Analyzing the Binding of Nevirapine and Rilpivirine to HIV-1 Reverse Transcriptase through Computationally Derived Charge Optimization

Mona Minkara

April 2009

© 2009 Mona Minkara

Abstract

Research was conducted to computationally study the binding site of wild-type and mutated HIV-1 reverse transcriptase (RT) complexed with the non-nucleoside inhibitors nevirapine (NVP) and rilpivirine (TMC278). This research was conducted to analyze and to understand the electrostatic determinants of tight binding in these systems. Our ultimate goal is to design RT inhibitors with improved binding to wild-type and mutant RT, which could lead to more effective therapeutics to treat HIV-1. First, a point-charge approximation to the charge distribution of NVP and TMC278 was obtained by using a restrained fit to the electrostatic potential generated by its quantum mechanical electron density.[3] Then, using the linearized Poisson-Boltzmann equation, we calculated the $\Delta G_{electrostatic}$ of binding between each ligand and the HIV-1 reverse transcriptase target. Electrostatic charge optimization theory[4-6] was then applied to the system to analyze how well the non-nucleoside RT inhibitors bind and to look for potential ways to improve drug-target interactions. A set of optimal point charges was obtained and the optimal $\Delta G_{electrostatic}$ of binding was found for each system. It was determined that nevirapine is not electrostatically optimized for binding to RT in both the wild type and variants. Nevirapine is very polarized, which is not ideal for the relatively hydrophobic binding site. This data is in agreement with clinical data that show that patients treated with nevirapine are very susceptible to resistance mutations. In contrast, rilpivirine appears to be electrostatically optimized for wild-type RT, except for the cyano and the cyanovinyl group at the edges of the molecule. These data support available experimental results that show that rilpivirine is up to 20 times more effective than earlier RT drugs.

Nevertheless, results show that a strong hydrogen bond between a nitrogen on the central pyrimidine ring of rilpivirine is lost in the L100I/K103N complex, which reduces somewhat the optimality of charge distribution in this mutant. Additional work is necessary to identify the atoms for which changes in atomic charge would result in the greatest change in Gibbs free energy of binding. This analysis can inform next steps in the design of novel drug molecules that yield higher efficacy.

Acknowledgements

First and foremost, I would like to thank my advisor, Professor Mala Radhakrishnan. Without her constant support, guidance, and motivation this thesis would not be possible. I am so grateful for the opportunity to work in your lab. I have learned a great deal and have enjoyed the ride.

I would also like to thank Pam Davis, who has been with me through all the highs and lows during my Wellesley career. Pam, I am incredibly grateful for having met you and I could not have written this thesis without you. Thank you for the car rides, late nights, and exploring the world of computational chemistry with me. You have had a positive and profound effect on me, and I am better person for it.

To my thesis committee members: Flick Coleman, Nolan Flynn, and Louise Marlow—thank you for all your help and taking time out of your busy schedules to be on my committee.

Professor Christopher Arumainayagam has been in constant support of me during my academic career. Professor Arumainaygam, thank you for always believing in me, supporting me, inspiring me, and exposing me to new aspects of the wondrous world of physical chemistry.

The chemistry department at Wellesley has been my home for the past four years. Thank you to all those in the department who have welcomed me and kept their doors open to me. I could not have found a better niche at Wellesley. I would also like to thank Irene Laursen who has been patient and always willing to lend a helping hand.

To Andrea Johnston, thank you for helping me through this process. We suffered together and have finally reached the end. Also, to my lab mates past and present:

Maggie Chidote, Morgan Carr-Markell, Ilene Tsui, Bilin Zhuang, Naomi Sedani, Vic Abrenica, Kathy Chen, and Emma Nechamkin, thank you for your laughter and friendship.

I would like to thank Anh Tran for helping me write my acknowledgements. I would also like to thank Katie Wagner for aiding me in developing some of the images in my thesis. Finally, I would like to thank Kate Pawlowski for helping me read through and edit my thesis. I would also like to thank everyone who has helped me during my four years at Wellesley. Without their help, I would not have been able to get to the point of writing a thesis.

To my dearest friends: thank you for your support and friendship. Last, but not least, I would like to thank my parents. Baba and Mama, thank you for keeping me on track and checking in. I owe a big thank you to my sister, Sara, for checking up on me and keeping me sustained during this crazy time. And, finally, to Ibrahim, thank you for being such a cool brother.

