UNIVERSITY OF CALIFORNIA, SAN DIEGO

Computational Studies in Diffusion and Drug Discovery

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy

in

Chemistry

by

Patricia M. Bauler

Committee in charge:

Professor J. Andrew McCammon, Chair
Professor Michael Holst
Professor Katja Lindenburg
Professor Francesco Paesani
Professor Susan Taylor

2012

The Dissertation of Patricia M. Bauler is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2012

DEDICATION

I must first thank my family. You have all been so supportive and proud, taking joy in my accomplishments even when I did not. It was enough to keep me going through the rough patches and help me finish what I started. I want thank my sister for listening to me gripe and bellyache and for putting up with my nonsense. I want to thank my parents for helping me to keep skating. It truly saved my sanity and gave me purpose when I felt I had none. They have helped me achieve so many goals I thought I could not, and for that I am eternally grateful.

I would also like to thank the many friends who have helped me along the way. Alix, for deciding to make me her "new project," and the entire Warhammer group for including me in their shenanigans. Jennifer and Krystal for caring, despite time and distance keeping us apart. Alpha, Paul, RJ, and Brian for reminding me about the good times. I want to thank Meagan for being able to talk to me in times of need. I want to thank Maggie for being my grad school buddy. I wish that you had been able to stay, and I admire you for being able to leave. I wish you all possible happiness in your life. And I want to thank Bryan Woo. So many conversations on the road helped me remember what is good in life and why we keep struggling. I wish you the best of luck in achieving all of the goals you set before yourself.

I want to thank all of the teachers, professors, and students who have given me skills and knowledge that I will use the rest of my life.

I want to thank Robert Jordan and Brandon Sanderson. Although distant influences, they have been strong and inspirational influences on my life none the less. They gave me friends when I needed them, and escape when everything else felt impossible. Though they may never know, I will forever be grateful for their work.

Finally, I want to thank Bob Leonard for being a consistently positive influence in my life. For standing by me in good times and bad, and teaching me life

lessons I could not have learned elsewhere. For understanding my goals and giving me the technique, grace, knowledge, and love for the sport necessary to try and reach them. For being understanding about my circumstances and dealing with them as best as possible. For displaying dignity, diplomacy, and honor in all situations. I can only hope to emulate these qualities in all aspects of my life.

EPIGRAPH

You can never know everything, and part of what you know is always wrong. Perhaps even the most important part. A portion of wisdom lies in knowing that. A portion of courage lies in going on anyway.

-a'Lan Mandragoran, Winter's Heart, Chapter 32: A Portion of Wisdom

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Professor J. Andrew McCammon for his support as the chair of my committee and as my dissertation advisor. His advice, guidance, occasional prodding, and generous help have been incredibly valuable and have helped me accomplish many goals throughout my graduate school career.

I would like to acknowledge the many McCammon group members who have given me help, support, and encouragement along the way. Special acknowledgements must be given to Gary Huber without whom none of the diffusion work would have been accomplished, Bill Sinko for helping me learn so much about drug discovery procedures and protocols, Levi Pierce for helping with molecular dynamics simulations and computer issues. I also want to thank Pat Blachly for talking to me about other projects along the way, Joe Zhou for the weekly encouragements, Sara Nichols for help with the docking software, Nuo Wang for the fun discussions after class, and Elaine Thompson who shared tidbits of wisdom that saved me countless hours. Also an especially important acknowledgement to Patti Craft, without whom I know we would all be completely lost.

I would like to acknowledge the other members of my committee: Professor Katja Lindenburg, Professor Francesco Paesani, Professor Susan Taylor, and Professor Michael Holst for offering valid criticisms and excellent advice about my work and progress throughout the graduate school process.

I would like to acknowledge our collaborator Professor Thomas Leyh of The Albert Einstein College of Medicine for his help with the work on substrate channeling.

I would like to acknowledge our collaborator Professor Eric Oldfield and his research group from University of Illinois, Urbana-Champaign for his help with the work on dihydropteroate synthase.

Chapter 1, in full, is a reprint of the material as it appears in The Journal of Physical Chemistry Letters 2010. Bauler, Patricia; Huber, Gary A.; Leyh, Thomas; McCammon, J. Andrew, American Chemical Society 2010. The dissertation author was the primary investigator and author of this paper.

Chapter 2, in full, is a reprint of the material as it appears in The Journal of Chemical Physics 2012. Bauler, Patricia; Huber, Gary A.; McCammon, J. Andrew, American Institute of Physics 2012. The dissertation author was the primary investigator and author of this paper.

VITA

2008	Bachelor of Science, University of California, Irvine
2008-2009	Teaching Assistant, Department of Chemistry and Biochemistry, University of California, San Diego
2010	Master of Science, University of California, San Diego
2009-2012	Research Assistant, University of California, San Diego
2012	Doctor of Philosophy, University of California, San Diego

PUBLICATIONS

Bauler, P.; Huber, G. A.; Leyh, T.; McCammon, J. A. Channeling by Proximity: The Catalytic Advantages of Active Site Colocalization Using Brownian Dynamics. *J. Phys. Chem. Let.*, **2010**, *1*, 1332-1335.

Bauler, P.; Huber, G. A.; McCammon, J. A. Hybrid finite element and Brownian dynamics method for diffusion-controlled reactions. *J. Chem. Phys.*, **2012**, *136*, 164107-1 - 164107-7.

FIELDS OF STUDY

Major Field: Chemistry (Physical)

Studies in Computational Chemistry
Professor J. Andrew McCammon

SCHOLARSHIPS AND AWARDS

2008	Hypercube Scholar Award, University of California, Irvine
2010-2011	Chancellor's Interdisciplinary Collaboratories Fellowship, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Computational Studies in Diffusion and Drug Discovery

by

Patricia M. Bauler

Doctor of Philosophy in Chemistry

University of California, San Diego 2012

Professor J. Andrew McCammon, Chair

Computational methods are becoming increasingly important for studying complex biochemical systems. As computers have become faster and more powerful, it is possible to perform calculations of larger and more complex systems than has been possible in the past. As these systems increase in complexity, it becomes increasingly important to develop methods that efficiently utilize the available computational power, and use new tools and methods to search for meaningful results. The work presented in this dissertation uses computational methods for studies in two major areas of computer aided chemistry research - diffusion and drug discovery.

Diffusion is often a rate determining step in many biochemical processes, therefore being able to study multi-step diffusion reactions is important for learning about how diffusion regulates reactions in complex biological systems. The diffusion work presented in the first chapter of this dissertation utilizes simplified spherical

models to perform diffusion studies of a two-step reaction model. The results give insight into how the relative locations of reaction targets are important in determining the success of a multi-step diffusion reaction. The second chapter of this dissertation proposes a new hybrid method for diffusion studies that capitalizes on the benefits of two popular calculation methods, Brownian dynamics simulations and finite element methods. The hybrid method utilizes the speed of finite element method calculations for a majority of the diffusion system, but allows for atomistic detail of the target and diffusing particles in the region of interest near the site of the reaction. Basic one-dimensional test cases are presented to demonstrate that the method is accurate and viable for future development. Though the systems presented in both diffusion studies are basic test cases, the results provide a useful basis for future studies of more complicated biological systems.

Another major area of computational chemistry research is that of computer aided drug discovery. The third chapter of this dissertation uses molecular dynamics simulations and molecular docking tools to search for potential drug compounds to inhibit the bacterial enzyme dihydropteroate synthase. Visualization tools are utilized to identify structures of interest, and several hundred compounds are identified for further testing.

INTRODUCTION

Computers have become increasingly important tools in scientific research. As they develop in speed and capacity, they are being used to study increasingly complex physical phenomena. As the size and scale of these systems increase, it becomes important to develop more efficient methods of calculation, and to utilize newly developed tools to analyze results. Computers can be used to study many types of physical phenomena, but two important areas of computational chemistry research deal with calculating the motion of particles in diffusion studies and in searching for new drug compounds.

One important area of computational research is in the study of diffusion. Diffusion is the spontaneous movement of particles from an area of higher concentration to an area of lower concentration. Although the overall motion of the diffusing particles will predictably be moving from higher to lower concentration, the movement of each individual particle can be described as a random walk and does not have a prescribed direction. Diffusion is often a rate-determining step in biological pathways, and thus learning how various environments affect the diffusion process can give insight into the timescales of these biological reaction pathways. Two of the most popular computational methods for studying diffusion are Brownian dynamics simulations and finite element method calculations. Brownian dynamics simulations use stochastic equations to follow the random walk of individual diffusing particles. They can be useful for studying the pathway that an individual particle takes, especially when in an unusual environment. Because the Brownian dynamics simulation only follows one possible path of a diffusing particle, many simulations must be run to obtain statistically significant results. This can be considerably time consuming, especially as a system gets more complex. Alternatively, the finite element method

can be used to calculate the overall motion of the diffusion of a collection of particles. This method works by breaking up a region into smaller sections called simplicies, with calculations occurring at each simplex, and being combined to find the solution over the entire region of diffusion. This is a relatively fast computation, but it does not allow for the visualization of the pathway of individual particles. In order to learn more about diffusion in biological pathways it is important to study the behavior of particles in more complex systems. Additionally, as the systems being studied continue to increase in complexity, it will become increasingly important to develop computational methods that can efficiently calculate the motions of the diffusing particles.

Chapter 1 presents a Brownian dynamics study that examines the efficiency of substrate transfer in a multi-step reaction. It expands upon early work[15,16,17,18] in the field of computational Brownian dynamics studies by using simple spherical models to represent diffusing substrates and the enzymes with which they will react. The distances and angles of orientation between the active site zones of two enzyme spheres are changed in order to observe how strongly the effects of relative distances influence the rates of reaction for a diffusing substrate. In this work it is shown that smaller distances between the active site zones produce a more efficient transfer of a substrate intermediate between the two enzyme spheres, but that if the enzyme spheres are too close, then the efficiency of the initial diffusion reaction is impeded.

Chapter 2 presents the development of a new method for computational studies of diffusion. It creates a hybrid method that combines two of the most popular methods for diffusion calculations the finite element method and Brownian dynamics simulations. The hybrid method utilizes the relative speed of the finite element method to calculate diffusion over most of the system space. It uses a Brownian dynamics description of the system near the target of interest, thereby gaining the

ability to use structures of the target and diffusing particles in more detail. This new method is used to study diffusion in a simple one-dimensional model for both linear and radially symmetric systems. The results from the test systems indicate that this method produces accurate results and should be further developed for use with more complex systems.

Another major use of computers in chemistry research is the area of drug discovery. Searching for new drug compounds is a lengthy and expensive process, but using computational methods as a preliminary step can reduce the amount of testing that needs to be done in the laboratory. As the number of determined protein crystal structures grows, it becomes easier to search for compounds that will interact with very specific drug targets. With the protein crystal structure, molecular dynamics can be used to study the motions of the protein and important structures can be identified. Molecular docking software can then be used to test libraries of small molecules with each of the isolated structures in order to find leads for possible drug compounds. The libraries can contain hundreds of thousands of compounds, but the use of the computer allows for a much more efficient and cost-effective screening than would be possible if each compound had to be tested manually. Laboratory testing of the highest ranked compounds then offers insight into which compounds should be pursued for further development.

Chapter 3 presents a computer aided drug discovery study of dihydropteroate synthase. Dihydropteroate synthase (DHPS) is a key enzyme in the folate production pathway in many infectious bacteria. It has two active sites, one of which has previously been the successful target of anti-bacterial sulfa drugs, but mutations have made these drugs increasingly ineffectual. The second active site is highly conserved among DHPS enzymes from many different bacterial species,[5,6] thus making it an intriguing target for future drug design. This study uses computational methods to

search for potential drug compounds to be used as inhibitors in this second, highly conserved active site. The chapter demonstrates how molecular dynamics simulations can be used to generate representative structures of the drug target and how these structures can be used in computer aided docking studies. It also uses recently developed visualization tools, such as POVME (the POcket Volume MEasurer)[7] to identify potentially significant structures for inclusion in the docking studies. The results yielded several hundred compounds for testing in a laboratory environment.

0.1 References

[1] Ermak, D.L.; McCammon, J.A. Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys.* **1978**, *69*, 1352-1360.

[2] Northrup, S.H.; Allison, S.A.; McCammon, J.A. Brownian dynamics simulation of diffusion-influenced bimolecular reactions. *J. Chem. Phys.* **1984**, *80*, 1517-1524.

[3] Zhou, H.-X. On the calculation of diffusive reaction rates using Brownian dynamics simulations. *J. Chem. Phys.*, **1990**, *92*, 3092-3095.

[4] Luty, B.A.; McCammon, J.A.; Zhou, H.-X. Diffusive reaction rates from Brownian dynamics simulations: Replacing the outer cutoff surface by an analytical treatment. *J. Chem. Phys.*, **1992** *97*, 5682-5686.

[5] Hevener, K.E.; Yun, K.; Qi, J.; Kerr, I.D.; Babaoglu, K.; Hurdle, J.G.; Balakrishna, K.; White, S.W.; Lee, R.E. Structural Studies of Pterin-Based Inhibitors of Dihydropteroate Synthase. *J. Med. Chem.* **2010**, *53*, 166-177.

[6] Vinnicombe, H.G.; Derrick, J.P. Dihydropteroate Synthase from *Streptococcus pneumoniae*: Characterization of Substrate Binding Order and Sulfonamide Inhibition. *Biochem. Biophys. Research Comm.* **1999**, *258*, 752-757.

[7] Durrant, J.D.; de Oliveira, C.A.; McCammon, J.A. POVME: An algorithm for measuring binding-pocket volumes. *J. Mol. Graph. Model.* **2011**, *29*, 773-776.

CHAPTER 1

Channeling by Proximity - The Catalytic Advantages of Active Site Colocalization Using Brownian Dynamics

1.1 Abstract

Nature often co-localizes successive steps in a metabolic pathway. Such organization is predicted to increase the effective concentration of pathway intermediates near their recipient active sites and to enhance catalytic efficiency. Here, the pathway of a two-step reaction is modeled using a simple spherical approximation for the enzymes and substrate particles. Brownian dynamics are used to simulate the trajectory of a substrate particle as it diffuses between the active site zones of two different enzyme spheres. The results approximate distances for the most effective reaction pathways, indicating that the most effective reaction pathway is one in which the active sites are closely aligned. However, when the active sites are too close, the ability of the substrate to react with the first enzyme was hindered, suggesting that even the most efficient orientations can be improved for a system that is allowed to rotate or change orientation to optimize the likelihood of reaction at both sites.

1.2 Introduction

Nature frequently co-localizes linked catalytic functions - this is seen in single polypeptides that catalyze multiple consecutive steps in a metabolic pathway,[1,2,3] and in large assemblies of non-covalently associated biomolecules that carry out complex cellular processes with remarkable fidelity.[4,5,6,7] A consequence of such or-

ganization is that pathway intermediates that enter solution are precisely positioned near the active center waiting to receive them. Thus, for a period of time, the intermediate and its recipient binding site remain in close proximity, which increases the effective concentrations of these interactors and leads to enhanced binding and catalytic efficiency. Intermediates that do not enter bulk solution, but remain bound to the biomolecule, are channeled between binding sites via tunnels and electrostatic grooves whose conformational states are responsive to the positioning of the ligand.[8,9,10,11,27] It is conceivable that intermediates released from properly positioned active sites can be transferred between the sites with efficiencies that approach those of channeling systems. While the effects of high local concentrations are much anticipated,[13,14] quantitative estimates of the magnitudes of these effects are lacking.

1.3 Methods

Here, Brownian dynamics simulations are used to study particle trajectories using short time solutions of the Smoluchowski equation. The trajectories are used to obtain particle collision probabilities, and to predict the likelihood of reaction.[15,16,17,18] The BrownDye[17] software was used to simulate trajectories and collect collision probabilities. The system has a temperature of 298K and the solvent has the viscosity of water. The enzyme particles were modeled as two separate spheres with either 4Å or 8Å radii, each with a spherical active site zone of 5Å radius centered at a point on the surface of the sphere (Figure 1.1). The enzyme spheres were given a +1 charge located either in the center of the sphere, or at the center of the active zone. The large spheres were held at a constant distance during each simulation, and the distance between the reactive zone centers was varied between 5Å and 50Å in 5Å intervals over several simulations. Each simulation consisted of 10,000 trajectories, which provided enough successful reactions to draw conclusions while keeping the

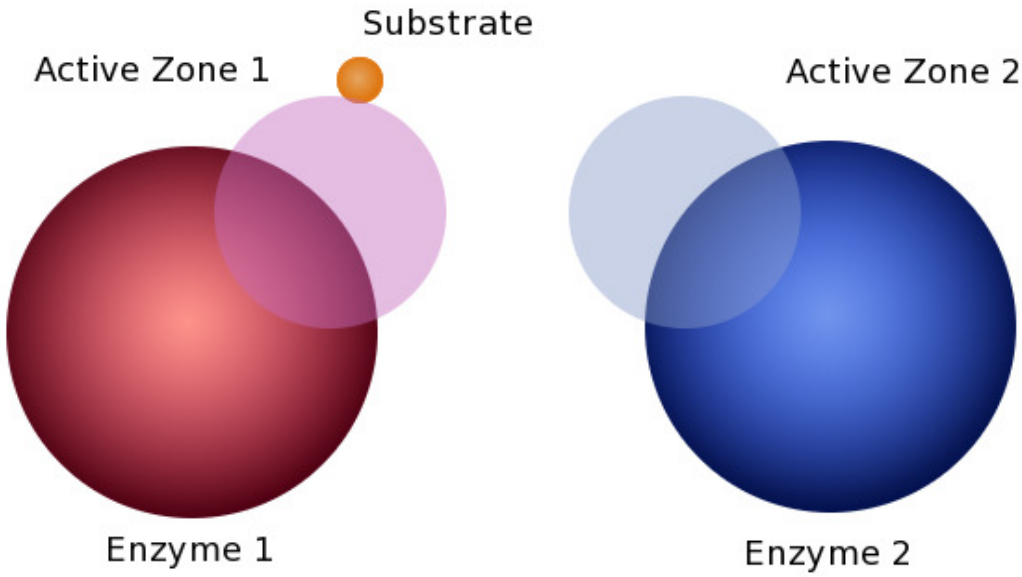


Figure 1.1: Schematic of experimental set-up. The substrate sphere (orange) must diffuse to the active zone of the first enzyme (pink) and then to the active zone of the second enzyme (blue). Shown are the enzyme spheres with the 8 radius and the active zones in the 45 degrees orientation. The center of the substrate sphere must encounter the active zone in order for a reaction to occur.

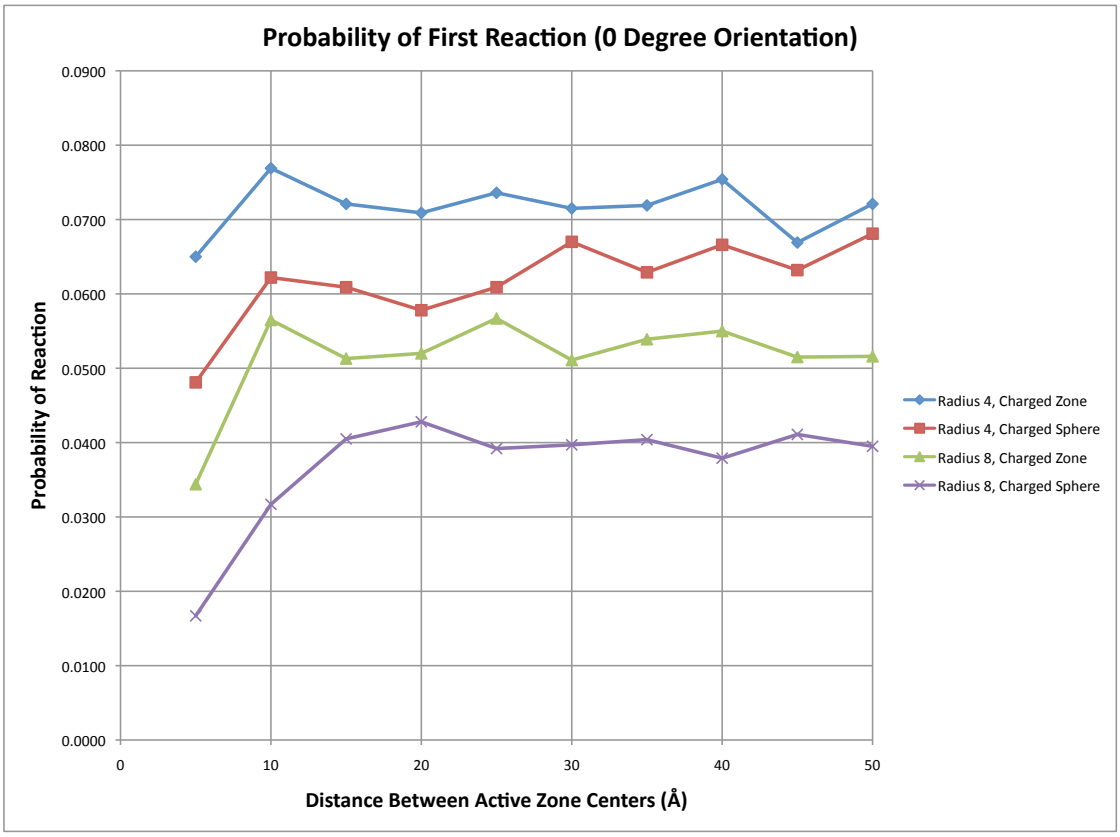
calculation time manageable. The relative orientation of the reactive zones was also varied. The original starting position (0 degree orientation) consisted of the reactive zones directly facing each other. The zones were rotated in opposite directions by 45, 90, 135, and 180 degrees (Figure 1.3(b), inset). The substrate particle was modeled as a sphere with a 1Å radius and a -1 charge. The sizes of the enzyme and substrate spheres were chosen to give as simple a system as possible, while having size ratios of 1:4 and 1:8. The substrate trajectory was modeled with BrownDye, where the translational and rotational diffusivities of the spheres follows Stokes Law. The substrate particle started from an orientation in which its center was separated by 50Å from the center of the first enzyme spheres active zone. A collision was recorded

whenever the substrate particle center contacted the reactive zone. It was assumed that each collision leads to a reaction, as the most efficient enzymes will react with every substrate they encounter. In order for a reaction to be considered complete, the substrate had to first diffuse to and react with the first enzyme sphere, then diffuse from that position to react with the second enzyme sphere. The number of collisions for the first and second interactions was recorded separately, so the reaction probability for each interaction can be calculated independently, as well as the total probability for the overall reaction pathway.

