Crystallographic Studies on mutants of *Trichomonas vaginalis* Purine Nucleoside Phosphorylase

Honor Thesis

Presented to the College of Agriculture and Life Sciences

Biological Sciences Program of Cornell University

in Partial Fulfillment of the Requirements for the

Research Honors Program

by

Siyao Xing

May 2008

Steven Ealick

INTRODUCTION

Trichomonas vaginalis is a protozoan human parasite that causes trichomoniasis, one of the most common sexually transmitted diseases in the world. Metronidazole is used to treat the disease, but due to its toxicity, especially for pregnant women (1), effort has been put into finding a new chemotherapeutic agent against the parasite. The purine salvage pathway is well known for its importance in organisms, and key enzymes in the protozoan pathway exhibit important differences from those in humans (2). Purine nucleoside phosphorylase (PNP) is an important enzyme in the purine salvage pathway, and there are two major categories of PNP: a low-molecular-mass homotrimeric form and a high-molecular-mass homohexameric form (3). Human PNP belongs to the first class, which recognizes inosine and guanosine as the substrates, while TvPNP is of the other class, which additionally recognizes adenosine (15). The differences in quaternary structure and substrate preference are the basis for the hypothesis that inhibitors could be identified that affect TvPNP but not the human pathway, as a new class of chemotherapeutic agents against the *T. vaginalis* parasite. A previous study of TvPNP identified that 2-fluoro-2'-deoxyadenosine (F-dAdo) is a subversive substrate for *T. vaginalis* (4), and serves as a prodrug that produces the highly cytotoxic 2-fluoro-adenine (F-Ade) when cleaved. Since F-dAdo is non-toxic and not a substrate for human PNP (5), it may represent a new therapeutic strategy for treating trichomoniasis.

TvPNP catalyzes a reversible reaction (Fig. 1) in which the glycosidic bond of a purine

nucleoside or deoxynucleoside is cleaved after nucleophilic attack by a phosphate. A phosphate-ribose and a purine ring are released as the products. Phosphorolysis is favored *in vivo* (3) and the cleaved purine bases from humans are necessary for *T. vaginalis* to synthesize its own DNA and RNA. In order to determine the basis for substrate binding and selectivity, the native structure of TvPNP was determined by Zhang Y. *et al.* (4). Enzyme structures with several ligands were solved, including adenosine, inosine, FdAdo, and formycin A (FMA). These enzyme complexes revealed interactions contributing to substrate binding through a purine binding site and as well as a sugar binding site (4). Asp-204 was observed in the otherwise hydrophobic binding pocket of the purine ring. Interestingly, it is proposed to be protonated during catalysis and donates a hydrogen bond to the N-7 atom of the purine base to augment the buildup of negative charge during the reaction. However, this hydrogen bond was not observed in any of the previous crystal structures except for the FMA complex (4). Since the hypothesized interaction between the ligands and the residue 204 based on the reaction mechanism of TvPNP was not observed, our goal is to further understand the active site of TvPNP, particularly with respect to Asp-204 in terms of substrate specificity.

Furthermore, our collaborator pointed out some interesting behaviors of mutants of TvPNP, which included D204A, D204N and D204E. While the native protein prefers adenosine over inosine as its substrate, D204A and D204N mutants lost activity on adenosine but retained activity on inosine (Table I), and D204E retained some lower activity on both adenosine and inosine. In order to learn more information about the

binding environment, the nature of TvPNP substrate selectivity, and especially the role of Asp-204, the crystallization studies of the D204N TvPNP mutants with adenosine and inosine were undertaken. This information would be useful as a preface to identifying inhibitors that bind more effectively to TvPNP but not to human PNP, and can be employed as new chemotherapeutic drugs against trichomoniasis.