Table of Contents

Abstract …….. 1

Acknowledgements …… 3

1. Introduction …… 7

2. Theory …… 16

2.1. Introduction ……...…… 16

2.2. Determining partial atomic charges through RESP fitting …… 17

2.3. Fundamentals of electrostatics ……. 18

2.4. Charge optimization …….…. 22

3. Methodology …… 28

3.1. Structure preparation ……. 28

3.2. Charge optimization …….. 29

4. Results ……..……. 31

4.1. Wild-type reverse transcriptase …….. 31

4.1.1. Nevirapine …… 31

4.1.2. Rilpivirine …… 34

4.2. Variants of HIV-1 reverse transcriptase ……..…… 35

4.2.1. Nevirapine …… 35

4.2.2. Rilpivirine …… 36

4.3. Sensitivity analysis ……. 44

4.3.1. Sensitivity figures……. 44

4.3.2. Eigenvalue and eigenvector sensitivity analysis…..……. 50

5. Discussion …….. 56

5.1. Nevirapine ……. 56

5.2. Rilpivirine …….. 57

5.3. Limitations and extensions of charge optimization ……... 58

6. Appendices …… 60

7. References ……. 76

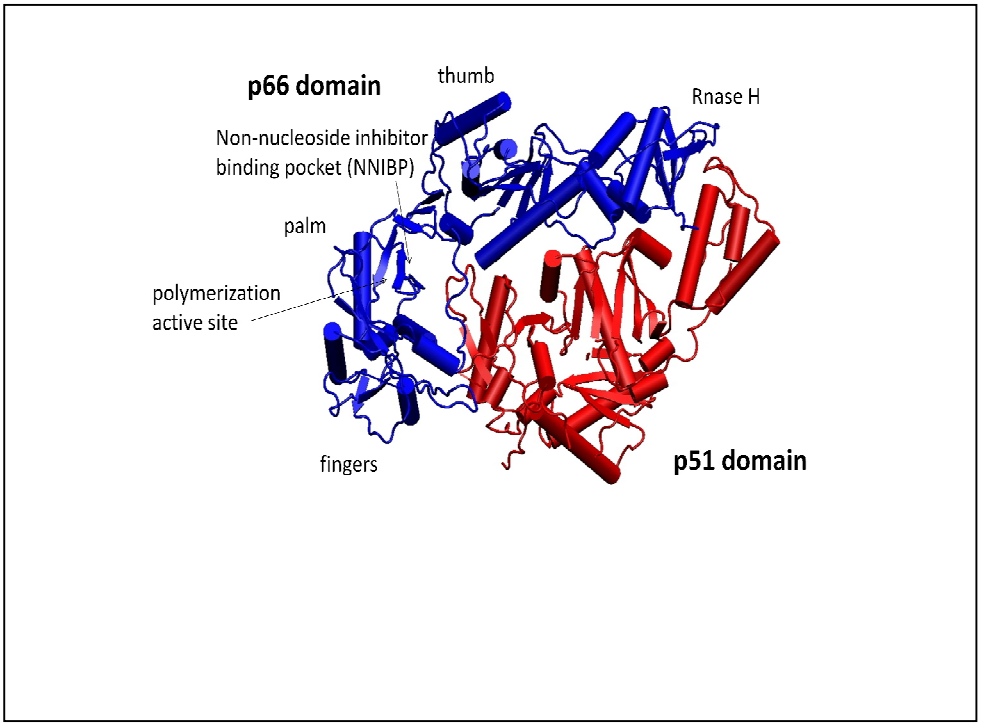
1. INTRODUCTION

By the end of 2007, 33 million people worldwide were living with the human immunodeficiency virus (HIV), the virus that causes acquired immune deficiency syndrome (AIDS) with 2.7 million new patients and 2.0 million deaths in 2007 alone.[7] The vast majority of people infected with the virus live in impoverished countries. One of the main challenges in successfully combating HIV is the resistance that inevitably occurs when patients are administered drugs. Mutation-based resistance can arise within one week of beginning treatment[8] and is the major challenge in designing effective HIV drug therapies. Resistance occurs rapidly because of the frequent mistakes made by the viral enzyme reverse transcriptase (RT) and cellular RNA polymerase II—a rate of one mutation per viral replication cycle (1 base change in 10,000 RNA nucleotides).[9] In a group of patients treated with protease inhibitors, the average HIV-1 production was 10.3×10^9 new virons per day, offering billions of opportunities daily for resistant mutants to occur.[10]

More than 30 drugs have been approved by the US Food and Drug Administration to treat HIV.[11] They fall into classes, each of which targets, or inhibits, a phase of the HIV life cycle, including entry into the host cell, assembly of the virus, and transcription of the viral genome. HIV's proclivity to mutate makes for an enormous challenge in designing drugs that achieve a balance between selectivity and promiscuity. Drugs must be selective enough to specifically target HIV enzymes without inhibiting other enzymes essential to patient health. At the same time, due to the rapidly changing structure of the virus, viable drugs must be promiscuous enough to target not only the wild type but also its many and constantly evolving mutant variants. The first approved HIV medication,

azidothymidine (AZT) [12] initially showed good results for selectively targeting reverse transcriptase and slowing the progress of the disease, but AZT became ineffective when administered alone due to mutations of the RT protein [13], such as D67N, K70R, and T215F or Y.[13] As additional drugs were developed and approved, the highly active antiretroviral therapy (HAART) approach was developed to overcome the problem of resistance by administering several drugs simultaneously to improve patient outcome .[14] HAART achieves the maximum anti-viral activity possible while preserving the patient's immune system.[15] HAART has become a standard treatment protocol for AIDS patients.[11] It is worth noting that AIDS medications do not eradicate the HIV virus—they are only able to slow the progress of the disease.

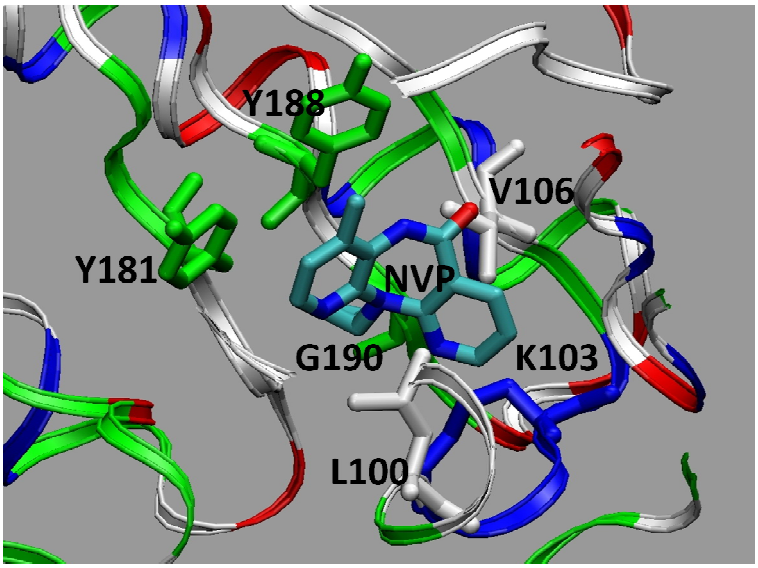
This study is concerned with reverse transcriptase inhibitors, which comprise more than half of the currently approved drugs. Reverse transcriptase is the enzyme that is responsible for the transcription of DNA from the viral RNA, which is transported



throughout the body in virons. HIV RT is a heterodimer consisting of a polypeptide subunit with a mass of 66 kDa (p66) and a second polypeptide subunit of 51 kDa (p51). The p66 subunit has 5 domains. Reverse transcriptase is typically described as a right hand with a thumb, palm and fingers regions.[16]

There are two classes of RT inhibitors. Neucleoside/nucleotide inhibitors work at the polymerase active site of the target reverse transcriptase. They are analogs of naturally occurring nucleosides that lack a 3' OH moiety on their ribose or ribose mimic moiety. Thus they act as a chain terminator [17] and work by being incorporated into the RT to inhibit polymerization. Non-nucleoside reverse transcriptase inhibitors (NNRTI) have a completely different mode of action. They are non-competitive inhibitors and act by interfering with the process of polymerization. NNRTIs bind to a hydrophobic binding site near the polymerase active site known as the non-nucleoside inhibitor binding pocket (NNIBP), which is at the cleft of the thumb and palm of the p66 subdomain.[1] This binding pocket exists only when the NNRTI is present; the region is plastic enough to make the conformational changes necessary for the drug.[18] The binding pocket slightly disaligns the template:primer of the catalytic site and prevents a nucleotide from being incorporated into the growing DNA chain, thereby terminating polymerization.[19]

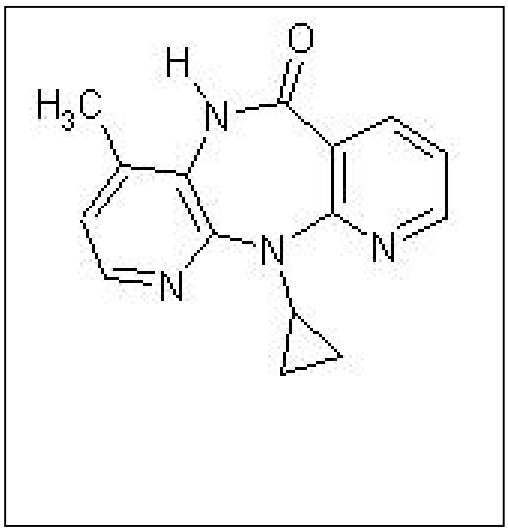
The most common mutations that make NNRTI's ineffective include L100I, K103N, V106A, Y181C, Y188L, and G190A.[20] According to Yin *et al.*, there are three mutation mechanisms that result in drug resistance. First, amino acid substitutions may cause resistance by steric hindrance so that the drug cannot bind at the binding site. An example is the substitution of leucine for isoleucine in the L100I mutation.[21] Another form of resistance is the reduction of aromatic amino acids leading to a loss of