1.4 Results

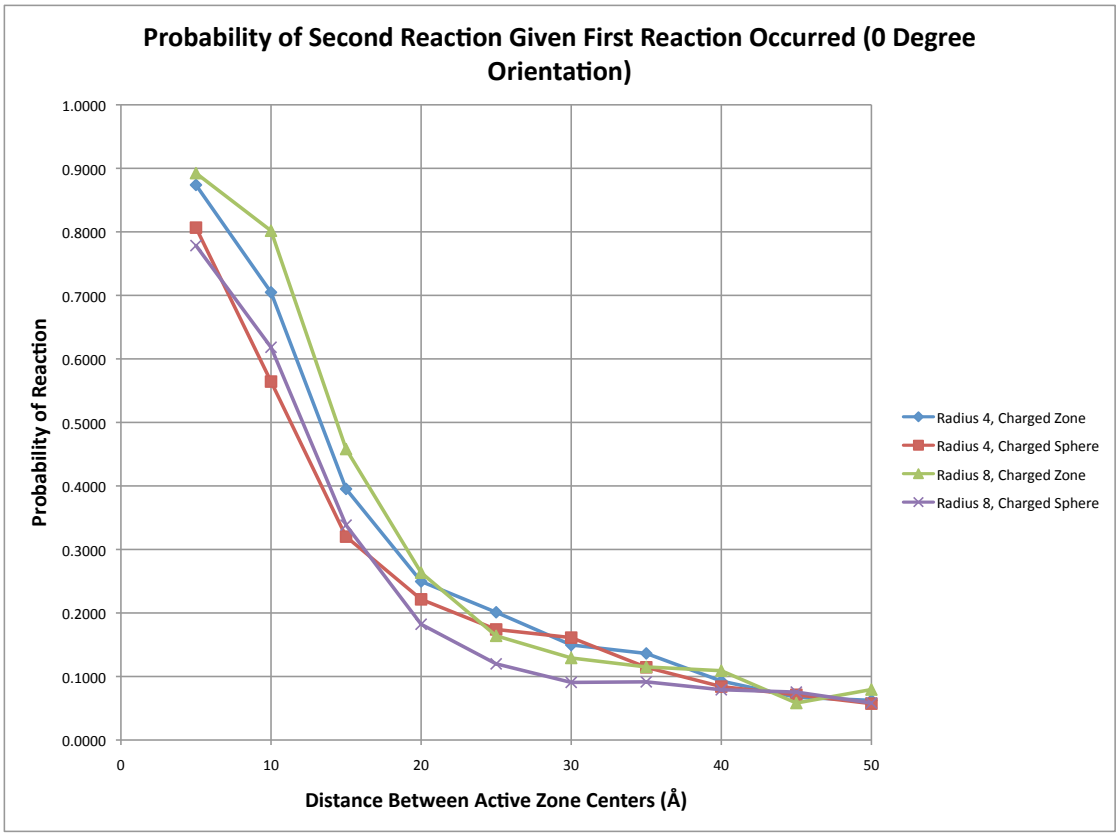
After running all of the simulations, the reaction probabilities were compared. The reaction of the substrate with the first reactive zone should not depend strongly on the location of the active zone or the distance between the zones, and so similar reaction probabilities would be expected (Figure 1.2(a), Figure 1.3(a)). The only difference is the localization of the charge in the active zone or at the center of the enzyme sphere. Thus the average from the four sets of simulations was calculated; the reaction with the first zone has an average probability of reaction of 0.0713 ± 0.0063 for 4Å spheres with charged active zones, 0.0632 ± 0.0076 for 4Å spheres where the charge is centered in the enzyme, 0.0536 ± 0.0076 for 8Å spheres with charged active zones, and 0.0403 ± 0.0090 for 8Å spheres where the charge is centered in the enzyme. The results indicate that spheres with less buried active zones are more likely to have initial reactions. In addition, the spheres with the charged active zones are slightly more likely to have an initial reaction than those with the charge centered in the large sphere, which can be explained by the charge acting to guide the substrate sphere to the specific location of reaction, rather than just generally toward any point on the large sphere. Interestingly, when the reactive zones are in the 0 degree

orientation (facing each other) and the large spheres are at a 5Å distance from each other, the initial probability of reaction for the 4Å spheres is 0.0650 for the charged active zones and 0.0481 for the charged enzyme spheres and for the 8Å spheres is 0.0344 for the charged active zones and 0.0167 for the charged enzyme spheres (Figure 1.2(a)). These are all significantly lower (more than one standard deviation) than the average, though the effect is more pronounced in the 8Å spheres. This may be because the enzyme spheres shield each other to some extent. Similarly, when the active zones are in the 45 degrees orientation and 5Å distance, the probabilities of the first reactions are slightly lower than the average probability for the first reaction. All of the other active zone orientations (90, 135, and 180 degrees) did not show this lowered probability for the closer spheres. At the larger distances (10Å - 50Å), the reaction probabilities are about the same for all active zone orientations.



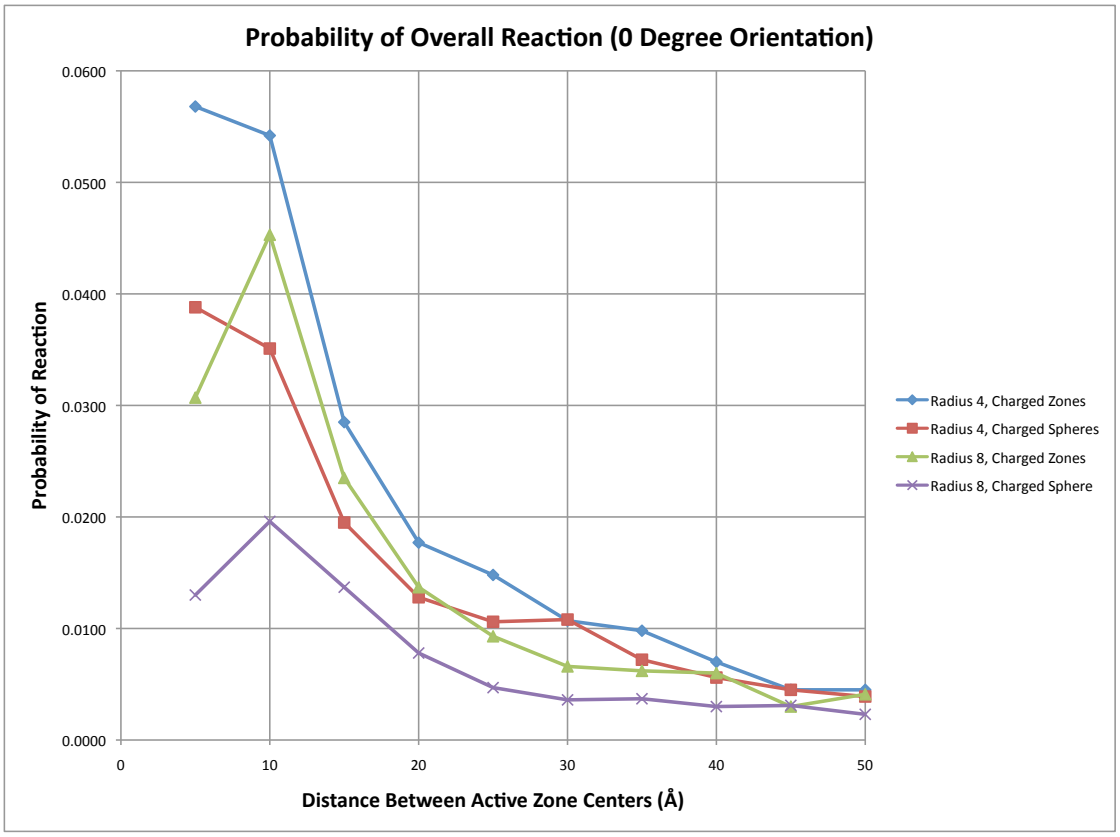
(a)

Figure 1.2: Probabilities of reactions at changing distances when active zones are in 0 degree orientation. (a) Probability of the first reaction occurring. The probability is fairly constant except when the active zone centers are only 5Å apart. The close proximity of the enzyme spheres may hinder the substrate ability to encounter the active site. (b) Probability of second reaction, given that first reaction occurred. Close proximity of the active zones leads to effective reaction pathway, with the charged active zones being more efficient until about 25Å separation. (c) Probability of second (overall) reaction. The effect of shielding the first active zone can clearly be seen. In general, the closer active zones lead to a more effective pathway, with the charged active zones being significantly more efficient at the closer distances.



(b)

Figure 1.2: Continued



(c)

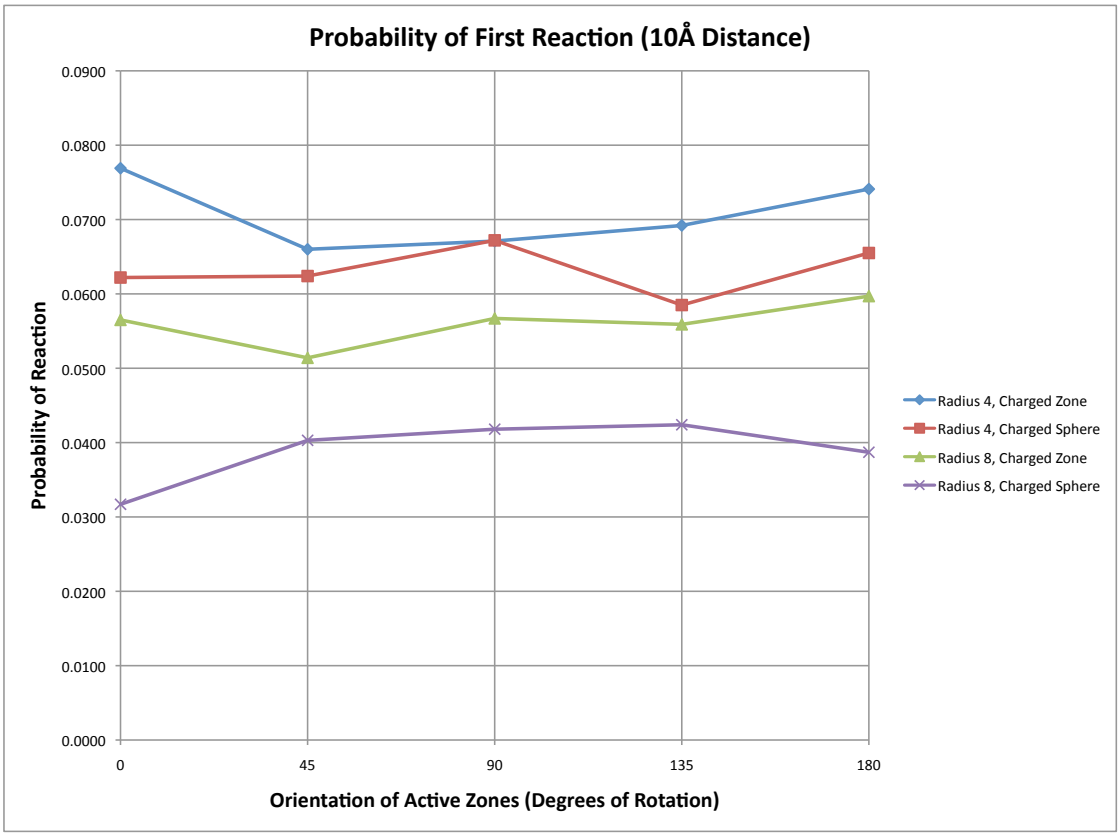
Figure 1.2: Continued

The probability of the second reaction depends on the orientation of the active zones and the distance between the zones, as well as the location of the enzymes charge. The probability can either be examined as the probability of the second reaction given that the first reaction occurred (Figure 1.2(b), Figure 1.3(b)), which demonstrates the effects of the active zone orientation and distances in the system, or as the overall probability of the second reaction (Figure 1.2(c), Figure 1.3(c)), which is the probability that the product from the complete reaction pathway will be produced. As expected, the largest number of successful reactions occurs when the two active zones are facing each other in the 0 degrees orientation at 5Å distance

with the charges localized on the active zones. The efficiency of these configurations gives such a large probability to the second reaction occurring if the first has already taken place, that the overall probability of the completed reaction is much larger than for other orientations and farther zone distances. The probability of reaction completion decreases as the active zones are rotated away from each other and as the distance between the zones increases. The charged zones have a higher reaction probability until the separation between the active zones reaches 30Å, after which the reaction probabilities are much more similar. The charged active zones also have higher reaction probabilities when comparing the orientations of the zones, the probabilities being similar when the zones are in the 180 degrees orientation (facing in opposite directions). One interesting result is that the 45 degrees orientation at 5Å for the charged active zones has a larger probability (0.0453) for the overall second reaction than does the 0 degree 5Å orientation when the charge is centered in the enzyme sphere (probability 0.0388).

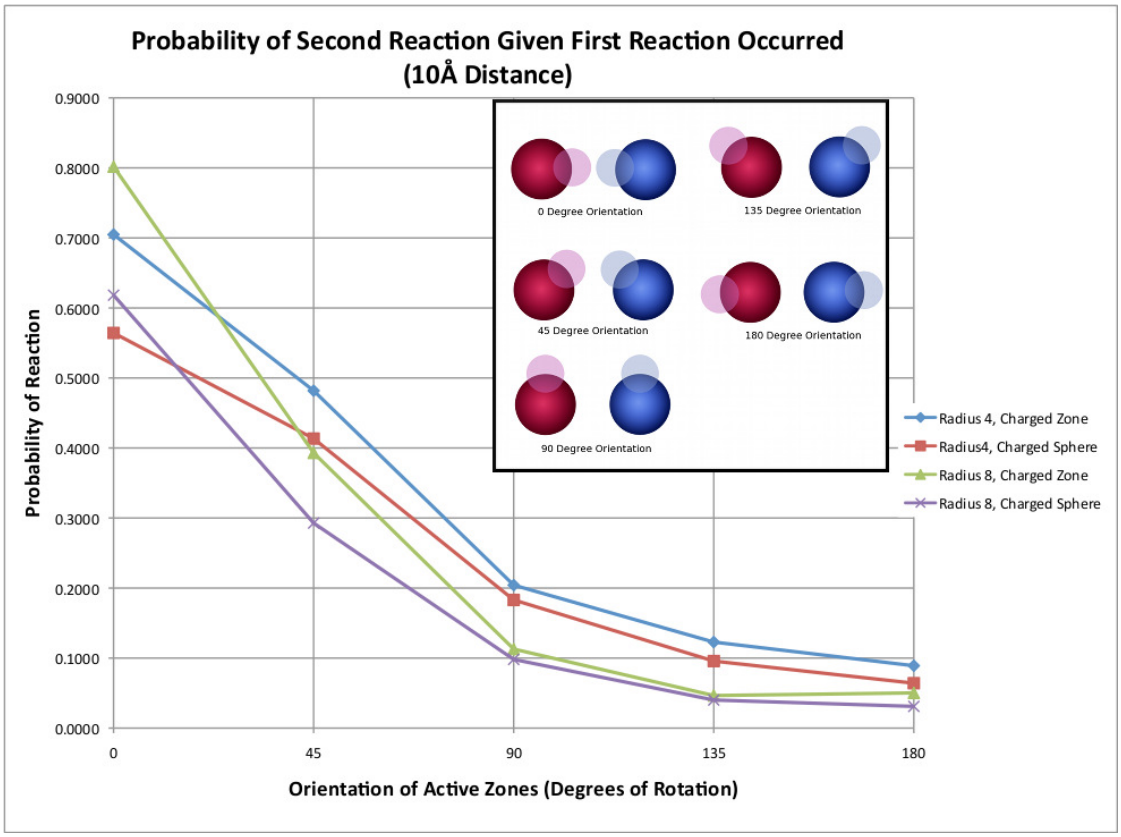
The probability of the second reaction depends on the orientation of the active zones and the distance between the zones, as well as the location of the enzymes charge. The probability can either be examined as the probability of the second reaction given that the first reaction occurred (Figure 1.2(b), Figure 1.3(b)), which demonstrates the efficiency of the active zone orientation and distances in the system, or as the overall probability of the second reaction (Figure 1.2(c), Figure 1.3(c)), which is the probability that the product from the complete reaction pathway will be produced. As expected, for the 4Å spheres the largest number of successful reactions occurs when the two active zones are facing each other in the 0 degrees orientation at 5Å distance with the charges localized on the active zones. However, for the 8Å spheres, the effect of shielding the active site of the first reaction at the 5Å distance can clearly be seen to hinder the success of the overall reaction (Figure

1.2(c)). Although this hinders the success of the overall reaction, it is still the most efficient configuration for passing the substrate sphere from the first to the second active zone, perhaps because this shielding effect prevents the substrate sphere from escaping once it has interacted with the active sites (Figure 1.2(b)). In general, the probability of reaction completion decreases as the active zones are rotated away from each other and as the distance between the zones increases. The charged zones have a higher reaction probability at the 0 degree orientation until the separation between the active zones reaches 25Å, after which the size of the enzyme sphere seems to be a more important factor in determining reaction success (Figure 1.2(b), Figure 1.2(c)). For the other active zone orientations, the size of the enzyme sphere seems to determine the probability of reaction more strongly than does the position of the charge (Figure 1.3(b), Figure 1.3(c)).



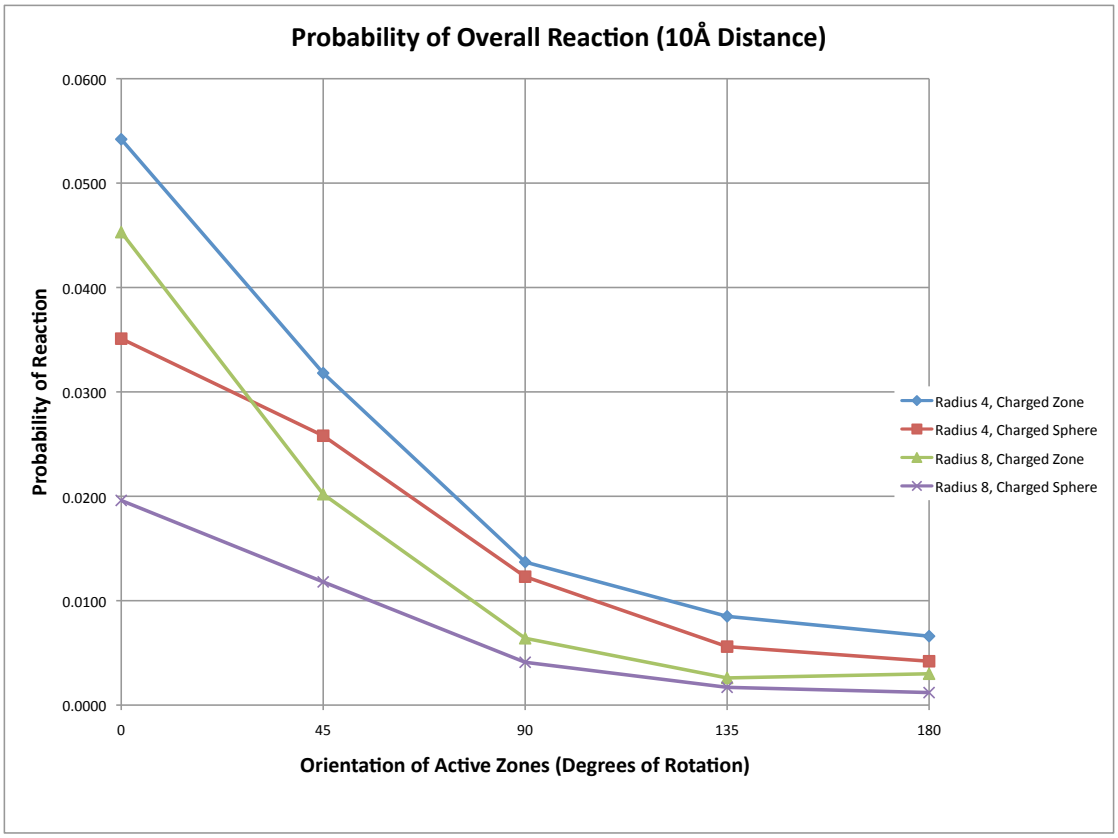
(a)

Figure 1.3: Probabilities of reactions at rotated active zone orientations when active zones are at 10Å distance. (a) Probability of the first reaction occurring. The probability is fairly constant as expected. (b) Probability of second reaction, given that first reaction occurred. As the active zones rotate away from each other the reaction probability decreases considerably, and the effect of enzyme sphere size becomes more important. Inset shows active zone orientations. (c) Probability of second (overall) reaction. The active zones that face each other demonstrate a more effective pathway. Again, other than the 0 degree orientation, enzyme size is more important than location of the charge.



(b)

Figure 1.3: Continued



(c)

Figure 1.3: Continued

1.5 Discussion

Clearly, the orientation of the active zones and the distance between the zones as well as the size of the active zone relative to the enzyme sphere are important in determining the success of a reaction. For enzymes of the same size, the overall reaction probability is greater when the charge is localized on the active zone, and this effect is more significant when the two active zones are closer together. Once the zones have been rotated away from the original 0 degree orientation or separated by a distance of greater than 25Å, the charge localization has much less impact

than the size of the enzyme spheres. However, when the active zones are too close to each other the enzyme spheres can hinder initial access of the substrate to the active zones. Although this provides a more efficient substrate transfer, it makes the overall reaction less effective. This behavior could potentially argue for a mechanism in which the enzymes either move or rotate in order to control the efficiency of both initial substrate uptake and substrate transfer. These and other more realistic models will be the subject of future studies.

1.6 Acknowledgements

Work at UCSD was supported in part by NIH, NSF, HHMI, CTBP, and NBCR. Work at The Albert Einstein College of Medicine was supported by NIH grant R01GM054469.

Chapter 1, in full, is a reprint of the material as it appears in The Journal of Physical Chemistry Letters 2010. Bauler, Patricia; Huber, Gary A.; Leyh, Thomas; McCammon, J. Andrew, American Chemical Society 2010. The dissertation author was the primary investigator and author of this paper.

1.7 References

[1] Charles, I.G.; Keyte, J.W.; Brammar, W.J; Smith, M.; Hawkins, A.R. The isolation and nucleotide sequence of the complex AROM locus of Aspergillus nidulans. *Nucleic Acids Res.* **1986**, *14*, 2201-2213.

[2] Dunn, M.F.; Niks, D.; Ngo, H.; Barends, T.R.; Schlichting, I. Tryptophan synthase: the workings of a channeling nanomachine. *Trends Biochem. Sci.* **2008**, *33*, 254-264.

[3] Sun, M.; Andreassi, J.L.; Liu, S.; Pinto, R. Triccas, J.A.; Leyh, T.S. The trifunctional sulfate-activating complex (SAC) of Mycobacterium tuberculosis. *J. Biol. Chem.* **2005**, *280*, 7861-7866.

[4] Schmeing, T.M.; Ramakrishnan, V. What recent ribosome structures have revealed about the mechanism of translation. *Nature* **2009**, *461*, 1234-1242.

