Structural basis for CARM1 inhibition by indole and pyrazole inhibitors

John Sack, Sandrine Thieffine, Tiziano Bandiera, Marina Fasolini, Gerald J Duke, Lata Jayaraman, Kevin F Kish, Herb E Klei, Ashok Purandare, Pamela Rosettani, et al.

Structural Basis for CARM1 Inhibition by Indole and Pyrazole Inhibitors

John S. Sack[*], Sandrine Thieffine[†], Tiziano Bandiera[†‡], Marina Fasolini[†], Gerald J. Duke[*], Lata Jayaraman[*], Kevin F. Kish[*], Herbert E. Klei[*], Ashok V. Purandare[*], Pamela Rosettani[†], Sonia Troiani[†], Dianlin Xie[*] and Jay A. Bertrand [†1]

[*]Bristol-Myers Squibb Research & Development, P.O. Box 4000, Princeton, New Jersey 08543-4000, USA
[†]Nerviano Medical Sciences, Via Pasteur 10, 20014 Nerviano, Italy
[‡]Current Address: Italian Institute of Technology, Via Morego 30, 16163 Genova, Italy
[1]Corresponding author: Jay A. Bertrand, Nerviano Medical Sciences, Via Pasteur 10, 20014 Nerviano, Italy, Phone: +39 0331 58 1395, Fax: +39 0331 58 1360; Email: jay.bertrand@nervianoms.com

Short Title: Indole and Pyrazole Inhibitors of CARM1

Keywords: Coactivator-associated arginine methyltransferase 1 (CARM1), drug discovery, indole, Protein Arginine Methyl Transferases (PRMTs), pyrazole, selective enzyme inhibition.

Abbreviations: BSA, bovine serum albumin; CARM1, coactivator-associated arginine methyltransferase 1; CMPD-1, 2-{4-[3-fluoro-2-(2-methoxyphenyl)-1H-indol-5-yl]piperidin-1-yl}-N-methylethanamine; CMPD-2, N-(3-{5-[5-(1H-indol-4-yl)-1,3,4-oxadiazol-2-yl]-3-(trifluoromethyl)-1H-pyrazol-1-yl}benzyl)-L-alaninamide; CMPD-3, 1-{3-[(L-alanylamino)methyl]phenyl}-N-(naphthalen-1-ylmethyl)-3-(trifluoromethyl)-1H-pyrazole-5-carboxamide; GAR, glycine-arginine rich; GST, glutathione S-transferase; ITC, isothermal titration calorimetry; MBP, maltose-binding protein; NHR, nuclear hormone receptor; PRMT, protein arginine methyl transferase; RMSD, root mean square deviation; SAH, S-adenosyl-L-homocyseine; SAM, S-adenosyl-L-methionine; SNF, sinefungin; TCA, trichloroacetic acid.

The coordinates and structure factors for the two crystal structures have been deposited in the Protein Data Bank with the accession codes 2Y1W and 2Y1X.

SYNOPSIS

Coactivator-associated arginine methyltransferase 1 (CARM1) is a protein arginine methyl transferase (PRMT) family member that catalyzes the transfer of methyl groups from S-Adenosyl-L-Methionine to the sidechain of specific arginine residues of substrate proteins. This post-translational modification of proteins regulates a variety of transcriptional events and other cellular processes. Moreover, CARM1 is a potential oncological target due to its multiple roles in transcription activation by nuclear hormone receptors and other transcription factors like p53. Here we present crystal structures of the CARM1 catalytic domain in complex with cofactors (S-adenosyl-L-homocyseine or Sinefungin) and indole or pyazole inhibitors. Analysis of the structures reveals that the inhibitors bind in the arginine-binding cavity and the surrounding pocket that exists at the interface between the N- and C-terminal domains. In addition, we show using isothermal titration calorimetry that the inhibitors bind to the CARM1 catalytic domain only in the presence of the cofactor S-adenosyl-L-homocysteine. Furthermore, sequence differences for select residues that interact with the inhibitors may be responsible for the CARM1 selectivity versus PRMT1 and PRMT3. Together, the structural and biophysical information should aid in the design of both potent and specific inhibitors of CARM1.

INTRODUCTION

Arginine methylation is a common post-translational modification that in eukaryotes is catalyzed by the Protein Arginine Methyl Transferases (PRMTs), a family of proteins that transfer methyl groups from S-Adenosyl-L-Methionine (SAM) to the side chain of specific arginine residues [1]. The covalent modification of proteins by the addition of a methyl group to arginine residues can modulate their binding interactions and, as a result, provides a means of regulating their physiological function. The PRMT family, which consists of at least 9 members (PRMTs 1-9), can be classified into two main classes depending on whether they generate asymmetric (type I) or symmetric (type II) dimethyl arginine residues on substrate proteins.

Coactivator-associated arginine methyltransferase 1 (CARM1), also known as PRMT4, was originally identified in a yeast two hybrid screen for proteins that associate with the p160 steroid receptor coactivator, GRIP1 [2]. CARM1 is recruited to the nuclear hormone receptor transcriptional complex by GRIP1 and this recruitment results in the methylation of other coactivators in the complex such as p300/CBP and AIB1. The recruitment of CARM1 to promoter regions of genes by the NHR complex or by transcription factors such as p53, β-catenin [3] or NF-kB [4] results in the methylation of specific arginine residues in the N-terminal tails of Histone H3. The direct consequence of these methylation events is the enhancement of gene transcription. [2, 5, 6]. Additional roles for CARM1 have been suggested in muscle differentiation [7] as well as protein repair, chromatin regulation, mRNA stabilization [8] and gene splicing [9].

While full length CARM1 (608 residues) is required for co-activator function [10], the central 350 residue core region that contains both the SAM and activator binding sites is able to methylate substrate peptides *in vitro*. Crystal structures have been determined for the core region of CARM1 as well as for several other members of the arginine methyltranferases family that include PRMT1, PRMT3 and Hmt1 [11-15]. These proteins share a similar fold consisting of a two-domain structure; an N-terminal domain containing a mixed α/β Rossmann fold that is often referred to as the SAM binding domain and a C-terminal β-barrel-like substrate-binding domain.

The CARM1 N-terminal domain corresponds to the region of highly conserved sequence homology (50-60%) among the arginine methyltransferases and has been identified as responsible for binding SAM and for catalysis of the methyl transfer [14]. Three of the four PRMT signature sequences (motifs I, II and III) belong to the N-terminal domain (Figure 1). The C-terminal domain is the more structurally diverse region of the core domain with only 20-40% homology across the PRMT family. There are several divergent loops in this domain that may be involved in substrate specificity. This domain also contains the highly conserved "THWxQ" loop (motif IV) that forms part of the substrate-binding groove. The β-barrel is interrupted after the first strand by a 39-residue two-helix bundle (residues 300-338) capped by a hydrophobic tip (residues 316-322). This dimerization arm or 'antenna' makes extensive interactions with a second molecule to form the CARM1 dimer. CARM1 is unique among the PRMTs in that it also contains a C-terminal extension from the core region [16].

