

# Journal of Medicinal Chemistry

Subscriber access provided by American Chemical Society

- Copyright permission to reproduce figures and/or text from this article

[View the Full Text HTML](#)



**ACS Publications**  
High quality. High impact.

Journal of Medicinal Chemistry is published by the American Chemical Society,  
1155 Sixteenth Street N.W., Washington, DC 20036

# Quantitative Three Dimensional Structure Linear Interaction Energy Model of 5'-O-[N-(Salicyl)sulfamoyl]adenosine and the Aryl Acid Adenylyating Enzyme MbtA

Nicholas P. Labello,<sup>†</sup> Eric M. Bennett,<sup>†</sup> David M. Ferguson,<sup>\*,†,‡</sup> and Courtney C. Aldrich<sup>†</sup>

Center for Drug Design, Academic Health Center, University of Minnesota, Minneapolis, Minnesota 55455, and Department of Medicinal Chemistry, University of Minnesota, Minneapolis, Minnesota 55455

Received June 3, 2008

MbtA (a salicyl AMP ligase) is a key target for the design of new antitubercular agents. On the basis of structure–activity relationship (SAR) data generated in our laboratory, a structure-based model is developed to predict the binding affinities of aryl acid–AMP bisubstrate inhibitors of MbtA. The approach described takes advantage of the linear interaction energy (LIE) technique to derive linear equations relating ligand structure to function. With only two parameters derived from molecular dynamics simulations, good correlation ( $R^2 = 0.70$ ) was achieved for a set of 31 inhibitors with binding affinities spanning 6 orders of magnitude. The results were applied to understand the effect of steric and heteroatom substitutions on bisubstrate ligand binding and to predict second generation inhibitors of MbtA. The resulting model was further validated by chemical synthesis of a novel inhibitor with a predicted LIE binding affinity of 1.6 nM and a subsequently determined experimental  $K_i^{app}$  of 0.7 nM.

## Introduction

Iron is an essential nutrient for almost all known organisms; however, its concentration in serum and human body fluids is approximately  $10^{-24}$  M, which is far too low to support bacterial colonization and growth.<sup>1</sup> In order to overcome this iron limitation, many bacterial pathogens such as *Mycobacterium tuberculosis*, the causative agent of tuberculosis (TB<sup>4</sup>), synthesize small molecule iron chelators termed siderophores for iron acquisition.<sup>2</sup> Consequently, siderophore biosynthesis has emerged as a promising biochemical pathway for antibiotic development.<sup>3</sup> The availability of detailed structural and biochemical information for aryl acid adenylation enzymes (AAAEs), which catalyze the first step in biosynthesis of aryl-capped siderophores, has made these enzymes attractive targets.<sup>4–6</sup> 5'-O-[N-(Salicyl)sulfamoyl]adenosine (**1**, Sal-AMS, Figure 1), a rationally designed bisubstrate analogue, is the prototypical AAAE inhibitor that has been shown to potently inhibit AAAEs from numerous organisms.<sup>7–9</sup> As shown in Figure 1, Sal-AMS comprises four modules: aryl, linker, glycosyl, and nucleobase. We have systematically and methodically examined the impact of each module on enzyme inhibition of the AAAE known as MbtA from *M. tuberculosis* as well as antitubercular activity to provide comprehensive structure–activity-relationships (SAR).<sup>10–15</sup>

The ability to predict the binding affinity of new compounds can be of substantial benefit during the optimization phase of drug development. While the difference in free energy of binding could be calculated exactly for two related molecules it is in practice an intractable problem to consider large numbers of

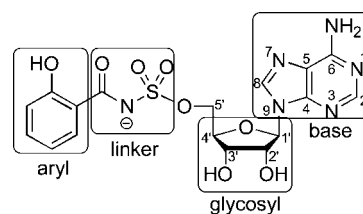


Figure 1. Structure of Sal-AMS (**1**).

ligands in this manner. The free energy perturbation method (FEP), for instance, requires dozens of converged molecular dynamics (MD) simulations for each ligand. On the other hand, the literature contains a number of reports where a predictive binding model is constructed using experimental binding affinities to weight theoretical interaction energies.<sup>16–19</sup> Within such linear interaction energy (LIE) approximations, binding affinities are estimated after only one ligand–receptor and one ligand–solvent simulation for each additional compound. The resulting models often display high correlation and, in contrast to some activity relationship models, have the advantage of being structure-based and therefore serve as interpreters and guides for rational drug design. While most such studies utilize crystal structures, there are examples in the literature where homology models have been substituted with good results.<sup>20,21</sup>

We present a structure based analysis and linear LIE model for the Sal-AMS scaffold with emphasis on providing a quantitative model for predicting binding affinities and a grounded physical interpretation of the SAR to guide future synthesis. Modifications of the nucleobase are of particular interest, as this moiety represents the best opportunity for improving potency and increasing specificity and lipophilicity. The linker and glycosyl regions are also examined, as variation of these moieties may be required to modify the number of hydrogen bond donors and acceptors or otherwise tune pharmacokinetic properties.

## Computational Methods

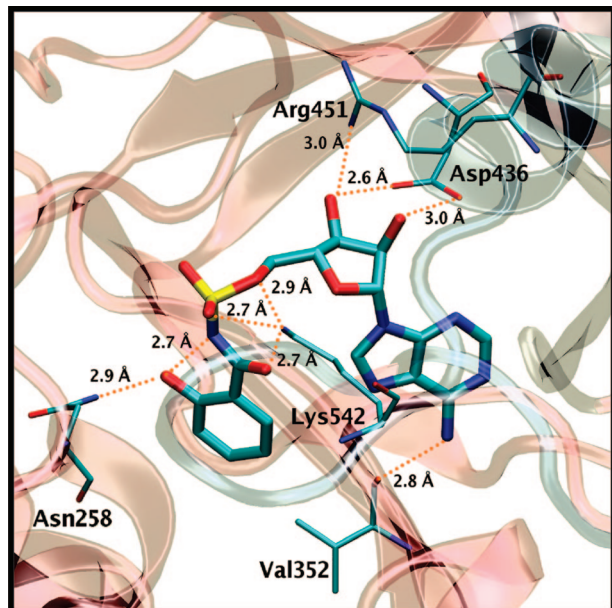
**Receptor and Ligand Starting Structures.** Although a crystal structure for MbtA is not yet available, we have detailed the

\* To whom correspondence should be addressed. Address: Department of Medicinal Chemistry, University of Minnesota, 308 Harvard Street SE, 8-101 WDH, Minneapolis, MN 55455. Telephone: 612-626-2601. Fax: 612-624-0139. E-mail: ferguson@umn.edu.