[5] Leibundgut, M.; Maier, T.; Jenni, S.; Ban, N. The multienzyme architecture of eukaryotic fatty acid synthases. *Curr. Opin. Struct. Biol.* **2008**, *18*, 714-725.

[6] Jørgensen, K.; Rasmussen, A.V.; Morant M.; Nielsen, A.H.; Bjarnholt, N.; Zagrobelny, M.; Bak, S.; Møller, B.L. Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. *Curr. Opin. Plant Biol.* **2005**, *8*, 280-291.

[7] Shaw-Reid, C.A.; Kelleher, N.L.; Losey, H.C; Gehring, A.M.; Berg, C.; Walsh, C. T. Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of E. coli EntF catalyzes both elongation and cyclolactonization. *Chem. Biol.* **1999**, *6*, 385-400.

[8] Miles, E.W.; Rhee, S.; Davies, D.R. The molecular basis of substrate channeling. *J. Biol. Chem.* **1999**, *274*, 1219312196.

[9] Elcock, A.H.; Potter, M.J.; Matthews, D.A.; Knighton, D.R.; McCammon, J.A. Electrostatic Channeling in the Bifunctional Enzyme Dihydrofolate ReductaseThymidylate Synthase. *J. Mol. Biol.* **1996**, *262*, 370374.

[10] Elcock, A.H.; McCammon, J.A. Evidence for Electrostatic Channeling in a Fusion Protein of Malate Dehydrogenase and Citrate Synthase. *Biochemistry* **1996**, *35*, 1265212658.

[11] Elcock, A.H.; Huber, G.A.; McCammon, J.A. Electrostatic Channeling of Substrates between Enzyme Active Sites: Comparison of Simulation and Experiment. *Biochemistry* **1997**, *36*, 1604916058.

[12] Cheng, Y.; Chang, C.A.; Yu, Z.; Zhang, Y.; Sun, M.; Leyh, T.S.; Holst, M.J; McCammon, J.A. Diffusional Channeling in the Sulfate Activating Complex: Combined Continuum Modeling and Coarse-grained Brownian Dynamics Studies. *Biophys. J.* **2008**, *95*, 4659-4667.

[13] Conrado, R. J.; Varner, J. D.; DeLisa, M.P. Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. *Curr. Opin. Biotechnol.* **2008**, *19*, 492-499.

[14] Erban, R.; Chapman, S. J. Stochastic modeling of reaction-diffusion processes: algorithms for bimolecular reactions. *Phys. Biol.* textbf2009, *6*, 46001.

[15] Ermak, D.L.; McCammon, J.A. Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys.* **1978**, *69*, 1352-1360.

[16] Northrup, S.H.; Allison, S.A.; McCammon, J.A. Brownian dynamics simulation of diffusion-influenced bimolecular reactions. *J. Chem. Phys.* **1984**, *80*, 1517-1524.

[17] Zhou, H.-X. On the calculation of diffusive reaction rates using Brownian dynamics simulations. *J. Chem. Phys.*, **1990**, *92*, 3092-3095.

[18] Luty, B.A.; McCammon, J.A.; Zhou, H.-X. Diffusive reaction rates from Brownian dynamics simulations: Replacing the outer cutoff surface by an analytical treatment. *J. Chem. Phys.*, **1992** *97*, 5682-5686.

[19] Huber, G.A.; McCammon, J.A. Browndye: A software package for Brownian dynamics. *Comp. Phys. Comm.* **2010**, *181*, 1896-1905.

CHAPTER 2

Hybrid finite element and Brownian dynamics method for diffusion-controlled reactions

2.1 Abstract

Diffusion is often the rate determining step in many biological processes. Currently, the two main computational methods for studying diffusion are stochastic methods, such as Brownian dynamics, and continuum methods, such as the finite element method. This paper proposes a new hybrid diffusion method that couples the strengths of each of these two methods. The method is derived for a general multidimensional system, and is presented using a basic test case for 1D linear and radially symmetric diffusion systems.

2.2 Introduction

Diffusion is an important factor is many biological processes, including protein-substrate reactions, protein-protein interactions,[1,2,3] calcium channeling in cardiac myocytes,[4,5], the neuromuscular junction,[6,7,8,9] and the myelin sheath gap.[10] Currently, there are two main methods used to study diffusion: stochastic methods, such as Brownian dynamics,[11,12,13,14,15] and continuum methods, such as the finite element method.[16] Each of these methods has its strengths and weaknesses, and can be more or less beneficial in different simulation situations.

The use of Brownian dynamics in computer simulations has been quite prevalent, with many available computational packages.[17,18,19,20,21,22] This method has the benefit of providing trajectories of the diffusing particle as it interacts with a target,

and this method can be especially useful for looking at protein-protein interactions. However, since the usefulness of this method is based on statistics, many simulations must be run in order to have meaningful results. However, these calculations can take a significant amount of computational resources, especially when hundreds of thousands of simulations are required for sufficient statistical sampling. Some methods use multiple diffusing particles,[21,22,23] and some consider only one moving particle at a time.[17,18,19,20]

Whereas Brownian dynamics focuses on the movement of particular particles, the continuum methods can be used to study overall diffusion patterns exhibited by a collection of particles. One of the most popular continuum methods for biological systems is the finite element method.[7,8,10,24,25,26,27] In this method, the region over which diffusion occurs is broken up into smaller simplicies, and then the diffusion equation is solved over each simplex. Combining the results from the simplicies will result in the overall solution. The calculations for this type of model are relatively fast, once the initial set up of the system is complete. In addition, the finite element method works well for arbitrary geometries and the solutions are usually fairly accurate. This method has the benefit of simulating the ensemble behavior of many diffusing particles, instead of focusing on the motions of just a single particle. On the other hand, the trajectories of individual particles cannot be observed, which could be useful for understanding how a diffusing particle will find and interact with a target. The other major drawback of the finite element method is the need to create a grid or mesh of the diffusion domain by breaking it up into simplicies. While this process can be fairly easy for a simple domain, it is not a trivial task when dealing with even a single protein at atomic resolution, because of the large number of small elements needed to resolve the complex shape of the molecular surface, and this difficulty would be compounded with larger systems such as subcellular

compartments.

The goal of this paper is to present a new hybrid diffusion model that combines the best of both the Brownian dynamics and finite element methods. This model places a target particle at the center of the model, surrounded by a Brownian dynamics model with multiple diffusing particles. The Brownian dynamics region is itself surrounded by a finite element region. Although only the most basic of cases are presented here, this model could potentially be useful in combing the best of both methods, by allowing atomistic detail of the target and the nearby diffusing particles, while allowing the speed of the finite element method to handle the calculations for a majority of the region of diffusion. This will also eliminate the problems of only being able to study one diffusing particle at a time and the need to generate a quality mesh of the protein target.

This work continues the work of Gorba, Geyer, and Helms,[22,28,29] who have devised an interface between the continuum diffusion equation and Brownian dynamics for both steady-state and non-steady-state systems. Previous work also includes that of Im, Seefeld, and Roux,[30] who have coupled Brownian dynamics with a particle reservoir described by a grand-canonical ensemble. Like the previous works, this work presents systems that are mathematically 1D in the continuum domain, but the algorithm formulation in terms of finite elements can be directly applied to higher dimensional systems. For formulating the finite element equations, we draw upon finite element solutions of non-steady-state diffusion equations in biophysical settings.[7,8,25]

2.3 Theoretical Background

The goal of this method is to solve the diffusion equation

$$\frac{\partial c}{\partial t} = -\nabla \cdot (\frac{1}{k_B T}\mathbf{D} \cdot \mathbf{F}c) + \nabla \cdot (\mathbf{D} \cdot \nabla c), \tag{2.1}$$

where c is the concentration, T is temperature, $\mathbf{D}$ is diffusion matrix, and $\mathbf{F}$ is force on particles as function of position.

This equation can be described as a cloud of particles moving according to the stochastic differential equation given by

$$d\mathbf{x} = \frac{dt}{k_B T}\mathbf{D} \cdot \mathbf{F} + \sqrt{2dt\mathbf{D}} \cdot \mathbf{w} + \nabla \cdot \mathbf{D}dt, \tag{2.2}$$

where $\mathbf{w}$ is a vector of independent Gaussian random variables with a zero mean and unit variance.

The solution to Eq. 2.1 can also be described by dividing the spatial domain into simplicies and defining hat functions on each node between the simplicies. A hat function has a value of 1 at the node and 0 at all other nodes, and varies linearly across the simplex. Assuming that the force and the diffusivity vary only with position and not time, and denoting the hat function centered at node i as u_i, the diffusion equation can be multiplied by u_i and with the use of Green's theorem, the *weak formulation* is obtained:

$$\begin{aligned}\int_V \frac{\partial c}{\partial t} u_i d\mathbf{x} = \int_V \mathbf{F} \cdot \mathbf{D} \cdot \nabla u_i \cdot \nabla c d\mathbf{x} - \int_V \nabla u_i \cdot \mathbf{D} \cdot \nabla c d\mathbf{x} \\ - \oint_S \mathbf{F} \cdot \mathbf{D} \cdot \mathbf{n} u_i c dS + \oint_S \nabla c \cdot \mathbf{D} \cdot \mathbf{n} u_i dS.\end{aligned} \tag{2.3}$$

Next, it is assumed that concentration can be represented using the hat functions u_i:

$$c(\mathbf{x}, t) = \sum_i c_i(t) u_i(\mathbf{x}) \tag{2.4}$$

Eq. 2.3 can then be written as a set of ordinary differential equations with coefficients c_i:

$$\begin{aligned}\sum_j \frac{dc_j}{dt} \int u_j u_i d\mathbf{x} = \sum_j c_j \int_V \mathbf{F}\cdot\mathbf{D}\cdot\nabla u_i u_j d\mathbf{x} - \sum_j c_j \int_V \nabla u_i \cdot \mathbf{D}\cdot\nabla u_j d\mathbf{x} \\ - \sum_j c_j \oint_S \mathbf{F}\cdot\mathbf{D}\cdot\mathbf{n} u_i u_j d\mathbf{x} + \sum_j c_j \oint_S \nabla u_j \cdot \mathbf{D}\cdot\mathbf{n} u_i d\mathbf{x}.\end{aligned} \tag{2.5}$$

The integrals involving the pairs of hat functions can be computed before the simulation. The results can be expressed analytically if the simplicies are small enough so that the diffusivity and force can be assumed to be linear across each simplex, since this involves integration of polynomials over simplicies. Dirichlet, Neumann, and Robin boundary conditions can be expressed as linear equations in terms of the coefficients c_i. Equation 2.5 in conjunction with the boundary conditions given above can be expressed in matrix form as

$$\mathbf{A}\cdot\frac{d\mathbf{c}}{dt} = \mathbf{B}\cdot\mathbf{c} - \mathbf{b}, \tag{2.6}$$

where $\mathbf{b}$ is a constant vector that depends on the boundary conditions. Because each hat function has local support and overlaps only a few neighbors, the matricies $\mathbf{A}$ and $\mathbf{B}$ are sparse.

In order to step this equation forward in time, the backward Euler method can be used:

$$\mathbf{A}\cdot(\mathbf{c}_{n+1} - \mathbf{c}_n) = \triangle t\mathbf{B}\cdot\mathbf{c}_{n+1} - \triangle t\mathbf{b}, \tag{2.7}$$

where the index n represents progress through time. This equation can be re-written as:

$$\mathbf{c}_{n+1} = (\mathbf{A} - \triangle t\mathbf{B})^{-1}\cdot(\mathbf{A}\cdot\mathbf{c}_n - \triangle t\mathbf{b}), \tag{2.8}$$

which can be used to calculate the diffusion over the finite element region at each time step in the simulation. The sparsity of the matricies allows rapid solution of Eq. 2.8.

2.4 Coupling Method

Now that we have the theoretical framework, the goal is to couple the stochastic model, given by Eq. 2.2, and the continuum model, which can be solved using Eq. 2.8. The first case that will be examined is diffusion in a 1D linear system. In order to couple these methods, the diffusion space must be divided into three regions: the continuum region (C), the particulate region (P), and a buffer region (B) 2.1. The continuum region C consists of finite element simplicies. The particulate region P consists of a region occupied only by discrete particles governed by the stochastic diffusion equation. The buffer region lies between the continuum and particulate regions and consists of a relatively small number of finite element simplicies, and also contains particles. The nodes which border only simplicies in region B are denoted as set $\mathcal{B}$, and the subset of $\mathcal{B}$ of nodes that directly border the particulate region is denoted $\partial\mathcal{B}$.

In order to transfer the particle concentration from the continuum region to the particulate region, a Dirichlet boundary condition of zero concentration is imposed on the nodes of $\partial\mathcal{B}$. Before each time step, the concentration of the nodes in B is set to zero. Then, using a suitable time step Δt, the continuum domain and particles are stepped forward in time using Eqs. 2.8 and 2.2, respectively. If the model describing the particulate region has any features that result in the removal of particles, such as a reactive region or absorbing boundary, any particles that meet such criterial for removal are removed at this time.

After the forward time step, some particles may have returned to the con-

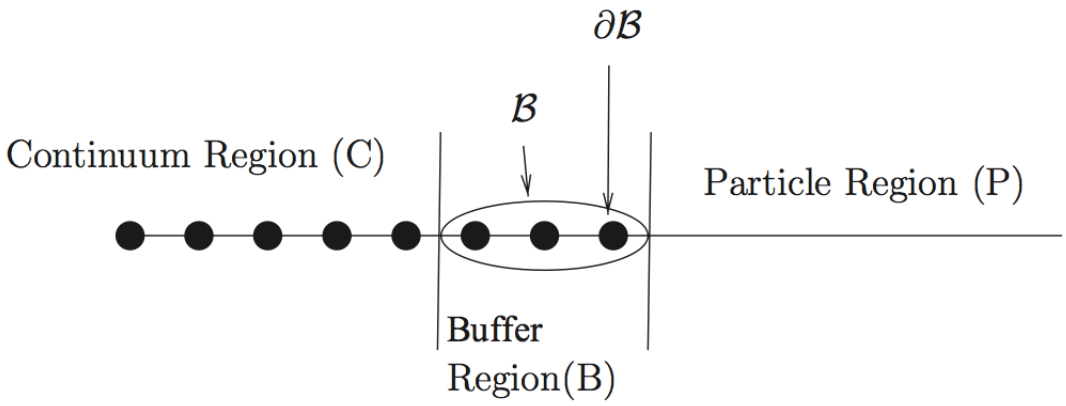


Figure 2.1: Schematic diagram of the continuum, buffer, and particle regions. The nodes are represented as dots on the line.

tinuum region C. These are reabsorbed into the continuum hat functions, with the coefficients increased accordingly. Once the particle is reabsorbed, its weight is divided among the N+1 nodes that border the simplex containing the particle in a manner that best preserves the original location of the particle. Before the simulation starts, the zeroth and first moment for each hat function is computed and stored:

$$p_i = \int u_i(\mathbf{x}) d\mathbf{x}, \tag{2.9}$$

$$\mathbf{m_i} = \int \mathbf{x} u_i(\mathbf{x}) d\mathbf{x}. \tag{2.10}$$

If N+1 weights w_i are added to the $N+1$ surrounding hat functions, then the first moment induced by the weights is given by:

$$\mathbf{m} = \frac{\sum_{i=1}^{N-1} w_i \mathbf{m_i}}{\sum_{i=1}^{N-1} p_i}. \tag{2.11}$$

Using the rule that the moment in Eq. 2.11 ref be equal to the position of the particle,

$$\mathbf{x} = \mathbf{m} \tag{2.12}$$

provides N equations. One more equation is needed to determine $N + 1$ weights; this is provided by the condition that the weights sum to 1:

$$\sum_{i=1}^{N+1} w_i = 1. \tag{2.13}$$

Equations 2.12 and 2.13 are solved to obtain weights, the coefficient at each node is augmented by its corresponding weight, and the particle is removed.

Once the continuum and the particles have been stepped forward, the coefficients at the nodes of $\mathcal{B} - \partial\mathcal{B}$ are no longer zero. Each hat function, weighted by its coefficient, is converted into zero or more particles. To do this, the number of particles generated by the hat function is computed by

$$n_f = \alpha c_i p_i. \tag{2.14}$$

The factorα allows one to create "particles" that represent fractional amounts of material. Using a large factor and this a large number of particles can be useful for reducing the amount of noise introduced by using the stochastic simulation. From a physical viewpoint, this is acceptable because the particles do not interact with each other in this model. In a more physically realistic system in which each particle corresponds to a physical particle, this parameter must be consistent with the units for c_i. If the concentration is expressed in terms of number of particles per unit volume, then α is 1.

The particle creation step generally results in a non-integer number, but this number is then "stochastically rounded" to avoid fractional particles. The number

n_f is rounded by computing the fractional part f, and rounding up with probability f or rounding down otherwise. Any particles that are generated must then be placed in the system. This is done by treating the normalized hat functions u_i/p_i as a probability distribution function and sampling it, usually by using the rejection method. The node coefficients are then set to zero.

One possible source of error could occur during the stepping forward of the continuum equations and the absorption of material at the boundary defined by $\partial\mathcal{B}$. This causes a net loss of material from the system and violates the conservation of matter. Therefore, this must be minimized by constructing the buffer region to have several layers of simplicies between the particulate and continuum regions. Since the node coefficients c_i are zero at the start of the time step, using a time step that is small compared to the time scale of diffusion across the smallest simplex ensures that the loss of material is negligible. This is demonstrated in the Appendix by a simple model.

2.5 1D Linear Example Problem

For the example problem solved by the hybrid method, a 1D linear domain was considered, with x ranging from 0 to 1. At $x = 0$, the concentration was fixed to be 1. The other side of the domain acts as an absorbing boundary; the concentration at $x = 1$ was fixed at 0. The initial conditions had a concentration of 0 throughout the domain, with diffusivity constant $D = 1$. The continuum region C occupied the interval $[0, \frac{9}{20}]$, and was subdivided into 9 elements, each of length $\frac{1}{20}$. The particulate region P is occupied the interval $[\frac{9}{20}, 1]$. The buffer region B is comprised three elements, each of length $\frac{1}{20}$. For this simulation, the time step size was 0.000625, and the factor α was 1000. The time step size was calculated using the equation

$$\Delta t = \frac{h^2}{4D} \tag{2.15}$$

where h is the width of each element (in this case $\frac{1}{20}$). Basing the time step size on the element width size ensures that the concentration from the continuum region will not advance too quickly and be absorbed into the buffer region at dB.

For the particles, the time step size was reduced when they were near the absorbing boundary. Each time step, obtained from Eq. 2.15, was subdivided into intervals less than or equal to a time step from

$$\Delta t_{sub} = \frac{x^2}{4D}, \tag{2.16}$$

where x is the distance from the boundary. The particles that were close enough so that Δt_{sub} was less than Δt were individually stepped forward using the smaller time step. This helped resolve the absorbing boundary more precisely and led to more accurate fluxes. It must be noted that in physically realistic systems, the use of very small time steps violates he assumption behind the derivation of the Brownian dynamics equation (Eq. 2.2) from molecular systems with inertia,[11] and using such small time steps could be controversial. In the models under consideration here, though, these small steps are used merely to solve a model mathematical problem.

The analytical series solution to this problem can be shown to be[31]

$$\mathcal{N} = t - 2\sum_{k=1}^{\infty} \frac{(-1)^k}{\pi^2 k^2} \left[\exp\left(-\pi^2 k^2 t\right) - 1\right], \tag{2.17}$$

where$\mathcal{N}$ is accumulated flux.

Figure 2.2 shows good agreement between the calculated flux and series solution.

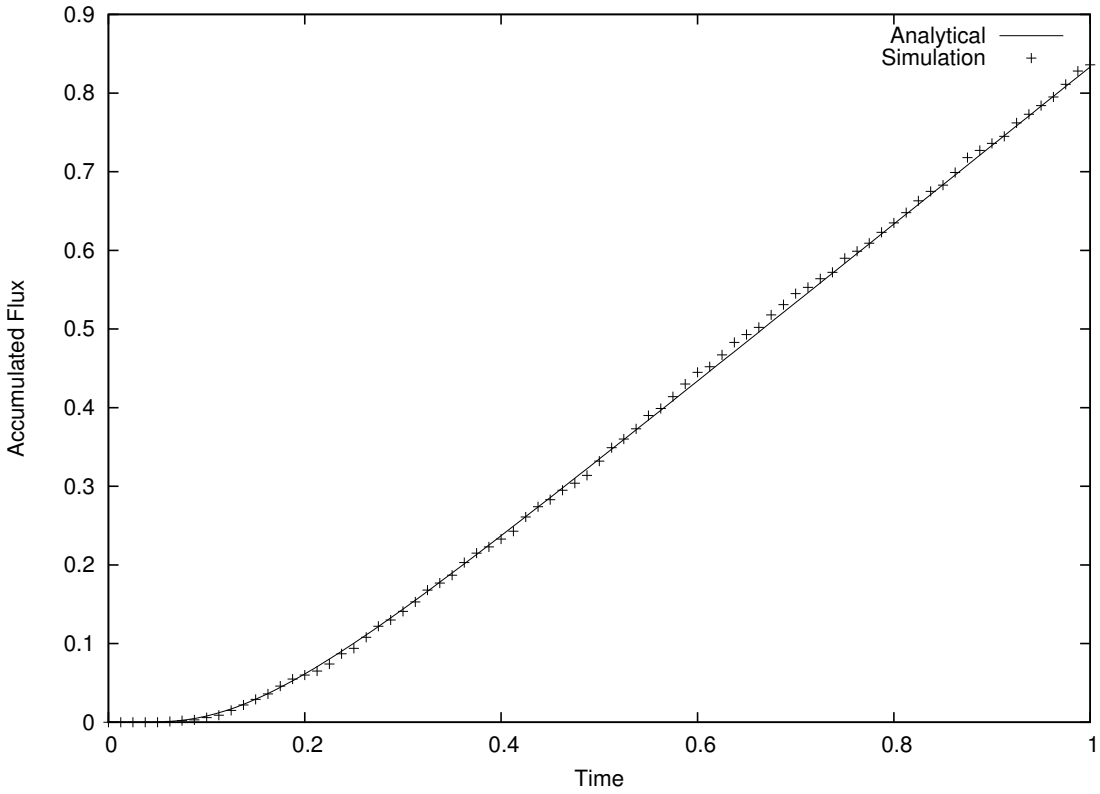


Figure 2.2: Plot of accumulated flux vs. time. The dashed line is the analytical solution and the data points represent the computational calculations.

2.6 1D Radial Hybrid Method

The method of the radial hybrid calculations is entirely similar to that for the 1D linear system, but the concerning equations are different. Thus, the calculations are reformatted below.