While the full extent of CARM1's role in cellular regulation is not yet known, it is clear that its multiple roles in transcription activation by nuclear hormone receptors [17] and other transcription factors like p53 [18] make it a likely target for the treatment of cancer. In fact, the over-expression of CARM1 has been observed in both breast and prostate cancers [19, 20]. To date, there have been a limited number of publications describing CARM1 inhibitors [21-26] and, of these, only our own efforts yielded inhibitors that were selective versus PRMT1 and PRMT3 [23, 24, 26]. Thus, in order to understand the basis for selectivity and to provide a starting point for structure-based design we determined crystal structures of the CARM1 catalytic domain in complex with cofactors and representative inhibitors of two different classes. Here, we report two crystal structures of CARM1 in complex with cofactors and inhibitors (Figure 2); one is the complex with sinefungin (SNF) and an inhibitor of the indole class (CMPD-1) and the other is the complex with S-adenosyl-L-homocyseine (SAH) and an inhibitor of the pyrazole class (CMPD-2). In addition, we characterized the binding of inhibitors from each class to CARM1 using Isothermal Titration Calorimetry (ITC). The structures of these complexes and the information obtained by ITC will help in the subsequent rounds of inhibitor design to obtain high affinity, specific inhibitors of CARM1.

EXPERIMENTAL

Chemical Synthesis

CMPD-1 of the indole class was synthesized following the procedures reported for the analogous benzo[d]imidazole class [26] and additional information can be found in the Patent Application US 2006/0235037, 19 October 2006. CMPD-2 and CMPD-3 of the pyrazole class were synthesized following the procedures reported by Purandare and coworkers [23, 24].

Protein Expression, Purification and Crystallization

The CARM1 catalytic domain from residues 134 to 483 was cloned into a pMAL vector (New England Biolabs) obtaining by consequence a protein fused with MBP. This plasmid was transformed into the MM294(DE3) strain of E. coli. Inoculum was grown from a single colony at 37 °C for 16h in LB2x media supplemented with 50 μg/ml of carbenicillin. The culture was grown at 37 °C until OD_{600} of 0.8, the temperature was reduced to 20° C and expression was induced using 1 mM IPTG. Cells were harvested by centrifugation after 20h and stored at –80 °C. For purification, cell pellets were resuspended in Buffer A (50 mM Tris-HCl, 100 mM NaCl, 5 mM DTT, 2 mM EDTA, pH 7.5) supplemented with protease inhibitor tablets (Roche Biochemicals) and lysed by sonication. The clarified lysate was loaded onto an Amylose column and washed with Buffer B (25 mM Tris-HCl, 100 mM NaCl, 5 mM DTT, 2 mM EDTA, pH 7.5). MBP-CARM1 was eluted with maltose buffer, diluted 10 fold with 25 mM Tris-HCL, pH 7.5 and loaded onto an Ion exchange column. Next, the protein was eluted with a NaCl gradient and immediately loaded onto an Amylose column. MBP was cleaved by enterokinase overnight at 20 °C on the column. After overnight cleavage, the eluted protein was loaded onto a gel filtration column pre-equilibrated with 25 mM Hepes, 100 mM NaCl, 0.5 mM TCEP, pH 7.5. Final yields were ~ 1.5 mg/L of culture. The purified CARM1 catalytic domain was ~ 95% pure as judged by SDS PAGE and electrospray mass spectrometry analysis revealed that spontaneous processing at the N-termini removes the first

6 amino acids. Thus, the resulting protein starts with Ser136-Val137-Phe138 and ends with Thr481-Pro482-Ser483.

Crystals of the CARM1 catalytic domain in complex with SAH were grown using the hanging-drop vapor diffusion method from a solution of 20-30% PEG MME 2000, 0.1 M Tris-HCl pH 8.5 and 0.2 M Trimethylamine N-oxide dihydrate at 4°C. Before crystallization the CARM1 was concentrated to 2.4 mg/mL and SAH was added to the protein solution to give final concentrations of 0.25 mM. Diffraction quality crystals appeared within one week. To provide crystals with CMPD-2, crystals of CARM1 in complex with SAH were soaked in a solution of 20-30% PEG MME 2000, 0.1 M Tris-HCl pH 8.5 and 0.2 M Trimethylamine N-oxide dihydrate with 2 mM of CMPD-2 for 5 days. Crystals of the complex with sinefungin (SNF) and CMPD-1 were obtained by following the procedure described above for the crystals with SAH with some modifications. Specifically, 0.25 mM of SNF and 1 mM of CMPD-1 were added to the concentrated protein before setting up the crystallization trials. For data collection, the crystals were transferred to drops containing the equivalent mother liquor with 12.5% glycerol.

Data Collection, Structure Determination and Refinement

Crystals belong to the space group $P2_12_12$ (a =75.572, b = 98.637, c = 207.761; $\alpha = \beta = \gamma = 90$) with four molecules in the asymmetric unit. Diffraction data were collected at the European Synchrotron Radiation Facility (Grenoble, France) on beamlines ID14-4 and ID14-1 for CARM1•SNF•CMPD-1 and CARM1•SAH•CMPD-2, respectively. Indexing, integration and scaling were performed using HKL2000 [27]. Both CARM1 structures were solved by molecular replacement using the Molrep software [28]. Search models for both structures were an in-house structure of CARM1 in complex with SAH and a proprietary inhibitor that was solved by SeMet MAD. Model building was done using Coot [29] and refinement was done with CNX [30]. Crystallographic data is reported in Table 1 and SigmaA weighted difference omit maps of the inhibitors are shown in Figure 4. Structural images have been generated with PyMol [31]. Superposition of the two CARM1 structures gave RMSD values ranging from 0.2 to 0.9 Å for the 342 structurally equivalent Cα atoms.

Enzymatic Activity Assay

The methyltransferase activity of CARM1 was determined as described elsewhere [24] using a methylation based filter assay. Briefly, methylation reactions were performed for 60-90 minutes using full-length GST-tagged CARM1 (6 nM), the substrate Histone H3 (1 μM) and tritiated SAM (0.05 μM) in methylation buffer (20 mM Tris-HCl, pH 8.0, 200 mM NaCl, 0.4 mM EDTA) both with and without inhibitors. Subsequently the reaction was stopped by the addition of TCA, the reaction mixture was precipitated with BSA overnight and the resulting mixture was filtered and washed before being read in a Top Count after the addition of MicroScint-20. PRMT1 and PRMT3 specific methylation assays were performed as described above for CARM1 but using as enzyme/substrate the combinations GST-PRMT1 (8 nM)/ Histone H4 (0.74 μM) or GST-PRMT3 (14 nM)/ GST-GAR (0.27 μM). The GST-fusion proteins (GST-CARM1, GST-PRMT1, GST-PRMT3 and GST-GAR) were expressed and purified from bacteria and the Histones H3 and H4 were both purchased from Roche Applied Science. Purity assessments using the Odyssey Imaging system with SDS-PAGE gels revealed that GST-CARM1 and GST-PRMT1 were 80 % pure and GST-PRMT3 was 85 % pure.

Peptide mass fingerprinting was used to verify the identity of the visible contaminants in the gel; all three proteins contained truncated versions of the target proteins, GST-CARM1 also contained endogenous GST (Spodoptera frugiperda) and GST-PRMT1 also contained DNAK (E. coli, at ~ 76 kDa) and 60 kDa chaperonin (E. coli, at ~ 65 kDa). Protein concentrations were measured with the Coomassie Plus (Pierce) assay using BSA as a reference.