<sup>†</sup> Center for Drug Design, University of Minnesota.

<sup>‡</sup> Department of Medicinal Chemistry, University of Minnesota.

<sup>4</sup> Abbreviations: AAAE, aryl acid adenylyating enzyme; ATP, adenosine triphosphate; FEP, free energy perturbation;  $K_i^{app}$ , apparent inhibition constant; LIE, linear interaction energy; MD, molecular dynamics; MM, molecular mechanics; PDB, Protein Data Bank; QM, quantum mechanics; QSAR, quantitative structure–activity relationship; rmsd, root-mean-square deviation; SAR, structure–activity relationships; Sal-AMS, 5'-O-[N-(salicyl)sulfamoyl]adenosine; TB, tuberculosis.



**Figure 2.** Sal-AMS (**1**) displayed in tube representation bound in a homology model of the MbtA binding site.<sup>10</sup> The N-terminal domain residues are presented in red ribbon, and the C-terminal domain residues are in blue. Residues that make important electrostatic contacts with the ligand are presented in tube representation. The orange dotted lines are possible hydrogen bonds. Hydrogens are not shown for clarity.

construction of a homology model in a previous publication.<sup>10</sup> Our homology model is based on the cocrystal structure of Dhbe with an adenylated 2,3-dihydroxybenzoic acid (2,3-DHB).<sup>22</sup> Dhbe shares 42% sequence identity with MbtA but almost absolute conservation of active site residues. Thus, 16 of the 21 residues within 4 Å of the adenylated ligand are identical in MbtA and the remaining 5 changes are conservative. Three of these variations in MbtA (Y236F, S240C, V337L) map to the aryl acid substrate binding pocket and are responsible for conferring selectivity to the native substrate salicylic acid (Sal) over 2,3-DHB. The predicted binding conformation of Sal-AMS within the homology model is shown in Figure 2.

The ligand Sal-AMS (**1**) assumes a relatively compact form when bound to the receptor compared to the extended conformations that are possible in solution phase. Sal-AMS forms an internal hydrogen bond between the 2-hydroxy and negatively charged nitrogen atom of the acylsulfamate linker, which enforces a coplanar arrangement of the aryl moiety and linker carbonyl. The aryl binding pocket is largely nonpolar, and only a single hydrogen bond between the carboxamide side chain of Asn258 and the aryl hydroxyl group of **1** is predicted. The linker moiety of **1** interacts with conserved Lys542 (protonated) via a network of three hydrogen bonds and electrostatically with the negatively charged sulfamate as shown in Figure 2. The glycosyl moiety adopts a 3'-endo pucker, and both 2'- and 3'-hydroxyl groups are predicted to form a bidentate hydrogen bond with Asp436. Deletion of either hydroxyl group in **8** and **9** (Table 1) does not disturb binding affinity, while deletion of both oxygen atoms in analogue **10** results in an approximate 10-fold loss in binding affinity. The interactions of the nucleobase are largely hydrophobic with a single predicted hydrogen bond between the exocyclic N-6 amino group and Val352, which is further supported by the 238-loss in binding affinity when the N-6 amino group is replaced with an oxygen atom. Significantly, no electrostatic interactions of the purine N-1 and N-3 atoms are observed, and the N-7 shows poor hydrogen bonding geometry with Gly330 and Gly354. A small pocket vicinal to the N-6 amino group enables binding of no more than three carbon atoms, while the interdomain region of the receptor adjacent to the C-2 purine position is flexible as predicted by molecular dynamics simulations and can accommodate bulky substituents. A major consequence of the significant and relatively rigid hydrogen bonding structure is

**Table 1.** LIE Coefficients for Various Receptor and Ligand Groups<sup>a</sup>

model	$A = \alpha\Delta G_{vdw} + \beta\Delta G_{Ei} + \gamma$			no. of comps	$R^2$
	$\alpha$	$\beta$	$\gamma$		
I	X	$6.49 \times 10^{-3}$	-8.22	31	0.66
N I	$1.04 \times 10^{-2}$	$1.31 \times 10^{-2}$	-8.36	25	0.71
L I	$1.58 \times 10^{-2}$	$4.98 \times 10^{-3}$	-8.23	6	0.99
II	$1.37 \times 10^{-2}$	$7.48 \times 10^{-3}$	-8.48	31	0.70
BII	$1.64 \times 10^{-2}$	$1.14 \times 10^{-2}$	-8.66	21	0.81

<sup>a</sup> N, L, and B refer to nucleoside, linker, and base compound sets, respectively. I and II refer to truncation procedures 1 and 2, respectively.  $\alpha$  and  $\beta$  have units of 1/(kJ/mol) so that multiplication by an interaction energy in kJ/mol results in a pure number.  $\gamma$  is unitless.  $A$  is an estimation of  $\log(K_i^{app})$ . X indicates that the coefficient was statistically insignificant.

that the core of the ligand scaffold can be considered anchored by these contacts and has little room for adjustment. For substituents added to the core to have the desired effect of improving potency, they must readily access the available binding pockets from their attachment to the core without disrupting the structure apparent in Figure 2. Compounds **16** and **28** are representative ligands that do exactly that, accessing the N-6 and C-2 binding clefts, respectively.