The radial diffusion equation to be solved is given by

$$\frac{\partial c}{\partial t} = \left(\frac{D}{r^2}\right) \frac{\partial}{\partial r} \left(r^2 \frac{\partial c}{\partial r}\right). \tag{2.18}$$

For the sample problem being solved, the boundary conditions are given by

$$c\left(r_1, t\right) = 0, \tag{2.19}$$

$$c\left(r_2, t\right) = c_b,$$

where r_1 is the inner absorbing boundary, r_2 is the outer boundary of the system, and c_b is the value of the fixed concentration at the outer boundary. As above, the entire problem is divided into a finite element, buffer, and particle regions, as shown in Figure 2.3.

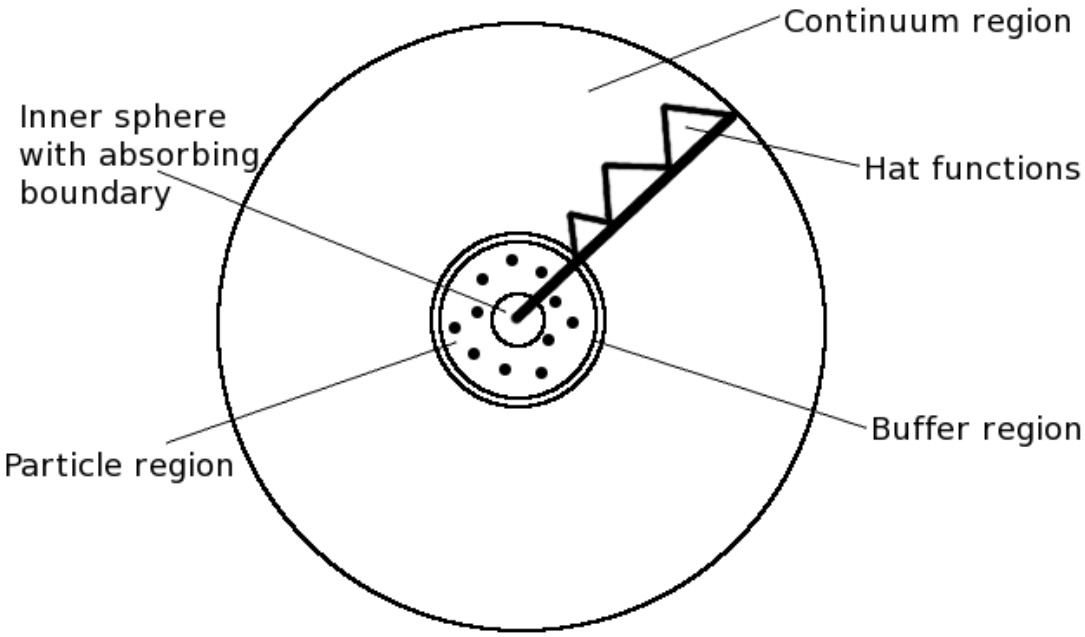


Figure 2.3: Schematic diagram of the 1-D finite element and particle hybrid model. The innermost region is a sphere with absorbing boundary conditions. The particle region, the buffer region, and the continuum region surround the sphere. This diagram also displays a depiction of the hat functions used to calculate diffusion in the particle region.

For the particle region, the particles are simulated in three dimensions using Eq. 2.2.

The finite element region is solved in a similar manner to the 1D solution, as shown in Sec. 2.3 of the paper, using the Galerkin method:

$$\int_{r_1}^{r_2} \frac{\partial c}{\partial t} r^2 dr = \int_{r_1}^{r_2} D \left(\frac{\partial^2 c}{\partial r^2} + \frac{2}{r} \frac{\partial c}{\partial r} \right) r^2 dr. \tag{2.20}$$

Through the use of the chain rule and integration by parts, and once again using u_i to represent the hat functions, it can be shown that this simplifies to

$$\int_{r_1}^{r_2} u_i \frac{\partial c}{\partial r} r^2 dr = -D \int_{r_1}^{r_2} \frac{\partial u_i}{\partial r} \frac{\partial c}{\partial r} r^2 dr. \tag{2.21}$$

Representing the concentration using hat functions

$$c(x,t) = \sum_j c_j(t)\, u_j(x). \tag{2.22}$$

Eq. 2.21 can now be written as

$$\sum_j \frac{dc_j}{dt} \int_{r_1}^{r_2} u_i u_j r^2 dr = -D \sum_j c_j \int_{r_1}^{r_2} u_i' u_j' r^2 dr, \tag{2.23}$$

where

$$\begin{aligned} (u_i, u_j) &= \int_{r_1}^{r_2} u_i u_j r^2 dr, \\ \left(u_i', u_j'\right) &= \int_{r_1}^{r_2} u_i' u_j' r^2 dr. \end{aligned} \tag{2.24}$$

Thus Eq. 2.23 can be written as

$$\sum_j \frac{dc_j}{dt} (u_i, u_j) = -D \sum_j c_j \left(u_i', u_j'\right). \tag{2.25}$$

This can be written in matrix form as given in Eq. 2.7. Once the integrals given in Eq. 2.24 have been solved analytically, this calculation proceeds exactly as demonstrated in the 1D linear case.

2.6.1 Results

The sample problem for this case was solved over a 1D spherically symmetric domain where $r_1 = 1$ and $r_2 = 10$. At r_2, the concentration was fixed to be 1 and at the absorbing boundary r_1 the concentration is fixed at 0. Once again the initial conditions have a concentration of 0 throughout the domain, with diffusivity constant $D = 1$.

In this case, the nodes are spherical surfaces, and the finite elements are spherical shells with a hat profile. The continuum region C is subdivided into 39 elements with 0.18 length units between each node. In this case the buffer region B is comprised three nodes. The time step was calculated using Equation 2.15 above, and the factor α was set to 10. The analytical solution for the flux across the the boundary at time t is given by[31]

$$\mathcal{N} = J_{ss}\left[t + \frac{2}{\pi^2}\frac{(r_2 - r_1)^2}{D}\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}\left(1 - \exp\left(-\left(\frac{\pi n}{r_2 - r_1}\right)^2 Dt\right)\right)\right], \quad (2.26)$$

where J_{ss} is the steady-state flux:

$$J_{ss} = \frac{4\pi c_b D r_1 r_2}{r_2 - r_1}. \quad (2.27)$$

Figure 2.4 shows good agreement between the analytical solution and calculated accumulated flux.

2.7 1D Spherical Example with Unevenly Spaced Symmetric Elements

With the goal of eventually using this method to study more complicated realistic systems, the spherical simulation was altered to allow for a non-uniform grid of the diffusion domain. For this sample problem, the diffusion domain was much larger, with $r_1 = 1$ and $r_2 = 100$. In this case the initial conditions set the

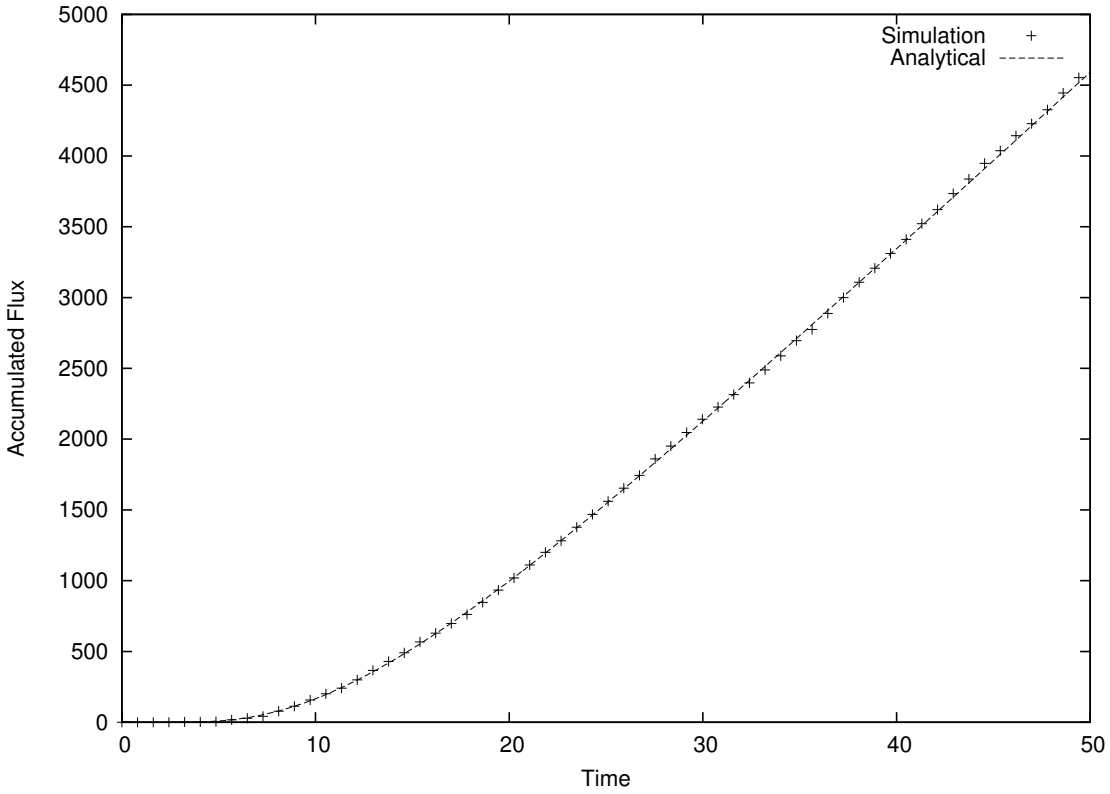


Figure 2.4: Plot of accumulated flux vs. time for the first spherically symmetric case. The dashed line is the analytical solution and the data points represent the computational calculations.

initial concentration at $c = 1$ throughout the diffusion domain. The node positions were assigned according to the positions described in Figure 2.5 in which the nodes are numbered starting at 1 with the closest one to the absorbing sphere.

In this case there were three elements in the buffer region, B. The elements were generated by starting with the interval (1, 100), and assigning values to the endpoints from the known steady-state solution $c = 1 - r_1/r$. A linear interpolation was constructed on the interval, and if the maximum relative difference between the interpolation and the steady-state solution was greater than 0.01, the interval was divided in half, with a value from the steady-state solution assigned to the new point. The new intervals were recursively divided in the same manner until no more

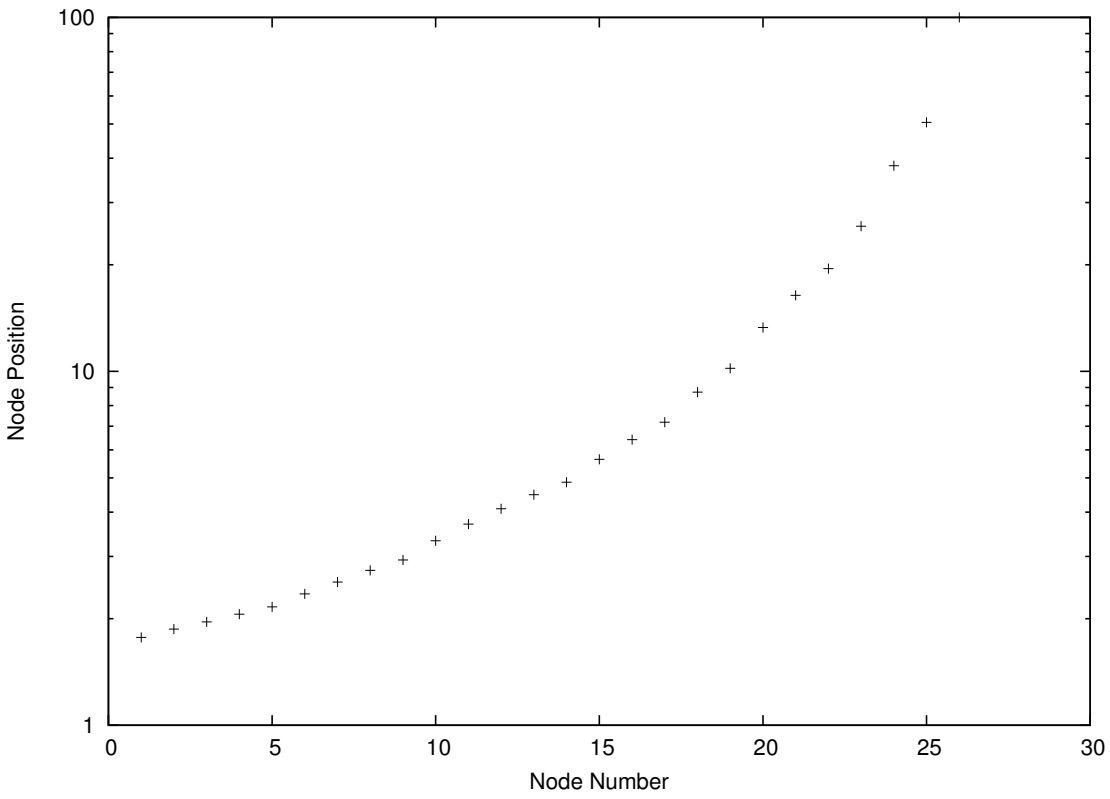


Figure 2.5: Graph showing node positions for unevenly spaced nodes.

divisions occurred. This procedure assigns smaller finite elements to regions where the concentration is expected to vary more strongly across distance. Three of the intervals, encompassing the distance about 1.8-2 length units from the center of the sphere, were designated as the buffer region, and the intervals closer to the sphere than the buffer region were removed from consideration. The time step size from Eq. 2.15 was based on the most narrow finite element, and the factor α was set to 100.

This case approximatly solves the unsteady-state Smoluchowski problem, in which a reactive sphere is placed in an infinite, uniform sea of diffusing substrate points. Because the outer boundary is very far compared to the sphere radius, the solution obtained should be very similar to that with a boundary at infinity.

The analytical solution for this case is[31]

$$\mathcal{N} = 4\pi r_1 D c_b \left[t + 2r_1 \sqrt{\frac{t}{\pi D}} \right]. \tag{2.28}$$

The results for the flux calculations can be seen in Figure 2.6.

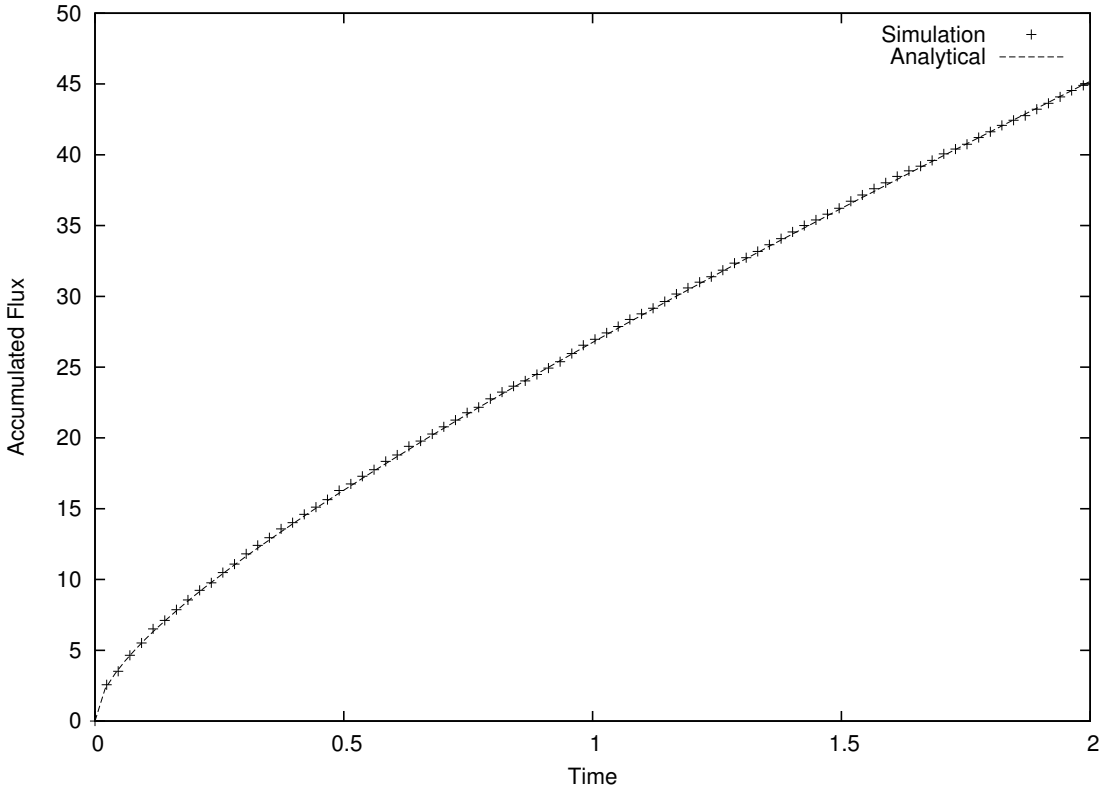


Figure 2.6: Plot of flux vs. time for the case of unevenly spaced nodes. The dashed line show the analytical solution and the data points represent results from the calculation.

2.8 Discussion and Conclusion

As demonstrated by the examples above, the results from this method closely match the analytical solutions of the time course of consumed material. Although the flux has temporal fluctuations due to the discrete particles, the fluctuations

are smoothed out by the integration of the flux. The method's limitations include the approximation and limitations inherent in both finite element and Brownian dynamics methods. The only additional approximation introduced is the small loss of material at one of the interfaces between the continuum and particulate regions, and this approximation can be controlled as described in the Appendix. Even though this method can be used as a multiscale method, with finite elements describing the courser domains and Brownian dynamics describing the finer domains, namely, the motion of the individual particles and the finite elements in the buffer region. Additional methodology development using stiff time integrators[32] might allow this limitation to be overcome.

Finally, further development will be needed to study high concentrations of interacting particles. Currently, one can include interparticle interactions in both finite element computations[10,26] and Brownian dynamics simulations, as long as they are not run together as in the proposed method. One can also include external forces on the particles, as seen in the equations above. The missing link is a method to compute interparticle interactions between the particulate and continuum regions; this will be the subject of future work.

Although the examples provided in this paper are quite basic, this methodology will be further implemented to study much more complex systems. One goal is to use receptor crystal structures with specified active sites at the center of the diffusive system, surrounded by a small stochastic region which is then surrounded by a much larger continuum region on the cellular level. The diffusion domain can have arbitrary complexity in its geometry and boundary conditions; it is not limited to spherically symmetric geometries such as those used in current Brownian dynamics software packages for determining reaction rates.[17,18,19,20] One system of great interest if that of calcium ion diffusion in the cardiac myocyte,[4,5,33] the dif-

fusion domain can be constructed from light and electron microscopy data,[33] while the calcium transporters and receptors can include in atomic detail for Brownian dynamics.[34] Other systems of interest include diffusion of acetylcholine in the neuromuscular junction[6,7,8,9] and the diffusion of sodium and potassium ions in the node of Ranvier.[10] Because such systems have multiple length and time scales, hybrid methods that can combine different algorithms and take advantage of their respective strengths will be useful for predicting and studying physiologically significant rates, fluxes, and currents.

2.9 Appendix: Demonstration of Negligible Loss of Material

In order to demonstrate that a negligible amount of material is lost in this method, we set up a 1D interval of L units wide, with concentration c held at c_0 at the left side ($x = 0$) and zero at the right side $x = L$. Since our examples have used buffer regions that are three elements wide, we assume three intervals, each $L/3$ units wide. Using Eq. 2.15 gives a time step of $\Delta t = L^2/(36\mathcal{D})$, where $\mathcal{D}$ is the constant diffusivity. Using separation of variables gives the concentration as a function of position and time:[31]

$$c(x,t) = \frac{c_0(L-x)}{L} - \frac{2c_0}{L}\sum_{n-1}^{\infty}\frac{1}{n}\exp(-\frac{\pi^2 n^2}{L^2}\mathcal{D}t)\sin(\frac{\pi n x}{L}). \tag{2.29}$$

We compute the amount of matter lost at the right boundary between zero time and the time step value relative to the amount of matter that remains in the interval after the time step.

$$f = \frac{-\int_0^{\Delta t}\mathcal{D}\frac{\partial c}{\partial x}|_{x=L}dt}{\int_0^L c(x,\Delta t)dx}. \tag{2.30}$$

Applying this to the expression for the concentration and plugging in the value for Δt give the fraction 1.2×10^{-5}. Thus, this approximation is controllable

by having a sufficiently wide buffer region.

The question then arises: why not eliminate the loss altogether by using a no-flux boundary condition between the buffer and particulate regions? With our present geometry, this would probably cause only a very small distortion in the average position of the material. However, of the boundary effect were to become significant, for example, because of large forces of long-running equilibrium simulations, it would be necessary to devise a method for converting absorbed material into particles at the boundary. It is not as clear how one would handle a no-flux condition. This is our justification for currently using an absorbing boundary rather than a no-flux boundary condition.

2.10 Acknowledgements

The work at UCSD was supported in part by NIH, NSF, HHMI, CTBP, NBCR, and the UCSD Chancellor's Interdisciplinary Collaboratories Fellowship.

Chapter 2, in full, is a reprint of the material as it appears in The Journal of Chemical Physics 2012. Bauler, Patricia; Huber, Gary A.; McCammon, J. Andrew, American Institute of Physics 2012. The dissertation author was the primary investigator and author of this paper.

2.11 References

[1] Ourporov, I.V; Knull, H.R.; Huber, A.; Thomasson, K.A. Brownian Dynamics Simulations of Aldose Binding Glyceraldehyde 3-Phosphate Dehydrogenase and the Possibility of Substrate Channeling. *Biophys. J.* **2001**, *80*, 2527-2535.

[2] Spaar, A.; Dammer, C.; Gabdoulline, R.R.; Wade, R.C., Helms, V. Diffusional Encounter of Barnase and Barstar. *Biophys. J.* **2006**, *90*, 1913-1924.

[3] Northrup, S.H; Jeffrey, A.L.; Boles, J.O.; Reynolds, J.C.L. Brownian dynamics simulation of protein association. *J. Computer-Aided Molecular Design* **1987**, *1*, 291-311.

[4] Hake, J.; Lines, G.T. Stochastic Binding of Ca^{2+} Ions in the Dyadic Cleft; Continuous versus Random Walk Description of Diffusion. *Biophys. J.* **2008**, *94*, 4184-4201.

[5] Lu, S.; Michailova, A.P; Saucerman, J.J.; Cheng, Y.; Yu, Z.; Kaiser, T.H.; Li, W.W.; Bank, R.E.; Holst, M.J., McCammon, J.A.; Hayashi, T.; Hoshijima, M.; Arzberger, P.; McCulloch, A.D. Multiscale Modeling in Rodent Ventricular Myocytes: Contributions of Structural and Functional Heterogeneities to Excitation-Contraction Coupling. *IEEE Eng. Med. Bio. Mag.* **2009**, *28*, 46-57.