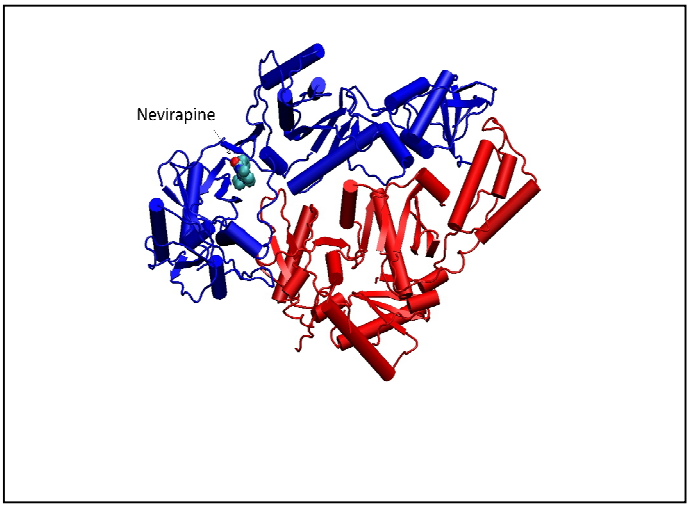


hydrophobic interactions such as in the Y181C and Y188L mutations. The K103N mutation represents a third type of resistance. It prevents the binding of the NNRTI as the asparagine forms a hydrogen bond with the phenoxyl group of tyrosine on Y188, making a sort of "gate" in the binding pocket.[20] All three of the first generation NNRTI's are ineffective against this mutation.[1]

This study examines nevirapine (NVP), the first clinically approved NNRTI,[22] and rilpivirine (TMC278), a new and very promising drug that is still in clinical trials.[9] NVP was introduced as a commercial drug in 1995 under the market name of Viramune by Boehringer Ingelheim Pharmaceuticals. Nevirapine was developed in the early 1990's using a manual screening process that was common at the time.[22-24] Unlike today, the screening process was able to process only about 100 compounds a week. After screening 600 molecules, the research team identified a lead molecule from which they developed NVP.[25] The butterfly structure of the compound (Figure 1) was a common feature of first generation NNRTI's. The main problem with NVP treatment was its lack of efficacy in the presence of drug-resistant mutations. It has been shown that resistance mutations to

the drug can occur even after single-dose therapy.[26] It is notable that many residues of the binding pocket share two important features: they are in direct contact with the drug and are also susceptible to mutations. These mutations cause disruptions to van der Waals interactions that occur between the enzyme and the drug.[27] Some mutations either directly or indirectly change the shape of the binding pocket. For instance, the side chains for Y188 and Y181 form a wall in the binding pocket against which the drug rests.[27] In both instances, the tyrosine is mutated to cysteine and steric hindrance occurs due to the larger size of the sulfur atom on the cysteine residue.[27] Due to NVP's rigid confirmation, it does not have the ability to conformationally readjust to an evolving binding pocket in order to bind to different variants of RT.

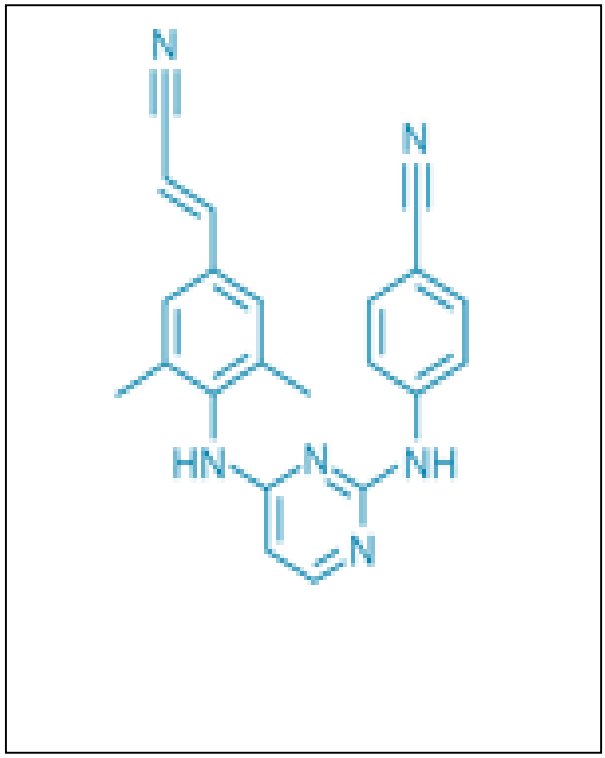




In spite of the prevalence of resistance mutations, NNRTI's have great value in HIV/AIDS therapy. In general, NNRTI's have many fewer side effects that earlier drugs. However, NVP still may cause serious side effects. The most common side effect is skin reactions that can range from mild rash to Stevens-Johnson syndrome, a potentially fatal

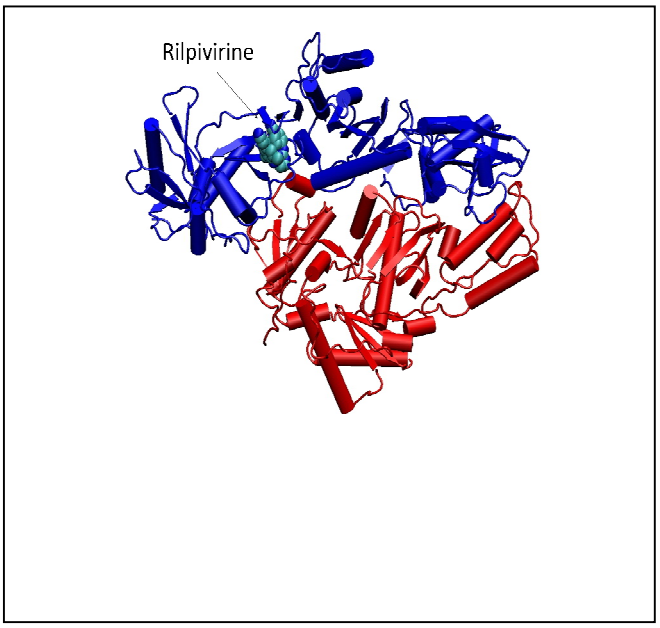
swelling of the skin and mucous membranes.[28] The drug can also cause sometimes severe liver toxicity in the first few weeks after treatment has begun.[29]

A new member of the class of diarylpyrimidine, or DAPY, NNRTI's, known as rilpivirine (also TMC278 and R278474) was developed in 2001 and announced in 2005.[9] It is being developed by the Belgian pharmaceutical company Tibotec. Rilpivirine began phase III clinical trials in April 2008 that will be completed in August 2010. Rilpivirine was designed with several structural changes to improve efficacy toward drug-resistant mutants. Rilpivirine has the ability to "wiggle" and "jiggle".[30] Wiggling refers to torsional rotation of the drug's subgroups relative to the each other and jiggling refers to the plasticity of the drug and the binding site. This conformational flexibility allows this class of drugs to be more promiscuous than any NNRTI predecessors.



Rilpivirine has been shown to be highly effective against wild-type and single and double mutant forms.[21] It binds in the RT binding pocket in a horseshoe configuration. Rilpivirine has a central pyrimidine ring and two methyl-ethyl side rings. The drug has amino groups connecting the three aromatic rings and a cyanovinyl group at one end. Key side chains in the binding pocket interact with the bound NNRTI including Y181, Y188, Y229, K103, K101, L100, L234, and Y318. For instance, the nitrogen on one of

the amino groups forms a hydrogen bond with RT's K101 residue. Rilpivirine is able to overcome the K103N mutation that causes resistance to NVP and most other NNRTI's by binding to the asparagines, thus preventing the asparagine's interaction with tyrosine 188, creating a gate.[20] The cyanovinyl group fits into a tunnel formed by residues Y188, F227, W229, and L234 that connects the NNRTI binding pocket to the nucleic acid-binding cleft. W229 is a functionally important and highly conserved residue.[31] The drug's interactions with this residue may explain why rilpivirine is the most potent DAPY designed to date.[32] Early results for clinical trials show that rilpivirine could be 10 to 20 times more effective than the earlier generation of NNRTI drugs.



