COMPUTATIONAL INSIGHTS INTO SERINE AND DHMA BINDING TO THE TSR CHEMORECEPTOR OF *ESCHERICHIA COLI*

An Undergraduate Research Scholars Thesis

by

ASUKA ORR

Submitted to the Undergraduate Research Scholars program
Texas A&M University
in partial fulfillment of the requirements for the designation as an

UNDERGRADUATE RESEARCH SCHOLAR

Approved by
Research Advisor: Dr. Phanourios Tamamis

May 2016

Major: Chemical Engineering

TABLE OF CONTENTS

ABSTRACT

Computational Insights into Serine and DHMA Binding to the Tsr Chemoreceptor of *Escherichia Coli*

Asuka Orr
Department of Chemical Engineering
Texas A&M University

Research Advisor: Dr. Phanourios Tamamis
Department of Chemical Engineering

3,4-Dihydroxymandelic Acid (DHMA) has been reported to attract *Escherichia coli* by binding to the serine-binding sites of the bacteria's Tsr chemoreceptor. Signaling via the Tsr chemoreceptor may promote the growth and virulence of *E. coli* and other bacteria with serine sensing receptors. In this study, we aimed to elucidate the binding conformation of DHMA in the Tsr chemoreceptor through implicit-solvent molecular dynamics (MD) simulations and free energy calculations. Our findings give insight to how DHMA binds to Tsr and suggests receptor residues that may be key to binding and signaling. We intend to continue this study by docking DHMA in one binding pocket at a time, rather than having both binding pockets of Tsr occupied, and using additional explicit-solvent MD simulations. These results provide impetus for new computational and experimental studies.

CHAPTER I

INTRODUCTION

The digestive tract is home to a large, complex community of microorganisms, most of which are bacteria.[1] While the majority of these bacteria have communalistic or mutualistic relationship with their host, the introduction of pathogenic bacteria can lead to severe detriments to the host's health.[2] The effects of environmental conditions such as temperature and pH on bacterial growth and virulence have been well established.[3,4] However, recent studies indicate that pathogenic bacteria use the hormones released from the host's intestinal walls to navigate to preferred sites of infection and cause disease through chemotaxis.[5,6] The chemotaxis of *E. coli* to one such hormone, norepinephrine, requires norepinephrine to be converted to DHMA before it can be sensed by the chemoreceptor Tsr.[7] The binding mechanism of DHMA to the Tsr chemoreceptor has not been elucidated. Computational methods can be introduced to understand the DHMA:Tsr complex and, in combination with new experiments, shed light onto the key ligand and receptor residues associated with binding and/or signaling. Understanding the binding mechanism of DHMA to Tsr will allow us to proceed in designing molecules that block the Tsr binding sites to block signaling and thus prevent bacteria from navigating to the intestinal walls and causing disease.

CHAPTER II

METHODS

Generating docking poses

Serine (Ser) and DHMA were independently docked into the serine-binding sites of the Tsr chemoreceptor of *Escherichia coli*. For the initial positioning of Ser, each serine-binding pocket of Tsr was occupied by a Ser molecule. For the initial positioning of DHMA, each serine-binding pocket of Tsr was occupied by a DHMA molecule. The initial position and conformation of Ser was in line with the X-ray structure[8]. The initial position and conformation of DHMA was guided by the X-ray structure[8] of Ser in complex with Tsr; the ShaEP[9] algorithm was used to superimpose DHMA with Ser. Four systems were created for both Ser and DHMA independently. In each of the four systems, for each binding pocket independently, the center of mass of the two molecules was simultaneously allowed to deviate 0.5, 1.5, 2.5, and 3.5 Angstroms from the initial center of mass. In each system, twenty "docking" MD simulation runs were performed. Every two picoseconds, a rotation of 120 degrees was attempted on the ligand docked into chain A and the ligand docked into chain B simultaneously about two separate random axes, which were randomized; the final two picosecond structures were saved for further analysis. The rotations aimed at allowing the ligands to explore different poses within the two binding sites. Two hundred steps of two picoseconds were attempted per system; thus, simulations consisting of approximately four hundred picoseconds were performed. In the end, approximately four thousand snapshots of the ligand in the binding site of chain A and four thousand snapshots of the ligand in the binding site of chain B per system were produced. Fifty steps of steepest descent minimization were then performed on the resulting snapshots, and the

interaction energy was evaluated. For each pocket and for each system we extracted three complexes with the lowest interaction energy, which correspond to 24 conformations of Ser:Tsr and 24 conformations of DHMA:Tsr. In addition to the 24 DHMA structures produced, 9 conformations of DHMA in complex with Tsr were produced using AutoDock Vina[10].