EXPERIMENTAL PROCECURES

Crystallization—Purified recombinant mutant proteins were produced with the previously described procedure and dialyzed into a buffer containing 10 mM HEPES pH 7.5, 3 mM $MgCl_2$, 5 mM β-mercaptoethanol and 30% glycerol (v/v) by our collaborators (6). The proteins were concentrated to 25 mg/ml by using a 10-kDa cutoff Microcon concentrator (Amicon) and then stored at -80 °C. Prior to crystallization, the protein was exchanged into 20 mM Tris pH 7.1 using Bio-Spin 6 columns, with 4 washes. Different protein concentrations were tried from 8 mg/ml to 25 mg/ml to set up the crystallization. Hampton I, Hampton II, Wizard I, and Wizard II sparse matrix screens (Hampton Research, Emerald Biostructures) were used to set up the initial crystallization experiments. Crystals were grown at 295 K using the hanging drop vapor diffusion method (400 μl reservoir, 2 μl drop that includes 1 μl sample and 1 μl reservoir solution). The most promising conditions were further refined to produce better crystals. The optimized crystallization condition contained 0.1 M acetate pH 4.3, 2.0 M ammonium sulfate, and 0.2

M NaCl and it was different from the native enzyme crystallization condition, which was 50% polyethylene glycol 400, 12% ethylene glycol, 200 mM $MgCl_2$, and 100 mM Tris-HCl pH 8.5-8.7 (4). The crystals reached a size of 100-150 μm in seven days. Before X-ray data collection, the crystals were transferred to a synthetic mother liquor solution comprising the crystallization growth conditions and one of the substrate analogs at saturating concentration. After 1 h soaking, the crystal complexes were dipped into a cryoprotection solution, which was the reservoir solution with supplemented 25% glycerol. Then the crystals were flash frozen in liquid nitrogen. The crystals belong to the space group R3, which differs from the space group of the native enzyme, $P4_132$ (4). Each asymmetric unit contains six protomers arranged as three homodimer pairs. Applying three-fold crystallographic symmetry to the three homodimers generates three disc-shaped hexamers. On the basis of similarity to the other high-molecular weight bacterial PNPs (4, 6, 16), as well as gel filtration performed on TvPNP (6), these hexamers are believed to be equivalent biological units. The Matthews coefficient for the crystals was approximately 2.4 $Å^3$/Da, which corresponds to a solvent content of 49%. The crystals diffracted to 1.80 Å resolution.

Data collection and processing—Data were collected at the NE-CAT 24-ID-C and ID-E beamlines of the Advanced Photon Source at Argonne National Laboratory (Argonne, IL) using a Quantum 315 detector (Area Detector Systems Corp.), with 1 second exposures per 1° oscillation step. The crystal to detector distance was 250 mm. The HKL 2000 program suite (7) was used for indexing, integration and scaling. Data

processing statistics are summarized in Table II.

Twinning detection and detwinning—The possibility that the crystals were twinned was identified during truncation of experimental diffraction intensities to amplitudes, and was confirmed using the program SFCheck as part of the CCP4 suite of programs (8) as well as CNS (9, 10). The partial twin fractions measured for the crystals are summarized in Table III. Data were detwinned in CNS as a preface to solving the structures.

Structure determination—Two protomers from the previous TvPNP structure (PDB code 1Z38) were used as the search model for molecular replacement in CNS employing the detwinned reflection data. A total of three dimers were placed in sequential translation searches. Iterative rounds of positional and b-factor refinement were performed using CNS, and the structural model was then manually adjusted in Coot (11), using composite omit electron density maps, and F_o-F_c difference maps. Ligands were built in manually based on clearly recognizable electron density. Sulfate ions were added to fill clear density in the active site at the position occupied by phosphate in the *E. coli* PNP structures (6). Water molecules were first picked automatically using CNS, followed by manual addition in Coot. Current refinement statistics are listed in Table IV.

Figure Preparation—Figures of structures were prepared using Pymol (12) and ChemDraw (13).

RESULTS AND DISCUSSION

Twinning Problem—Twinning is a phenomenon where a crystal comprises more than one lattice. There are several kinds of twinning according to the number of dimensions that are coincident between the twinned lattices (10). The TvPNP crystals reported here exhibited merohedral twinning, in which case the lattices coincide in three dimensions. The problem of twinning was not detected in the process of integration and scaling, except it was noted that the highly twinned data sets (i.e. >40% twin fraction) scaled almost equally well in R3 and R32 based on the chi^2 and linear R-factor values (data not shown). Visually, the diffraction patterns looked normal, but in fact the intensities were the convoluted sum of more than one reflection. As defined by Yeates (10), the twinning fraction, α, represents the fraction of molecules that adopt the alternative twinned orientation in the crystal. A non-twinned crystal has α=0%, while a perfect twin has α=50%. The various crystals reported here were estimated to have α=12% - 46%. Notably, crystals from the same drop have different twinning fraction, and therefore the twinning problem was not solely related to the crystallization condition. (Table III)