Interestingly, computational methods played a crucial role in the integrated drug development process for rilpivirine. X-ray crystallography of several HIV-RT complexes with possible analogues formed the basis of a structure-based design approach.[9] Crystallographic evidence suggested that optimizing hydrogen bonding between the ligand and the surrounding side chains would improve the drug potency. Computational modeling studies showed that the extension of one of the "ends" of the horseshoe conformation in the direction of the W229 residue would improve binding and reduce the opportunity for mutation. The modeling studies showed that the NH2 moiety on the central ring of etravirine (rilpivirine's parent drug) would be

improved with a hydrogen atom and a spacer group between the cyano group and the trisubstituted phenyl ring.[33] A variety of conformations for a given analogue were tested for binding energy in complex with RT. Binding energies took into consideration the electrostatic interactions, van der Waals potentials, and hydrogen bonding. Analogues with the lowest binding energies predicted the best virologic activity.[9]

The computationally-integrated approach to the development of rilpivirine as compared to nevirapine reflects an evolution in technology that made for increasingly sophisticated modeling software.[34] Whereas computational studies of nevirapine depended on first having crystallization studies of the complex, computational modeling now precedes and even predicts crystal structure.[35] In the mid–1990's, Smith and others[34, 36] used molecular modeling studies to examine binding of a number of inhibitors to specific residues and compared it with patient data to see if there were agreement with specific mutations associated with mutant-based resistance. Focus evolved over the last decade to specifically address the binding between inhibitors and conserved residues, that is, amino acids that do not tend to mutate.[34, 37-40] New modeling techniques evolved including Monte Carlo/molecular dynamics simulations, a method which combines Monte Carlo sampling for the chemical space of the system with the molecular dynamics method for generating a set of coordinates[35], PROFEC (pictorial representation of free energy changes) analysis, which produces contour maps to show favorable regions for the inhibitor to target for improved binding free energy,[39, 41] and sophisticated docking programs.[42, 43] These tools allowed Das *et al*. to develop the diarylpyrimidine analogues of which rilpivirine is the most potent example[30] and a new class of NNRTI's called indolyl aryl sulfones.[42] It seems likely that computational techniques will continue to play

an increasingly important role in drug development because as the modeling precision improves, computational methods will tend to decrease the time to develop and reduce the associated costs.

This study examines the electrostatic binding free energy of nevirapine and rilpivirine complexed with HIV-1 wild-type (WT) and selected mutant reverse transcriptase variants with a goal of providing further insight into electrostatics component of these drugs' interactions with the receptor. In this work, we have chosen to study binding free energy using the charge optimization methodology introduced by Lee and Tidor and Kangas and Tidor.[4, 6] This method has been used successfully to study charge optimization in many kinds of systems.[44] This research examines complexes that have been studied before using other computational methods, but to our knowledge, this is the first time that these systems have been analyzed using charge optimization techniques. It is possible to compare our results with those of other researchers to look for areas of convergence or inconsistency. While the other components of ΔG must be included in a comprehensive analysis of ligand-receptor interactions, a careful focus on the electrostatics of binding offers the opportunity to understand characteristics of specific residue-ligand interactions that are necessary for overall best binding toward both wild type and mutants. In this study we show that nevirapine is not electrostatically optimized for binding to RT in both the wild type and variants. NVP is highly polarized, making it suboptimal for the relatively hydrophobic binding site. However, rilpivirine appears to be nearly optimal for tight electrostatic binding with WT RT, and close to optimal for mutants as well, although important deviations from optimality are noted toward specific mutants.

2. THEORY

2.1 Introduction

The binding between two molecules is studied by measuring the change in Gibbs free energy of the system. In particular, we are interested in the change in Gibbs free energy between the bound and the unbound states of the system. The change in Gibbs free energy of binding has multiple components as seen in the following equation:

$$\Delta G_{total} = \Delta G_{elec} + \Delta G_{vdW} + \Delta G_{SASA} + \Delta G_{Geom}$$

ΔG_{vdW} measures the changes in van der Waals intermolecular interactions; ΔG_{SASA} represents the free energy change due to the hydrophobic effect and is modeled as being proportional to the solvent-accessible surface area (SASA) buried upon complexation; ΔG_{Geom} measures the change in conformational energy of the ligand and receptor (in this case, we assume rigid binding and so $\Delta G_{Geom} = 0$) and ΔG_{elec} measures the change in electrostatic interactions. In this work, we are concerned only with the last component of ΔG, ΔG_{elec}.

There are many ways of modeling the electrostatic interactions between molecules. These include the generalized Born model, which approximates the exact (linearized) Poisson-Boltzmann equation[45] and Ewald summation, which replaces summation of electrostatic interaction in real space with Fourier space.[46] Of course, the most accurate way to model electrostatic interactions is by solving the Schroedinger Equation for the molecular electronic wave function and integrating Coulomb's law over the electron density. However, the resulting probability density is too computationally demanding to be practical and the Schroedinger Equation has only been solved exactly

for one-electron systems. Hence, partial atomic charges (described below) are a crucial approximation to facilitate the electrostatic analysis of systems.

2.2 Determining partial atomic charges through RESP fitting

Although it is theoretically possible to determine absolute electrostatic properties from the ill-defined electron cloud around an atom, it is extremely impractical to do so when studying a large molecule due to the complexity of the computations. Partial atomic charges model atomic charge distributions as single point charges in order to simplify the calculation of a molecule's physical properties. For example, we can look at hydrogen fluoride which has the very electronegative fluorine atom. The electron density is pulled mostly toward the fluorine so we can assign a negative partial atomic charge of -0.45 and assign the hydrogen a +0.45 partial atomic charge. (These values were obtained using Mullikan population analysis.) This allows us to model the electron density as a point charge at the center of each atom. In this study, we use partial atomic charges to model the electrostatic interactions between the ligand and the receptor. There are many ways to devise a set of partial atomic charges to approximate the true charge distribution on a molecule, including those derived from Mullikan population analysis. However, this method is flawed by its dependency on the basis set that is chosen.[3] We used the standard two-stage restrained electrostatic potential (RESP) fitting method developed by Bayly *et al*.[3] In this method, quantum mechanical *ab initio* calculations are used to find the potential at four shells of the Merz-Kollman surface of a molecule.[47, 48] A least squares fitting procedure is then used to find the set of charges at the centers of atoms that best replicates the potential generated by the quantum mechanical wave function at the Merz-Kollman surface. A hyperbolic restraint in the form of a penalty function is placed on the

atoms in the center of the molecule because they have less of an impact on the potential. The least squares objective function is highly insensitive to the values of these buried charge centers and so they can vary tremendously without affecting the fit. The two-stage RESP process has been shown to be a much more accurate method to devise partial atomic charges as compared to electrostatic potential (ESP) methods.[3]

2.3 Fundamentals of electrostatics

With a set of partial atomic charges representing the charge distribution of each binding partner, one can calculate the electrostatic component of the ΔG of binding from potentials obtained by solving the linearized Poisson–Boltzmann equation. The linearized Poisson–Boltzmann Equation is derived from the most fundamental equation of electrostatics—the Poisson equation[49]:

$$\nabla^2\phi(r) = -\frac{\rho}{\varepsilon_0}$$

∇^2 is the LaPlacian operator,

$$\frac{\delta^2}{\delta x^2} + \frac{\delta^2}{\delta y^2} + \frac{\delta^2}{\delta z^2},$$

the sum of the partial second derivatives with respect to each spatial coordinate; $\phi(r)$ is potential; ρ is the charge distribution; and ε_0 is the permittivity of free space (8.854 x 10^{-12} C^2 $J^{-1}m^{-1}$).