Symmetrizing structures

To prepare the 33 structures of DHMA and 24 structures of serine for MD simulations and association free energy calculations, the ligands in each binding site were made symmetrical. The proteins making up the Tsr chemoreceptor consisted of two chains: chain A and chain B. Chain A originally included all residues between Asn39 and Tyr189. Chain B originally contained residues Glu38 to Ser188. Before we were able to symmetrize the proteins, we first had to make the chains have the same residues. In chain A, Tyr189 was removed. In chain B, Glu 38 was removed. This was done to ensure that the proteins were truncated symmetrically and to allow our protein alignment software to superimpose the proteins onto each other. For the structures containing binding poses from the binding pocket of chain A, chain A was superimposed onto chain B to create the ligand in the binding pocket of chain B. For the structures containing binding poses from the binding pocket of chain B, chain B was superimposed onto chain A to create the ligand in the binding pocket of chain A. Superimposing the proteins were achieved using LovoAlign[11]. Subsequently, CHARMM[12] was introduced to introduce symmetry with between the two symmetric pockets. After symmetrizing the structures, MD simulations of the 33 DHMA and 24 Ser dockings were initiated.

Implicit solvent molecular dynamics simulations

During the simulation, the complexes were heated to 75 K, 150 K, 225 K, and finally 300 K. During the heating process, the heavy atoms of the ligand, backbone atoms of the protein, and heavy sidechain atoms of the protein were constrained with a 5 kcal/mol/Å^2, 5 kcal/mol/Å^2, and 3.5 kcal/mol/Å^2 harmonic force respectively. After the system reached 300 K, equilibration runs were initiated. In these equilibration runs, harmonic constraints on the backbone and sidechain atoms were reduced to 3 kcal/mol/Å^2, 2 kcal/mol/Å^2, and 2 kcal/mol/Å^2 in the ligand, 3 kcal/mol/Å^2, 2 kcal/mol/Å^2, and finally 1 kcal/mol/Å^2 in the backbone atoms, and to 1.5 kcal/mol/Å^2, 0.75 kcal/mol/Å^2, and 0.3 kcal/mol/Å^2 in the sidechain atoms. These constraints were then removed in the final stages of the simulation. The CHARMM36 force field[12] was used for the topology and parameters of the complex. DHMA was parametrized using CGENFF[13, 14]. The lengths of bonds to hydrogen atoms were constrained using the SHAKE command. Each stage consisted of 100,000 steps with a duration of 1.5 fs per step making the duration of each equilibration stage 200 ps. The duration of the production runs was equal to 5.25 ns. During the production run, the only harmonic constraint imposed was a light (0.2 kcal/mol/Å$^{2)}$ best-fit constraint on the backbone atoms of the two N- and C- residues per chain (mention which residues). This constraint was introduced to avoid any deformation of the protein due to the lack of the rest of the protein residues, which were also missing in the X-ray structure[8]. *The aqueous environment was taken into account implicitly using GBMV II*[15]. These simulations produced 350 snapshots per binding mode, of which all were used for calculations and analysis.

Association free energy

To identify the most stable conformations of DHMA and Ser in Tsr, the association free energy of the complexes were evaluated. The association free energies of the structures were calculated using the Molecular Mechanics Generalized Born Surface Area (MM GBSA) approximation, in which the association free energy, ΔG, is determined by subtracting the total free energy of the free protein, G_P, and the free ligand, G_L, from the total free energy of the complex, G_{PL}, as shown in the equation below[16, 17, 18] (1):

$$\Delta G = G_{PL} - G_P - G_L \qquad (1)$$

In this calculation-approximation, the entropy contribution from the ligand and protein were assumed to be negligible, as in previous ligand:protein complexes. Entropy contributions from the solvent are taken into account implicitly[19]. The generalized Born with a simple switching (GBSW) model was introduced to calculate the polar and non-polar solvation free energies[20]. The non-polar surface tension coefficient used was 0.03 kcal/(mol Å^2). The total free energy of the free protein, G_P, is taken from the conformation of the bound Tsr protein while excluding the Ser/DHMA molecule. The association free energy for each molecule was calculated and averaged over 350 snapshots.

Interaction free energy

To analyze the binding modes, the interaction free energies between the proteins and each ligand of the structures with the lowest MM GBSA association free energies were calculated using the following equation[16, 17, 18] (2):

$$\Delta G_{RL}^{inte} = \left\{ \sum_{i \in R} \sum_{j \in L} (E_{ij}^{Coul} + E_{ij}^{GB}) \right\}^{(Polar)} + \left\{ \sum_{i \in R} \sum_{j \in L} E_{ij}^{VdW} + \sigma \sum_{i \epsilon R,L} \Delta S_i \right\}^{(Nonpolar)} \quad (2)$$

The first and second components of equation above describe the polar and nonpolar interactions between R and L respectively. For our study, R corresponds to a given residue in the Tsr protein. L corresponded to either the ligand, Serine or DHMA, in the binding pocket of chain A or the ligand in the binding pocket of chain B. These calculations were performed for all 350 snapshots for each molecule and averaged for analysis.

