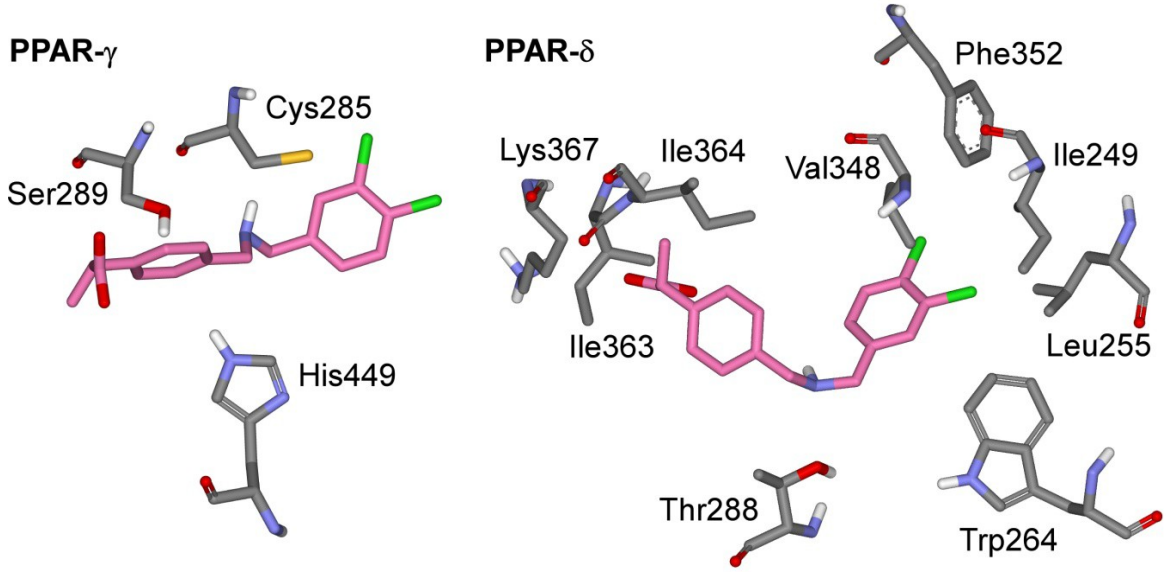


Figure S11. Compound **4-23** bound to the active site of PPAR-γ (left) and PPAR-δ (right) with key residues shown.

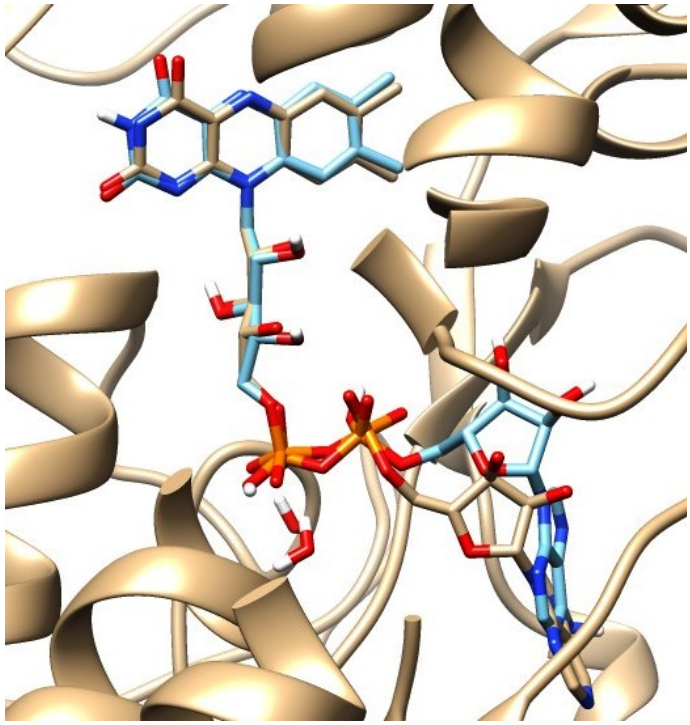


Figure S12. Experimental crystal structure with bound FAD (beige) and the predicted pose for FAD from the docking calculations (cyan).

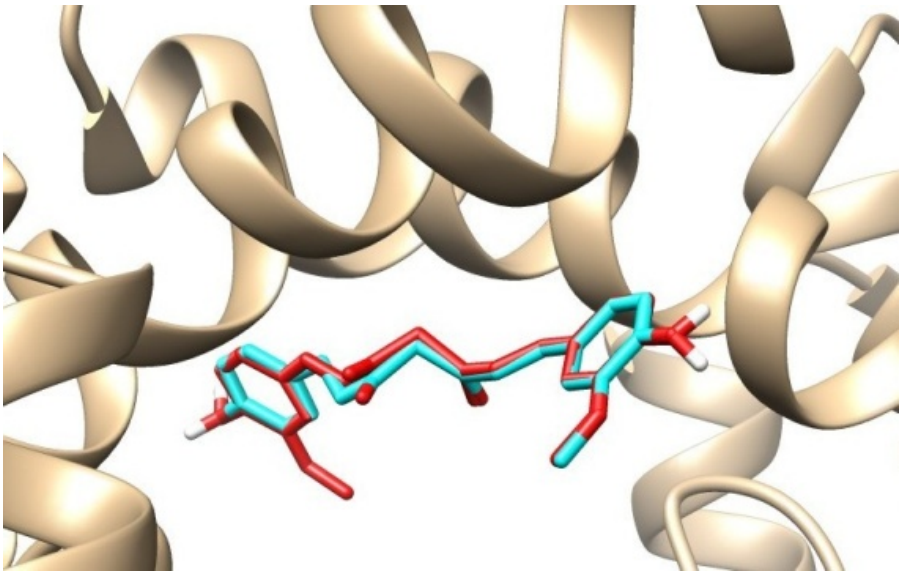


Figure S13. *Pf*TrxR bound to DMC (blue) and curcumin (red) as predicted by the docking calculations.