The electrostatic energy of a system is equal to:

$$E = q\phi$$

Solving the Poisson Equation for a single, isolated point charge yields the familiar Coulomb's Law:

$$\phi(r) = \frac{kq}{r}$$

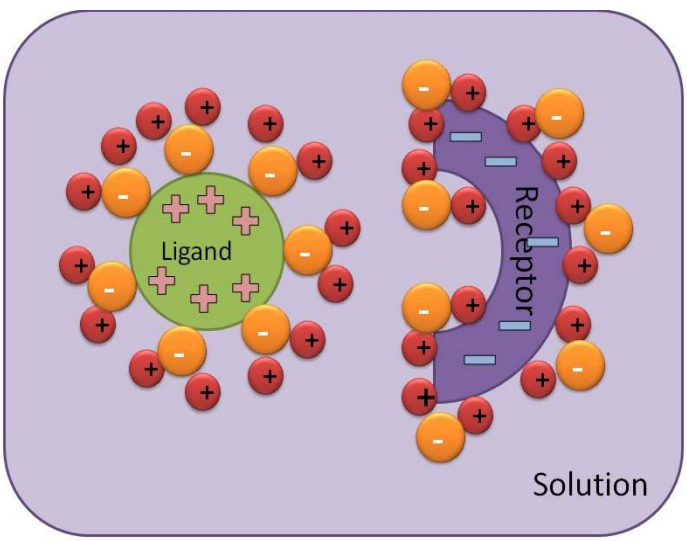
k is a constant $\frac{1}{4\pi\varepsilon_0}$, q is the charge value, and r is the distance in space between a point in space and the point charge generating the potential, chosen here to be located at the origin.

The interaction energy for a pair of point charges *in vacuo* is as follows:

$$E = q\phi = \frac{kq_1q_2}{r_{12}}$$

Here q_1 and q_2 are the charge values and r_{12} is the distance between the two charges. This solution to the Poisson Equation in three dimensions is fundamentally correct when describing a system of interacting point charges.

Recall that in the system under study the ligand and the receptor are considered in the bound and the unbound state. If the system is in a vacuum and the receptor has a single charge of -1, the ideal design for the ligand would be a charge of +∞, resulting in ΔG of -∞, which would produce the best binding. The reality is that the ligand and the receptor are not in a vacuum, but rather in aqueous solution.



Water is a very polar medium, therefore an infinitely charged positive ligand will interact strongly with

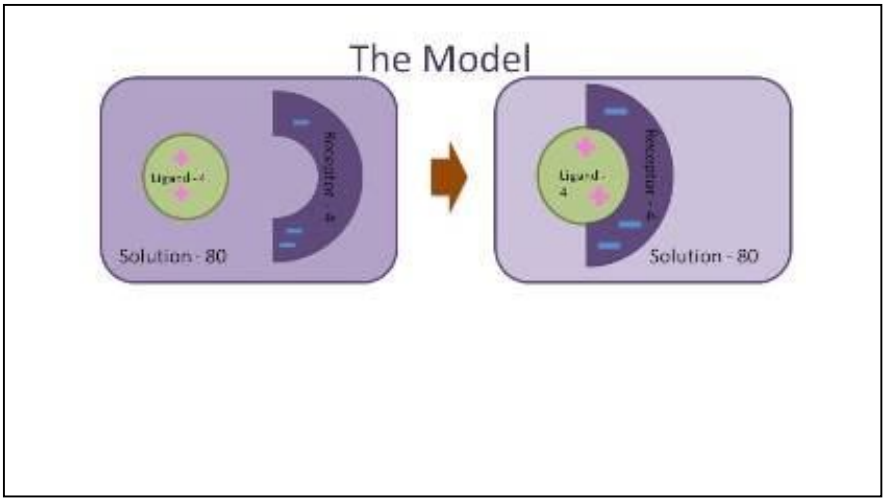
the negatively-charged oxygen atoms in water. As a consequence, the ligand cannot overcome its attraction with the waters to bind to the receptor. It is necessary to balance the interactions with the solvent and target in order for the target and ligand to readily bind.

The interactions among the ligand, receptor, and the water can be evaluated using the summation of Coulomb's law for every pair of interactions as follows:

$$\sum_{q\ on\ R} \sum_{q\ on\ L} \frac{kq_1q_2}{r}$$

The large number of interactions between pairs of atoms makes this approach very computationally intensive. In addition, one needs to model the entropy due to the polarization/reorientation of the water molecules, which requires dynamic simulations that are very time-intensive.

Assumptions are therefore necessary to measure the electrostatic interactions among the ligand, receptor, and the water. Each component is assumed to be an electrostatically polarizable continuum and is assigned a dielectric constant relative to its polarity. The ligand and receptor biomolecules are assigned a dielectric constant of 4.



Water is assigned a dielectric constant of 80. Additionally, partial atomic charges are assigned to atomic

centers of the ligand and receptor.

It is possible then to use a variation of the Poisson equation that is dependent on a spatially varying dielectric constant $D(r)$. The Poisson equation changes to be the following:

$$-\nabla \cdot |D(r)\nabla\phi(r)| = \frac{\rho(r)}{\varepsilon_0}$$

This equation, which implicitly models the solvent, automatically accounts for the polarization/reorganization of solvent such that the solvent generates a reaction field in response to the presence of a point charge.

Another factor that goes into determining the electrostatic interaction is the presence of ions in the solvent containing the ligand and the receptor. Na^+ and Cl^- ions and other electrolytes are present in biological aqueous solution. The presence of sodium chloride and other electrolytes provides an additional effect on the potential. Consider a positively charged amino acid such as lysine in the presence of a solution that has sodium and chloride ions. The chloride will be attracted to the positively charged lysine resulting in a buildup of negative potential to partially cancel out the positive potential around the amino acid. This redistribution of ions in solution comes from the Debye-Huckel theory and its concentration can be described using the Boltzmann factor as follows using chloride as the example ion:

$$c_{Cl}(r) = c_{Cl,bulk}e^{-\beta\emptyset qCl}$$

where c_{Cl} represents the local concentration of ions and $c_{Cl,bulk}$ represents the bulk concentration and $\beta = \frac{1}{k_B T}$., where k_B is the Boltzmann factor and T is the temperature.

The presence of ions changes the electrostatic interaction and results in a modified form of the Poisson equation known as the Poisson-Boltzmann equation that takes into account the summation of N types of ions with charge q_i and concentration c_i.

$$-\varepsilon_0 \nabla \cdot [D(r) \nabla \phi(r)] = \rho(r) + \sum_{i=1}^{N} q_i c_{i,bulk}(r) e^{-\beta q_i \phi_i(r)}$$

This non-linear form of the equation includes the Boltzmann factor. This equation is difficult to interpet in this form, and therefore it is commonly approximated. Under certain conditions including monovalent electrolytes, weak source charges, and high ionic strengths,[50] the exponent is near zero and hence the Poisson-Boltzmann equation can be linearized. First we write a Taylor expansion of the Boltzmann factor only using the first term of ϕ. The resulting linearized Poisson-Boltzmann equation is:

$$-\varepsilon_0 \nabla \cdot [D(r) \nabla \phi(r)] = \rho(r) + \varepsilon_0 D(r) \kappa^2(r) \phi(r)$$

$$\text{where } \kappa^2 = \frac{\beta}{D\epsilon_0} \sum_{i=1}^{N} c_{i,bulk} q_i^2 .$$

As a result of this linearization, interaction of the weak source charges can be described using linear response theory.