The polar component of the total interaction free energy was determined by setting the charges of all atoms, except for those of the ligand and residue in question, to zero. The non-polar component of the total interaction free energy is determined by the van der Waals interactions between the residue and the ligand and the change in the non-polar solvation free energy due to binding.

CHAPTER III
RESULTS

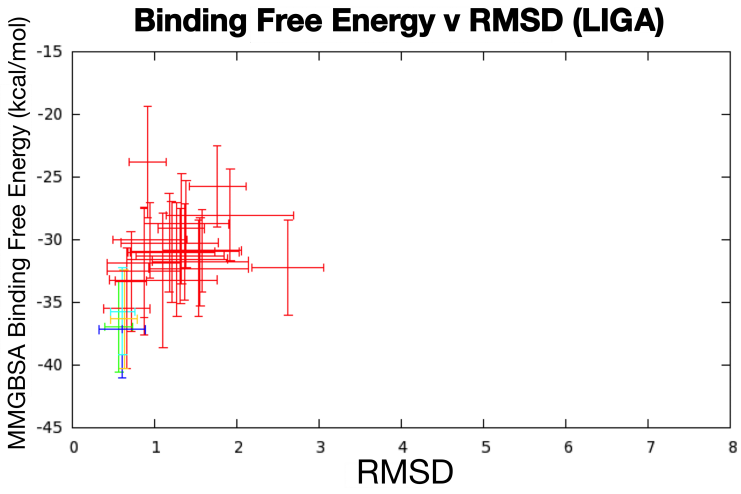
Reproduction of the Serine:Tsr complex

To validate the protocol used to determine the conformation of DHMA in Tsr, the protocol was first introduced to test if it can reproduce serine:Tsr binding with regard to the experimentally derived X-ray structure[8]. Thus, we ran MD simulations on the 24 serine binding modes we generated, calculated their MM GBSA association free energies and compared them to the X-ray structure[8]. The comparison was done by taking the average root-mean-square deviation (RMSD) of serine in our structures across 350 snapshots to serine in the crystal structure[8]. Before taking the RMSDs, the structures were aligned with respect to the crystal structure's binding pocket residues (residues within 9 Å of serine in the X-ray structure[8]).

From our comparison, we found that the binding modes of serine with the lowest MM GBSA association free energies generally had the lowest RMSDs as shown in Figure 1. We also found that the binding modes with the lowest MM GBSA association free energies in one binding pocket had the lowest MM GBSA association free energies in the other binding pocket.

As the association free energy of the binding modes increased, the RMSD of the binding mode also increased. Thus, the biding modes with the lowest association free energies had conformations that were closer to the conformation of serine in the crystal structure[8]. By selecting the binding mode with the lowest MM GBSA association free energy, we were able to isolate a structure with an RMSD of 0.612 Å from the crystal structure[8].

(a)



(b)

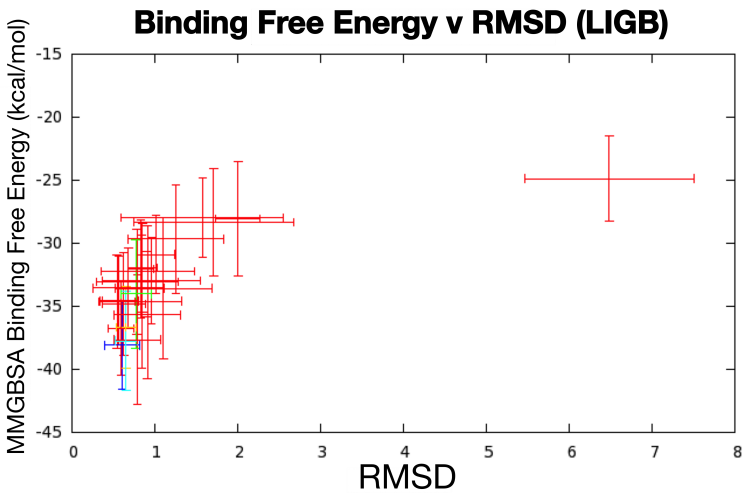


Figure 1: The calculated association free energy (called binding free energy in the figures) with respect to the RMSD of docked Ser to the crystal structure of Ser using Ser binding to the binding pocket of (a: upper panel) chain A and (b: bottom panel) chain B. The colored points show that the best binding modes in one binding pocket are also the best binding modes in the other binding pocket. In both chain A and chain B, the binding modes with the lowest MMGBSA association free energy had the lowest RMSD to the X-ray crystal structure[8].

In addition to having a conformation nearly identical to that of the crystal structure[8], the binding mode with the lowest binding free energy also exhibited strong interactions to the residues identified as key residues for ligand specificity and sensing in Tsr[8]. To fairly compare the

interactions, the X-ray structure[8] of the serine:Tsr complex underwent the same implicit-solvent MD simulations. With both the binding mode and the crystal structure[8] simulated, the interaction free energies were calculated. In both structures, the serine molecules interacted with the same Tsr residues with nearly the same intensity. This shows that, in addition to predicting the same binding conformation of serine in Tsr, we are able to predict which residues are critical to serine binding.