While the model was found to be somewhat sensitive to starting structures (because of the relatively rigid conformation of the core scaffold discussed above), most ligands have only one conformation of interest. The active conformation of the ligand Sal-AMS was constructed in the homology model by mutation of Dhbe's adenylated natural substrate 2,3-DHB and subsequent minimization. Ligands with large adenine C-2 substituents result in significant changes of nearby residues such as Lys332 and Val448. A second homology model differing from the Dhbe crystal structure but allowing for the exceptional potency of compounds **25–31** was used to model these ligands. Briefly, the largest difference involves the loss of a hydrogen bond between the two residues mentioned above and relaxation of the receptor's tertiary structure around the bulkier ligands. MD simulations indicate that the C-2 position, despite having little obvious room for substitution in the Dhbe crystal structure, is located in an extremely flexible region of the protein at the seam of the two domains and linker and thus very amenable to substitution.<sup>10</sup> As this work progressed, it was found that ligands with relatively small modifications of **1** could be studied effectively with the C-2 pocket residues optimized in either conformation.

**Molecular Dynamics.** Two molecular dynamics simulations are necessary to estimate the binding affinity of a single ligand (ligand in receptor, ligand in solvent). Two truncation procedures were applied to the homology model before MD simulation. In the first, all residues within 10 Å of the ligand were included in the calculation, and 21 complete residues, those with any atom 3 Å or closer to Sal-AMS (**1**) after a frozen-protein minimization, were allowed to move freely through the course of the simulation. The outer shell was frozen. After minimization and a 25 ps equilibrium run, data were sampled every 25 fs and averaged over the entire simulation. In the second procedure, all residues within 14 Å of the ligand were included in the simulation. The 33 residues with any atom within 4 Å of **27** were allowed to move freely while the outer shell was frozen. To eliminate possible differences due to the conformation of side chains near the C-2 pocket, the second truncation procedure was only applied to the more permissive homology model, which was used with all of the ligands. After an identical equilibrium run data averaged over the entire simulation were collected. In every case a GB/SA<sup>23</sup> implicit solvation model was used to approximate the effects of solvent. The running average of the force field terms was collected and examined at 10, 25, and 50 ns for each ligand. As the terms changed very little between the 25 and 50 ns simulations, no longer calculations were completed.

**LIE Terms.** The MD-LIE terms collected in this work differ only by which truncation procedure was used and length of simulation. All of the MD models reported utilize values from the 50 ns simulation of the ligand in solution (implicit solvation) and the electrostatic and van der Waals terms from increasingly lengthy simulations of the complexes to yield the  $\Delta G_{vdw}$  and  $\Delta G_{Ei}$  terms

of eq 1. MD calculations were completed with MacroModel<sup>24</sup> 9.1 using the OPLS force field.<sup>25</sup>

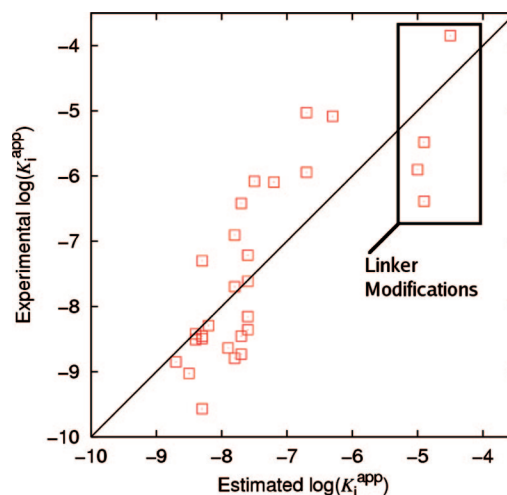
## Results and Discussion

The most basic LIE model is given below in eq 1 where  $A$  is a free energy,  $IC_{50}$ ,  $K_D$ , or some other experimentally measured indication of binding affinity,  $\Delta G_{vdW}$  and  $\Delta G_{EI}$  are the force field determined van der Waals and electrostatic interaction energies of the receptor–ligand complex, and  $\alpha$ ,  $\beta$ , and  $\gamma$  are constants usually derived through multivariate regression to best fit the available data.

$$A = \alpha\Delta G_{vdW} + \beta\Delta G_{EI} + \gamma \quad (1)$$

The  $\Delta G_{vdW}$  term approximates nonpolar binding contributions, while  $\Delta G_{EI}$  accounts for electrostatic free energy differences of the ligand–solvent and ligand–receptor complexes. Each represents the difference in interaction energy of the ligand in (a) the receptor environment and (b) the solvent environment. While the constants are typically unique for each receptor and force field, once  $\alpha$ ,  $\beta$ , and  $\gamma$  are established, an affinity prediction for a new compound can be determined with minimal effort, requiring only the relevant interaction energy calculation. We chose to transform the interaction energies calculated from the MD simulations ( $\Delta G_{vdW}$  and  $\Delta G_{EI}$ ) into values relative to **1** (the parent compound) before further processing. As a result, the ideal value of  $\gamma$  is the experimental  $\log(K_i^{app})$  of compound **1** (−8.3). Variation in the binding affinity estimation,  $A$ , for each compound is determined by the first two terms. For example, desolvating an additional hydrophobic group relative to **1**, as in the case of **28**, results in a negative  $\Delta G_{vdW}$  and correspondingly lower (better binding) estimation of  $A$ . An additional term fitting the difference in solvent accessible surface area (SASA) is sometimes included to further improve correlation; however, we found no significant improvements when including a SASA term possibly because of the polar nature of the ligands (a dipole moment of 16.8 D, for instance, was calculated for a compact solution phase conformation of **1** at the B3LYP<sup>26,27</sup> level of theory and a basis set<sup>28,29</sup> designed for calculating dipole moments) or the van der Waals issues with some ligands growing too large for the pocket discussed below. More recently, specific receptors have been studied with linear response methods that replace the fitted van der Waals and electrostatic terms with a QM/MM interaction energy.<sup>30,31</sup> The QM/MM interaction energy was calculated for each ligand at both minimized and time averaged single points by treating the protein with the OPLS force field and the ligand at the B3LYP/6-31G\*<sup>32</sup> level of theory. While trends were noted for specific ligand sets, the QM/MM interaction energy was incapable of describing all or even most of the ligand binding in even a qualitative sense. We concluded that in this case evaluation of a relatively large ensemble is more important for estimating binding affinities than highly accurate studies on a single or smaller number of structures. Furthermore, attempts to correlate binding affinity with short 10–50 ps MD simulations featuring the entire unrestricted MbtA protein resulted in worse correlation than longer simulations (150–300 ps) in a truncated receptor.