A dataset collected from a crystal soaked with adenosine, with a measured partial twinning fraction of 26%, was detwinned prior to solving the structure by molecular replacement. After detwinning, the partial twin fraction measured was less than 1%. Clearly recognizable composite omit electron density maps were generated for all three

dimers placed during molecular replacement. Subsequent rigid-body, b-factor and positional refinement brought the R factor and R_{free} values to below 25% each. A dataset collected from a crystal soaked with inosine exhibited a partial twinning fraction of 31% and was also solved using the same method, but not as successfully as that of adenosine. The R factor and R_{free} were 25.6% and 28.6% after three rounds of refinement (Table IV). While the electron density for the ligand was clearly visible (Fig. 5), the electron density for nearby side chains, notably that of the position 204, was not clear enough to make conclusions for the ligand-binding site interaction. Additionally, unexplained electron density was observed around the ligand.

Overall Structure—The overall structure of D204N TvPNP resembles the native structure very well with an RMSD of 0.93 Å calculated for an all-atom superposition of the D204N structure with adenosine and the native structure with the same ligand. This suggests that pH does not affect the global fold or quaternary structure of this enzyme since the structures reported here were from pH 4.5 and the previous native structures were from pH 8.5. The enzyme is a homohexamer, containing six active sites, which are located at the interfaces of each pair of 2-fold related monomers (4). Each protomer has 235 residues, consisting of eleven β strands and seven α helices with an overall αβα sandwich fold. The β sheets form a barrel as the enzyme core, flanked by three and four α helices at each side (Fig. 2, 3).

Active Site and Asn-204—All six protein chains and bound ligands present in the adenosine complex showed clear electron density from composite omit maps (Fig. 4).

There were no significant differences in the conformation of active-site amino acids or the bound adenosine ligands between the different chains. As described previously, the ligand binding site of TvPNP comprises two parts (Fig. 5, 6), a purine binding site and a sugar binding site (4). The purine binding site contains four hydrophobic residues Phe-159, Val-178, Met-180, Ile-206 and hydrophilic Thr-156 (4). The same environment was observed in the mutant complex structure presented here. Due to the interest in subversive substrates with C-2 substituents (e.g. F-dAdo), the distance between C-2 of the purine and Cγ of the Thr-156 is important and this distance in the mutant structure is not significantly different from the wild type enzyme structures complexed with the ligands without C-2 substituents. In the sugar pocket Met-64, Met-180, Glu-181 and His-4* were observed (* indicates residues from an adjacent protomer). As previously observed, Glu-181 interacts with the sugar by accepting hydrogen bonds from the 2' and 3' hydroxyl groups, and His-4* engages in a hydrogen bond with the 5' hydroxyl group (4). Additionally, Arg-43* was observed very close to the ribose and forms hydrogen bond with 5' hydroxyl group.

A sulfate ion was placed in electron density present at the active site in a position occupied by phosphate in previous structures of *E. coli* PNP (6). The crystallization condition contains 2.0 M ammonium sulfate and no phosphate. Our structure supports the prediction that the phosphate forms hydrogen bonds with the charged residues Arg-87, Thr-90, Arg-43* and Gly-20 backbone nitrogen.

The most apparent difference between the mutant and the native structures of TvPNP