[6] Bartol, T.; Land, B.; Salpeter, E.; Salpeter, M. Monte Carlo simulation of Miniature endplate current generation in the vertebrate neuromuscular junction. *Biophys. J.* **1991**, *59*, 1290-1307.

[7] Smart, J.; McCammon, J.A. Analysis of Synaptic Transmission in the Neuromuscular Junction Using a Continuum Finite Element Model. *Biophys. J.* **1998**, *75*, 1679-1688.

[8] Tai, K.; Bond, S.D.; MacMillan, H.R.; Baker, N.A.; Holst, M.J.; McCammon, J.A. Finite Element Simulations of Acetylcholine Diffusion in Neuromuscular Junctions. *Biophys. J.* **2003**, *84*, 2234-2241.

[9] Antosiewicz, J.; Gilson, M.K.; Lee, I.H.; McCammon, J.A. Acetylcholinesterase: Diffusional Encounter Rate Constants for Dumbbell Models of Ligands. *Biophys. J.* **1995**, *68*, 62-68.

[10] Lopreore, C.L.; Bartol, T.M.; Coggan, J.S.; Keller, D.X.; Sosinsky, G.E.; Ellisman, M.H.; Sejnowski, T.J. Computational Modeling of THree-Dimensional Electrodiffusion in Biological Systems: Application to the Node of Ranvier. *Biophys. J.* **2008**, *95*, 2624-2635.

[11] Ermak, D.L.; McCammon, J.A. Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys.* **1978**, *69*, 1352-1360.

[12] Northrup, S.H.; Allison, S.A.; McCammon, J.A. Brownian dynamics simulation of diffusion-influenced bimolecular reactions. *J. Chem. Phys.* **1984**, *80*, 1517-1524.

[13] Zhou, H.-X. On the calculation of diffusive reaction rates using Brownian dynamics simulations. *J. Chem. Phys.* **1990**, *92*, 3092-3095.

[14] Luty, B.A.; McCammon, J.A.; Zhou, H.-X. Diffusive reaction rates from Brownian dynamics simulations: Replacing the outer cutoff surface by an analytical treatment. *J. Chem. Phys.* **1992**, *97*, 5682-5686.

[15] Bauler, P.; Huber, G.; Leyh, T.; McCammon, J.A. Channeling by Proximity: The Catalytic Advantages of Active Site Colocalization Using Brownian Dynamics. *J. Phys. Chem. Letters* **2010**, *1*, 1332-1335.

[16] Song, Y.; Zhang, Y.; Shen, T.; Bajaj, C.L.; McCammon, J.A.; Baker, N.A. FInite Element Solution of the Steady-State Smoluchowski Equation Rate COnstant Calculations. *Biophys. J.* **2004**, *86*, 2017-2029.

[17] Huber, G.A.; McCammon, J.A. Browndye: A software package for Brownian dynamics. *Comp. Phys. Comm.* **2010**, *181*, 1896-1905.

[18] Gabdoulline, R.; Wade, R. Simulation of the Diffusional Association of Barnase and Barstar. *Biophys. J.* **1997**, *72*, 1917-1929.

[19] Northrup, S.; Laughner, T.; Stevenson, G. *MACRODOX macro-molecular simulation program*, Department of Chemistry, Tennessee Technological University, Cookeville, TN, 1997; online at iweb.tntech.edu/macrodox/macrodox.html.

[20] Madura, J.; Briggs, J.; Wade, R.; Davis, M.; Luty, B.; Ilin, A.; Antosiewicz, J.; Gilson, M.; Bagheri, B.; Scott, L.; McCammon, J. Electronics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program. *Comput. Phys. Commun.* **1995**, *91*, 57-95.

[21] Długosz, M.; Zieliński, P.; Trylska, J. Software News and Updates Brownian Dynamics Simulations on CPU and GPU with BD_BOX. *J. Comput. Chem.* **2011**, *32*, 2734-2744.

[22] Geyer, T. Many-particle Brownian and Langevin Dynamics Simulations with the Brownmove package. *BMC Biophys.* **2011**, *4*.

[23] McGuffee, S.R.; Elock, A.H. Diffusion, Crowding & Protein Stability in a Dynamic Molecular Model of the Bacterial Cytoplasm. *PLoS Comp. Bio.* **2010**, *6*.

[24] Cheng, Y.; Suen, J.K.; Zhang, D.; Bond, S.D.; Zhang, Y.; Song, Y.; Baker, N.A.; Bajaj, C.L.; Holst, M.J.; McCammon, J.A. Finite Element Analysis of the Time-Dependent Smoluchowski Equation for Acetylcholinesterase Reaction Rate Calculations. *Biophys. J.* **2007**, *92*, 3397-3406.

[25] Lu, B.; Zhou, Y.C.; Huber, G.A.; Bond, S.D.; Holst, M.J.; McCammon, J.A. Electrodiffusion: a continuum modeling framework for biomolecular systems with realistic spatiotemporal resolution. *J. Chem. Phys.* **2007**, *127*, 135102-1 - 135102-17.

[26] Zhou, Y.C.; Lu, B.; Huber, G.A.; Holst, M.J.; McCammon, J.A. Continuum Simulations of Acetylcholine Consumption by Acetylcholinesterase: A Poisson-Nernst-Planck Approach *J. Phys. Chem. B* **2008**, *112*, 270-275.

[27] Cheng, Y.; Chang, C.A.; Yu, Z.; Zhang, Y.; Sun, M.; Leyh, T.S.; Holst, M.J.; McCammon, J.A. Diffusional Channeling in the Sulfate-Activating Complex: Combined Continuum Modeling and Coarse-Grained Brownian Dynamics Studies. *Biophys. J.* **2008**, *95*, 4659-4667.

[28] Geyer, T.; Gorba, C.; Helms, V. Interfacing Brownian Dynamics Simulations. *J. Chem. Phys.* **2004**, *120*, 4573-4580.

[29] Gorba, C.; Geyer, T.; Helms, V. Brownian Dynamics SImulations of Simplified Cytochrome c Molecules in the Presence of a Charged Surface. *J. Chem. Phys.* **2004**, *121*, 457-464.

[30] Im, W.; Seefeld, S.; Roux, B. A Grand Canonical Monte Carlo-Brownian Dynamics Algorithm for Simulating Ion Channels. *Biophys. J.* **2000**, *79*, 788-801.

[31] Crank, J. *The Mathematics of Diffusion*, 2nd ed; Clarendon Press: Oxford,1975.

[32] Hairer, E.; Wanner, G. Stiff differential equations solved by Radau methods. *J. Comput. App. Math.* **1999**, *111*, 93-111.

[33] Cheng, Y.; Yu, Z.; Hoshijima, M.; Holst, M.J.; McCulloch, A.D.; McCammon, J.A.; Michailova, A.P. Numerical Analysis of Ca^{2+} Signaling in Rat Ventricular Mocytes with Realistic Transverse-Axial Tubular Geometry and Inhibited Sarcoplasmic Reticulm. *PLoS Comput. Biol.* **2010**, *6*.

[34] Lindert, S.; Kekenes-Huskey, P.M.; Huber, G.; Pierce, L.; McCammon, J.A. Dynamics and Calcium Association to the N-Terminal Regulatory Domain of Human Cardiac Troponin C: A Multiscale Computational Study. *J. Phys. Chem. B* **2012**, *29*, 8449-8459.

CHAPTER 3

Computer aided drug discovery studies on dihydropteroate synthase

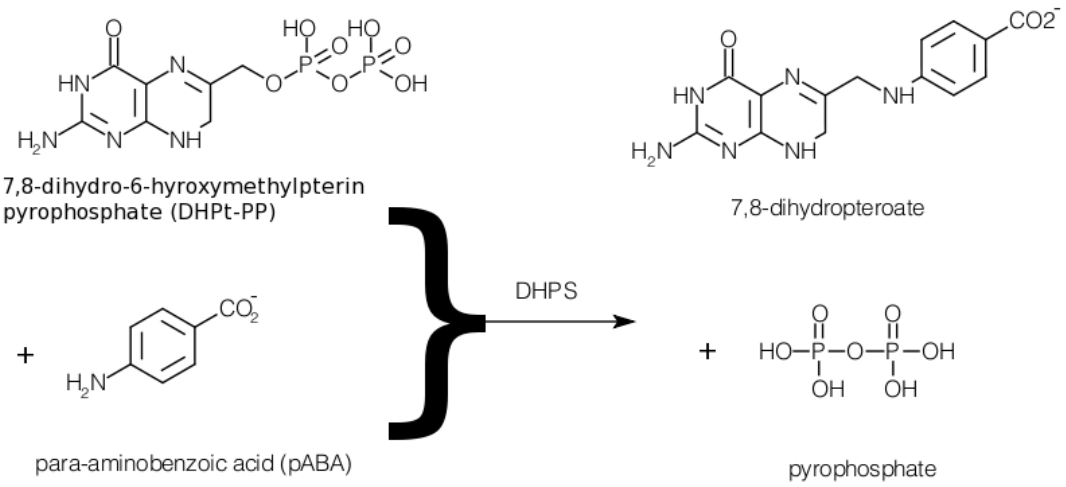
3.1 Abstract

Dihydropteroate synthase (DHPS) has long been of interest as a drug target due to its key role in the folate production pathway. However, many strains of infectious bacteria have become resistant to current treatments, and new inhibitors must be developed. Computational methods have been proven effective in streamlining the search for new drug compounds. This study makes use of molecular dynamics, virtual screening, and visualization tools to search for new potential inhibitors for the DHPS enzyme. The molecular dynamics simulations provided a range of conformations to be used in the virtual screening, as well as providing a visualization of how the DHPS motions may aid in its catalytic functions. In the end, the top 230 scoring compounds from two different databases of compounds are presented for further studies.

3.2 Introduction

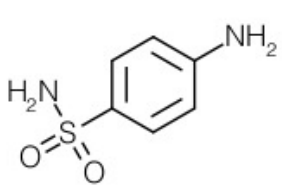
DHPS is one of the key enzymes in folate synthesis in bacteria. It catalyzes the reaction between *p*-aminobenzoic acid (pABA) and 7,8-dihydro-6-hydroxymethylpterin pyrophosphate (DHPt-PP) to form 7,8-dihydropteroate and pyrophosphate.[1,2] In mammals, folate is obtained through food and transported to cells via membrane-associated folate transport proteins. In plants and most microorganisms, including many infectious bacteria and some fungi, folate must be synthesized *de novo*. As folate aids in DNA and RNA production in bacteria, it is essential for survival,

and disrupting this pathway could be an excellent way to target infectious bacteria without interrupting functions in the host organism. Since DHPS is one of the key enzymes in the folate production pathway, it is considered an excellent potential drug target, and finding an inhibitor could aid in fighting many diseases including malaria, tuberculosis, and staph infections.[3] One of the main reasons that DHPS is considered to be one of the best targets in the folate production pathway is because it has already been shown that inhibiting this enzyme is effective for bacterial death. DHPS has long been a target of the sulfonamide class of antibacterials (often called sulfa drugs), which are structural analogs of pABA and act as competitive inhibitors.[4] Sulfa drugs work by creating dead end sulfa-pterin products that halt the final step in the folate production pathway (Figure 3.1).

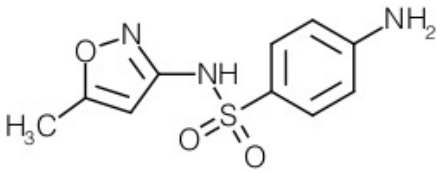


(a)

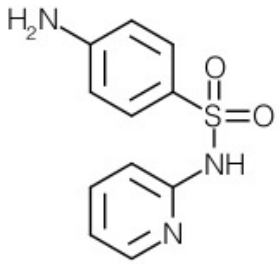
Figure 3.1: DHPS reaction mechanism and current inhibitors. (a) The top figure shows the ligands and products catalyzed by the DHPS enzyme. (b) The bottom figure shows a few examples of sulfa drug inhibitors.



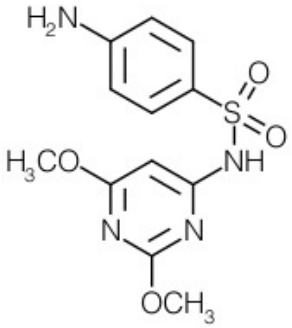
Sulfanilmide



Sulfamethoxazole



Sulfapyridine



Sulfadimethoxine

(b)

Figure 3.1: Continued

Unfortunately, new strains of bacteria are showing increasing resistance to the currently used sulfa drugs, and this combined with toxic side effects in patients has lead to a decrease in potency and use.[5,6,7,8] There have been other recent studies that attempt to find alternatives to the currently available sulfa drugs by searching for compounds that will act in the same manner as competitive inhibitors to pABA.[9] In this study, however, the focus is on searching for potential competitive inhibitors for the other substrate, DHPt-PP. Although the only known inhibitors for DHPS work by binding in the pABA active site, the DHPt-PP active site is more highly

conserved across species, so looking for a competitor of DHPt-PP could be more effective in a wider variety of diseases.[10] In addition, studies indicate that the order of substrate binding in DHPS is DHPt-PP followed by pABA, so the hope is that if a suitable inhibitor could be found, it could potentially block both the DHPt-PP and pABA from binding and prevent any catalytic reaction from taking place at all.[11]

This study will employ the use of computer-aided drug discovery techniques to simulate and examine the motions of the DHPS binding pocket and search for potential inhibitors. Though there are many structures of DHPS available in the RCSB Protein Data Bank, this study will specifically look at DHPS from *Staphylococcus aureus*. One of the main reasons to look at this target specifically is that methicillin-resistant *Staphylococcus aureus* (MRSA) is becoming increasingly prevalent in hospitals with fewer effective antibiotics.[12] In addition, the *S. aureus* DHPS crystal structure has been determined in both apo and holo forms, with the holo form containing a 6-hydroxymethylpterin-diphosphate ligand (DHP) bound in the active site of interest. The presence of this DHP ligand, which is coordinated to the enzymes manganese metal ion, will allow for a more accurate look at the dynamics of the active site in the presence of a ligand, which will be useful in searching for other potential ligands to act as inhibitors.

3.3 Molecular Dynamics Simulations

The goals of this project are to use molecular dynamics simulations to sample the conformation space of the DHPS drug target. From there, additional computational tools will be used to examine the active site and to search for potential drug compounds.

The first step of this process is to have a complete model of the protein that can be used for the molecular dynamics simulations. Molecular dynamics is

a method for studying the motions of a system by calculating the forces on all of the atoms within that system and simulating the resulting movements. In the case of *S. aureus* DHPS enzymes, there were two crystal structures available from the PDB. Both were crystallized as homo-dimers, but both were missing sections of their structure. The apo structure (PDBID: 1AD1) had a complete A chain, but was missing residues 15-24 and 50-55 on the B chain. The holo structure (PDBID: 1AD4) was missing residues 13-24 on the A chain, as well as residues 14-24 and 50-56 on chain B. DHPS is part of the TIM-barrel group of proteins with an $(\beta/\alpha)_8$-folding topology that contains a β-barrel active site.1,13 The missing segments in the crystal structure are located on the loop regions at the top of the β-barrel. It is thought that these regions are extremely flexible, especially A chain residues 13-24, which are part of the proposed pABA binding site.[14] Although the apo structure is more complete, the presence of the ligand and manganese metal ions in the holo structure makes it more useful for examining the dynamics of the active site. The missing segments were modeled in by hand by copying the complete segments from the apo structure into the holo structure. VMD (Visual Molecular Dynamics) and Avogadro software programs were used to create a reasonable starting structure for the molecular dynamics simulations.[15]

Once the necessary changes were made to complete the structure, the ligand was parameterized and the protein was solvated. The entire system consisted of 83908 atoms. After a 5ns equilibration, the molecular dynamics of the system were run using AMBER to produce a 100ns trajectory. The system was run at a temperature of 300K using a 2ps timestep. The RMSD plot of the data (Figure 3.2) is fairly stable and indicated that the modeled loop sections did not cause distortions to the overall structure.

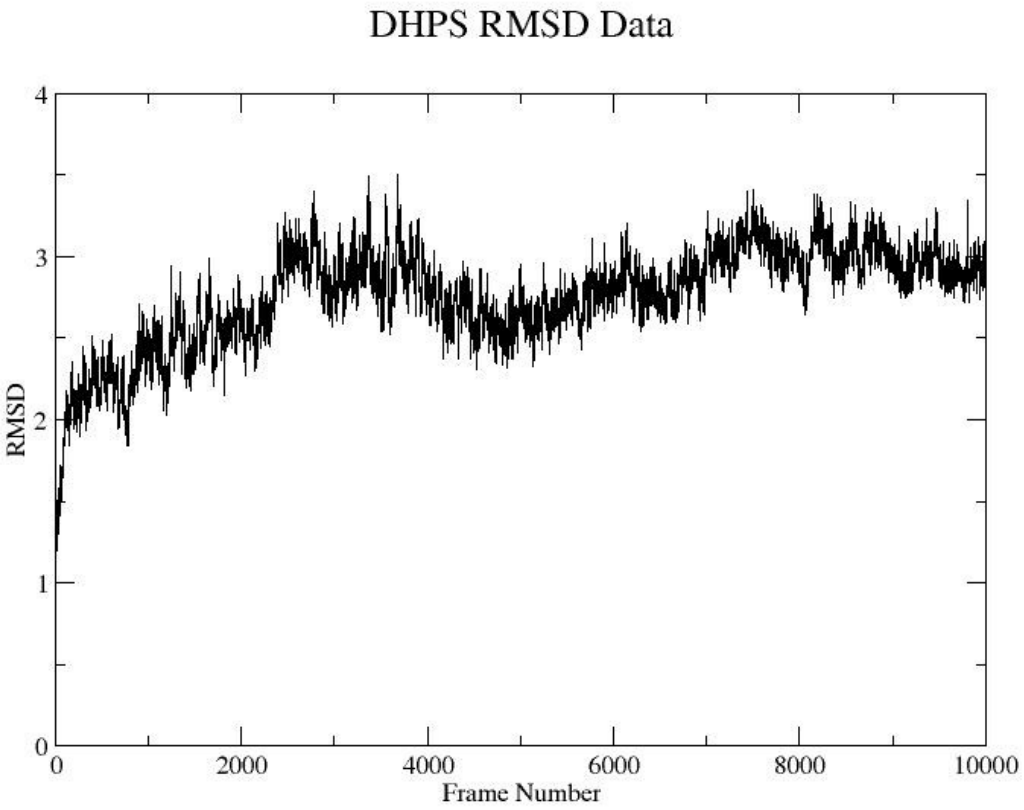
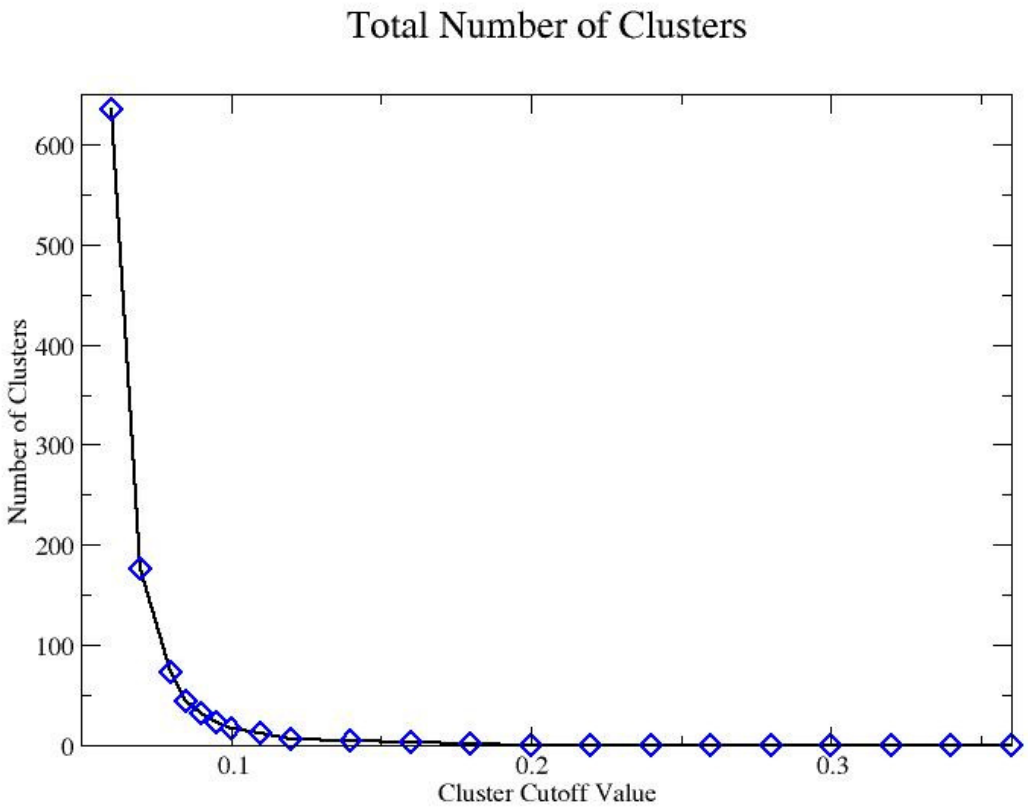


Figure 3.2: All-atom RMSD plot from molecular dynamics trajectory. The RMSD data shows that the overall structure is quite stable.