Initially, a poor correlation was noted when including all active ligands in a single model, and we opted to separate the inhibitors based on the modular structure of Sal-AMS. The analogues were separated into groups corresponding to modifications in the aryl, linker, glycosyl, and base moieties. Most of the error in a unified model resulted from compounds featuring modifications of the aryl ring. As a result of the conjugated nature of the salicyl moiety, much of the modulation



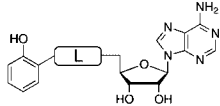
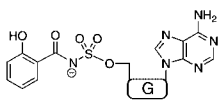
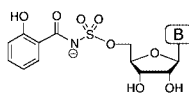




















**Figure 3.** Experimental vs estimated binding affinity of MbtA inhibitors, with LIE model of all compounds within the small receptor model. The boxed ligands all feature modifications of the linker such that the charged nitrogen is removed.

in potency caused by substitution of the aryl ring may be indescribable by classical physics, instead requiring electronic structure and polarization effects. For example, substituting the para-position of the aryl ring with the trifluoromethyl group results in a 380-fold difference in potency, suggesting force fields based methods may be incapable of describing the phenomena without highly specific reparameterization. Regardless, almost every modification (>20) on the ring has resulted in a significant loss of potency and the focus of this work is on the adenosine binding site.<sup>22</sup> The linker is a critically important region, as it must be metabolically stable and correctly position the aryl and nucleoside moieties in their respective binding pockets.<sup>13</sup> A systematic series of six compounds differing only in the linker feature widely varying potencies ranging from 3.8 nM to 143  $\mu$ M.<sup>9,11,13–15</sup> The complete final set of compounds that we investigated additionally includes 25 inhibitors with modifications of the glycosyl and nucleobase groups. All  $K_i^{app}$  values were determined using an ATP pyrophosphate exchange assay under identical assay conditions, and tight-binding inhibition observed for many of the most potent compounds was accounted for using the analysis of Morrison.<sup>14</sup> Therefore, differences in binding affinity as measured by the assay are representative of differences in free energy. The binding affinities of the inhibitors span 6 orders of magnitude with  $K_i^{app}$  values ranging from 270 pM to 143  $\mu$ M.

The smaller of the two receptor models (truncation procedure 1) has the distinct advantage of being much more computationally efficient allowing longer simulation times while simultaneously converging in fewer time steps because of the fewer degrees of freedom. Evaluation of eq 1 using 150 ps simulations of all ligand–receptor complexes results in an  $R^2$  value of 0.66. Experimental  $\log(K_i^{app})$  are plotted against those predicted by the equation in Figure 3. The resulting coefficients necessary to predict binding affinity are presented in Table 1. The structures of these compounds are reported in Table 2 along with the  $K_i^{app}$  and  $\log(K_i^{app})$  used to construct the model.

Examination of the coefficients reveals that the intercept constant,  $\gamma$ , maintains a very reasonable value of −8.22. As the interaction energies of eq 1,  $\Delta G_{vdW}$ , and  $\Delta G_{EI}$ , were transformed into values relative to Sal-AMS (**1**) before further statistical analysis, the ideal value of  $\gamma$  is the  $\log(K_i^{app})$  of Sal-AMS (the relative  $\Delta G$  and  $\Delta G_{EI}$  terms are always zero for this compound). The  $K_i^{app}$  values of the additional compounds are

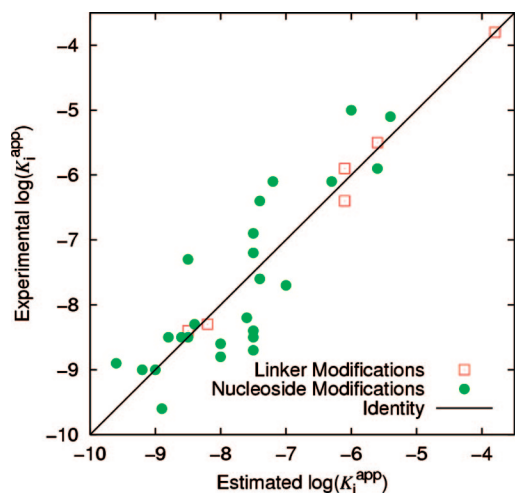
**Table 2.** Experimental and Calculated  $\log(K_i^{\text{app}})$  Values<sup>a</sup>

										
Entry	$K_i^{\text{app}}$ (nM)	$\log(K_i^{\text{app}})$	Calculated $\log(K_i^{\text{app}})$	Entry	$K_i^{\text{app}}$ (nM)	$\log(K_i^{\text{app}})$	Calculated $\log(K_i^{\text{app}})$			
L =		5.1 <sup>b</sup>	-8.3	-8.5	G =		5.1 <sup>b</sup>	-8.3		
		3.8 <sup>b</sup>	-8.4	-8.4			2.3 <sup>b</sup>	-8.6	-8.3	
		410 <sup>c</sup>	-6.4	-4.9			3.5 <sup>b</sup>	-8.5	-8.3	
		3,300 <sup>d</sup>	-5.5	-5.0			3.2 <sup>b</sup>	-8.5	-8.3	
		1,250 <sup>e</sup>	-5.9	-6.5			830 <sup>b</sup>	-6.1	-7.6	
		143,000 <sup>c</sup>	-3.8	-5.1			61 <sup>b</sup>	-7.2	-7.3	
										