Isothermal Titration Calorimetry

The CARM1 catalytic domain was extensively dialyzed against buffer containing 25 mM Hepes pH 7.5, 100 mM NaCl, 0.5 mM TCEP. Experiments were carried out at 20°C with a VP-ITC titration calorimeter (MicroCal Inc., Northampton, MA) with protein in the cell and SAM, SAH or inhibitors in the syringe. Each titration experiment consisted of a first (5 μL) injection followed by 29 injections of 10 μL and the experiments with inhibitors were performed both in the presence and absence of 0.2 mM SAH. The SAH titration was run with 15 μM of CARM1 catalytic domain and 230 μM of SAH and all of the others with 20 μM of CARM1 catalytic domain and 200 μM of SAM, CMPD-1 or CMPD-3. Protein concentrations were measured with the Coomassie Plus (Pierce) assay using BSA as a reference after confirming that these values correspond well with those using the calculated extinction coefficient.

RESULTS

CMPD-1 and CMPD-2 are potent inhibitors of CARM1. A high-throughput screening effort using human full-length CARM1 in a methylation assay identified a variety of 'hits' with modest activity and a subset of these were selected for further expansion to improve the *in vitro* properties [23, 24, 26]. This led to the indole and pyrazole inhibitor classes for which CMPD-1 and CMPD-2 (Figure 2), respectively, are representatives. Using a methylation based filter assay, IC50 values for the two compounds are 0.030 ± 0.015 and 0.027 ± 0.009 μM, respectively, for CARM1 and > 10 μM for PRMT1 and PRMT3.

Indole and Pyrazole inhibitors bind to CARM1 in the presence of SAH. ITC experiments were done to confirm that the compounds identified and characterized with full-length CARM1 also bind to the catalytic domain under consideration for crystallographic studies. It is worth mentioning that, at the time, structures were available of PRMT1, PRMT3 and HMT1 but none of CARM1. Thus the idea was to try and crystallize a catalytic domain construct that was designed using sequence alignments in combination with structural data from the homologous proteins. In preparation for experiments with compounds, initial ITC titrations of CARM1 were done with SAM and SAH to control the integrity of the protein. SAM binds to CARM1 with a dissociation constant (Kd) of 0.781 μM, a ΔH of -13.8 kcal/mol and a ΔS of –19.1 cal/mol and SAH binds to CARM1 with a dissociation constant (Kd) of 6.90 μM, a ΔH of –11.0 kcal/mol and a ΔS of –13.8 cal/mol (Supplementary Figure S1). Experiments with CMPD-1 were done by titrating the compound into CARM1 both in the absence and presence of the cofactor SAH (Figure 3). Analysis of the resulting binding isotherm reveals that CMPD-1 binds to CARM1 with a dissociation constant of 0.465 μM, a ΔH of -9.0 kcal/mol and a ΔS of -1.8 cal/mol in the presence of 0.2 mM SAH and no binding is observed in the absence of SAH. ITC experiments with CMPD-2 were not feasible due to the poor aqueous

solubility so, instead, experiments were done with CMPD-3 (Figure 2), an analog of CMPD-2 with an IC50 of 0.128 ± 0.017 μM in the methylation based filter assay. Analysis of the binding isotherms reveals that CMPD-3 only binds to CARM1 in the presence of 0.2 mM SAH and, interestingly, fitting of the data with a single-site model gave poor results (data not shown) while fitting with a two-site model gave quite good results (as shown in Figure 3). Thus, CMPD-3 binds to CARM1 with a first dissociation constant (Kd1) of 0.008 μM, a ΔH of 7.9 kcal/mol, a ΔS of 64.1 cal/mol and a stoichiometry of 0.16 and second dissociation constant (Kd2) of 0.020 μM, a ΔH of -1.0 kcal/mol, a ΔS of -0.6 cal/mol and a stoichiometry of 0.90. It's worth mentioning that although the data fits with a two-site model the combined stoichiometry equals 1.06. So, a logical explanation would be that CMPD-3 binds to a single site with two different affinities.

CMPD-1 and CMPD-2 occupy the substrate-binding site. The ternary complex of CARM1, CMPD-1 and SNF, a stable analog of SAM, was co-crystallized and the structure determined to 2.1 Å resolution and CMPD-2 was soaked into co-crystals of CARM1 with SAH and the structure determined to 2.4 Å resolution. Analysis of the resulting electron density maps reveals that both inhibitors bind in a rectangular-shaped pocket that exists at the interface between the two domains (Figures 4, 5 and S2). The sides of the pocket are formed by helices X and Y and the loops connecting β4 to αD, β11 to β12 and β13 to β14 (Figure 5C). At one end of the pocket, near His415, there is a deep cavity, hereafter referred to as the Arg-binding cavity, that extends towards the site of methyl transfer. Interestingly, residues from three of the PRMT signature motifs (motifs I, III and IV) help to form the pocket and cavity (Figure 1).

In the case of CMPD-1, the inhibitors N-methylethanamine group is directed towards the bottom of the Arg-binding cavity and the piperidine moiety is positioned at the mouth of the cavity (Figure 6A). Interestingly, the pucker of the 6-member ring 'bends' the inhibitor towards Ser146 and allows the indole ring system to make additional hydrophobic interactions with the residues lining the pocket. The amine nitrogen of the inhibitor makes several polar interactions with CARM1 residues in the Arg-binding cavity that include one of the side chain oxygens of Glu258, the carbonyl oxygen of Met260 and the nitrogen of SNF that mimics the methyl group of SAM. At the mouth of the cavity, the piperidine nitrogen hydrogen bonds with Nε2 of His415. Outside of the cavity, a bridging water molecule (W3) makes the only polar interaction between the protein and inhibitor, linking the indole nitrogen and Nδ2 of Asn266. Interestingly, two other water molecules help form the CMPD-1 binding-pocket; one by filling the space underneath the indole ring (W2) and the other positioned above the mouth of the Arg-binding cavity (W1).

For CMPD-2, the terminal L-alaninamide is directed towards the bottom of the Arg-binding cavity and the benzyl ring is positioned at the mouth of the cavity (Figure 6B). The pyrazole moiety sits above the imidazole ring of His415 and the attached substituents point in opposite directions. Specifically, the trifluoromethyl group extends towards the solvent, passing between the side chains of Asn162 and Tyr417, while the 1,3,4-oxadiazole crosses back over the mouth of the Arg-binding cavity. The attached indole group is positioned above the Glu267 side chain and packs against the side chains of Tyr262, Pro473 and Phe475. The L-alaninamide of CMPD-2 makes several polar interactions with CARM1 in the Arg-binding

cavity. At the bottom of the cavity, the terminal amino nitrogen interacts with one of the side chain oxygens of Glu258 and the carbonyl oxygens of Glu258 and Met260. Then, towards the middle of the cavity the carbonyl oxygen of the L-alaninamide hydrogen bonds with Nε2 of His415 while the adjacent nitrogen hydrogen bonds with one of the side chain oxygens of Glu267. Outside of the Arg-binding cavity, the oxadiazole makes the only other polar interactions with CARM1. These include a hydrogen bond between the oxadiazole oxygen and the hydroxyl oxygen of Tyr262 and a bridging water molecule (W4) that interconnects the oxadiazole nitrogen in position 3 with Nε2 of Gln159. It's also worth noting that the three water molecules that were observed in the structure with CMPD-1 are also present in the structure with CMPD-2.