3.4 Relaxed Complex Scheme Docking

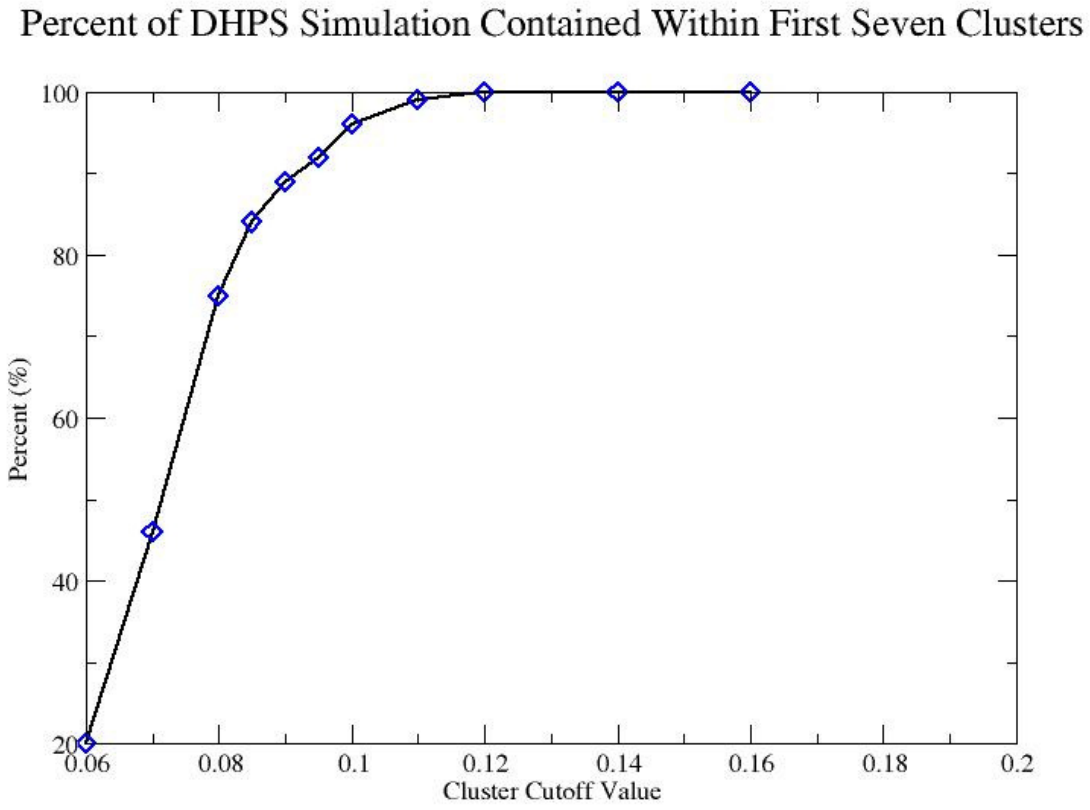
The Relaxed Complex Scheme is a method of searching for potential drug compounds.[16] It uses snapshots from the molecular dynamics simulations for docking studies so that multiple conformations can be tested. In order to find these representative structures, the structures from each frame of the molecular dynamics trajectory are clustered into representative groups. Ideally, the clustering should provide enough groupings that structures can be sufficiently differentiated, but not so many that structural similarities are ignored. Additionally, the clusters chosen

for study should be representative of a majority of the conformation space so that prevalent structures are not ignored during the docking. Cluster cutoff values above 0.20 nm produced only one cluster, and cutoff values below 0.08 nm produced hundreds of clusters (Figure 3.3(a)). Cutoff values of 0.09-0.16 nm were examined to find optimum balance between number of clusters and sufficient representation of conformation space. It was decided that at least 90% of the simulation trajectory should be located within the seven most populated clusters, in order to provide sufficient sample space but a reasonable number of structures to be used for docking (Figure 3.3(b)).



(a)

Figure 3.3: Generating the clusters. (a, top) This graph shows the number of clusters at each cutoff value. All of the values above 0.20 produced only one cluster, and so were not used for further analysis. (b, bottom) This shows the percent of the simulation contained within the seven most populated clusters at each cutoff value. We wanted something at or above 90%, so the possible cluster cutoff range is 0.09-0.16. After looking at the number of clusters produced by each cutoff value, it was decided that 0.09 would provide the best balance between representing a large section of the sample space and providing a sufficiently diverse set of structures that can be subject to docking calculations in a reasonable amount of time.



(b)

Figure 3.3: Continued

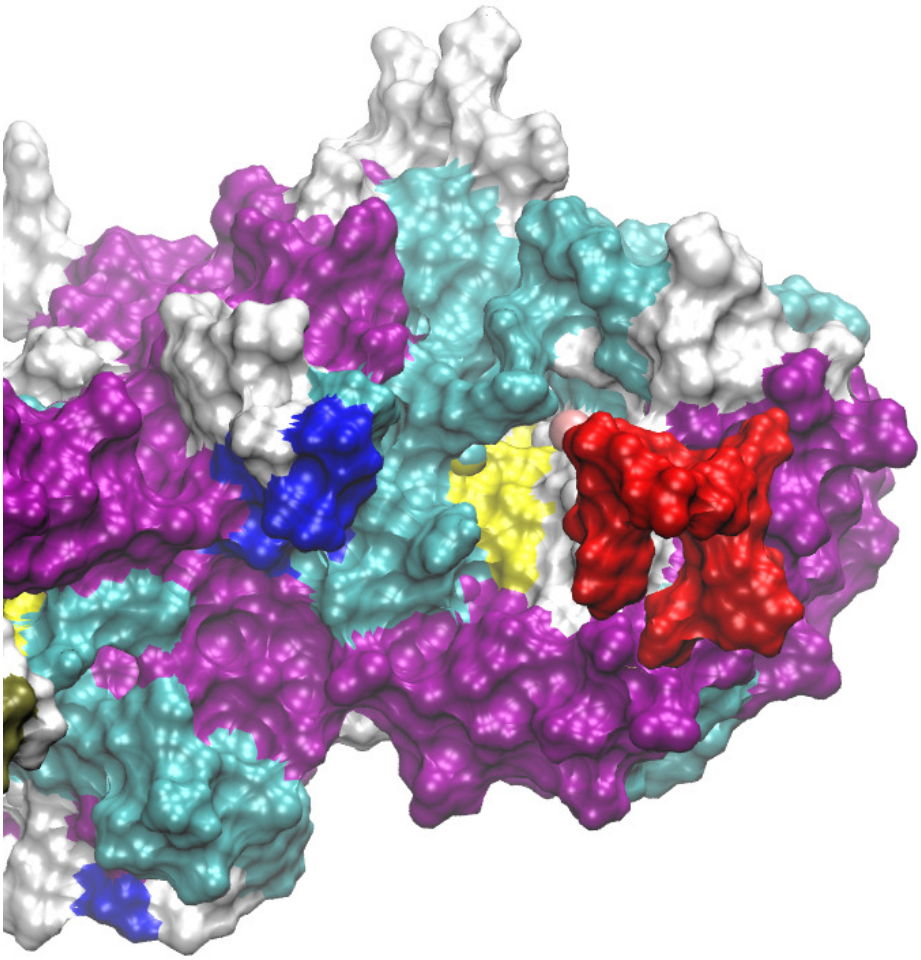
It was decided to use 0.09 nm as a cluster cutoff value. This produced a set of 32 clusters, with sufficiently different representative structures. The seven most populated structures were used as targets for docking studies. The docking studies were performed using the Schrodinger Maestro software package. The NCI Diversity Set III was used as test compounds and was docked in each of the seven representative structures. As the NCI Diversity Set III is a relatively small compound library, docking results for all compounds within the data set were obtained. In order to obtain a more manageable number of compounds for testing, a cutoff value based

on docking score was used. Compounds were then ranked based on docking score and duplicate compounds were removed from the test pool, which resulted in 65 unique compounds from the NCI Diversity Set III. (See Appendix 1 for compound structures).

3.5 POVME

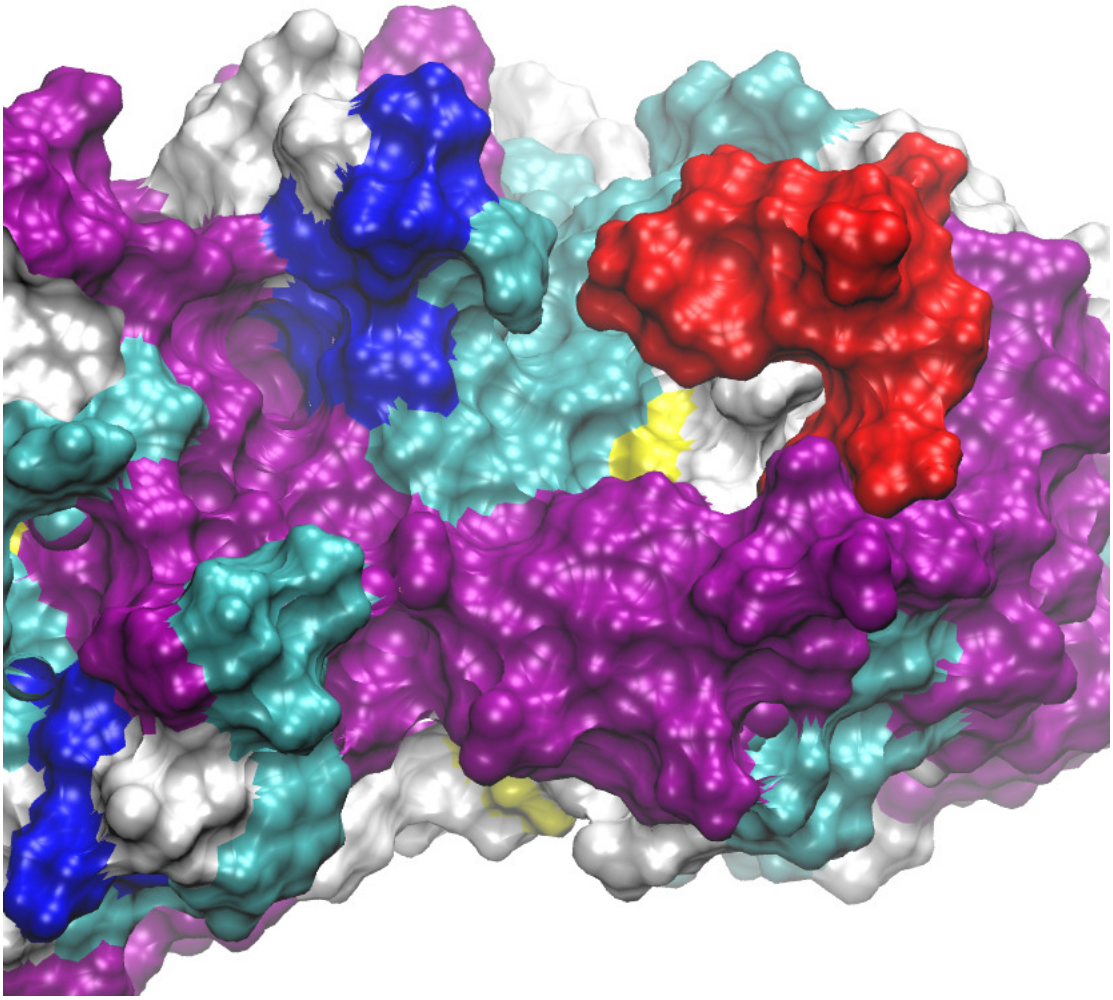
Further examination of active site conformations was done by utilizing the POVME (POcket Volume MEasurer) software.[17] POVME is a useful tool for visualizing the changes in pocket size during a molecular dynamics simulation. The user defines a volume of grid points in a region of interest (in this case, the active site of the DHPS). The POVME software deletes grid points that are too close to protein atoms, and also removes discontiguous points. The remaining points are used to calculate the volume of the protein pocket, and can also be used to easily visualize the shape of the pocket. This can be done for each frame in the molecular dynamics trajectory, which provides easy visualization of the changes in the shape of the active site as the trajectory progresses.

Throughout the trajectory the modeled flexible loop region (chain A residues 13-24) shows a lot of movement, and opening and closing motions of a possible pocket in the proposed pABA binding region can be seen (Figure 3.4).



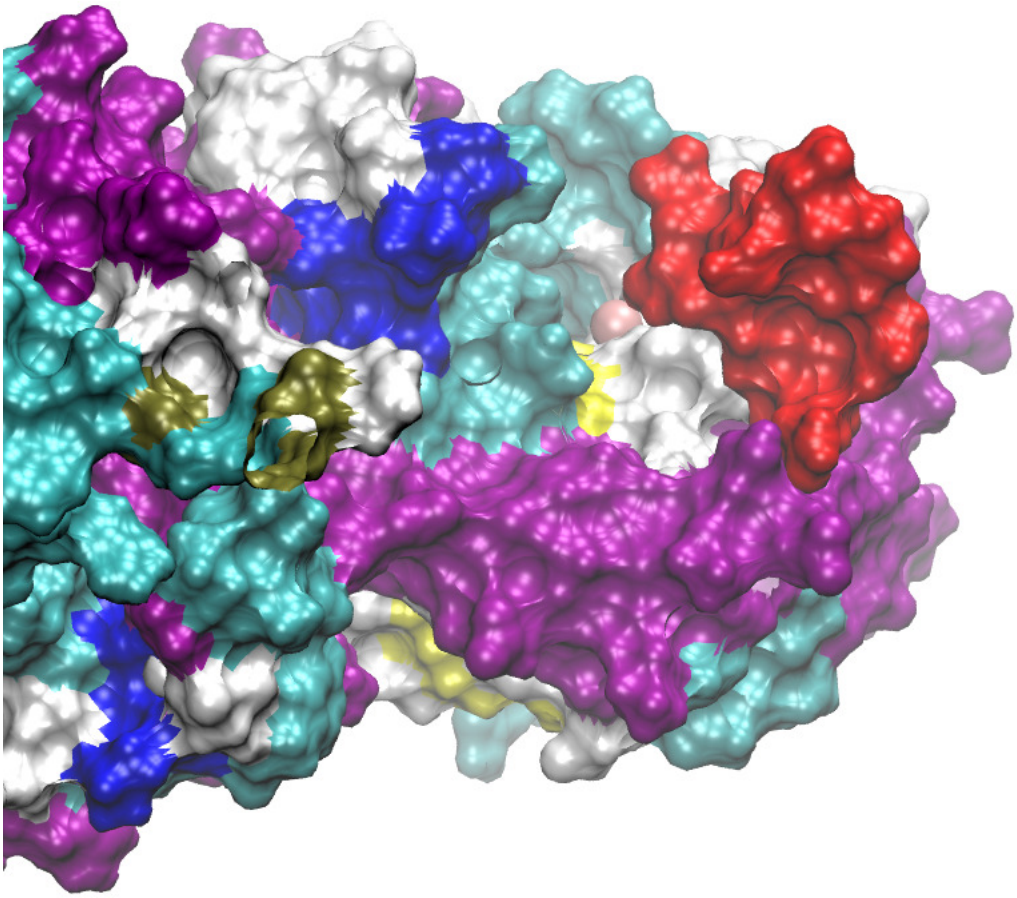
(a)

Figure 3.4: Flexible loop motions of pABA active site. (a) The top image shows the structure of the protein early in the simulation (step 3000). The flexible loop region (shown in red) is open and shows clear access to the DHPt-PP active site. The manganese metal ion is visible (light pink sphere) but the ligand has been hidden in these images for visual clarity. (b) The middle image is taken from the middle of the trajectory (step 6000). In this image the flexible loop region has folded itself over the opening to the active site. There is still a small pocket visible that could be the binding site for pABA. (c) By the end of the trajectory (step 8000), the flexible loop region has re-opened with full access to the DHPt-PP active site once again.



(b)

Figure 3.4: Continued

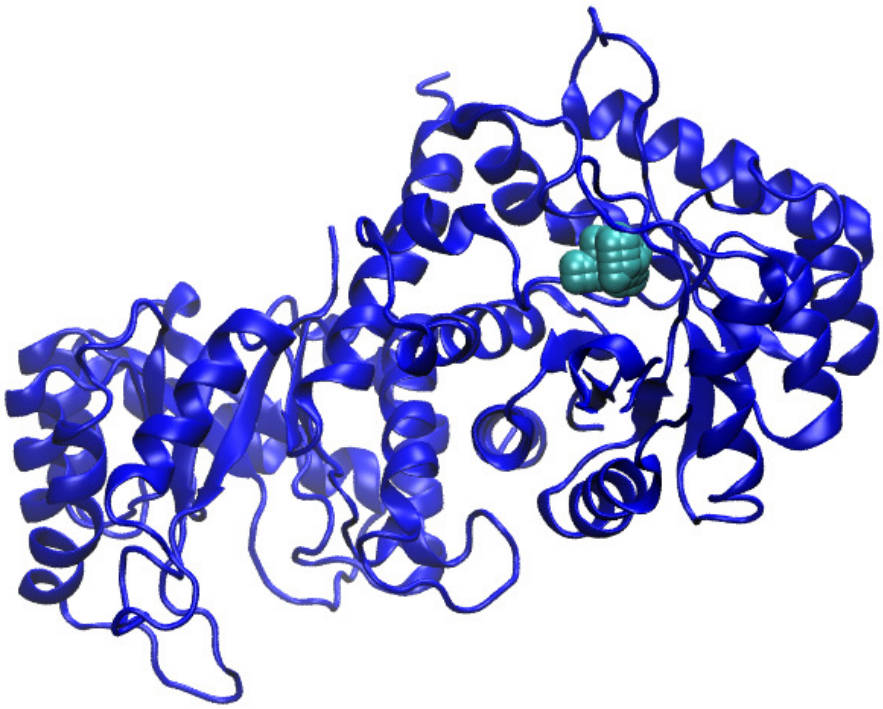


(c)

Figure 3.4: Continued

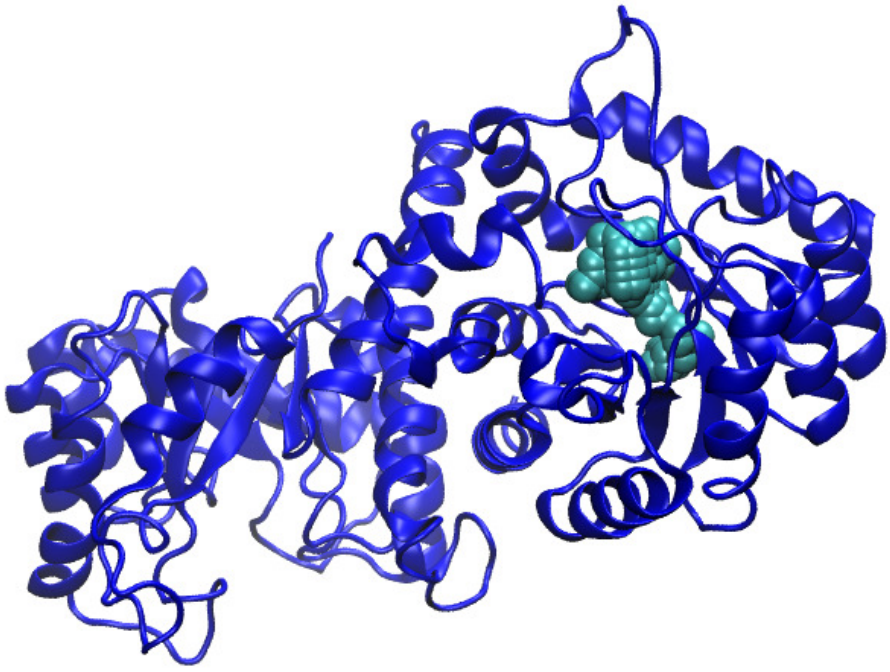
This motion has been proposed by previous studies and the movement of this loop may provide a hint at how the DHPS enzyme functions to bring the pABA substrate in contact with the DHPt-PP and keep it in position during the catalytic reaction.[14] The DHPt-PP binding pocket remains fairly consistent in size throughout most of the trajectory, but during the middle of the simulation, it can be seen that

the DHPt-PP pocket widens and deepens (Figure 3.5).



(a)

Figure 3.5: Change of active site shape. (a) The top image shows the DHPS protein (blue) with the active site filled by the grid points used by POVME to estimate pocket volume (teal). (b) The bottom image shows that the binding pocket widens and deepens during the middle of the trajectory.



(b)

Figure 3.5: Continued

This pocket opening occurs after the loop closing motion, so it is possible that these motions are correlated in the mechanism of this enzyme. The deepening of the DHPt-PP pocket is a fairly rare occurrence during the simulation, and none of these structures were located in the most populated clusters that were originally chosen for the docking studies. However, as these open conformations may be important for catalysis, they could provide additional results from docking studies that could be

relevant for further testing.[18] To choose additional structures, the POVME output was sorted by calculated pocket size, and the five largest pockets were chosen for use in further docking studies. With these new structures, an additional 44 structures from the NCI Diversity Set III were found that had not docked favorably in the originally chosen cluster structures. It is interesting to note that when all of the POVME and cluster docking results are combined and ranked by docking score, of the top 109 unique compounds (those found by using the cluster structures and POVME separately), only 29 of them are from the cluster structures. Thus it seems that the POVME structures provided a more favorable docking environment (most likely due to their more-open shape) than many of the cluster structures.

3.6 ZINC Database Docking

After the initial NCI Diversity Set III docking studies, the larger ZINC database was used to search for additional compounds. Because the ZINC database is much larger, the number of compounds was reduced at each phase of the docking process in order to complete the computations in a reasonable amount of time. This produced more unique results for each structure, as opposed to the NCI Diversity Set III, where most of the top scoring results were repeated for each tested structure. At the end of the docking, all of the ZINC results reported by the software were above the docking score cut-off used for the NCI Diversity Set III. The results from the cluster and POVME structures were combined and ranked by docking score. It was decided that the top 109 compounds would be reported (to be comparable in number to the amount of compounds chosen from the Diversity Set III) (See Appendix 2). Only 17 of these top scoring compounds came from docking results based on the cluster structures, with the majority coming from POVME structure docking results.

3.7 Conclusion

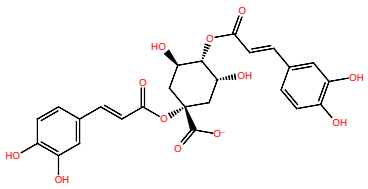
DHPS is a desirable drug target in the folate production pathway. Mutations have made current strains resistant to a majority of the treatments available currently. Due to the fact that the DHPt-PP binding site is highly conserved across species, finding a suitable inhibitor could be incredibly beneficial for treating a wide range of diseases. This study made use of a variety of computational tools and techniques to search for potential drug compounds that will act as competitive inhibitors in the DHPt-PP active site. Molecular dynamics was used to generate a trajectory of the DHPS enzyme with analog ligand DHP bound in the DHPt-PP active site. Using the Relaxed Complex Scheme to test several conformations in docking studies provided a set of testable compounds for further study. Utilizing the visualization tools of POVME, a widening and deepening of the DHPt-PP active site was seen during the middle of the molecular dynamics trajectory. It was also possible to see how the movement of the flexible loop region could act to bring the pABA substrate to the DHPt-PP substrate for catalysis. Based on the POVME results, structures of interest were chosen for further docking studies and additional compounds were chosen for future testing. The results from the docking studies using the POVME structures indicate that using additional tools to study the molecular dynamics trajectory and choose structures of interest can be helpful in producing additional docking results that would have been missed had only the more traditional cluster-based methods been used. Further studies will be needed to test the selected compounds and see if any of them have the potential to act as competitive inhibitors and develop those compounds into possible drug candidates. Further computational work should be done to compare the dynamics of the apo DHPS enzyme to the holo DHPS enzyme to see if the dynamics of the DHPt-PP active site or if the dynamics of the flexible loop pABA binding site change without a ligand present.

3.8 Acknowledgements

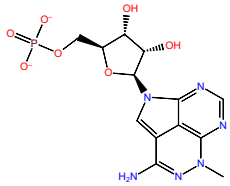
The work at UCSD was supported in part by NIH, NSF, HHMI, CTBP, NBCR, and the UCSD Chancellor's Interdisciplinary Collaboratories Fellowship. The author also wishes to acknowledge Levi Pierce, without whom the molecular dynamics simulations would have taken infinitely longer, and Bill Sinko for his help and suggestions about the docking software and procedures throughout the entirety of this project.