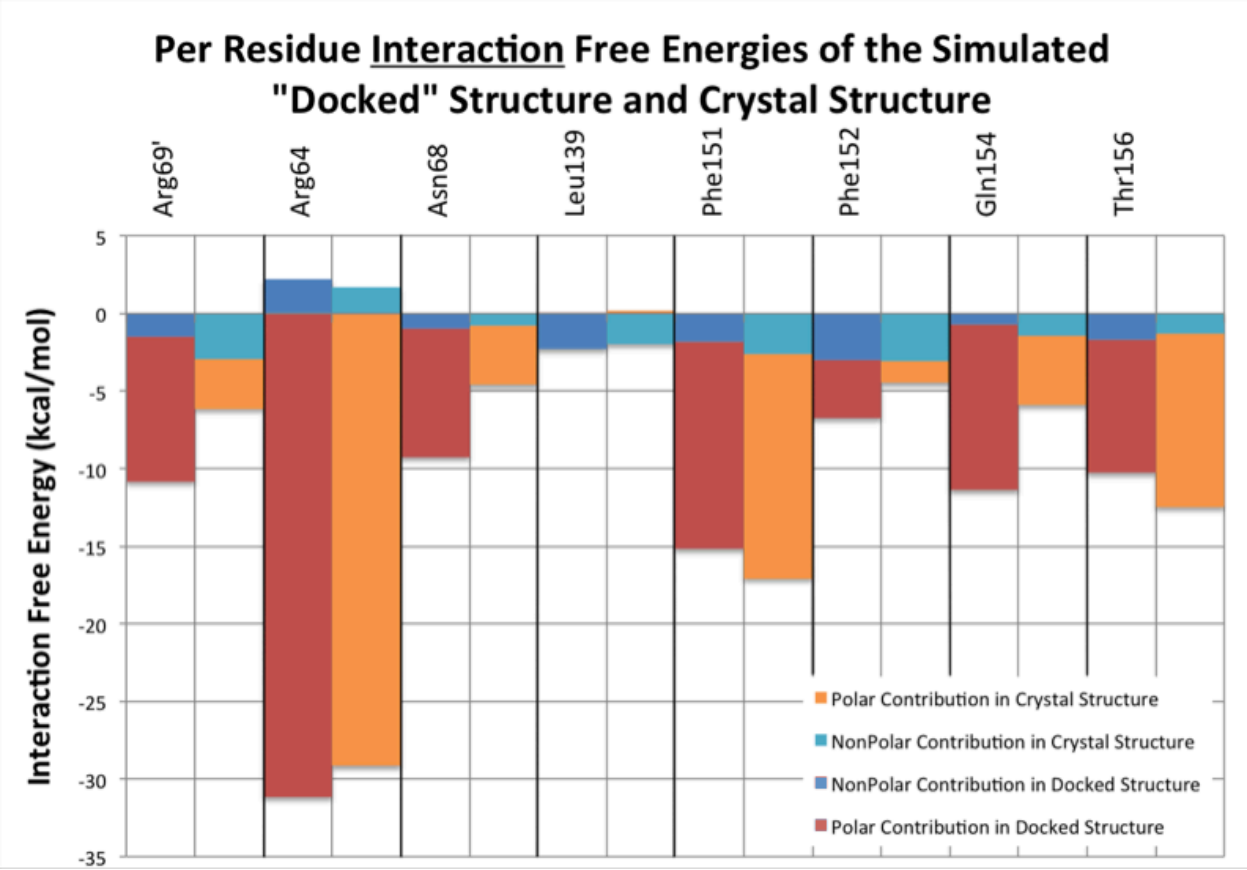


Figure 2: The per residue interaction free energy contribution from critical residues in Tsr of the crystal structure[8] and the binding mode with the lowest MMGBSA binding free energy. Both structures interact strongly with the same critical residues.

With both the direct correlation between the association free energy and RMSD to the X-ray structure[8], and the similarity between the interactions of the binding mode with the lowest association free energy and the interactions of the crystal structure[8], our protocol was validated. We were able to reproduce the serine:Tsr complex fairly accurately. Figure 2 shows the similarities in interaction free energies between the simulated crystal structure[8] and the binding mode with the lowest association free energy.

Now that the protocol was proven to work with serine, we applied the same methodology to predict the binding conformation of DHMA in Tsr.

Binding modes of DHMA in complex with Tsr

The protocol was repeated on the 33 dockings of DHMA. The resulting implicit-solvent MD simulations and association free energy calculations indicated that the most probable structures of the DHMA:Tsr complex were in the binding pocket of chain B. These binding modes are 05b2 and 25b1. However, with the following results, the most probable conformation cannot be clearly identified without further investigation.