Entry	$K_i^{\text{app}}$ (nM)	$\log(K_i^{\text{app}})$	Calculated $\log(K_i^{\text{app}})$	Entry	$K_i^{\text{app}}$ (nM)	$\log(K_i^{\text{app}})$	Calculated $\log(K_i^{\text{app}})$			
Compound, B =				Compound, B =						
	5.1 <sup>b</sup>	-8.3	-8.5		21, R = -CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>2</sub> -	1.85 <sup>f</sup>	-8.7	-7.8		
	12, X = N, Y = CH	24.3 <sup>f</sup>	-7.6		-8.3		22, R = -CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>2</sub> -	124 <sup>f</sup>	-6.9	-7.0
		13, X = Y = CH	20.1 <sup>f</sup>		-7.7		-8.1		23, R = -CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>2</sub> -	9,410 <sup>f</sup>
		14	802 <sup>f</sup>		-6.1	-8.0			24, R = -CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>2</sub> -	8,260 <sup>f</sup>
		15	380 <sup>f</sup>	-6.4	-7.9			25, R = -CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>2</sub> -	3.03 <sup>f</sup>	-8.5
		16, R = Me	1.6 <sup>f</sup>	-8.8	-8.6			26, R = -CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>2</sub> -	0.94 <sup>f</sup>	-9.0
		17, R = Et	4.4 <sup>f</sup>	-8.4	-8.0			27, R = -CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>2</sub> -	0.40 <sup>f</sup>	-9.4
		18, R = <i>n</i> -Pr	3.5 <sup>f</sup>	-8.5	-7.8			28, R = -CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>2</sub> -	0.27 <sup>f</sup>	-9.6
		19, R = <i>i</i> -Pr	7.0 <sup>f</sup>	-8.2	-7.9			29, R = -CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>2</sub> -	47 <sup>f</sup>	-7.3
		20, R = <i>i</i> -Bu	1140 <sup>f</sup>	-5.9	-6.2			30, R = -CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>2</sub> -	1.4 <sup>f</sup>	-8.9
								31, R = -CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>2</sub> -	0.99 <sup>f</sup>	-9.0

<sup>a</sup> Calculated using interaction energies from 150 ps MD simulations and truncation procedure II ( $\alpha = 1.37 \times 10^{-2}$ ,  $\beta = 7.48 \times 10^{-3}$ ,  $\gamma = -8.48$ ) described in text. <sup>b</sup> Reference 9. <sup>c</sup> Reference 12. <sup>d</sup> Reference 15. <sup>e</sup> Reference 11. <sup>f</sup> Reference 10.

then modulated by the  $\Delta G_{\text{vdW}}$  and  $\Delta G_{\text{El}}$  terms. The coefficients  $\alpha$  and  $\beta$  are both always positive indicating that the model responds intuitively to the  $\Delta G$  terms rewarding compounds that are predicted to interact more strongly with the receptor with greater binding affinities. Closer inspection, however, reveals

that  $\alpha$  is nearly statistically insignificant and that most of the correlation is due to the electrostatic term. The most important reason for this phenomenon appears to be due to compounds **15–24**, where a series of aliphatic and aromatic groups are appended onto the N-6 position of the purine. The removal of

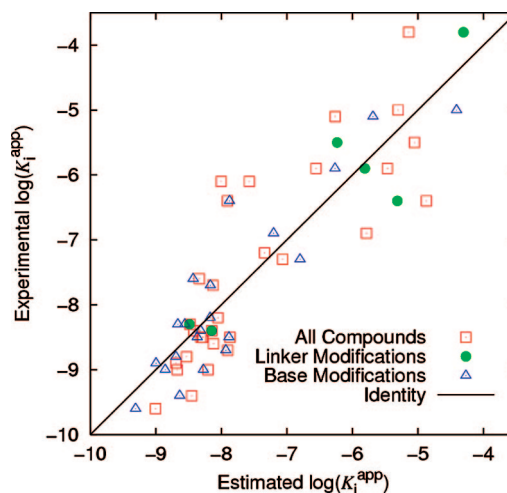


**Figure 4.** Plots of the experimental versus estimated binding affinities for MbtA inhibitors using model I after 150 ps of molecular dynamics simulations. The linker set (6 compounds, hollow red squares) and nucleoside set (25 compounds, filled green circles) are treated separately (see Table 1, LI and NI, respectively). The solid identity line represents the point where prediction is equal to the experimental value.

these hydrophobic groups from solution results in increasingly favorable  $\Delta G_{\text{vdW}}$  contributions. The binding affinity, however, shows a correlated decrease as the substituents grow larger than three carbon atoms. The isobutyl, cyclobutyl, cyclopentyl, and benzyl substituents are larger than the volume available in the N-6 pocket and significantly displace the nucleoside (rms of heavy base atoms 1.6–1.9 Å) making room for the large hydrophobic groups. As the more polar glycosyl and linker regions are displaced, an energetic penalty is partially expressed in the electrostatic term, recovering correlation though in a nonphysical manner resulting from unrealistic binding modes.