is the orientation of residue 204. In the structure of native enzyme complex, the Asp-204 side chain pointed away from the purine base in structures that contained bound adenosine, inosine, F-dAdo and deoxyinosine. One exception was FMA, which forms hydrogen bonds between the purine N-6 and N-7 atoms and the Asp-204 side chain with bond distances of 3.2 Å and 3.4 Å respectively (4). The explanation provided for this hydrogen bonding pattern was that the alkaline crystallization conditions (pH 8.5-8.7) deprotonated Asp-204, so that it could not donate a hydrogen bond to the N-7 atom of the ligands mentioned above. FMA, however, is protonated at physiological pH at the N-7 position and can interact with Asp-204 by donating a hydrogen bond. The D204N mutant structure further confirms the potential for a 204 position side chain simultaneously hydrogen bond to the N6 and N7 atoms of a purine ligand. In the mutant complex, Asn-204 is capable of serving as a hydrogen bond donor at physiological pH (even more so at the low pH in which the crystals reported here were grown), and the adenosine complex structure shows that the carboxyl group accepts a hydrogen bond from the N-6 atom and the amido group donates a hydrogen bond to the N-7 atom with average hydrogen bond distances of 3.2 Å (ranges from 3.1 Å to 3.4 Å in the six protomers) and 2.9 Å (2.8 Å-3.0 Å), respectively (Fig 5, Fig 6), making the asparagine mutation provide a complementary fit for the adenosine in the active site. This is in contrast to the kinetic data (Table I), which showed an unmeasurable K_m for adenosine and the D204N mutant. One possible explanation for this discrepancy is that saturating ligand concentrations used in crystallization may not be comparable to those

used for the kinetic studies. Additionally, the unmeasurable K_{cat} exhibited by the D204N mutant for adenosine is surprising, considering that the asparagine side chain was observed to donate a hydrogen bond the purine N7 atom that is believed to be important for catalysis. More information of other mutant structures and different ligand binding are needed to resolve these apparent discrepancies.

FUTURE WORK

We intend to treat the twinning problem of inosine complex with other methods, for example using Refmac twinning refinement as part of the CCP4 suite (8), and Phenix (14). Improving the electron density maps is necessary to help explain the role of residue 204 in catalysis and shed light on the explanation of the kinetic data. It will also be useful to solve the structures of D204A and D204E mutants with both ligands. For D204A, we expect the binding position to be slightly different from that with the native enzyme, since alanine should not form hydrogen bonds with any ligand. For D204E, glutamate has a very similar structure to aspartic acid but has an additional CH_2 group. It would be interesting to know whether it forms hydrogen bonds with ligands like aspartic acid does, and why D204E binds to both ligands tighter, but exhibits diminished catalytic efficiency based on the kinetic data (Table I).

Since the mutations are not biologically relevant, if we can correlate measured activity with the positioning of the ligand in the active site and the favorable or

unfavorable interactions with active site groups, we would gain insight into what modifications to potential inhibitors could reproduce the same effects in the native enzyme, and design inhibitors that can bind tightly to the enzyme but not trigger catalysis. Finally, it may be necessary to find other conditions to crystallize the mutants, since the twinning problem remains challenge to refine the structures.

Table I. Biochemical assays [a]

Enzyme	Substrate	$K_m(\mu M)$	$K_{cat}(s^{-1})$	$K_{cat}/K_m(\mu M^{-1} s^{-1})$
Wild type	adenosine	5.65±0.40	2.68±0.130	0.477±0.058
D204A	adenosine	---	---	---
D204N	adenosine	---	---	---
D204E	adenosine	3.95±0.33	0.086±0.002	0.021±0.002
Wild type	inosine	32.73±0.41	6.85±0.632	0.20±0.021
D204A	inosine	13.60±1.00	0.07±0.006	0.004±0.0001
D204N	inosine	7.78±0.72	1.62±0.020	0.20±0.020
D204E	inosine	44.30±5.43	2.71±0.050	0.06±0.006

[a] The table was kindly provided by our collaborator C. C. Wang at the University of California, San Francisco.

Table II. Data Processing Statistics

Mutant		D204N	D204N
Ligand		adenosine	Inosine
Resolution (Å)[a]		50.0-1.8 (1.87-1.80)	50.0-1.8 (1.86-1.80)
Total Reflections		470323	465252
Unique Reflections		125244	128454
Completeness (%)[a]		94.7 (72.1)	97.5 (89.7)
R_{sym} **(%)**[a,b]		0.050 (0.272)	0.084 (0.304)
I/σ[a]		23.86 (3.8)	15.01 (1.6)
Redundancy[a]		3.8 (3.5)	3.6 (2.7)
Wavelength (Å)		0.9792	0.9792
Space Group		R3	R3
Unit Cell Parameters	**a (Å)**	166.18	166.04
	b (Å)	166.18	166.04
	c (Å)	139.35	138.47
	α	90°	90°
	β	90°	90°
	γ	120°	120°

[a] Values for the highest resolution shell are given in parentheses.