To understand linear response theory, we will look at an example. Consider a molecule with a single point charge of +0.1. We can assign the reaction field that it produces a value of *x*. Consider another molecule with a point charge of +0.2. Its reaction field will be 2*x*. As long as the charge value is relatively small, there will be a linear response in the size of the reaction field.

2.4 Charge optimization

Recall that the goal of this work is to computationally evaluate the electrostatic interactions between various non-nucleoside reverse transcriptase inhibitors (NNRTI) and

common reverse transcriptase (RT) variants. We will evaluate the electrostatic binding of each ligand–receptor pair by comparing the actual electrostatic binding free energy to a hypothetical optimal electrostatic binding free energy, which is defined as the most negative ΔG_{elec} that is possible for a given ligand, receptor, and complex shapes, and existing charge center locations and receptor charge distribution. This optimal binding free energy is achieved via a hypothetical charge distribution that we will henceforth call the "optimal charge distribution". Charge optimization is the process of finding such an optimal charge distribution. It has been previously described in great detail,[6, 49] applied with success,[4, 5, 44, 51] and will be summarized below. A comparison between optimal and actual charge distributions for various NNRTI's in RT complexes can allow one to improve binding by modifying specific functional groups to better qualitatively match the optimal charge distributions.

Charge optimization assumes a continuum electrostatics framework. Due to the high polarity of water and the relative lack of polarity of biomolecules, the ligand and the receptor are modeled as dielectric-4 cavities with discrete point charges within a dielectric-80 solvent.

Assuming such a continuum electrostatic model and a linear response of the reaction field, calculations involving the linearized Poisson-Boltzmann equation allow us to express the electrostatic component of binding free energy in the following form:

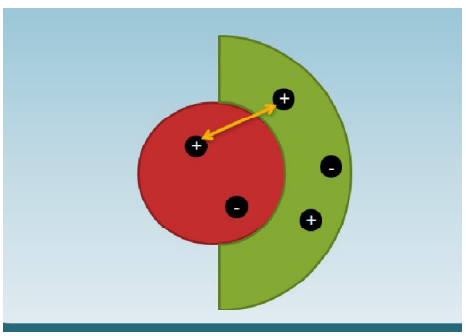
$$\Delta G_{elec} = ql'Lql + qr'Cql + qr'Rqr.$$

There are three terms to this equation: the *qr'Rqr* term represents the receptor desolvation penalty, the *ql'Lql* term represents the ligand desolvation penalty, and the *qr'Cql* term represents the complex interaction. *ql* and *qr* are vectors of the partial

atomic charges on the ligand and the receptor, respectively, and L, R, and C are matrices that will be described in further detail below.

The ligand desolvation penalty captures the difference in electrostatic free energy between the bound and the unbound states due to charges in the ligand, when neglecting the charges on the receptor. In the unbound state, the low-dielectric ligand is completely surrounded by water, which has a high dielectric of 80. The high dielectric of water has a favorable effect on the ligand charges because the highly polarizable water molecules align themselves according to the charges on the ligand. In moving to the bound state, there is a loss of interaction with the water as the ligand becomes desolvated by the low-dielectric cavity in the shape of the receptor. The ligand desolvation penalty is defined mathematically as $ql'Lql$, where ql is a vector of charges on the ligand, and L is the ligand desolvation matrix which describes the differences in the interactions of the charges on the ligand between the bound and unbound states. The matrix is a square with dimensions equal to the number of charges on the ligand. We obtain the matrix values by first setting each charge center on the ligand in turn to +1e and then solving the linearized Poisson-Boltzmann equation for the difference in potential between the bound and the unbound states at grid points in space. We then multiply the difference in potential at this point and at the other charged points on the ligand by +1 to get an energy per charge at each point on the ligand. Each ijth element is therefore a measure of the difference, between bound and unbound states, in the interaction energy between the ith and jth charge positions, assuming a charge of +1 at both positions. Each diagonal element of the desolvation matrix is the change in the potential per charge at point i between the unbound and bound states due to changes in the surrounding reaction field. Because each

off-diagonal interaction is duplicated within the matrix, it is necessary to divide the off-diagonal elements by two in solving for the desolvation penalty. The elements of the *L* matrix have units of energy per charge per charge. In order to get the true ligand desolvation penalty, we multiply the *L* matrix on either side by the actual charge distribution. The ligand desolvation penalty is always nonnegative because it is always energetically unfavorable to replace high-dielectric solvent with a low dielectric cavity. Therefore, all the eigenvalues of the *L* matrix are nonnegative, meaning that *L* is positive semidefinite.



The process for solving for the receptor desolvation penalty *qr'Rqr* is analogous to solving for the ligand desolvation penalty, except that *qr* represents the actual charges on the receptor and *R* represents the receptor desolvation matrix.

The complex interaction term, *qr'Cql*, measures the screened Coulombic interaction between the charges on the ligand and the charges on the receptor. The complex interaction can be either energetically favorable or unfavorable, depending on the charge distributions of the ligand and receptor. *C* is the complex interaction matrix. The matrix is rectangular with dimensions equal to the number of charges on the receptor by the number of charges on the ligand. We obtain the matrix values by setting each charge, in turn, on either the ligand or the receptor to +1e

and solving the linearized Poisson-Boltzmann equation for the difference in potential between the bound and unbound state at the charge centers of the other binding partner.

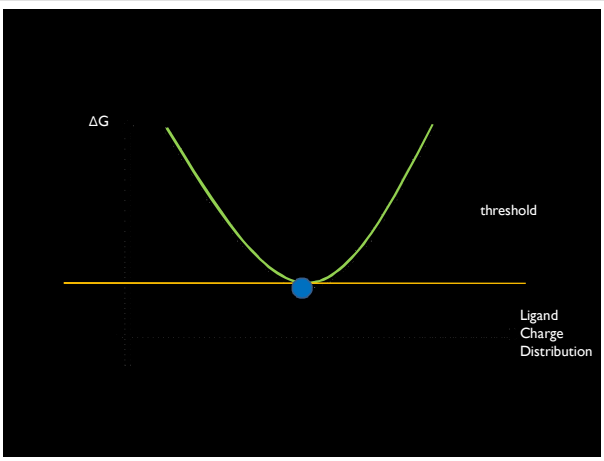
Together, the ligand desolvation penalty, the receptor desolvation penalty, and the complex interaction terms combine in the equation mentioned earlier,

$$\Delta G_{elec} = ql'Lql + qr'Cql + qr'Rqr.$$

In our research, we are trying to optimize the set of partial atomic charges on the ligand, so we treat ql as a variable.