Further examination of Figure 3 reveals that even for compounds without large substituents the electrostatic term alone is not an adequate predictor of activity. Four ligands included in the set modify the linker such that the negatively charged nitrogen is either removed from the scaffold or has neighboring atoms modified such that it is neutral within the receptor. Removing the compounds with linker modifications increases the  $R^2$  to 0.71 for the remaining set (Figure 4). While most of the correlation is still due to the electrostatic term,  $\alpha$  increases to a meaningful value of  $1.04 \times 10^{-2}$ . Increasing simulation time to 300 ps resulted in no significant changes to the model and only very slightly improved  $R^2$  to 0.72. Simulations longer than 300 ps did not improve correlation. The six compounds with varying linker groups (1–6), when treated separately from the nucleoside modifications, demonstrate an extremely high  $R^2$  of 0.99. While the correlation would undoubtedly decrease with a larger set of compounds, the physically sensible values of the coefficients ( $\alpha = 1.58 \times 10^{-2}$ ,  $\beta = 4.98 \times 10^{-3}$ ,  $\gamma = -8.23$ ) and individual  $R^2$  values resulting from  $\Delta G_{\text{vdW}}$  and  $\Delta G_{\text{El}}$  of 0.39 and 0.82, respectively, indicate that this small receptor model should be a reasonable starting point for interpretation of any similar future modifications. It also implies that differences in charge or protonation state may require reparameterizing the electrostatic coefficient or properly accounting for this effect by electrostatic continuum calculations.

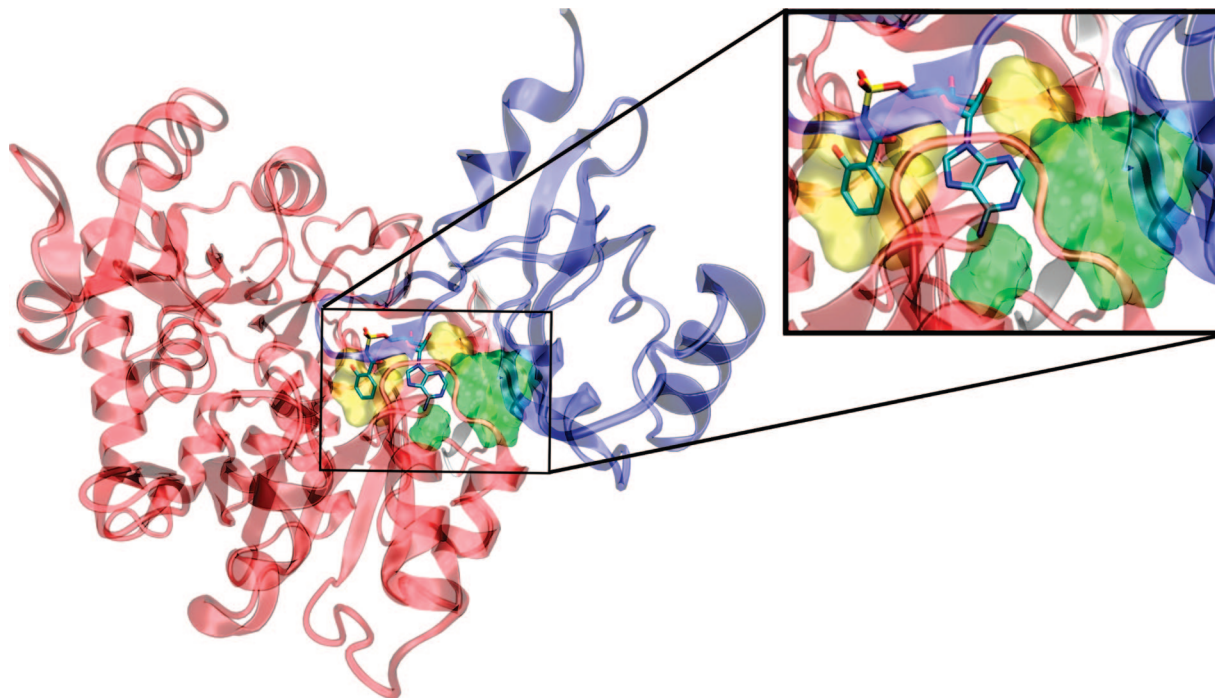
Several issues were noted that indicate that a larger, more flexible model may be important to provide a fully rational interpretation of the binding affinity data. Charge–charge interactions in particular are long-range, suggesting that a larger protein shell may be necessary if ligands with different charges



**Figure 5.** Experimental vs estimated binding affinity of MbtA inhibitors: experimental  $\log(K_i^{\text{app}})$  vs predicted activity. All compounds are included in a single set (hollow red squares), though improvement is still noted by clustering ligands into groups based on which region of the scaffold was modified (linker, filled green circles, and base modifications separately as hollow blue triangles).  $R^2$  (all compounds,  $\log(K_i^{\text{app}}) = 0.0137\Delta G_{\text{vdW}} + 0.00748\Delta G_{\text{El}} - 8.48$ ) = 0.70.

are to be included in a single model. Additionally, the relatively poor correlation of the van der Waals term and observation that several N-6 and C-2 substituents are large enough to extend well beyond the volume occupied by **1** indicate that it may be important to extend the shell of residues allowed to move. To address these deficiencies, a study similar to that described above was completed, though modeling all residues with any atom closer than 14 Å and allowing 33 residues in an expanded inner shell to move freely. Interestingly, many aspects of the model were improved (Figure 5). The van der Waals term is considerably more significant, demonstrating an  $R^2$  of 0.33. The electrostatic term continues to correlate well,  $R^2 = 0.56$ , and remains independent of the van der Waals term demonstrated by the cross correlation between the two of only 0.08. Furthermore, the improvement from using both terms over the entire set of 31 ligands is clear with an  $R^2$  of 0.70 and standard deviation of 0.85 log units. The leave-group-out (or leave-one-out) cross-validation method was run for 100 cycles with group sizes of 1, 2, 4, and 9. That is, 100 (or 31 for a group size of 1) randomly selected groups of molecules of the appropriate size were treated as a test set with the remainder serving as a training set. The average correlation coefficient ( $q^2$ ) of the test sets was generally near and did not fall below 0.60 in any of the validation tests. Furthermore, manually breaking the ligands into a trainer and test set of 21 and 10 ligands (chosen to represent modifications of the linker, glycosyl, and base subunits) yields  $R^2$  values of 0.69 and 0.59, respectively, in line with the leave-group-out analysis.