In binding mode 05b2, DHMA forms two salt bridges between the ligand's negatively charged carboxyl group and the positively charged residues Arg64 and Arg 69' (from the minority half of Tsr). Cation-π interactions between the Arg64 and Arg69' and the aromatic ring of DHMA are also formed. Additionally, DHMA forms hydrogen bonds to residues Leu67, Asn68, Phe151, Gln154, Thr156, and Gln157. The aromatic ring of DHMA forms π-π interactions with Phe151 and Phe152. The aromatic ring of DHMA also forms van der Waals interactions with Ile72',

Leu67, Asn68, Gly71, Leu139 and Pro155. These interactions are shown in Figure 3.

(a) (b)

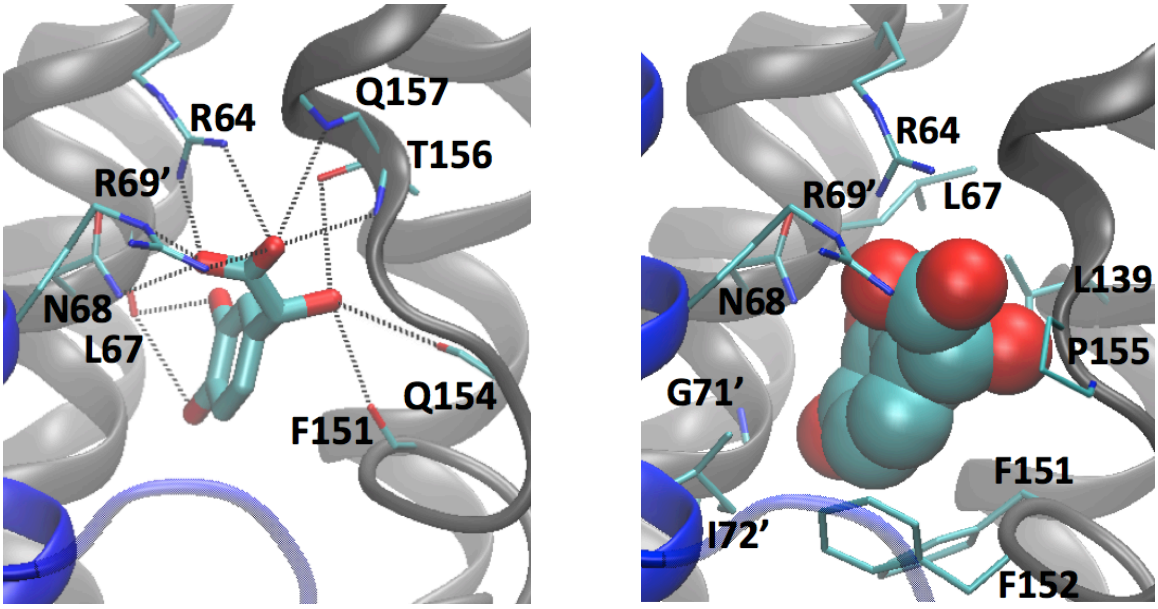


Figure 3: The key (a) polar and (b) non-polar interactions in binding mode 05b2 formed between DHMA and the residues of Tsr. Residues marked with ‘ are residues from the minority half of Tsr.

Similar to DHMA of binding mode 05b2, DHMA in binding mode 25b1 also forms two salt bridges between the negatively charged carboxyl group of DHMA and the positively charged residues Arg64 and Arg 69’. Unlike in the previously mentioned binding mode, DHMA in mode 25b1 exhibits a cation-π interaction with Arg73’ as well as with Arg69’ and Arg64. DHMA also forms hydrogen bonds with Ser85’, Phe151, Gln154, Thr156 and Gln157. While in the previously mentioned binding mode, 05b2, DHMA forms a hydrogen bond with Asn68, in binding mode 25b1 DHMA is not hydrogen bonded to Asn68. The hydrogen bond between DHMA and Ser85’ in binding mode 25b1 is not seen in 05b2. The aromatic ring of DHMA also participates in π-π interactions with Phe151 and Phe152. Van der Waals interaction between DHMA and residues Pro155, Thr156, Gln157, and Leu90’ of Tsr are also observed. The Van der

Waals interactions observed in binding mode 05b2 and 25b1 are different largely due to the different orientation of the aromatic ring of DHMA in the two binding modes. These interactions are shown in Figure 4.