[b] $R_{sym}=\Sigma\Sigma_i |I_i-\langle I\rangle| / \Sigma\langle I\rangle$, where $\langle I\rangle$ is the mean intensity of the N reflections with intensities I_i and common indices h, k, l.

Table III. Twinning Fraction Statistics

Mutant	Ligand	Condition			Partial Twinning Fraction (%)	Resolution (Å)
		pH	[ammonium sulfate] (M)	[protein] (mg/ml)		
D204N	adenosine	4.6	2.0	12	42	2.30
D204N	inosine	4.4	2.0	12	25	3.10
D204N	inosine	4.4	1.9	10	22	2.35
*[a] **D204A**	inosine	4.4	1.6	8	39	2.60
*[a] **D204A**	inosine	4.4	1.6	8	35	2.20
&[a] **D204N**	adenosine	4.3	2.0	12	12	1.88
&[a] **D204N**	inosine	4.3	2.0	12	31	1.80
#[a] **D204N**	adenosine	4.3	2.0	10	26	1.80
#[a] **D204N**	inosine	4.3	2.0	10	46	1.80

[a] The symbols at left indicate different crystals from the same drop.

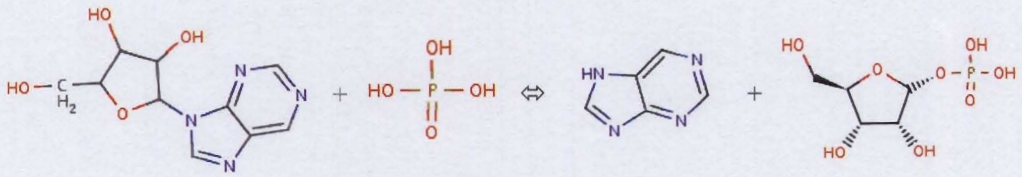
Table IV. Refinement Statistics

Mutant		**D204N**	**D204N**
Ligand		adenosine	Inosine
Resolution (Å)		1.8	1.8
No. of protein atoms		10814	10679
No. of ligand atoms		144	144
No. of water atoms		278	0
R factor[a] (%)		21.9	25.6
R free[b] (%)		24.9	28.6
Average B factor (Å^2)	**Proteins**	28.3	24.0
	Ligands/ions	33.5	24.5
	Waters	27.0	0.0
Rmsd[c] bond (Å)		0.006	0.006
Rmsd angle (°)		1.5	1.4

[a] R factor = $\Sigma_{hkl}||F_{obs}| - k\,|F_{calc}||\,/\Sigma_{hkl}|F_{obs}|$, where F_{obs} and F_{cal} are observed and calculated structure factors, respectively.

[b] For R_{free} the sum is extended over a subset of reflections that were excluded from all stages of refinement.

[c] Root mean square deviation.



Purine nucleoside + phosphate ⇔ purine + α-D-ribose 1-phosphate

Figure 1: Revertible reaction catalyzed by purine nucleoside phosphorylase.

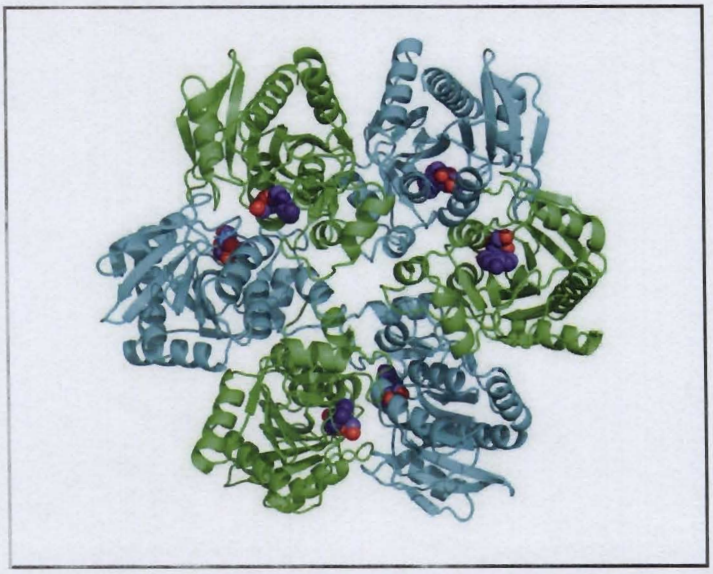


Figure 2: Overall structure of the TvPNP biological unit hexamer complex. The active sites are indicated by the purple and red colored ligands at the interface of the blue and green protomers.