Correlation can be increased further if the ligands are broken into sets that split away the four compounds with very different electrostatic properties. While all parameters are allowed to reoptimize freely, the van der Waals term encouragingly varies little between groups while the electrostatic coefficient adjusts significantly accounting for the difference in electrostatic interaction energies between charged and noncharged species. Finally, issues apparent in the smaller, less flexible receptor model have been addressed. The N-6 pocket appears to be more realistically modeled, as the substituents that crowd the small cleft no longer displace a large part of the ligand resulting in electrostatic penalties but instead see van der Waals energetic



**Figure 6.** Sal-AMS (1) docked in the MbtA homology model. The protein is represented as a ribbon diagram: N-terminal, red; C-terminal, blue. The ligand is shown in a tube representation and colored by element. The green surfaces expose volumes that, when filled with hydrophobic groups, improve potency and specificity. The larger of the two is the C-2 cleft (accessed by compounds **25–31**), while the smaller is the N-6 pocket (accessed by compounds **15–24**). Filling the yellow volumes results in a reduction in potency.

penalties consistent with steric crowding. The electrostatic term of the four neutral ligands does not fit well with the remainder of the set, but reduced significance of  $\Delta G_{\text{EI}}$  with respect to  $\Delta G_{\text{vdW}}$  resulting from the larger, more flexible receptor allows for the full set to be included in a unified model and still exhibit quite reasonable predictive power. For affinity predictions of new molecules, however, a model trained with only the most closely related compounds appears to be the wisest course of action. For example, to investigate dual modifications of the scaffold at the purine N-6 and C-2 positions a priori, we would consider only the 21 compound subset representing single modifications at these positions.

## Conclusions

We have presented the first computational analysis of MbtA and analogues based on 5'-O-[N-(salicyl)sulfamoyl]adenosine, a promising class of new antibacterial compounds with a novel mechanism of action. The LIE model provides a clear framework for both qualitatively predicting trends and quantitatively estimating binding affinities of related compounds within MbtA using only two short MD simulations. The improvements in predictive power realized with a larger, more flexible protein model speaks well for the applicability of structure based design and interpretation for MbtA inhibitors within our homology model. Moreover, it seems reasonable to expect that similar methods should be applicable to a number of homologous aryl acid adenylating enzymes such as BasE, YbtE, VibE, PchD, and EntE (from the pathogens *Acinetobacter baumannii*, *Yersinia pestis*, *Vibrio cholerae*, *Pseudomonas aeruginosa*, and *Escherichia coli*, respectively).

Figure 6 presents an amalgamation of experimental and theoretical work, summarizing the structure–activity relationship data used in the development of the LIE model. The schematic shows where the scaffold can be most easily modified to favorably access additional volume within MbtA and improve

potency and specificity. As a final check of the applicability of the model, we have applied the LIE analysis to predict the binding affinity of a compound not included in the initial training set that includes modifications to both the N-6 and C-2 positions. The binding affinity of N-6-cyclopropyl-2-phenyl-Sal-AMS was estimated using the LIE model to be 1.6 nM, indicating the activity would fall between that of **23** and **28**. This compound, the first dual modification of the scaffold, has been synthesized and found to have a  $K_i^{\text{app}}$  of 0.7 nM.<sup>33</sup> The analogue is approximately twice as potent as **23** but 3-fold less active than **28**, in concert with the LIE predicted result. Additional applications of the LIE model are currently underway to guide the synthesis of even more diverse Sal-AMS analogues, and on the basis of the results presented here, significant gains in potency is expected.

**Acknowledgment.** This research was supported by a grant from the NIH (Grant R01AI070219) and funding from the Center for Drug Design, Academic Health Center, University of Minnesota to C.C.A. We thank the Minnesota Supercomputing Institute for computing time.

**Supporting Information Available:** Complete details describing the synthesis, compound characterization, and HPLC purity of N<sup>6</sup>-cyclopropyl-5'-O-[N-(2-hydroxybenzoyl)sulfamoyl]-2-phenyladenosine sodium salt. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- (1) Raymond, K. N.; Dertz, E. A.; Kim, S. S. Enterobactin: an archetype for microbial iron transport. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3584–3588.
- (2) Crosa, J. H.; Walsh, C. T. Genetics and assembly line enzymology of siderophore biosynthesis in bacteria. *Microbiol. Mol. Biol. Rev.* **2002**, *66*, 223–249.
- (3) Monfeli, R. R.; Beeson, C. Targeting iron acquisition by *Mycobacterium tuberculosis*. *Infect. Disord.: Drug Targets* **2007**, *7*, 213–220.