DISCUSSION

SAH binding to CARM1 causes large structural changes near the catalytic site that may explain the lack of inhibitor binding in the absence of SAH. To date there are three apo-structures of CARM1 (PDB codes 3B3G, 3B3J and 2V7E) and two structures of the CARM1•SAH complex (PDB codes 3B3F and 2V74). Analysis of these structures by the authors [11, 12] reveals different types of structural changes that are associated with SAH binding that include major modifications to the region 147-179 that contains the PRMT signature motif I and minor modifications to the other three PRMT signature motifs (motifs II, III and IV). Of the three apo-structures of CARM1, 3B3J shows the largest degree of structural reorganization with SAH binding while the other two structures are more similar to those of the CARM1•SAH complexes. In 3B3J, the segment 148-152 forms a β-strand (β0) followed by a 3_{10} helical turn (residues 153-156) that leads into a long helix comprising residues 159-181. Thus, the transition associated with SAH binding causes β0 to be converted into αX and a kink is introduced in the long helix comprising residues 159-179 to form the helices αY and αZ [11]. In addition, several residues are repositioned to properly form the Arg-binding cavity and the surrounding pocket. Interestingly, many of these residues are the same as those described above that interact with the inhibitors CMPD-1 and CMPD-2. Thus, it is not surprising that the inhibitors fail to bind in the in the absence of SAH since cofactor binding induces large structural changes that lead to the proper formation of the Arg-binding cavity and the surrounding pocket.

Comparing the published CARM1 structures with SAH (PDB codes 3B3F and 2V74) and those reported here with cofactors and inhibitors reveals only slight modifications between the different structures. However, one region that varies in the different structures is the N-terminus that in our structures initiates with the helix αW that is followed by a short loop that subsequently leads into the helix αX (Figure 5C). In contrast, the structures 3B3F and 2V74 both lack the helix αW and, as a result, the ordered regions of the N-termini initiate with the helix αX [11, 12]. In our structures, αW packs against the adenine group of SAH shielding it from the solvent while in the structures 3B3F and 2V74 a cavity exists below the helix αX that exposes SAH to the solvent. Also, since there are a series of interactions between the loop connecting αW to αX and the one connecting β13 to β14 that effectively forms one end of the substrate-binding pocket, the disorder of the loop between αW to αX could impact substrate binding. This could be a partial explanation of why Yue and coworkers were unable to obtain well-ordered electron density for the CARM1 complex with SAH and the H3-peptide [12].

Actually in the CARM1 structure with SAH reported by Yue and coworkers (2V74) the loop connecting β13 to β14 is shifted 1.7 angstroms away from the substrate-binding pocket with respect to the corresponding loops in 3B3F and the structures reported here with inhibitors. A logical explanation for the shift of the β13 to β14 loop would be because the crystallization construct used by Yue and coworkers begins with Ala147 and thus lacks all of the residues of the αW to αX loop that stabilize the loop position.

Comparing the CARM1 structures with CMPD-1 and CMPD-2 reveals how the inhibitors use different regions of the Arg-binding cavity and the associated rectangular pocket that exists at the interface between the two domains (Figures 5D, 6A and 6B). Not surprisingly, both inhibitors exploit the Arg-binding cavity to make polar interactions; CMPD-1 interacts with Glu258, Met260 and His415 while CMPD-2 makes similar polar interactions plus an additional one with Glu267. The mouth of the Arg-binding cavity is a point of deviation between the two inhibitors since each one makes extensive hydrophobic interactions with a residue lining the mouth of the cavity (Glu267 and His415) but the two residues are situated on opposite sides of the mouth; CMPD-1 follows the mouth of the cavity on one side, wraps over Glu267 and extends towards Ser146 while CMPD-2 follows the mouth of the cavity on the opposite side, packs against His415 and extends towards the N-terminus of the helix αZ. Actually, the binding mode of CMPD-2 is slightly more complicated since the pyrazole ring effectively directs the inhibitor in two directions with the trifluoromethyl group extending towards the N-terminus of helix αZ and the oxadiazole and indole rings crossing back over the mouth of the cavity and extending towards the loop interconnecting β15 and β16. Together this gives the general effect that CMPD-2 exploits better the region of the pocket between the Arg-binding cavity and the N-terminus of the helix αZ while CMPD-1 exploits better the region between the Arg-binding cavity and Ser146.

The chemical expansions that started from the initial hits and progressed to CMPD-1 and CMPD-2 provide a wealth of information that can be revisited using the structures. For example, efforts were made to optimise the portion of the indole and pyrazole inhibitors that binds in the Arg-binding cavity and, in both cases, even minor modifications gave reductions in activity [24, 25, 26]. Presumably the Arg-binding cavity requires a specific shape in order for it to selectively produce asymmetric dimethyl arginine and the inhibitors optimally exploit the shape but leave little room for modifications. Outside of the Arg-binding cavity the tolerance for variation increases for both chemical classes, a result that correlates with the size of the pocket and the possibilities to expand in different directions. For example, multiple scaffolds were attempted before identifying the preferred fluorine substituted indole ‘core‘ of CMPD-1. In fact, the initial hit and the subsequent series that eventually led to CMPD-1 contained a methylated benzo[d]imidazole ‘core’ rather than the fluorine substituted indole [26]. During the optimisation of the original pyrazole hit efforts were made to improve the permeability of the molecule by reducing the number of hydrogen bond donors. This led to the replacement of the amide that was present in the original hit with a 1,3,4-oxadiazole ring, a modification that not only improved the permeability as measured by the PAMPA assay but also improved the in vitro potency [23]. It is worth noting that CMPD-3 contains the original amide bond while CMPD-2 contains the 1,3,4-oxadiazole ring that was introduced as an amide surrogate (Figure 2). Finally, the pyrazole expansion revealed that a reasonable number of variations are permitted for the aromatic groups that are attached to the 1,3,4-oxadiazole

ring or the amide linkage [23, 24]. The positioning of the indole ring in the structure with CMPD-2 correlates well with this result since it only uses a limited amount of the space that is available between the Arg-binding cavity and Ser146.