(a) (b)

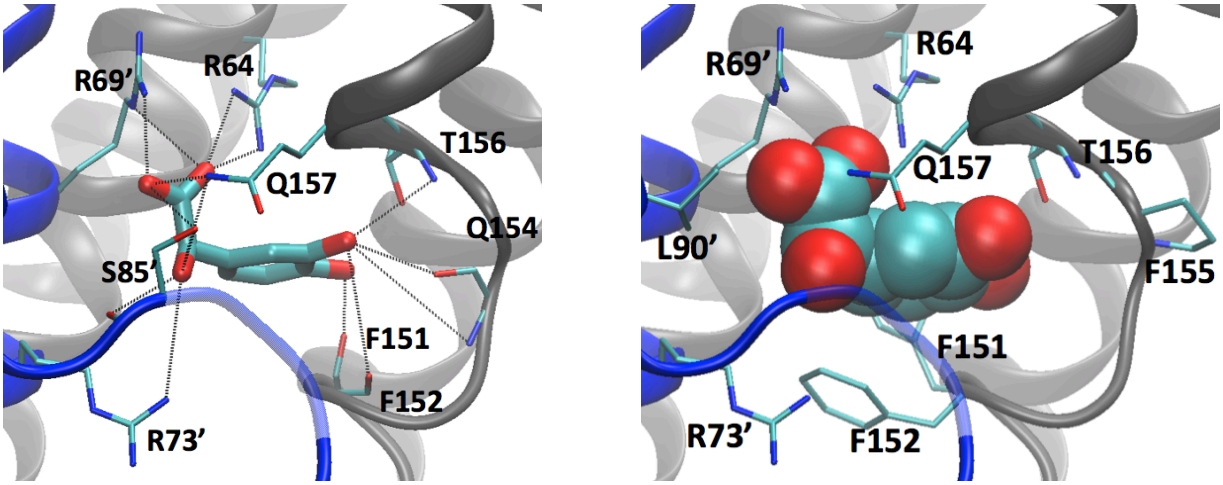


Figure 4: The key (a) polar and (b) non-polar interactions in binding mode 25b1 formed between DHMA and the residues of Tsr. Residues marked with ' are residues from the minority half of Tsr.

Interaction free energy analysis

To gain insights into the similarities and differences between the two selected binding modes, and to identify possible critical residues involved in DHMA binding to Tsr, we calculated the per residue interaction free energies between Tsr and DHMA for each binding mode. Figure 5 and Figure 6 show the average intermolecular interaction free energies between DHMA in binding pocket B to the residues of the Tsr protein decomposed into polar and non-polar components for modes 05b2 and 25b1, respectively.

Residues Arg64, Asn68, Thr156, and Arg69' were shown to be critical to serine recognition in

experimental experiments[8]. Furthermore, Arg69' and Thr156 were also determined to be critical residues to DHMA binding[7]. DHMA, in both modes 05b2 and 25b1, interacts strongly with Thr156, Arg64, and Arg69'. In mode 05b2, DHMA also interacts with Asn68.

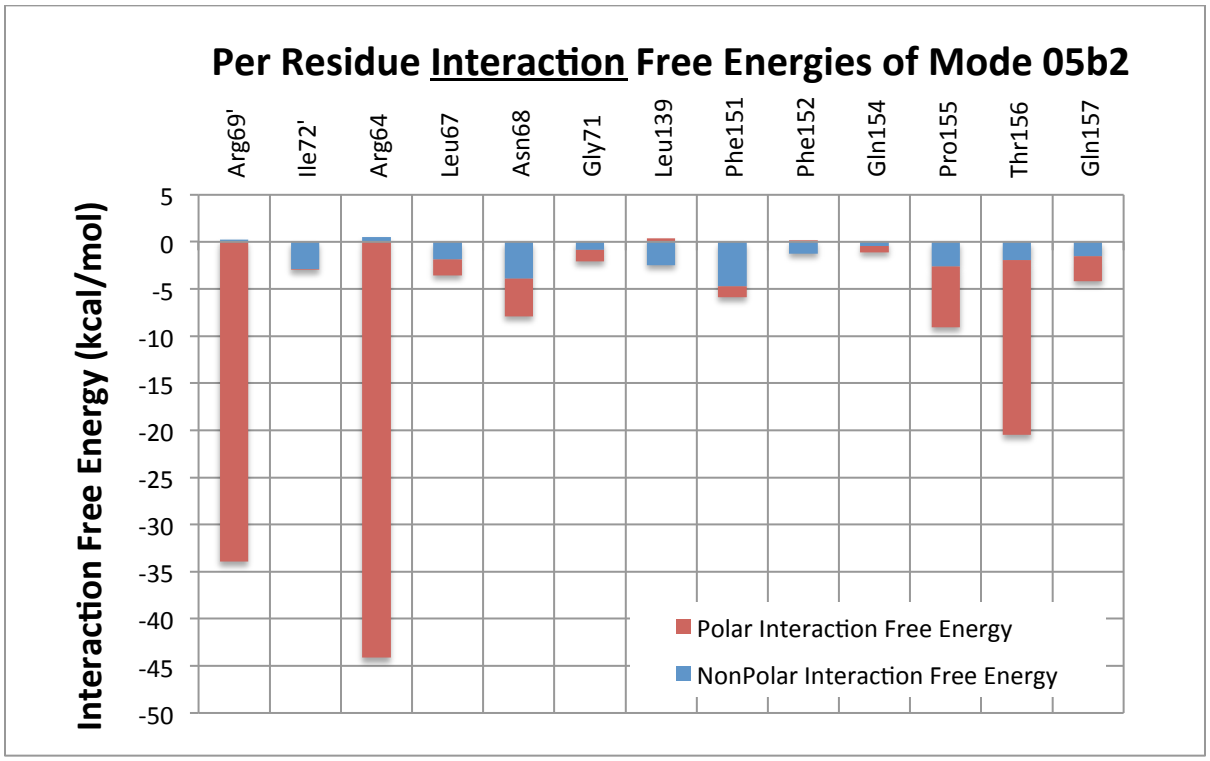


Figure 5: The per residue interaction free energy contributions from critical residues in Tsr of binding 05b2.