3.9 Appendix 1

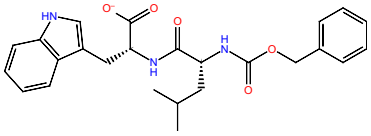
Table 3.1: Compound structures chosen from the NCI Diversity Set III based on cluster docking structures.



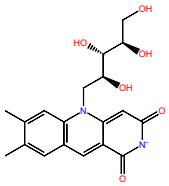
title: 91529
docking score: -14.786
QPlogPo/w: 0.471
QPlogKhsa: -0.796
RuleOfFive: 3



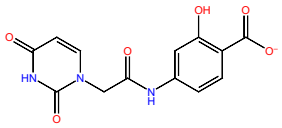
title: 280594
docking score: -13.187
QPlogPo/w: -1.147
QPlogKhsa: -1.378
RuleOfFive: 2



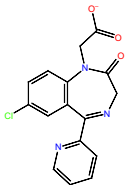
title: 335979
docking score: -12.329
QPlogPo/w: 5.083
QPlogKhsa: 0.409
RuleOfFive: 1



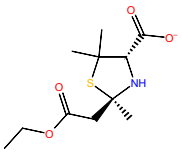
title: 275266
docking score: -12.066
QPlogPo/w: -0.192
QPlogKhsa: -0.616
RuleOfFive: 0



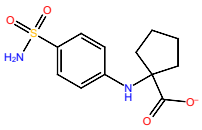
title: 93033
docking score: -11.748
QPlogPo/w: -0.023
QPlogKhsa: -0.692
RuleOfFive: 0



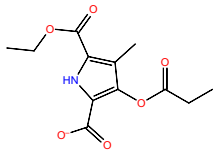
title: 157725
docking score: -11.639
QPlogPo/w: 2.243
QPlogKhsa: -0.392
RuleOfFive: 0



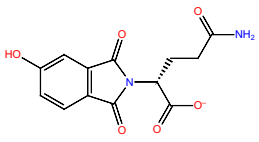
title: 117446
docking score: -11.607
QPlogPo/w: 0.085
QPlogKhsa: -0.203
RuleOfFive: 0



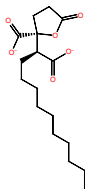
title: 105432
docking score: -11.553
QPlogPo/w: 0.472
QPlogKhsa: -0.738
RuleOfFive: 0



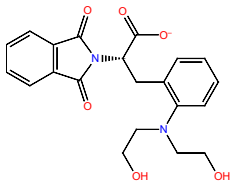
title: 13434
docking score: -11.446
QPlogPo/w: 1.45
QPlogKhsa: -0.521
RuleOfFive: 0



title: 78623
docking score: -11.396
QPlogPo/w: -0.868
QPlogKhsa: -1.134
RuleOfFive: 0

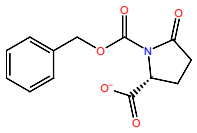


title: 244387
docking score: -11.361
QPlogPo/w: 2.643
QPlogKhsa: -0.634
RuleOfFive: 0

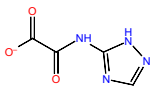


title: 76350
docking score: -11.352
QPlogPo/w: 2.033
QPlogKhsa: -0.555
RuleOfFive: 0

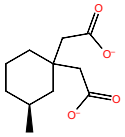
Table 3.1: Continued



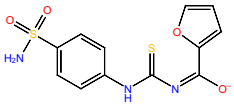
title: 156957
docking score: -11.267
QPlogPo/w: 1.999
QPlogKhsa: -0.313
RuleOfFive: 0



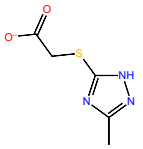
title: 177952
docking score: -11.227
QPlogPo/w: -0.897
QPlogKhsa: -1.067
RuleOfFive: 0



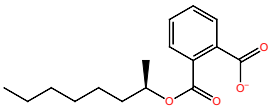
title: 3001
docking score: -10.358
QPlogPo/w: 1.949
QPlogKhsa: -0.639
RuleOfFive: 0



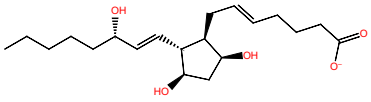
title: 205912
docking score: -10.333
QPlogPo/w: 0.391
QPlogKhsa: -0.637
RuleOfFive: 0



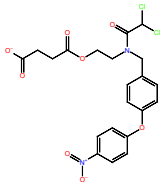
title: 78999
docking score: -10.202
QPlogPo/w: 0.414
QPlogKhsa: -0.851
RuleOfFive: 0



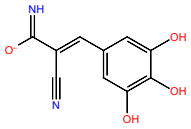
title: 524615
docking score: -11.079
QPlogPo/w: 3.897
QPlogKhsa: 0.119
RuleOfFive: 0



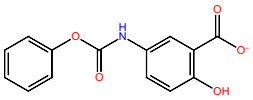
title: 196515
docking score: -10.726
QPlogPo/w: 3.167
QPlogKhsa: -0.287
RuleOfFive: 0



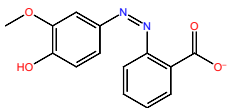
title: 107522
docking score: -10.699
QPlogPo/w: 3.441
QPlogKhsa: -0.188
RuleOfFive: 0



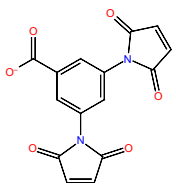
title: 638080
docking score: -10.673
QPlogPo/w: -0.882
QPlogKhsa: -0.885
RuleOfFive: 0



title: 68971
docking score: -10.671
QPlogPo/w: 2.127
QPlogKhsa: -0.333
RuleOfFive: 0

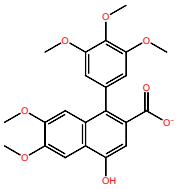


title: 153172
docking score: -10.262
QPlogPo/w: 1.778
QPlogKhsa: -0.47
RuleOfFive: 0

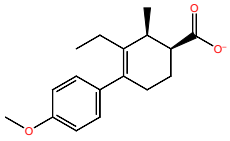


title: 44750
docking score: -10.218
QPlogPo/w: 0.262
QPlogKhsa: -0.649
RuleOfFive: 0

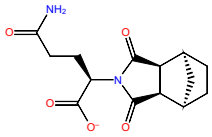
Table 3.1: Continued



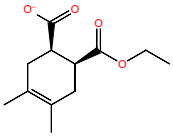
title: 41148
docking score: -10.163
QPlogPo/w: 3.733
QPlogKhsa: 0.102
RuleOfFive: 0



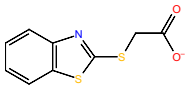
title: 19962
docking score: -10.071
QPlogPo/w: 4.084
QPlogKhsa: 0.329
RuleOfFive: 0



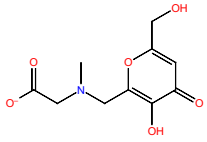
title: 241998
docking score: -10.781
QPlogPo/w: -0.192
QPlogKhsa: -0.968
RuleOfFive: 0



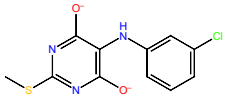
title: 144982
docking score: -10.687
QPlogPo/w: 2.303
QPlogKhsa: -0.311
RuleOfFive: 0



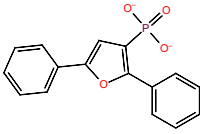
title: 11891
docking score: -10.508
QPlogPo/w: 2.457
QPlogKhsa: -0.494
RuleOfFive: 0



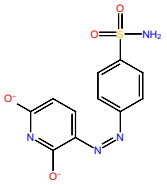
title: 114831
docking score: -10.462
QPlogPo/w: -2.982
QPlogKhsa: -1.098
RuleOfFive: 0



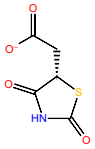
title: 59430
docking score: -10.753
QPlogPo/w: 2.746
QPlogKhsa: -0.032
RuleOfFive: 0



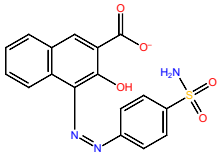
title: 43344
docking score: -10.727
QPlogPo/w: 3.313
QPlogKhsa: -0.385
RuleOfFive: 0



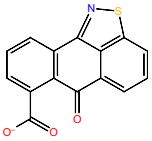
title: 134199
docking score: -10.62
QPlogPo/w: -0.627
QPlogKhsa: -0.644
RuleOfFive: 0



title: 60548
docking score: -10.361
QPlogPo/w: -0.232
QPlogKhsa: -0.929
RuleOfFive: 0

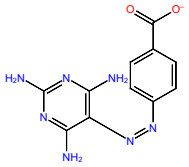


title: 134137
docking score: -10.329
QPlogPo/w: 0.883
QPlogKhsa: -0.497
RuleOfFive: 0

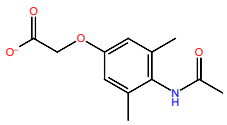


title: 39938
docking score: -10.221
QPlogPo/w: 2.204
QPlogKhsa: -0.349
RuleOfFive: 0

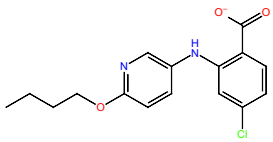
Table 3.1: Continued



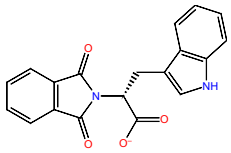
title: 408734
docking score: -10.218
QPlogPo/w: -1.101
QPlogKhsa: -0.962
RuleOfFive: 1



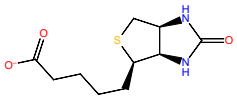
title: 38042
docking score: -10.199
QPlogPo/w: 1.68
QPlogKhsa: -0.485
RuleOfFive: 0



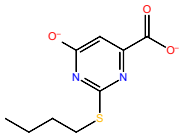
title: 60423
docking score: -11.051
QPlogPo/w: 4.778
QPlogKhsa: 0.398
RuleOfFive: 0



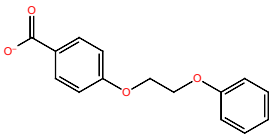
title: 401077
docking score: -11.009
QPlogPo/w: 3.13
QPlogKhsa: 0.019
RuleOfFive: 0



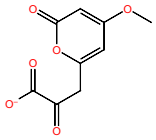
title: 63865
docking score: -10.978
QPlogPo/w: 1.157
QPlogKhsa: -0.58
RuleOfFive: 0



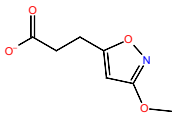
title: 42014
docking score: -10.798
QPlogPo/w: 1.677
QPlogKhsa: -0.516
RuleOfFive: 0



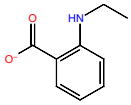
title: 120631
docking score: -10.605
QPlogPo/w: 3.292
QPlogKhsa: -0.118
RuleOfFive: 0



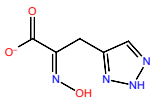
title: 174027
docking score: -10.226
QPlogPo/w: -0.344
QPlogKhsa: -1.153
RuleOfFive: 0



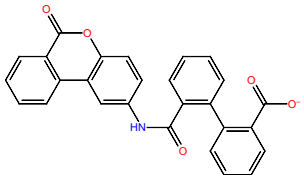
title: 303800
docking score: -10.131
QPlogPo/w: 0.975
QPlogKhsa: -0.712
RuleOfFive: 0



title: 16162
docking score: -10.045
QPlogPo/w: 1.993
QPlogKhsa: -0.409
RuleOfFive: 0

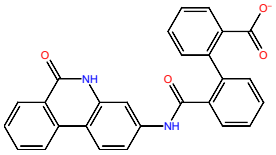


title: 117386
docking score: -10.014
QPlogPo/w: -0.951
QPlogKhsa: -1.124
RuleOfFive: 0

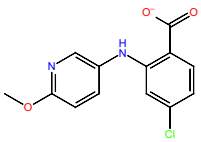


title: 128606
docking score: -10.405
QPlogPo/w: 4.328
QPlogKhsa: 0.386
RuleOfFive: 0

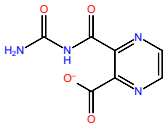
Table 3.1: Continued



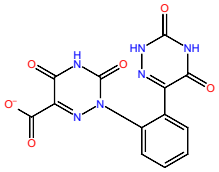
title: 127133
docking score: -10.135
QPlogPo/w: 4.218
QPlogKhsa: 0.364
RuleOfFive: 0



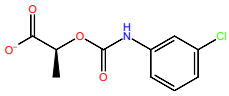
title: 60303
docking score: -10.915
QPlogPo/w: 4.093
QPlogKhsa: 0.098
RuleOfFive: 0



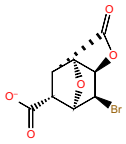
title: 117922
docking score: -10.867
QPlogPo/w: -1.417
QPlogKhsa: -1.188
RuleOfFive: 0



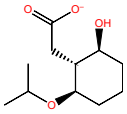
title: 362639
docking score: -10.658
QPlogPo/w: 0.63
QPlogKhsa: -0.375
RuleOfFive: 1



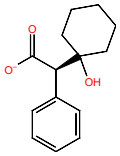
title: 73170
docking score: -10.614
QPlogPo/w: 1.887
QPlogKhsa: -0.456
RuleOfFive: 0



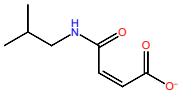
title: 6137
docking score: -10.458
QPlogPo/w: 0.0
QPlogKhsa: -1.066
RuleOfFive: 0



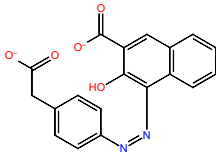
title: 25368
docking score: -10.432
QPlogPo/w: 1.375
QPlogKhsa: -0.651
RuleOfFive: 0



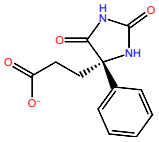
title: 16416
docking score: -10.299
QPlogPo/w: 3.491
QPlogKhsa: 0.132
RuleOfFive: 0



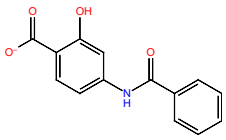
title: 55573
docking score: -10.103
QPlogPo/w: 1.012
QPlogKhsa: -0.767
RuleOfFive: 0



title: 156563
docking score: -10.063
QPlogPo/w: 2.829
QPlogKhsa: -0.269
RuleOfFive: 0

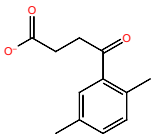


title: 38743
docking score: -11.163
QPlogPo/w: 0.849
QPlogKhsa: -0.592
RuleOfFive: 0

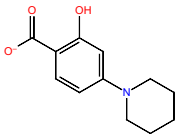


title: 159686
docking score: -11.01
QPlogPo/w: 2.39
QPlogKhsa: -0.311
RuleOfFive: 0

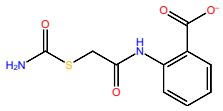
Table 3.1: Continued



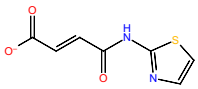
title: 215276
docking score: -10.789
QPlogPo/w: 2.013
QPlogKhsa: -0.403
RuleOfFive: 0



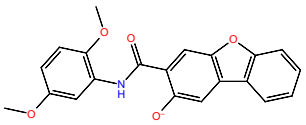
title: 23895
docking score: -10.7
QPlogPo/w: 2.474
QPlogKhsa: -0.095
RuleOfFive: 0



title: 13345
docking score: -10.651
QPlogPo/w: 0.561
QPlogKhsa: -0.72
RuleOfFive: 0

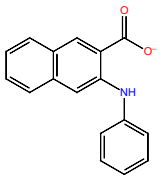


title: 80807
docking score: -10.343
QPlogPo/w: 0.277
QPlogKhsa: -0.95
RuleOfFive: 0

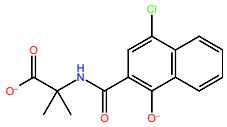


title: 50650
docking score: -10.267
QPlogPo/w: 4.206
QPlogKhsa: 0.541
RuleOfFive: 0

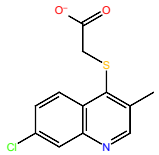
Table 3.2: Additional compound structures from the NCI Diversity Set III chosen based on docking results with the POVME structures.



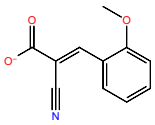
title: 40275
docking score: -11.565
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



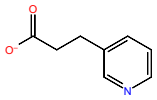
title: 270063
docking score: -11.559
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



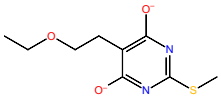
title: 13579
docking score: -11.552
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



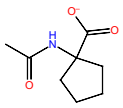
title: 638134
docking score: -11.419
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



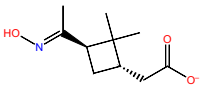
title: 89720
docking score: -11.347
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



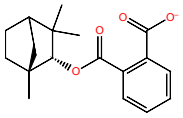
title: 66837
docking score: -11.345
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



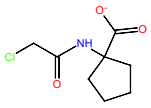
title: 22847
docking score: -11.277
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



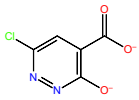
title: 127947
docking score: -11.11
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



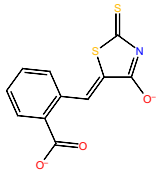
title: 26349
docking score: -11.046
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



title: 240502
docking score: -11.024
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

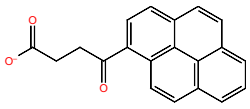


title: 338205
docking score: -10.952
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

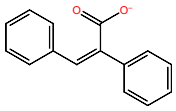


title: 403268
docking score: -10.938
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

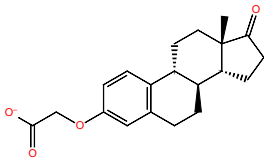
Table 3.2: Continued



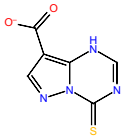
title: 407628
docking score: -10.886
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



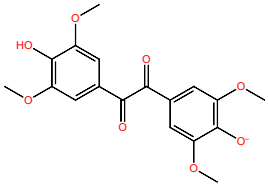
title: 40614
docking score: -10.879
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



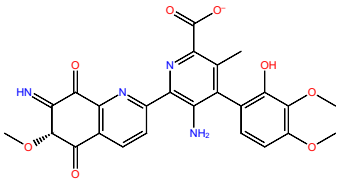
title: 86467
docking score: -10.869
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



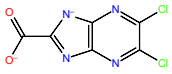
title: 173103
docking score: -10.86
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



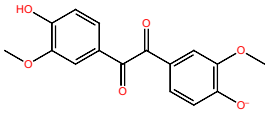
title: 16736
docking score: -10.828
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



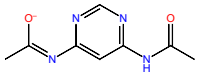
title: 83950
docking score: -10.828
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



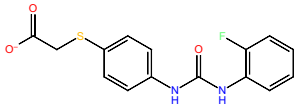
title: 182400
docking score: -10.827
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



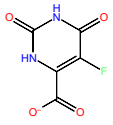
title: 16722
docking score: -10.759
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



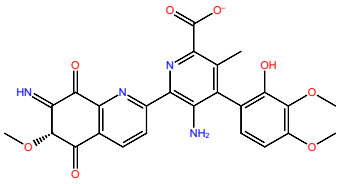
title: 57890
docking score: -10.74
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



title: 142277
docking score: -10.73
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

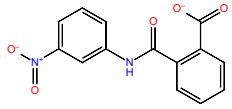


title: 31712
docking score: -10.722
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

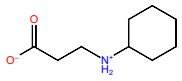


title: 45383
docking score: -10.721
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

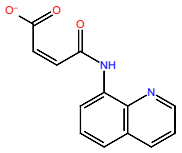
Table 3.2: Continued



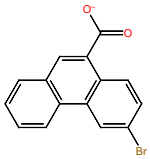
title: 408860
docking score: -10.631
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



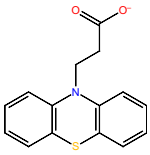
title: 62840
docking score: -10.611
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



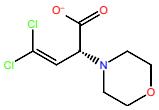
title: 144958
docking score: -10.596
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



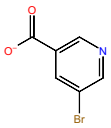
title: 151252
docking score: -10.581
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



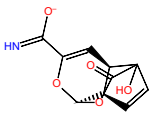
title: 525721
docking score: -10.563
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



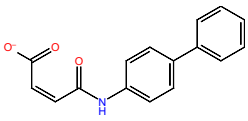
title: 53874
docking score: -10.549
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



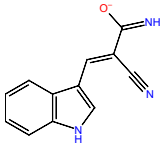
title: 9461
docking score: -10.524
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



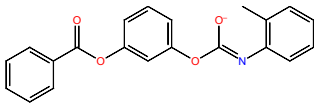
title: 357683
docking score: -10.513
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



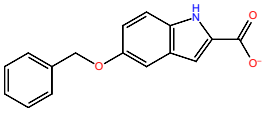
title: 68841
docking score: -10.476
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



title: 163144
docking score: -10.46
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

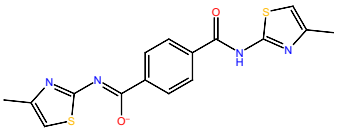


title: 191441
docking score: -10.408
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

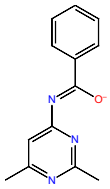


title: 30930
docking score: -10.361
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

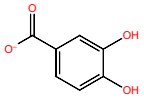
Table 3.2: Continued



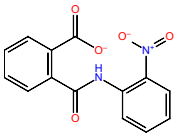
title: 110300
docking score: -10.331
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



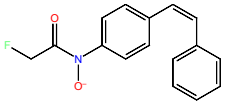
title: 23248
docking score: -10.288
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



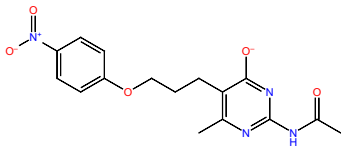
title: 16631
docking score: -10.269
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



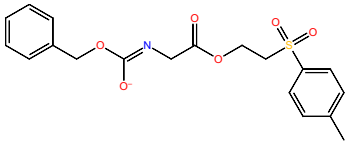
title: 41066
docking score: -10.224
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



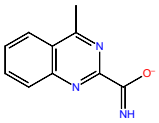
title: 112965
docking score: -10.219
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



title: 211356
docking score: -10.14
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



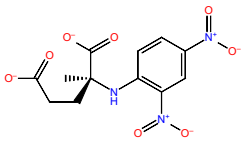
title: 126837
docking score: -10.106
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



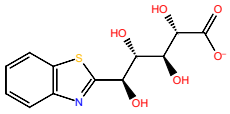
title: 13213
docking score: -10.098
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

3.10 Appendix 2

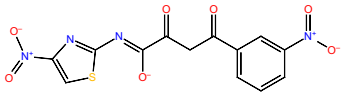
Table 3.3: Compound structures chosen from the ZINC docking library.