Several specific residues are believed to be responsible for the CARM1 selectivity of the inhibitors CMPD-1 and CMPD-2. These include residues coming from the helices αX and αY (Ser146, Gln149, Phe153, Gln159, Asn162), the loop between β4 and αD (Asn266), the loop between β11 and β12 (Tyr417), the loop between β13 and β14 (Arg446-Gln447-Ser448) and the C-terminal segment that contains β16 (Pro473, Phe475, Tyr477). Interestingly, the majority of these residues do not make polar interactions with the inhibitors but instead help to define the shape of the pocket (Figures 6A and 6B). For example, residues from the helices αX and αY (Ser146, Gln149, Phe153, Gln159, Asn162) form one side of rectangular-shaped pocket that exists at the interface between the two domains in CARM1. In PRMT1 and PRMT3, the corresponding residues differ in sequence thus modifying the shape of the pocket (Figures 1, 6C and 6D). For PRMT3, this difference in pocket shape can be observed with the structure 1F3L while for PRMT1 the analysis is complicated by the disorder at the N-terminus for all three of the published PRMT1 structures (1ORI, 1ORH, 1OR8). However, sequence differences between CARM1, PRMT1 and PRMT3 for helices αX and αY strongly suggest that the shape for this region of the PRMT1 pocket will be similar to that of PRMT3 and significantly different from that of CARM1. One noteworthy sequence difference that affects the shape of the pocket in this region is Phe153 in the CARM1 motif I which corresponds to Ser in both PRMT1 and PRMT3 (Ser 38 and Ser220, respectively). Since Phe153 is directed into the pocket where it makes hydrophobic interactions with the inhibitors the substitution of this residue by a hydrophilic serine will result in an overall loss of interactions. Asn266 in CARM1 is another noteworthy residue with respect to sequence differences that may affect inhibitor selectivity. In PRMT1 and PRMT3 the corresponding motif II residues are Tyr152 and Phe334, respectively, and the significant differences in residue size and character will clearly impact the shape of the pocket and the types of interactions that can be made with inhibitors. This point is highlighted by the water-mediated interaction that exists between Asn266 and the indole nitrogen of CMPD-1 since the aromatic residues of PRMT1 and PRMT3 would be unable to make the same type of interaction. Because of the different binding modes for CMPD-1 and CMPD-2, some of the specific residues may correlate with the selectivity of only one of the two inhibitors (Figures 6A and 6B). For example, the trifluoromethyl group of CMPD-2 is positioned between the side chains of Asn162 and Tyr417, two CARM1 specific residues, while none of the CMPD-1 atoms interact with these residues. In addition, since CMPD-2 extends farther out of the Arg-binding cavity than CMPD-1 it is able to make interactions with the C-terminal residues Pro473, Phe475 and Tyr477. In contrast, the methoxy-phenyl group of CMPD-1 is positioned at the other end of the rectangular pocket where it interacts with Ser146 and the loop residues Arg446-Gln447-Ser448.

A surprising result was the apparent difference in affinity between the methylation-based assay and ITC results for CMPD-1 (IC50 = 0.030 μM and Kd = 0.465 μM) and CMPD-3 (IC50 = 0.128 μM, Kd1 = 0.008 μM and Kd2 = 0.020 μM). The IC50 values show a 4-fold difference between the two compounds and CMPD-1 has the higher affinity while the Kd values have a significantly different pattern with a greater than 20-fold difference and CMPD-

3 has the higher affinity. Although a clear explanation concerning the source of these differences is beyond the scope of this work, several different factors could be responsible. One thing to keep in mind is that the two types of measurements are quite different since the assay measures the methylation of Histone H3 over an extended period of time (equilibrium conditions) while ITC measures inhibitor binding to CARM1 in the presence of the product cofactor SAH. Another relevant detail is that the measurements use different forms of CARM1, the IC50 measurements use full-length CARM1 while the ITC studies use the catalytic domain. Since the structural work revealed that the inhibitors bind in a cavity that is in close proximity to the N- and C-termini of the catalytic domain, it is possible that the inhibitors interact with residues from the N- and C-terminal domains of the full-length protein and these interactions impact the IC50 values. Hopefully additional insights will come from the structures of full-length CARM1 in complex with inhibitors and/or protein substrates.

In conclusion, this article reports the first crystal structures of the CARM1 catalytic domain in complex with inhibitors, thus enabling structure-based drug design for this attractive oncology target. Since the indole and pyrazole inhibitors bind in the Arg-binding cavity and, to date no other structures are available with something bound in this region, they provide valuable information concerning the types of interactions that could exist with the natural substrate. Finally, analysis of the structure identified zones of the active site that are likely to be responsible for the CARM1 selectivity of the indole and pyrazole inhibitors versus PRMT1 and PRMT3.

ACKNOWLEDGEMENTS

John S. Sack, Gerald J. Duke, Lata Jayaraman, Kevin F. Kish, Herbert E. Klei, Ashok V. Purandare, and Dianlin Xie are employees of Bristol-Myers Squibb and Sandrine Thieffine, T. Bandiera, Marina Fasolini, Pamela Rosettani, Sonia Troiani and Jay A. Bertrand are employees of Nerviano Medical Sciences. The authors thank James Bryson, James Tamura, Michael Wittekind, Changhong Wan, Hilary Gray, and Valentina Goldfarb of Bristol-Myers Squibb Company for their helpful suggestions on expression and purification of the enzyme.

REFERENCES

1. Lee, H.W., Kim, S. and Paik, W.K. (1977). S-Adenosylmethionine:Protein-Arginine Methyltransferase. Purification and Mechainism of the enzyme. Biochem. **16**, 78-85.

2. Chen D., Ma H., Hong H., Koh S.S., Huang S-M., Schurter B.T., Aswad D.W. and Stallcup M.R. 1999. Regulation of Transcription by a Protein Methyltransferase. Science **284**, 2174-2177.

3. Koh, S.S., Li, H., Lee, Y.-H., Widelitz, R.B., Chuong, C.-M. and Stallcup, M.R. (2002). Synergistic Coactivator Function by Coactivator-associated Arginine Methyltransferase (CARM) 1 and -b-Catenin with Two Different Classes of DNA-binding Transcriptional Activators. J. Biol. Chem. **277**, 26031-26035.

4. Covic, M., Hassa, P.O., Saccani, S., Buerki, C., Meier, N.I., Lombardi, C., Imhof, R., Bedford, M.T., Natoli, G., Hottiger, M.O. (2005). Arginine methyltransferease CARM1 is a promoter specific regulator of NF-kB dependent gene expression. EMBO J, **25**, 85-96.

5. Bauer, U-M., Daujat, S., Nielsen, S.J., Nightingale, K. and Kouzarides, T (2002). EMBO reports. **3**, 39-44.

6. Schurter, B.T., Koh, S.S., Chen, D., Bunick, G.J., Harp, J.M., Hanson, B.L., Henschen-Edman, A., Mackay, D.R., Stallcup, M.R. and Aswad D.W. (2001). Methylation of Histone H3 by Coactivator-Associated Arginine Methyltransferase 1. Biochem. **40**, 5747-5756.

7. Chen, S.L., Loffler, K.A., Chen, D., Stallcup, M.R. and Muscat, G.E.O. (2002). The Coactivator-associated Arginine Methyltransferase Is Necessary for Muscle Differentiation. J. Biol. Chem. **277**, 4324-4333.

8. Fugiwara, T., Mori, T., Chu, D.L., Koyama, Y., Miyata, S., Tanaka, H., Yachi, K., Kubo, T., Yoshikawa, H., and Tohyama, M (2006). CARM1 regulates proliferation of PC12 cells by methylating HuD. Mol Cell Biol **26**, 2273-2285.

9. Cheng, D., Côté, J., Shaaban, S. and Bedford MT. (2007) The arginine methyltransferase CARM1 regulates the coupling of transcription and mRNA processing. Mol Cell. **25**, 71-83.

10. Teyssier, C., Chen, D. and Stallcup, M.R. (2002). Requirement for Multiple Domains of the Protein Arginine Methyltransferase CARM-1 in Its Transcriptional Coactivator Function. J. Biol .Chem. **277**, 46066–46072.

11. Troffer-Charlier, N., Cura, V., Hassenboehler, P., Moras, D. and Cavarelli, J. (2007). Functional insights from structures of coactivator-associated arginine methyltransferase 1 domains. EMBO J. **26**, 4391–4401.

12. Yue, W.W., Hassler, M., Roe, S.M., Thompson-Vale, V. and Pearl, L. H. (2007). Insights into histone code syntax from structural and biochemical studies of CARM1 methyltransferase. EMBO J. **26**, 4402–4412.