In mode 25b1, DHMA does not strongly interact with Asn68. In mode 25b1, DHMA interacts with Arg73', Ser85', Phe152, Gln154, and Gln157 to a greater extent than it does in mode 05b2. The stronger interactions are primarily due to the different orientation of the aromatic ring.

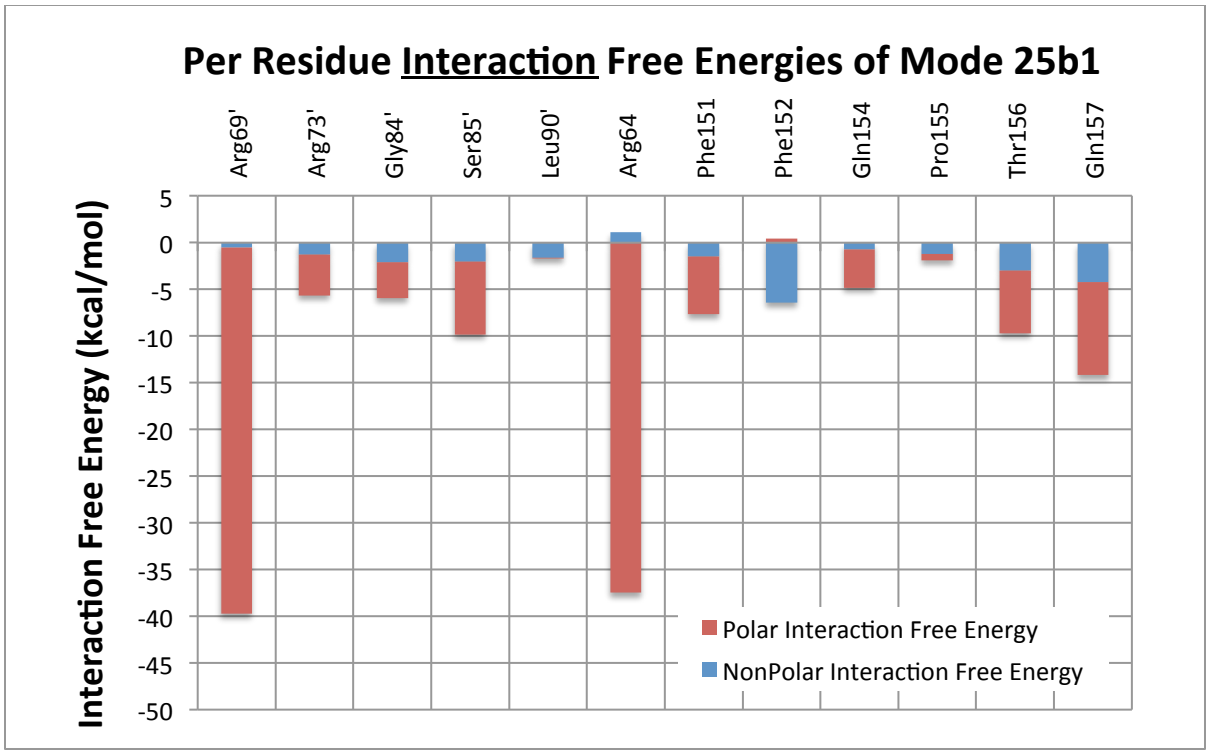


Figure 6: The per residue interaction free energy contributions from critical residues in Tsr of binding 25b1.

While in mode 25b1, DHMA exhibits strong interactions to some of the key residues, Arg64, Asn68, Thr156, and Arg69', only in mode 05b2, DHMA interact strongly with all of the aforementioned residues.

However, data from experiment suggest that Tsr has negative cooperativity, that is, the binding of DHMA in one binding pocket prevents another DHMA ligand from binding in the other pocket[7]. Additional simulations using explicit solvation with DHMA occupying one binding pocket rather than both binding pockets are needed to determine the naturally occurring binding mode. Further experiments are also needed to fully understand DHMA binding to Tsr. Additional, explicit-solvent simulations are currently being run to more accurately model the system. In these simulations, DHMA is docked in only one binding pocket of Tsr to account for

the negative cooperativity of Tsr. With these additional docking studies, we aim to produce results with higher accuracy.