- (4) Drake, E. J.; Nicolai, D. A.; Gulick, A. M. Structure of the EntB multidomain nonribosomal peptide synthetase and functional analysis of its interaction with the EntE adenylation domain. *Chem. Biol.* **2006**, *13*, 409–419.
- (5) Ehmann, D. E.; Shaw-Reid, C. A.; Losey, H. C.; Walsh, C. T. The EntF and EntE adenylation domains of *Escherichia coli* enterobactin synthetase: sequestration and selectivity in acyl-AMP transfers to thiolation domain cosubstrates. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2509–2514.
- (6) May, J. J.; Kessler, N.; Marahiel, M. A.; Stubbs, M. T. Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12120–12125.
- (7) Ferreras, J. A.; Ryu, J. S.; Di Lello, F.; Tan, D. S.; Quadri, L. E. Small-molecule inhibition of siderophore biosynthesis in *Mycobacterium tuberculosis* and *Yersinia pestis*. *Nat. Chem. Biol.* **2005**, *1*, 29–32.
- (8) Miethke, M.; Bissere, P.; Beckering, C. L.; Vignard, D.; Eustache, J.; Marahiel, M. A. Inhibition of aryl acid adenylation domains involved in bacterial siderophore synthesis. *FEBS J.* **2006**, *273*, 409–419.
- (9) Somu, R. V.; Wilson, D. J.; Bennett, E. M.; Boshoff, H. I.; Celia, L.; Beck, B. J.; Barry, C. E.; Aldrich, C. C. Antitubercular nucleosides that inhibit siderophore biosynthesis: SAR of the glycosyl domain. *J. Med. Chem.* **2006**, *49*, 7623–7635.
- (10) Neres, J.; Labello, N. P.; Somu, R. V.; Boshoff, H. I.; Wilson, D. J.; Vannada, J.; Chen, L.; Barry, C. E., 3rd; Bennett, E. M.; Aldrich, C. C. Inhibition of siderophore biosynthesis in *Mycobacterium tuberculosis* with nucleoside bisubstrate analogues: SAR of the nucleobase domain of 5'-O-[N-(salicyl)sulfamoyl]adenosine. *J. Med. Chem.* **2008**, *51*, 5349–5370.
- (11) Qiao, C. H.; Gupte, A.; Boshoff, H. I.; Wilson, D. J.; Bennett, E. M.; Somu, R. V.; Barry, C. E.; Aldrich, C. C. 5'-O-[(N-Acyl)sulfamoyl]adenosines as antitubercular agents that inhibit MbtA: an adenylation enzyme required for siderophore biosynthesis of the mycobactins. *J. Med. Chem.* **2007**, *50*, 6080–6094.
- (12) Qiao, C. H.; Wilson, D. J.; Bennett, E. M.; Aldrich, C. C. A mechanism-based aryl carrier protein/thiolation domain affinity probe. *J. Am. Chem. Soc.* **2007**, *129*, 6350–6351.
- (13) Somu, R. V.; Boshoff, H.; Qiao, C.; Bennett, E. M.; Barry, C. E., 3rd; Aldrich, C. C. Rationally designed nucleoside antibiotics that inhibit siderophore biosynthesis of *Mycobacterium tuberculosis*. *J. Med. Chem.* **2006**, *49*, 31–34.
- (14) Gupte, A.; Boshoff, H. I.; Wilson, D. J.; Neres, J.; Labello, N. P.; Somu, R. V.; Xing, C.; Barry, C. E., 3rd; Aldrich, C. C. Inhibition of siderophore biosynthesis by 2-triazole substituted analogues of 5'-O-[N-(salicyl)sulfamoyl]adenosine: antitubercular nucleosides effective against *Mycobacterium tuberculosis*. *J. Med. Chem.*, in press.
- (15) Vannada, J.; Bennett, E. M.; Wilson, D. J.; Boshoff, H. I.; Barry, C. E., 3rd; Aldrich, C. C. Design, synthesis, and biological evaluation of beta-ketosulfonamide adenylation inhibitors as potential antitubercular agents. *Org. Lett.* **2006**, *8*, 4707–4710.
- (16) Aqvist, J.; Luzhkov, V. B.; Brandsdal, B. O. Ligand binding affinities from MD simulations. *Acc. Chem. Res.* **2002**, *35*, 358–365.
- (17) Brandsdal, B. O.; Osterberg, F.; Almlof, M.; Feierberg, I.; Luzhkov, V. B.; Aqvist, J. Free energy calculations and ligand binding. *Adv. Protein Chem.* **2003**, *66*, 123–158.
- (18) Sham, Y. Y.; Chu, Z. T.; Tao, H.; Warshel, A. Examining methods for calculations of binding free energies: LRA, LIE, PDL-D-LRA, and PDL-D/S-LRA calculations of ligands binding to an HIV-protease. *Proteins* **2000**, *39*, 393–407.
- (19) Su, Y.; Gallicchio, E.; Das, K.; Arnold, E.; Levy, R. M. Linear interaction energy (LIE) models for ligand binding in implicit solvent: theory and application to the binding of NNRTIs to HIV-1 reverse transcriptase. *J. Chem. Theory Comput.* **2007**, *3*, 256–277.
- (20) Zhou, Z.; Bates, M.; Madura, J. D. Structure modeling, ligand binding, and binding affinity calculation (LR-MM-PBSA) of human heparanase for inhibition and drug design. *Proteins* **2006**, *65*, 580–592.
- (21) Ander, M.; Luzhkov, V. B.; Aqvist, J. Ligand binding to the voltage-gated Kv1.5 potassium channel in the open state—docking and computer simulations of a homology model. *Biophys. J.* **2008**, *94*, 820–831.
- (22) May, J. J.; Kessler, N.; Marahiel, M. A.; Stubbs, M. T. Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12120–12125.
- (23) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semi-analytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
- (24) *MacroModel*, version 9.1; Schrödinger LLC.; New York, 2005.
- (25) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.
- (26) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
- (27) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle–Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.
- (28) Labello, N. P.; Ferreira, A. M.; Kurtz, H. A. An augmented effective core potential basis set for the calculation of molecular polarizabilities. *J. Comput. Chem.* **2005**, *26*, 1464–1471.
- (29) Labello, N. P.; Ferreira, A. M.; Kurtz, H. A. Correlated, relativistic, and basis set limit molecular polarizability calculations to evaluate an augmented effective core potential basis set. *Int. J. Quantum Chem.* **2006**, *106*, 3140–3148.
- (30) Khandelwal, A.; Balaz, S. QM/MM linear response method distinguishes ligand affinities for closely related metalloproteins. *Proteins* **2007**, *69*, 326–339.
- (31) Khandelwal, A.; Lukacova, V.; Comez, D.; Kroll, D. M.; Raha, S.; Balaz, S. A combination of docking, QM/MM methods, and MD simulation for binding affinity estimation of metalloprotein ligands. *J. Med. Chem.* **2005**, *48*, 5437–5447.
- (32) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self consistent molecular orbital methods. XII. Further extensions of Gaussian-type basis sets for use in molecular orbital studies of organic molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261.
- (33) The synthesis, characterization, and biological activity of this compound are given in the Supporting Information.

JM800668U