13. Zhang, X. and Cheng, X. (2003). Structure of the Predominant Protein Arginine Methyltransferase PRMT1 and Analysis of its Binding to Substrate Peptides. Structure **11**, 509-520.

14. Zhang, X., Zhou, L. and Cheng, X. (2000). Crystal structure of the conserved core of protein arginine methyltransferase PRMT3. EMBO J. **19**, 3509-3519.

15. Weiss, V.H., Mcbride, A.E., Soriano, M.A., Filman, D.J., Silver P.A. and Hogle, J.M. (2000). The Structure and Oligomerization of the Yeast Arginine Methyltransferase, Hmt1. Nat. Struct. Biol. **7,** 1165-1171.

16. McBride, AE and Silver, PA. (2001). State of the arg: protein methylation at arginine comes of age. Cell **106**, 5-8.

17. Kraus, W.L. and Wong, J. (2002). Nuclear receptor-dependent transcription with chromatin. Is it all about enzymes?. Eur. J. Biochem. **269**, 2275-2283.

18. Woojin, A., Kim, J. and Roeder, R.G. (2004). Ordered Cooperative Functions of PRMT1, p300, and CARM-1 in Transcriptional Activation by p53. Cell **11**, 735-748.

19. 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. and Cheng, L. (2004). Aberrant expression of CARM-1, a transcriptional coactivator of androgen receptor, in the development of prostate carcinoma and androgen-independent status. Cancer **101**, 83-89.

20. El Messaoudi, S., Fabbrizio, E., Rodriguez, C., Chuchana, P., Fauquier, L., Cheng, D., Theillet, C., Vandel, L., Bedford, M. T. and Sardet, C. (2006) Coactivator-associated arginine methyltransferase 1 (CARM1) is a positive regulator of the Cyclin E1 gene. Proc. Natl. Acad. Sci. USA **103**, 13351-13356.

21. Allan, M., Manku, S., Therrien, E., Nguyen, N., Styhler, S., Robert, M.F., Goulet, A.C., Petschner, A.J., Rahil, G., Robert Macleod, A., Déziel, R., Besterman, J.M., Nguyen, H., and Wahhab, A. (2009). N-Benzyl-1-heteroaryl-3-(trifluoromethyl)-1H-pyrazole-5-carboxamides as inhibitors of co-activator associated arginine methyltransferase 1 (CARM1). Bioorg Med Chem Lett. **19**, 1218-1223.

22. Castellano, S., Milite, C., Ragno, R., Simeoni, S., Mai, A., Limongelli, V., Novellino, E., Bauer, I., Brosch, G., Spannhoff, A., Cheng, D., Bedford, M.T. and Sbardella, G.

(2010). Design, synthesis and biological evaluation of carboxy analogues of arginine methyltransferase inhibitor 1 (AMI-1). ChemMedChem. **5**, 398-414.

23. Huynh, T., Chen, Z., Pang, S., Geng, J., Bandiera, T., Bindi, S., Vianello, P., Roletto, F., Thieffine, S., Galvani, A., Vaccaro, W., Poss, M.A., Trainor, G.L., Lorenzi, M.V., Gottardis, M., Jayaraman, L. and Purandare, A.V. (2009). Optimization of pyrazole inhibitors of Coactivator Associated Arginine Methyltransferase 1 (CARM1). Bioorg Med Chem Lett. **19**, 2924-2927.

24. Purandare, A.V., Chen, Z., Huynh, T., Pang, S., Geng, J., Vaccaro, W., Poss, M.A., Oconnell, J., Nowak, K. and Jayaraman, L. (2008). Pyrazole inhibitors of coactivator associated arginine methyltransferase 1 (CARM1). Bioorg Med Chem Lett. **18**, 4438-4441.

25. Therrien, E., Larouche, G., Manku, S., Allan, M., Nguyen, N., Styhler, S., Robert, M.F., Goulet, A.C., Besterman, J.M., Nguyen, H. and Wahhab, A. (2009). 1,2-Diamines as inhibitors of co-activator associated arginine methyltransferase 1 (CARM1). Bioorg Med Chem Lett. **19**, 6725-6732.

26. Wan, H., Huynh, T., Pang, S., Geng, J., Vaccaro, W., Poss, M.A., Trainor, G.L., Lorenzi, M.V., Gottardis, M., Jayaraman, L. and Purandare, A.V. (2009). Benzo[d]imidazole inhibitors of Coactivator Associated Arginine Methyltransferase 1 (CARM1)--Hit to Lead studies. Bioorg Med Chem Lett. **19**, 5063-5066.

27. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode, Methods in Enzymology, Volume 276: Macromolecular Crystallography, part A, p.307-326, C.W. Carter, Jr. & R. M. Sweet, Eds.

28. Vagin, A.A. and Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. J. Appl. Cryst. **30**, 1022-1025.

29. Emsley, P. and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Cryst. **D60**, 2126-2132.

30. Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G. L. (1998) Crystallography and NMR System: a new software suite for macromolecular structure determination. Acta Cryst. **D54**, 905-921.

31. Delano, W. L. (2008) The PyMol Molecular Graphics System, DeLano Scientific LLC, Palo Alto, CA (http://www.pymol.org).

Table 1 Crystal structure data and refinement statistics		
CARM1 complex	CMPD-1	CMPD-2
PDB ID	2Y1W	2Y1X
Data collection		
Space Group	$P2_12_12$	$P2_12_12$
Cell Parameters (Å)		
a	74.896	75.558
b	98.471	98.757
c	207.184	207.456
X-ray source	ID14-4 ESRF	ID14-1 ESRF
Resolution (Å)	2.10	2.40
No. of observations		
Total	282,229	207,993
Unique	77,014	61,230
Completeness (%)	85.4 (30.0)	99.4 (99.9)
Rsym	0.080 (0.493)	0.116 (0.598)
I/σ I	14.1 (1.7)	9.4 (2.1)
Refinement		
Resolution (Å)	2.10	2.40
No. of reflections		
Working set (%)	73,115 (80.2)	58,154 (94.4)
Test set (%)	3,876 (5.0)	3,057 (5.0)
R_{cryst}/ R_{free}	0.205/ 0.244	0.208/ 0.264
R.m.s. deviations		
Bond lengths (Å)	0.012	0.010
Bond angles (°)	1.2	1.1
Ramachandran plot (%)		
Most favored	96.8	96.5
Allowed region	3.2	3.5
Outlier region	0	0

FIGURE CAPTIONS

Figure 1. Sequence alignment for human PRMT1, PRMT3 and CARM1.
Secondary structure elements corresponding to the CARM1 structure are shown below the sequences and the four characteristic PRMT motifs are shown above the sequences.

Figure 2. Chemical structures of CMPD-1, CMPD-2 and CMPD-3.

Figure 3. ITC results for titrations of CARM1 with CMPD-1 and CMPD-3.
Titrations with CMPD-1 and CMPD-3 were done in the absence and presence of 0.2 mM SAH and binding was observed only in the presence of SAH. From left to right there are the titrations of CMPD-1 without SAH, CMPD-1 with SAH, CMPD-2 without SAH and CMPD-2 with SAH, respectively.

Figure 4. Electron density associated with CMPD-1 and CMPD-2 in the CARM1 structures.
Diagrams showing the SigmaA weighted difference omit maps associated with the inhibitors CMPD-1 (Left Image) and CMPD-2 (Right Image) bound in the CARM1 active site contoured at the 3 σ level. Inhibitors are shown with light-blue carbon atoms and cofactors (SNF or SAH) are shown with orange carbon atoms.