CHAPTER IV
CONCLUSION

Bacterial chemotaxis to molecules in the intestines are linked to virulence of pathogens and infection. In our study, we examined possible binding conformations of DHMA in complex with the Tsr serine chemoreceptor of *E. coli* to determine how DHMA induces signaling in Tsr. Through MD simulations using implicit solvation and free energy calculations, we found two modes that likely represent the binding conformation of DHMA in complex with Tsr. The proposed modes are 05b2 and 25b1. We propose that additional explicit-solvent MD simulations using one ligand, rather than two, and experiments are needed to determine the binding modes that are found in nature. The results from this study allow us to understand the mechanism of DHMA binding to Tsr and pave the road for the development of methods to prevent bacterial infection in the intestines. We intend to (i) identify the DHMA:Tsr binding mode that is naturally occurring through explicit-solvent MD simulations in which the binding of one molecule, rather than two, will be investigated, and (ii) a possible future direction could be the computational design and/or discovery of molecules that inhibit DHMA binding to Tsr.

REFERENCES

[1] Simon GL, Gorbach SL. 1984. Intestinal flora in health and disease. Gastrenterology. 86:174-193.

[2] Guarner F, Malagelada JR. 2003. Gut flora in health and disease. Lancet. 361:512-519.

[3] Presser KA, Ross T, Ratkowsky DA. 1998. Modeling the Growth Limits (Growth/No Growth Interface) of Escherichia coli as a Function of Temperature, pH, Lactic Acid Concentration, and Water Activity. Appl. Environ. Microbiol. 64:1773-1779.

[4] Antoniou P, Hamilton J, Koopman B, Jain R, Holloway B, Lyberatos G, Svoronos SV. 1990. Effect of temperature and ph on the effective maximum specific growth rate of nitrifying bacteria. Elsevier. 24:97-101.

[5] Bansal T, Englert D, Lee J, Hegde M, Wood TK, Jayaraman A. 2007. Differential effects of epinephrine, norepinephrine, and indole on Escherichia coli O157:H7 chemotaxis, colonization, and gene expression. Infect. Immun. 75:4597-4607.

[6] Karavolos MH, Winzer K, Williams P, Khan CMA. 2013. Pathogen espionage: multiple bacterial adrenergic sensors eavesdrop on host communication systems. Mol. Microbiol. 87:455-465.

[7] Pasupuleti S, Sule N, Cohn WB, MacKenzie DS, Jayaraman A, Manson MD. 2014. Chemotaxis of Escherichia coli to Norepinephrine (NE) Requires Conversion of NE to 3, 4-Dihydroxymandelic Acid. J. Bacteriol. 196:3992-4000.

[8] Hirotaka Tajima et al. Dec 2011. Ligand Specificity Determined by Differentially Arranged Common Ligand-binding Residues in Bacterial Amino Acid Chemoreceptros Tsr and Tar. Journal of Biological Chemistry. 286: 42200-42210.

[9] M. J. Vainio, J. S. Puranen, M. S. Johnson. 2009. ShaEP: molecular overlay based on shape and electrostatic potential. Journal of Chemical Information and Modeling. 49(2): 492-502.

[10] Trott O1, Olson AJ. Jan 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 31(2): 455-61.

[11] L. Martinez, R. Andreani, J. M. Martinez. 2007. Convergent algorithms for protein structural alignment. BMC Bioinformatics. 8(306).

[12] B. R. Brooks et al. Jul 2009. CHARMM: The Biomolecular simulation Program. Journal of Computational Chemistry. 30: 1545-1615.

[13] K. Vanommeslaeghe et al. Mar 2010. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. Journal of Computational Chemistry. 31: 671-690.

[14] W. Yu et al. Dec 2012. Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. Journal of Compututational Chemistry. 33: 2451-2468.

[15] A. Suenaga et al. Mar 2009. Molecular Dynamics Simulations Reveal that Tyr-317 Phosphorylation Reduces Shc Binding Affinity for Phosphotyrosyl Residyes of Epidermal Growth Factor Receptor. Biophysical Journal. 96: 2278-2288.

[16] P. Tamamis, C. Floudas. Sep 2013. Molecular recognition of CXCR4 by a dual tropic HIV-1 gp120 V3 loop. Biophysical Journal. 105(6),:1502-1514.

[17] P. Tamamis et al. Aug 2014. Insights into the mechanism of C5aR inhibition by PMX53 via implicit solvent molecular dynamics simulations and docking. BMC Biophysics, 7:5.

[18] P. Tamamis et al. Sep 2010. Species specificity of the complement inhibitor compstatin investigated by all-atom molecular dynamics simulations. Proteins: Structure, Function, and Bioinformatics. 78: 2655-2667.

[19] J. Srinivasan et al. Sep 1998 Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices. Journal of the American Chemical Society. 120: 9401- 9409.

[20] W. Im, M.S. Lee, C.L. Brooks. 2003. 3rd Generalized Born model with a simple smoothing function. Journal of Computational Chemistry. 24: 1691–1702.