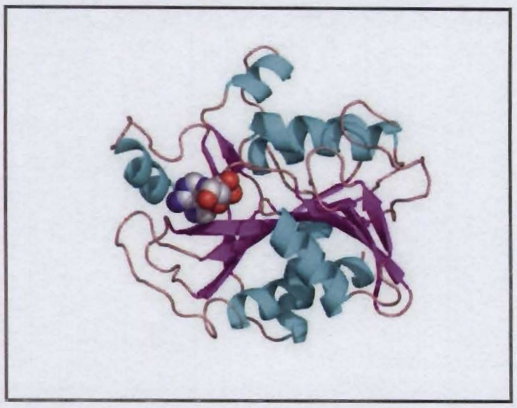
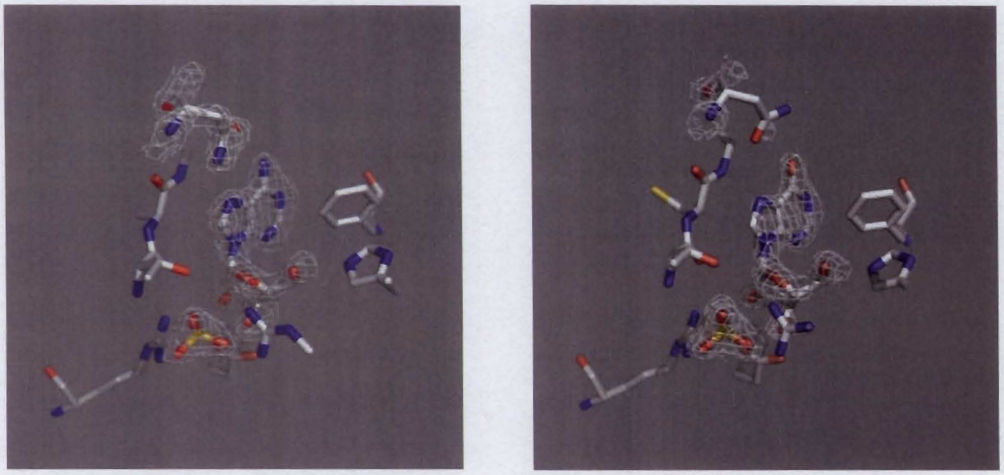


Figure 3: A ribbon diagram of a protomer complex showing an αβα sandwich structure. α-helices are shown in cyan and β-strands in purple. Ligand adenosine in sphere representation is colored with carbon atoms in grey, oxygen atoms inred and nitrogen atoms in blue.



A B

Figure 4: Composite Omit Electron Density Maps. A. Composite omit electron density map calculated for the adenosine complex contoured at 1.8 σ. B. Composite omit electron density map calculated for the inosine complex contoured at 1.8 σ. All residues are colored by element, with carbon atoms in grey, oxygen atoms in red, nitrogen atoms in blue and sulfur atoms in yellow.