Figure 5. Ribbon and surface diagrams of CARM1 with cofactors and inhibitors.
(A) Ribbon diagram showing the overall architecture of CARM1 with CMPD-1 (light-blue carbon atoms) and SNF (orange carbon atoms) bound in the active site. The N-terminal domain is shown in yellow and C-terminal domain is shown in green. (B) Ribbon diagram showing the overall architecture of CARM1 with CMPD-2 (light-blue carbon atoms) and SAH (orange carbon atoms) bound in the active site. (C) Close up view of CARM1•SNF•CMPD-1 with labels for select elements of secondary structure. The β4 strand is not visible in the diagram since it is hidden behind αB. (D) Comparison of the binding modes of CMPD-1 and CMPD-2 that result from the superposition of CARM1•SNF•CMPD-1 onto CARM1•SAH•CMPD-2. The CARM1 residues and the surface of the Arg-binding cavity are from the structure of CARM1•SAH•CMPD-2. For clarity, the front part of the protein has been removed to provide an unobstructed view of the Arg-binding cavity.

Figure 6. Views of CMPD-1 and CMPD-2 bound in the active site of CARM1.
(A) Close-up view of CMPD-1 (light-blue carbon atoms) and SNF (orange carbon atoms) bound in the active site of CARM1. Ordered water molecules are shown as red spheres and hydrogen bonds as dashed lines. (B) Close-up view of CMPD-2 (light-blue carbon atoms) and SAH (orange carbon atoms) bound in the active site of CARM1. (C) Diagram showing the active sites of PRMT1 (blue carbon atoms) and CARM1 (green carbon atoms) resulting from the superposition of PRMT1 (PDB code 1OR8) onto CARM1•SAH•CMPD-2. The residues shown are believed to be responsible for the CARM1 selectivity of CMPD-1 (cyan carbon atoms) and CMPD-2 (yellow carbon atoms). When present the labels indicate the identity of the corresponding PRMT1 residues. Several of the PRMT1 residues are partially or totally disordered in the 1OR8 structure. Ordered water molecules from the CARM1•SAH•CMPD-2

structure are shown as red spheres. (D) Diagram showing the active sites of PRMT3 (magenta carbon atoms) and CARM1 (green carbon atoms) resulting from the superposition of PRMT3 (PDB code 1F3L) onto CARM1•SAH•CMPD-2. The residues shown are believed to be responsible for the CARM1 selectivity of CMPD-1 (cyan carbon atoms) and CMPD-2 (yellow carbon atoms). The labels indicate the identity of the corresponding PRMT3 residues. Ordered water molecules from the CARM1•SAH•CMPD-2 structure are shown as red spheres.