Using the above equation as a starting point, we can solve for the ql that produces a minimum value of ΔG_{elec}—the optimal ligand charge distribution:

$$\Delta G_{elec} = xLx + qr'Cx + qr'Rqr.$$



This is a multivariate analog of a quadratic equation

$$y = ax^2 + bx + c.$$

In order to find the optimal value of x to this problem, we take the derivative with respect to ql and set it equal to 0. Taking the derivative of the original equation gives

$$0 = 2ql'L + qr'C.$$

Therefore,

$$ql_{opt} = -\tfrac{1}{2}L^{-1}\, qr'C$$

Recall that the L matrix is positive semidefinite, which means that a plot of the binding free energy versus ligand charge will result in an upward-facing n-dimensional

paraboloid, where n is the number of charges on the ligand. Therefore, the optimal ql_{opt} that we have attained must correspond to a well-defined minimum on the free energy hypersurface.

After obtaining the optimal charge distribution on the drug, one can obtain the optimal ΔG_{elec} by using the expression $\Delta G_{elec} = ql_{opt}'Lql_{opt} + qr'Cql_{opt} + qr'Rqr$.

3. METHODOLOGY

3.1 Structure preparation

Studies were initiated using a 2.2-Å resolution crystalline structure of wild-type HIV-1 Reverse Transcriptase (HIV RT) complexed with nevirapine (NVP) (PDB ID 1VRT),[1] a 2.6-Å resolution crystal structure of Y188C mutant HIV RT complexed with NVP (PDB ID 1JLF),[52] a 2.9-Å resolution crystal structure of K103N mutant HIV RT complexed with NVP (PDB ID 1FKP),[53] a 1.8-Å resolution crystal structure of wild-type HIV RT complexed with rilpivirine (TMC278) (PDB ID 2ZD1),[21] a 2.9- Å resolution crystal structure of L100I/K103N mutant HIV RT complexed with TMC278 (PDB ID 2ZE2),[21] and a 2.1-Å resolution crystal structure of K103N/Y181C mutant HIV RT complexed with TMC278 (PDB ID 3BGR).[21] In each structure, water molecules that were greater than 3.3 Å from either binding partner or that had fewer than three potential hydrogen bonding interactions were eliminated, as were any explicit ions that were sufficiently far ($\geq$ 20 Å) from the binding site. The amide groups of asparagine and glutamine residues and the imidazole group of each histidine were flipped based on examination of the possible hydrogen bonding with surrounding residues. For the Y188C RT/NVP complex**,** two oxygen atoms were deleted from the sulfinoalanine residue in order to change it into a cysteine residue. It was possible for us to make this change due to the fact that the residue was 31.75Å from the drug.

Hydrogens were modeled onto each structure using the hydrogen-building (HBUILD) facility of CHARMM,[54] using the CHARMm22 parameter set and force field.[55] CHARMm22 atom types were assigned to each atom in NVP or TMC278, and point charge magnitudes for NVP or TMC278 were computed by means of the two-stage

RESP method. [3] Using the Hartree-Fock level of theory and the 6-31g* basis set [56] within Gaussian 2003,[57] the geometry of each drug molecule was optimized and the electrostatic potential at the Merz-Kollman surface was calculated for use in the RESP fitting procedure. The 23 water molecules retained for the wild-type HIV-1/RT NVP complex and the six water molecules retained for the Y188C mutant HIV-1/RT NVP complex were assigned to the receptor based on proximity. There were no retained water molecules for the wild-type HIV-1/RT TMC278 complex, the K103N mutant HIV-1/RT NVP complex, the L100I/K103N mutant HIV-1/RT TMC278 complex, or the K103N/Y181C mutant HIV-1/RT TMC278 complex. All six crystal structures contained several areas of missing density, all of which were at least 29 Å from the binding site. The residues adjacent to these missing regions were patched accordingly with methylamide and acetamide groups.

Due to the geometric constraints on three-membered rings, placement of hydrogens on the three-membered ring of NVP was visually inspected, and the position of the single hydrogen on the carbon in the ring adjacent to the rest of NVP was manually adjusted after hydrogen building to be quantitatively similar to its position in the geometry-optimized NVP structure. It was not necessary to alter the position of any atoms on TMC278.

3.2 Charge optimization

Finite-difference methods were used to solve the linearized Poisson-Boltzman equation.[49, 58, 59] A multigrid locally-written finite-difference numerical solver of the linearized Poisson-Boltzmann Equation [60] was used to obtain the matrix and vector elements necessary for charge optimization calculations. PARSE radii were used for all

atoms in order to determine the dielectric boundary, and PARSE charges were used for receptor atoms.[61] For each structure, the calculations were run on a 225 x 225 x 225 grid resolution using a three-stage focusing procedure in which the structure occupied 23% of the grid, 92% of the grid, and 184% of the grid concentrating on the binding site. Each structure was rotated to minimize its maximal component along either the x, y, or z axis, and at the highest focusing, grid resolution was 3.7 grids/Å.

The optimal charge distribution and binding free energy were calculated using Matlab.[62] Sensitivity figures were created to show the sensitivity of each atom by displaying the radius of the atom proportionally based on the diagonal desolvation elements of the L-matrix. VMD [63] was used to generate all figures that are displayed in the Results section of this document.