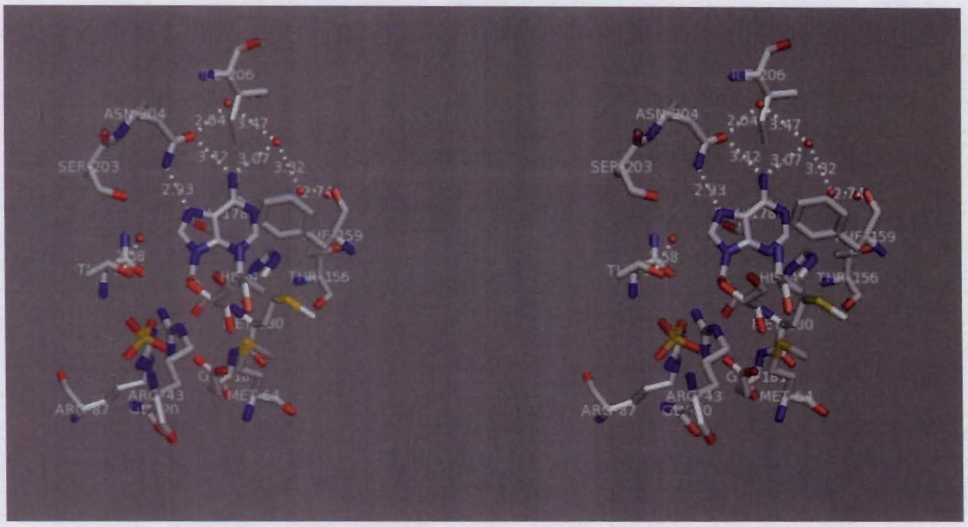


Figure 5: Stereoview of the active site of the adenosine complex. The hydrogen bond interactions are shown in dashed lines. Waters are shown as red spheres. Atoms are colored by element, with carbon atoms in grey, oxygen atoms in red, nitrogen atoms in blue and sulfur atoms in yellow. Residues and distance of the hydrogen bonds are labeled.

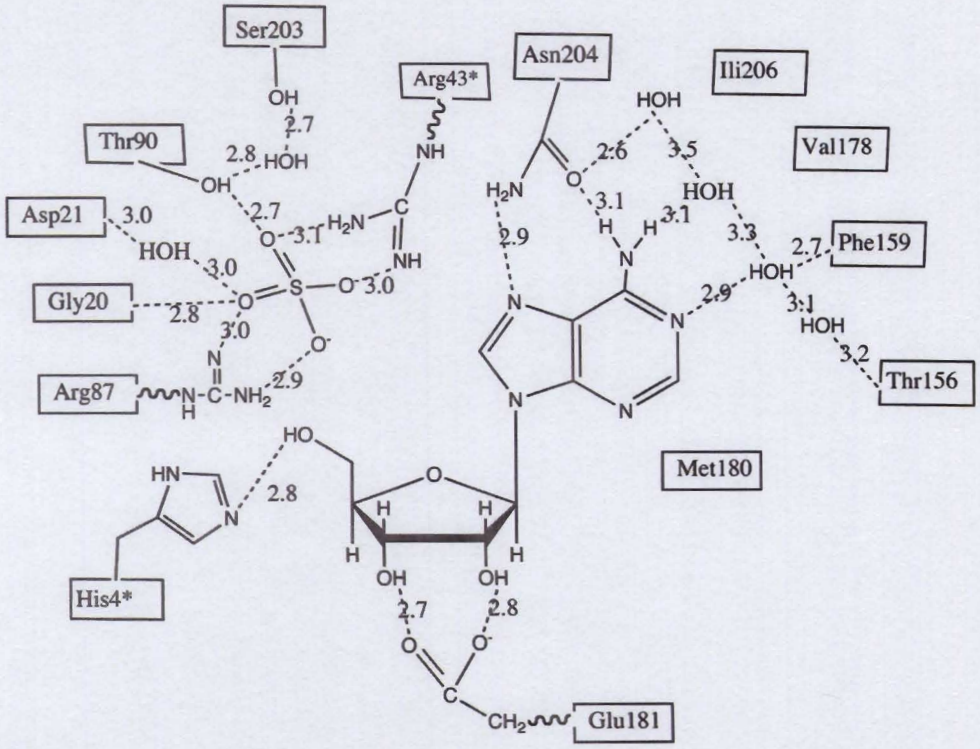


Figure 6: Schematic diagram of the active site of adenosine complex. Hydrogen bonds are shown as dashed lines with the distances labeled. The distances are all measured in one chain of the structure. * stands for the residues are from an adjacent protomer. Purine binding site, sugar binding site and phosphate binding site are shown. Water molecules interacting with the ligand are shown.

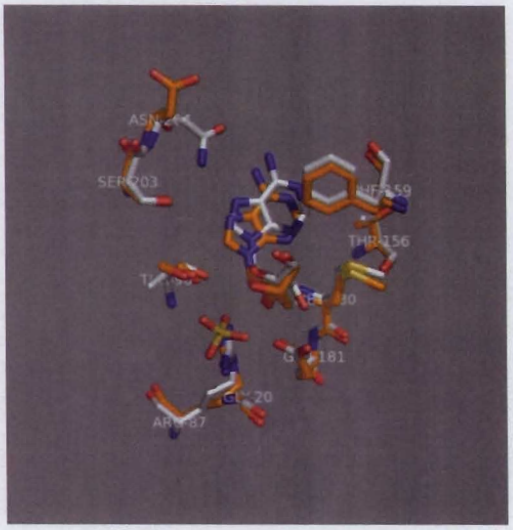


Figure 7: D204N adenosine complex superimposed with native enzyme adenosine complex. Carbon atoms for the mutant are grey and those for the native enzyme are orange. All oxygen atoms are red and nitrogen atoms are blue. Note the difference in the conformation of Asp-204.