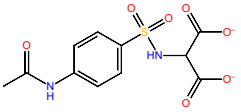
title: ZINC 08627245
docking score: -14.212
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



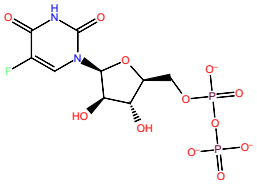
title: ZINC 03954014
docking score: -14.069
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



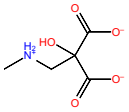
title: ZINC 05315348
docking score: -13.844
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



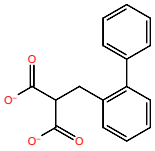
title: ZINC 01705179
docking score: -13.632
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



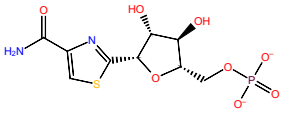
title: ZINC 17004707
docking score: -13.548
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



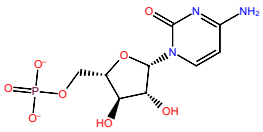
title: ZINC 01626889
docking score: -13.545
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



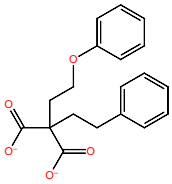
title: ZINC 01615121
docking score: -13.542
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



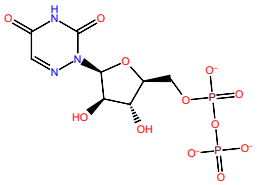
title: ZINC 16939760
docking score: -13.54
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



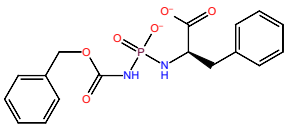
title: ZINC 03954230
docking score: -13.523
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



title: ZINC 01734085
docking score: -13.451
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

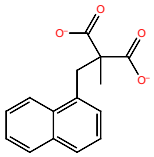


title: ZINC 16991540
docking score: -13.389
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

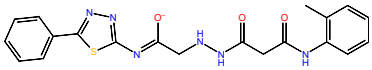


title: ZINC 01585252
docking score: -13.385
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

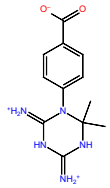
Table 3.3: Continued



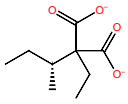
title: ZINC 01577308
docking score: -13.378
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



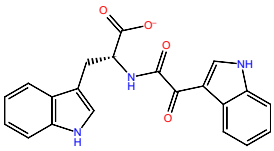
title: ZINC 01628506
docking score: -13.375
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



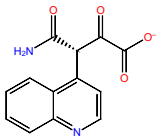
title: ZINC 08579847
docking score: -13.37
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



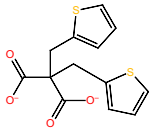
title: ZINC 01641611
docking score: -13.29
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



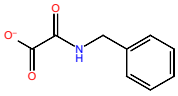
title: ZINC 01610321
docking score: -13.289
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



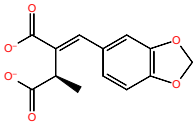
title: ZINC 13282956
docking score: -13.221
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



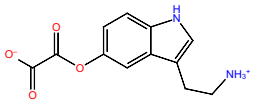
title: ZINC 01593046
docking score: -13.2
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



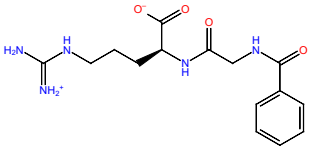
title: ZINC 01682608
docking score: -13.166
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



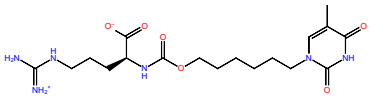
title: ZINC 01618671
docking score: -13.151
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



title: ZINC 01705197
docking score: -13.097
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

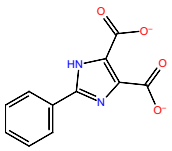


title: ZINC 02516109
docking score: -13.072
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

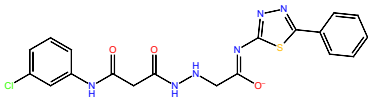


title: ZINC 05845654
docking score: -13.063
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

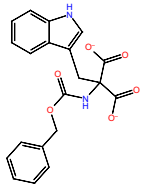
Table 3.3: Continued



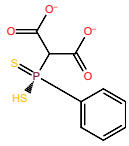
title: ZINC 00189552
docking score: -13.05
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



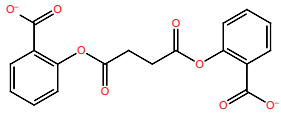
title: ZINC 01628507
docking score: -13.046
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



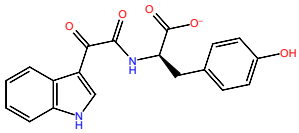
title: ZINC 01570969
docking score: -13.005
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



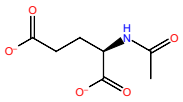
title: ZINC 01623800
docking score: -12.959
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



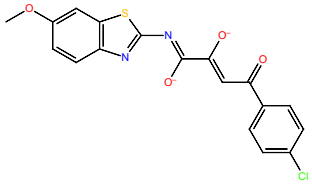
title: ZINC 04978944
docking score: -12.87
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



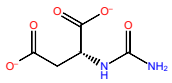
title: ZINC 01610320
docking score: -12.864
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



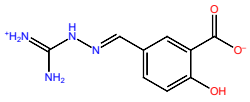
title: ZINC 01764662
docking score: -12.856
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



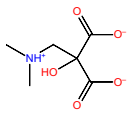
title: ZINC 05503279
docking score: -12.838
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



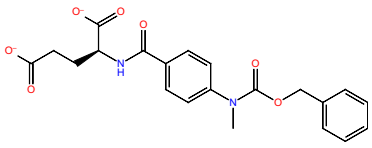
title: ZINC 01655440
docking score: -12.825
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



title: ZINC 01698576
docking score: -12.817
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

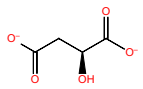


title: ZINC 01626890
docking score: -12.814
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

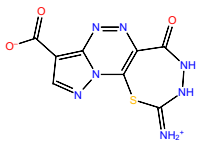


title: ZINC 05599799
docking score: -12.805
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

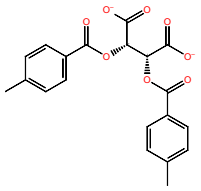
Table 3.3: Continued



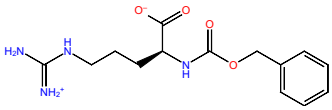
title: ZINC 00895074
docking score: -12.782
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



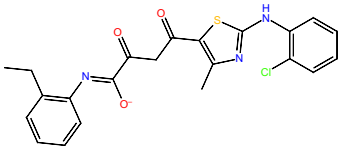
title: ZINC 16958136
docking score: -12.747
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



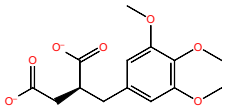
title: ZINC 01592488
docking score: -12.71
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



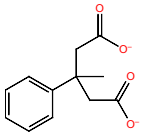
title: ZINC 01540612
docking score: -12.71
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



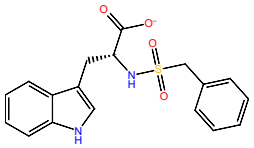
title: ZINC 05315859
docking score: -12.708
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



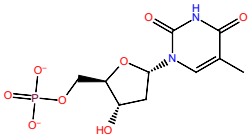
title: ZINC 02046072
docking score: -12.687
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



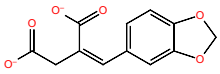
title: ZINC 01682368
docking score: -12.681
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



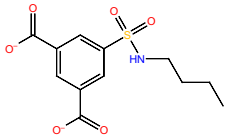
title: ZINC 06529645
docking score: -12.677
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



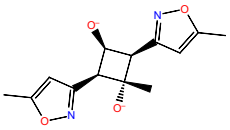
title: ZINC 06523446
docking score: -12.669
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



title: ZINC 02046075
docking score: -12.661
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

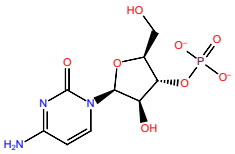


title: ZINC 01760558
docking score: -12.607
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

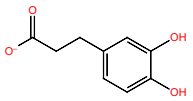


title: ZINC 05051990
docking score: -12.6
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

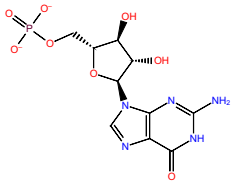
Table 3.3: Continued



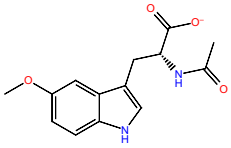
title: ZINC 16941352
docking score: -12.598
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



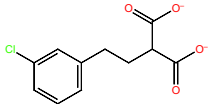
title: ZINC 00388044
docking score: -12.559
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



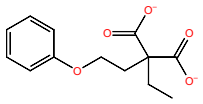
title: ZINC 12501415
docking score: -12.555
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



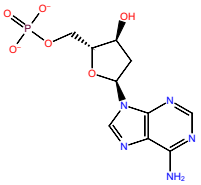
title: ZINC 01580371
docking score: -12.553
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



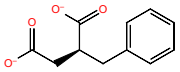
title: ZINC 01600982
docking score: -12.552
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



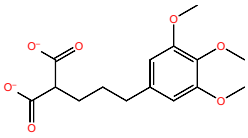
title: ZINC 01734078
docking score: -12.522
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



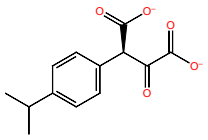
title: ZINC 13538982
docking score: -12.52
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



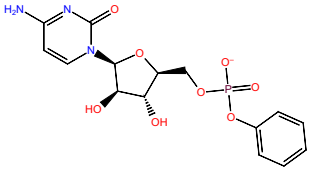
title: ZINC 00291740
docking score: -12.517
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



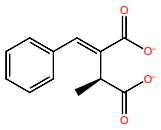
title: ZINC 01594178
docking score: -12.498
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



title: ZINC 03954164
docking score: -12.497
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

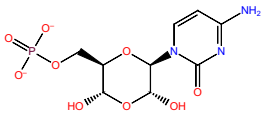


title: ZINC 17027397
docking score: -12.496
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

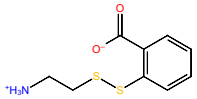


title: ZINC 01706044
docking score: -12.48
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

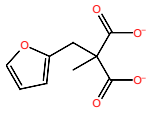
Table 3.3: Continued



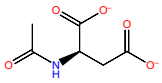
title: ZINC 16953048
docking score: -12.454
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



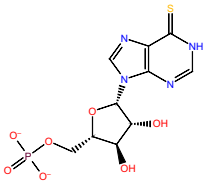
title: ZINC 01579484
docking score: -12.449
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



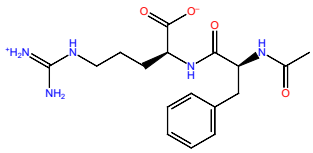
title: ZINC 01695060
docking score: -12.449
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



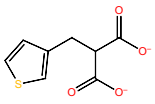
title: ZINC 00262217
docking score: -12.448
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



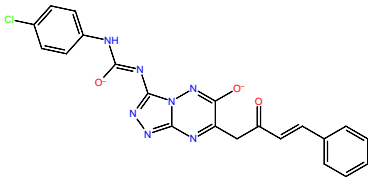
title: ZINC 16990100
docking score: -12.435
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



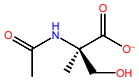
title: ZINC 05707434
docking score: -12.424
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



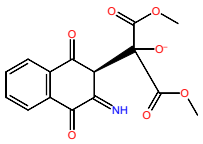
title: ZINC 01701425
docking score: -12.406
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



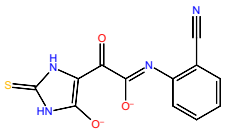
title: ZINC 05399696
docking score: -12.4
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



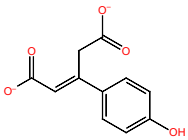
title: ZINC 01599667
docking score: -12.394
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



title: ZINC 04621907
docking score: -12.366
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

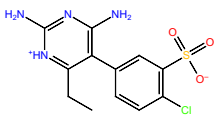


title: ZINC 01629832
docking score: -12.354
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

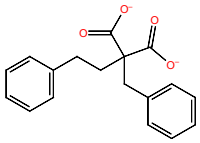


title: ZINC 01531867
docking score: -12.353
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

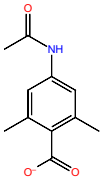
Table 3.3: Continued



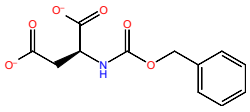
title: ZINC 06244515
docking score: -12.347
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



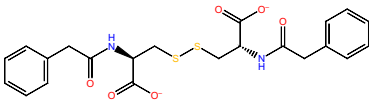
title: ZINC 01718236
docking score: -12.339
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



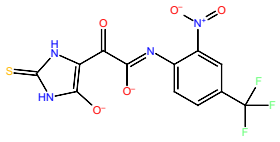
title: ZINC 01758502
docking score: -12.329
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



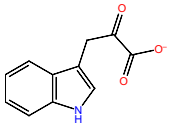
title: ZINC 01700285
docking score: -12.318
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



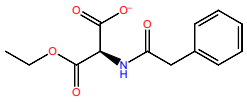
title: ZINC 16889888
docking score: -12.307
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



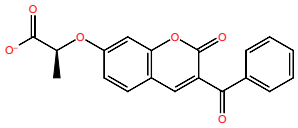
title: ZINC 08653423
docking score: -12.299
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



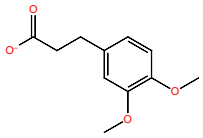
title: ZINC 01532617
docking score: -12.293
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



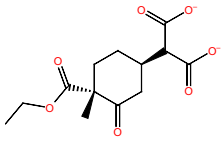
title: ZINC 03633183
docking score: -12.268
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



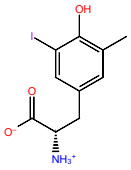
title: ZINC 01620110
docking score: -12.258
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



title: ZINC 00106688
docking score: -12.244
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

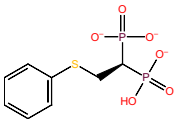


title: ZINC 01620201
docking score: -12.238
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

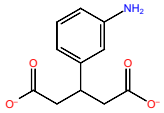


title: ZINC 01713868
docking score: -12.22
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

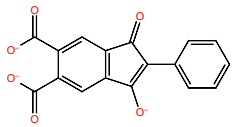
Table 3.3: Continued



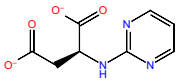
title: ZINC 01618740
docking score: -12.218
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



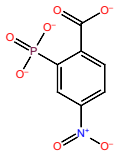
title: ZINC 01857386
docking score: -12.217
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



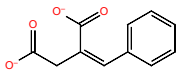
title: ZINC 04877185
docking score: -12.217
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



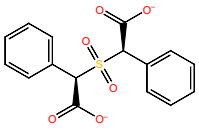
title: ZINC 17020900
docking score: -12.202
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



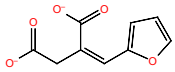
title: ZINC 01870296
docking score: -12.198
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



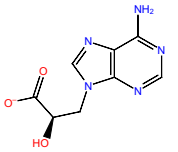
title: ZINC 00337086
docking score: -12.196
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



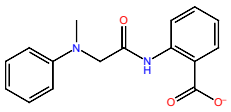
title: ZINC 01702011
docking score: -12.194
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



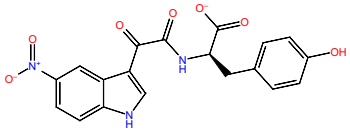
title: ZINC 02044053
docking score: -12.186
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



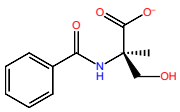
title: ZINC 01618076
docking score: -12.174
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



title: ZINC 01647196
docking score: -12.172
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

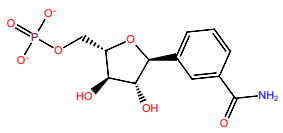


title: ZINC 01610325
docking score: -12.168
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

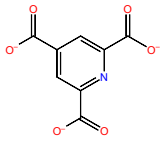


title: ZINC 01599636
docking score: -12.159
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

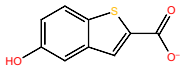
Table 3.3: Continued



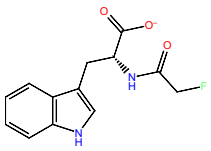
title: ZINC 16957949
docking score: -12.146
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



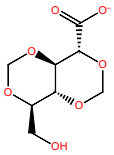
title: ZINC 01595538
docking score: -12.139
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



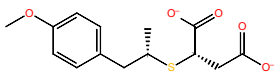
title: ZINC 01701101
docking score: -12.137
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



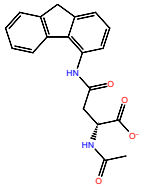
title: ZINC 02046623
docking score: -12.13
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



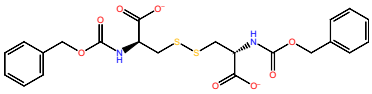
title: ZINC 04403777
docking score: -12.121
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



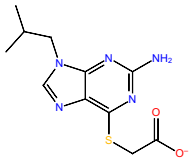
title: ZINC 01640990
docking score: -12.114
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



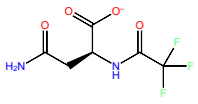
title: ZINC 05431458
docking score: -12.109
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



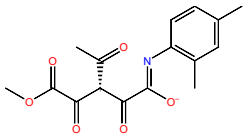
title: ZINC 04273474
docking score: -12.105
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



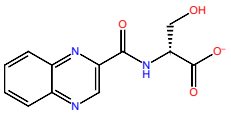
title: ZINC 01685086
docking score: -12.105
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



title: ZINC 05502126
docking score: -12.099
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

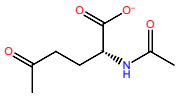


title: ZINC 05503982
docking score: -12.097
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None



title: ZINC 01626850
docking score: -12.092
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

Table 3.3: Continued



title: ZINC 01868377
docking score: -12.09
QPlogPo/w: None
QPlogKhsa: None
RuleOfFive: None

3.11 References

[1] Hampele, I.C.; D'Arcy, A.; Dale, G.E.; Kostrewa, D.; Nielsen, J.; Oefner, C.; Page, M.G.; Schönfeld, H.J.; Stüber, D.; Then, R.L. Structure and Function of the Dihydropteroate Synthase from *Staphylococcus aureus. J. Mol. Biol.* **1997**, *268*, 21-30.

[2] Dallas, W.S.; Gowen, J.E.; Ray, P.H.; Cox, M.J.; Dev, I.K. Cloning, Sequencing, and Enhanced Expression of the Dihydropteroate Synthase of *Escherichia coli* MC4100. *J. Bacteriology* **1992**, *174*, 5961-5970.

[3] Bermingham, A.; Derrick, J.P. The folic acid biosynthesis pathway in bacteria: evaluation of potential for antibacterial drug discovery. *BioEssays* **2002**, *24*, 637-648.

[4] Roland, S.; Ferone, R.; Harvey, R.J.; Styles, V.L.; Morrison, R.W. The Characteristics and Significance of Sulfonamides as Substrates for *Escherichia coli* Dihydropteroate Synthase. *J. Biol. Chem.* **1979**, *254*, 10337-10345.

[5] Sköld, O. Sulfonamide resistance: mechanisms and trends. *Drug Resistance Updates* **2000**, *3*, 155-160.

[6] Swedberg, G.; Castensson, S.; Sköld, O. Characterization of Mutationally Altered Dihydropteroate Synthase and Its Ability to Form a Sulfonamide-Containing Dihydrofolate Analog. *J. Bacteriology* **1979**, *137*, 129-136.

[7] Wise, E.M. Jr.; Abou-Donia, M.M. Sulfonamide resistance mechanism in *Escherichia coli*: R plasmids can determine sulfonamide-resistant dihydropteroate synthase. *Proc. Nat. Acad. Sci. USA* **1975**, *72*, 2621-2625.

[8] Giordanetto, F.; Fowler, P.W.; Saqi, M.; Coveney, P.V. Large scale molecular dynamics simulation of native and mutant dihydropteroate synthase sulphanilamide complexes suggest the molecular basis for dihydropteroate synthase drug resistance. *Phil. Trans. R. Soc. A* **2005**, *363*, 2055-2073.

[9] Namba, K.; Zheng, X.; Motoshima, K.; Kobayashi, H.; Tai, A.; Takahashi, E.; Sasaki, K.; Okamoto, K.; Kakuta, H. Design and synthesis of benzenesulfonanilides active against methicillin-resistant *Staphyloccocus aureus* and vancomycin-resistant *Enterococcus*. *Bioorg. Med. Chem.* **2008**, *16*, 6131-6144.

[10] Hevener, K.E.; Yun, K.; Qi, J.; Kerr, I.D.; Babaoglu, K.; Hurdle, J.G.; Balakrishna, K.; White, S.W.; Lee, R.E. Structural Studies of Pterin-Based Inhibitors of Dihydropteroate Synthase. *J. Med. Chem.* **2010**, *53*, 166-177.

[11] Vinnicombe, H.G.; Derrick, J.P. Dihydropteroate Synthase from *Streptococcus pneumoniae*: Characterization of Substrate Binding Order and Sulfonamide Inhibition. *Biochem. Biophys. Research Comm.* **1999**, *258*, 752-757.

[12] Al-Masaudi, S.B.; Day, M.J.; Russell, A.D. Antimicrobial resistance and gene transfer in *Staphylococcus aureus. J. Applied Bacteriology* **1991**, *70*, 279-290.

[13] Achari, A.; Somers, D.O.; Champness, J.N.; Bryant, P.K.; Rosemond, J.; Stammers, D.K. Crystal structure of the anti-bacterial sulfonamide drug target dihydropteroate synthase. *Nature Struc. Biol.* **1997**, *4*, 490-497.

[14] Levy, C.; Minnis, D.; Derrick, J.P. Dihydropteroate synthase from *Streptococcus pneumoniae*: structure, ligand recognition and mechanism of sulfonamide resistance. *Biochem. J.* **2008**, *412*, 379-388.

[15] Hanwell, M.D.; Curtis, D.E.; Lonie, D.C.; Vandermeersch, T.; Zurek, E.; Hutchison, G.R. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminformatics* **2012**, *4*, 17.

[16] Amaro, R. E.; Li, W.W. Emerging Methods for Ensemble-Based Virtual Screening. *Current Topics in Med. Chem.* **2010**, *10*, 3-13.

[17] Durrant, J.D.; de Oliveira, C.A.; McCammon, J.A. POVME: An algorithm for measuring binding-pocket volumes. *J. Mol. Graph. Model.* **2011**, *29*, 773-776.

[18] Robertson, J.G. Mechanistic Basis of Enzyme-Targeted Drugs. *Biochemistry* **2005**, *44*, 5561-5571.