Figure 1

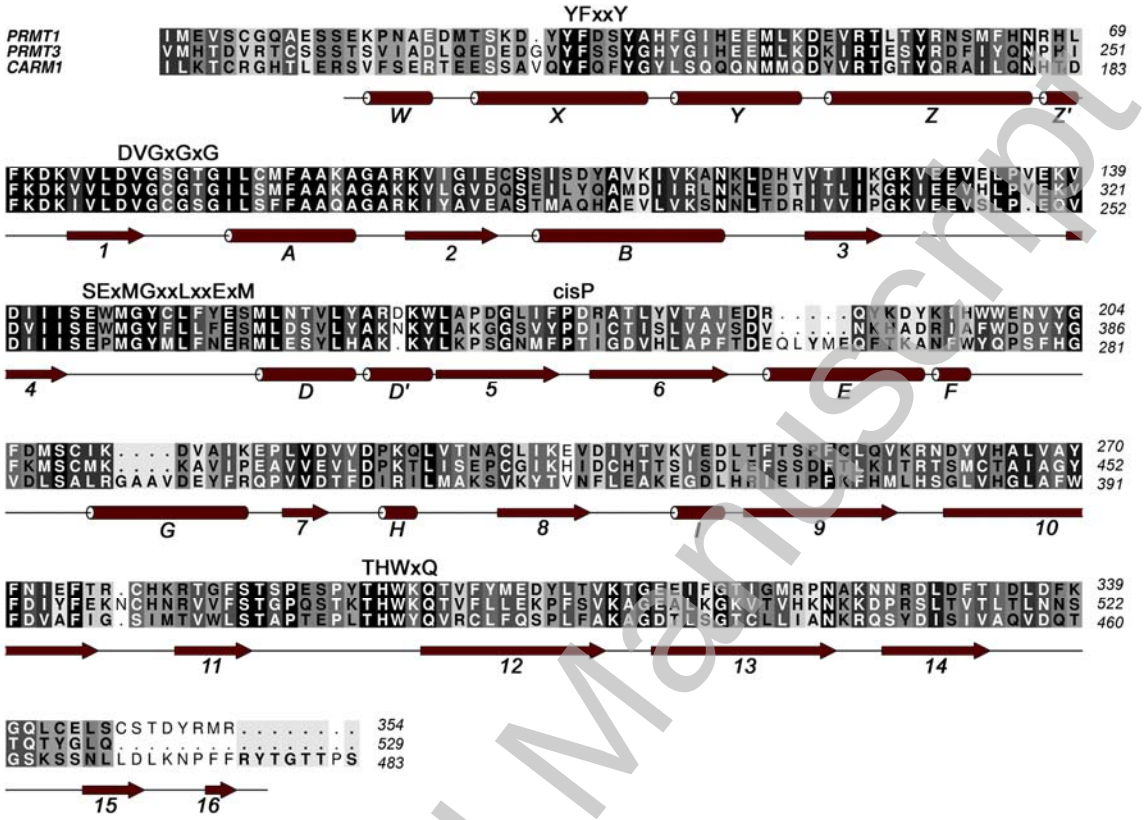

Figure 2

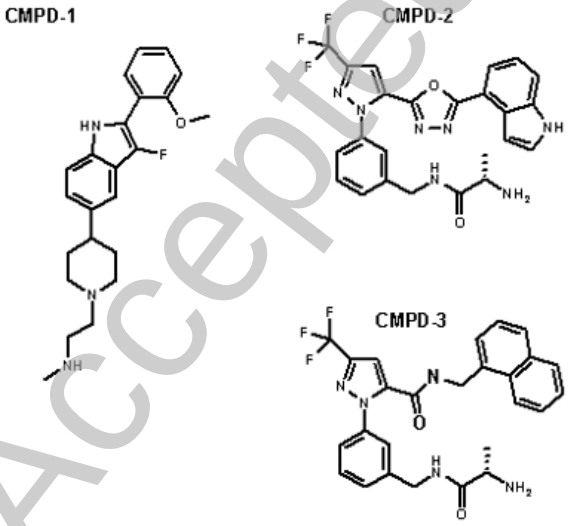

Figure 3

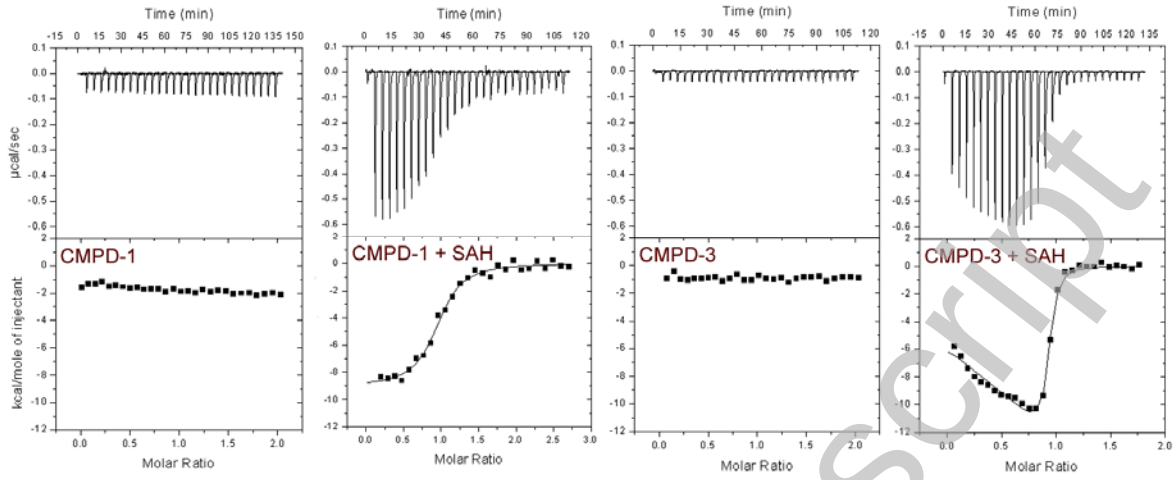

Figure 4

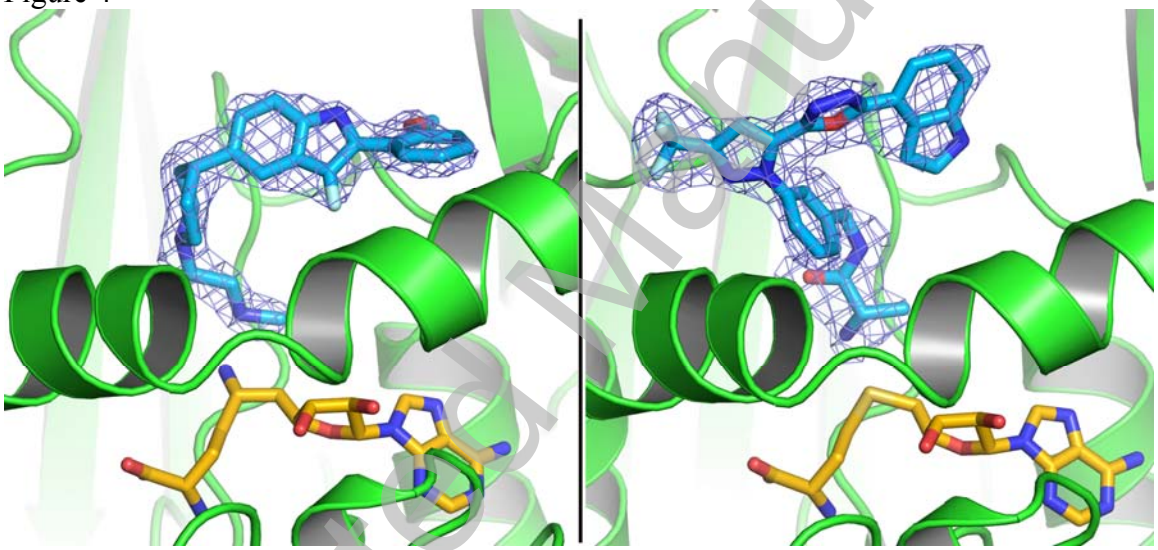

Figure 5

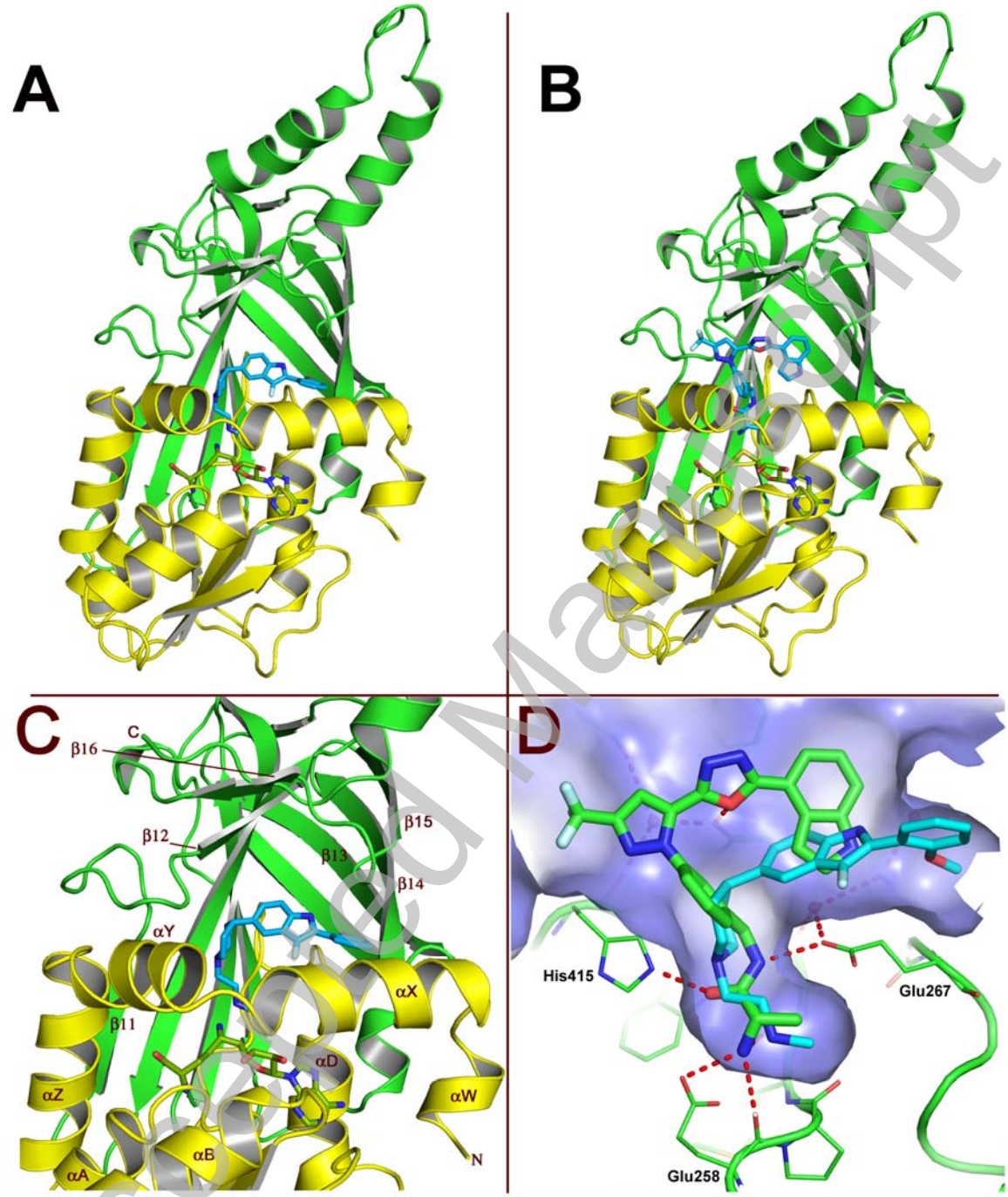

Figure 6