Acknowledgments— I am thankful for the great help from Andrew Torelli, Mariya Morar, and Simon Li for the supervision of the project. I thank Steven Ealick for the supervision and the provision of the research opportunity. I am grateful the help from all the people in Ealick's lab. I thank the Hughes Scholarship for the financial aid in summer 2008, and Cornell University for the opportunity of doing an honors thesis. I thank Deborah Campbell for her assistance in editing the manuscript.

REFERENCES

1. Meyer, EA: **Other Intestinal Protozoa and *Trichomonas Vaginalis*.** Medical *Microbiology (4th edition)* http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=mmed.chapter.4169
2. Wang CC: **Parasite enzymes as potential targets for antiparasitic chemotherapy.** *Journal of Medicinal Chemistry* 1984, **27**(1):1-9.
3. Bzowska A, Kulikowska E, Shugar D: **Purine nucleoside phosphorylases: properties, functions, and clinical aspects.** *Pharmacology & Therapeutics* 2000, **88**(3):349-425.

4. Zang Y, Wang W-H, Wu S-W, Ealick SE, Wang CC: **Identification of a subversive substrate of Trichomonas vaginalis purine nucleoside phosphorylase and the crystal structure of the enzyme-substrate complex**. *Journal of Biological Chemistry* 2005, **280**(23):22318-22325.
5. Secrist JA, Parker WB, Allan PW, Bennett LL, Waud WR, Truss JW, Fowler AT, Montgomery JA, Ealick SE, Wells AH *et al*: **Gene therapy of cancer: Activation of nucleoside prodrugs with E-coli purine nucleoside phosphorylase**. In*: 1999*: Marcel Dekker Inc; 1999: 745-757.
6. Munagala N, Wang CC: **The purine nucleoside phosphorylase from Trichomonas vaginalis is a homologue of the bacterial enzyme**. *Biochemistry* 2002, **41**(33):10382-10389.
7. Otwinowski Z, Minor W: **Processing of X-ray diffraction data collected in oscillation mode**. In: *Macromolecular Crystallography, Pt A.* vol. 276; 1997: 307-326.
8. Bailey S: **The CCP4 suite-programs for protein crystallography.** *Acta Crystallographica Section D-Biological Crystallography* 1994, **50**:760-763.
9. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS *et al*: **Crystallography & NMR system: A new software suite for macromolecular structure determination**. *Acta Crystallographica Section D-Biological Crystallography* 1998, **54**:905-921.
10. Yeates TO: **Detecting and overcoming crystal twinning**. *Macromolecular Crystallography, Pt A.* vol. 276; 1997: 344-358.
11. Emsley P, Cowtan K: **Coot: model-building tools for molecular graphics**. *Acta Crystallographica Section D-Biological Crystallography* 2004, **60**:2126-2132.
12. DeLano, WL *The PyMOL Molecular Graphics Systems*, Delano Publishing, San Carlos, CA.
13. Terry Brown: ChemBioDraw Ultra, Cambridgesoft
14. Adams PD, Grosse-Kunstleve RW, Hung L-W, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, and Terwilliger TC: **Phenix: building new software for automated crystallographic structure determination.** *Acta Crystallographica SectionD-Biological Crystallography* 2002, **58**:1948-1954.
15. Zhang Y, Parker WB, Sorscher EJ, Ealick SE: **PNP Anticancer Gene Therapy.** *Current Topics in Medicinal Chemistry* 2005, **5**:1259-1274.
16. Koellner G, Luic M, Shugar D, Saenger W, Bzowska A: **Crystal structure of the ternary complex of *E. coli* purine nucleoside phophorylase with formycin B, a structural analogue of the substrate inosine, and phosphate (sulphate) at 2.1 Å resolution.** *J. Mol. Biol.* 1998, **280**:153-166.