4. RESULTS

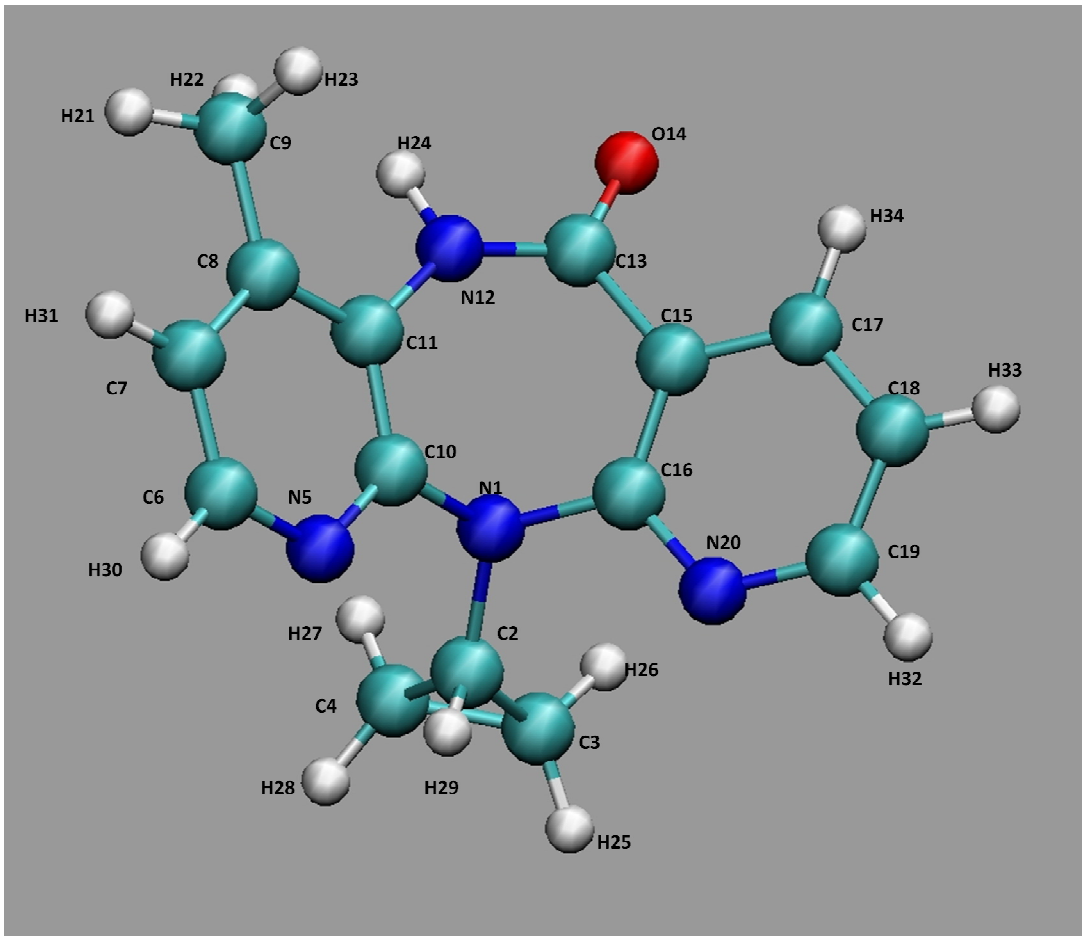
4.1 Wild-type reverse transcriptase

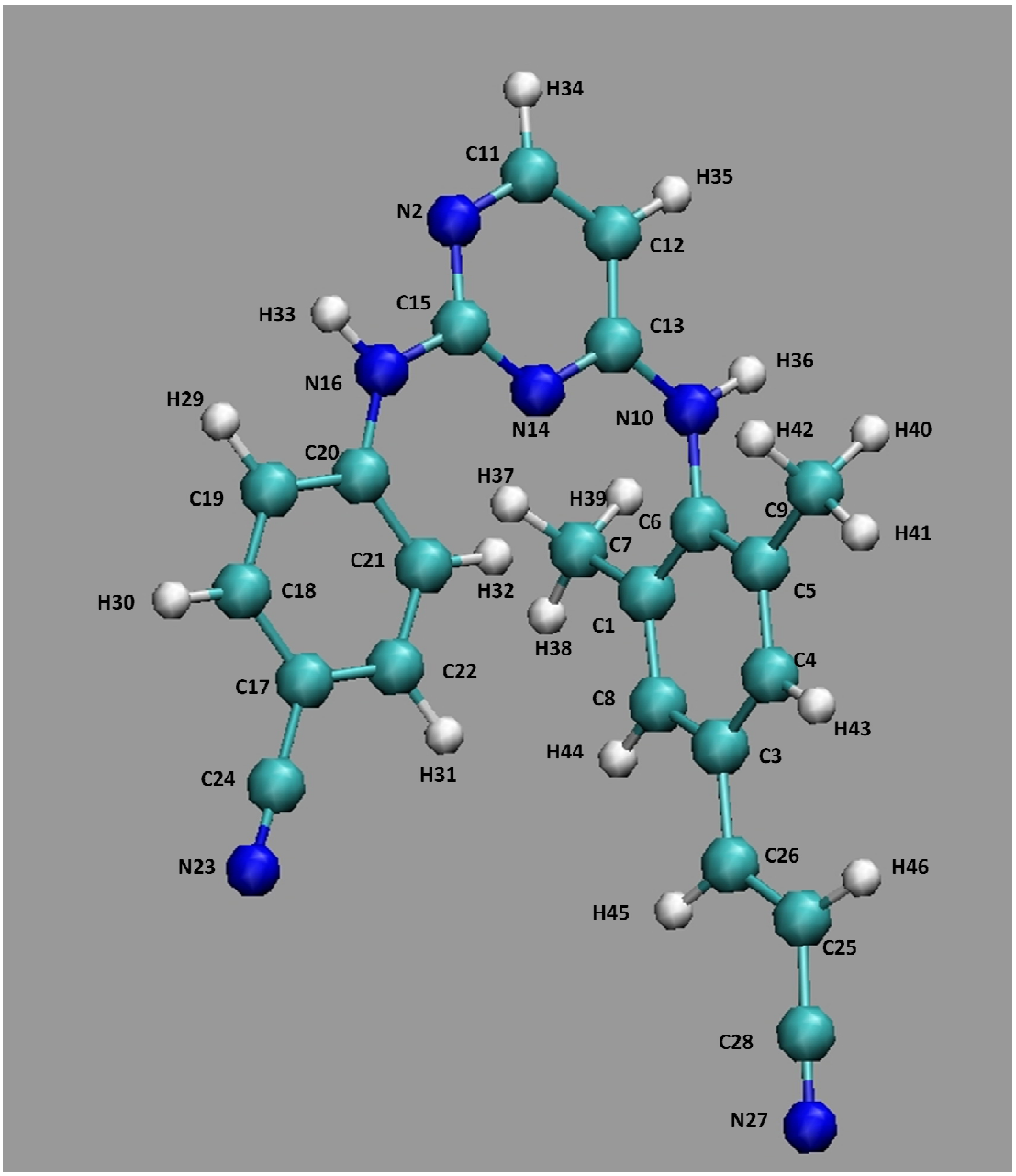
4.1.1 Nevirapine

The calculated charge distribution of NVP (fig 15a) , which is the set of charges obtained using the two-stage RESP model, shows that all nitrogens are negatively charged (from −0.27 to −0.52) and adjacent carbons atoms closest to the diazepine ring are positively charged (from 0.14 to 0.52). The C9 carbon in the methyl group is slightly negative. The carbonyl group on the diazepine ring shows expected charges with a negative oxygen and a positive carbon. C3 and C4 on the cyclopropyl ring and C17 and C18 on the pyrimidine rings have a slightly negative charge. Henceforth, we will refer to the calculated charge distribution as the "actual" charge distribution to facilitate comparison with the hypothetical, optimal charge distribution.

Figure 15b shows the optimal charges of NVP complexed with wild-type HIV-1 RT. As can be seen from the preponderance of white atoms in the figure, the optimal charge distribution is very hydrophobic. For example, these results indicate that the binding site prefers neutralization of the dipole in place of the carbonyl group. Nevertheless, there are some small dipoles in the optimal charge distribution: one at atoms C8 and C9, and another at atoms C15 and C17, and a third on the cyclopropyl group at atoms C3 and C4.

Figure 15c displays the charge difference between the optimal and the actual charge distribution of NVP complexed with HIV-1RT wild type. The figure illustrates the many changes necessary for the drug to achieve the hydrophobic optimal charge configuration. The colors on the figure indicate that nitrogen atoms N5,


H22
H23
H21
C9
H24
O14
H34
C8
C13
N12
H31
C15
C17
C11
H33
C7
C18
C10
C16
C6
N5
N1
N20
C19
H30
H27
C2
H26
H32
C4
C3
H28
H29
H25


H34
C11
H35
N2
C12
C15
H33
C13
H36
N16
N14
N10
H29
H42
H40
C20
H37
H39
C6
C9
C19
C7
H41
C21
H32
C5
H30
C18
C1
H38
C4
C22
C17
C8
H43
H44
C3
C24
H31
C26
H46
N23
H45
C25
C28
N27

N1, and N20 and oxygen O14 on the carbonyl are too negative. Carbons C7, C8, C11, C10, C13, C16, C18 and C19 are either too positive or too negative, especially those adjacent to nitrogens. Carbon C9 on the methyl group and C4 on the cyclopropyl ring are too negative. In general, these results indicate that NVP is not electrostatically optimized for tight binding because it contains too many polar groups.

4.1.2 Rilpivirine

The actual charge distribution of TMC278 (fig. 15d) complexed with wild-type HIV-1 RT appears largely neutral with only a few atoms showing a positive or negative charge. There is a set of fairly strong adjacent dipoles including C12 and C11, N2 and C15, and N16 and H33. The cyanovinyl and cyano groups each have a positively charged carbon and a negatively charged nitrogen. As is clear from the figure, this drug is more hydrophobic than nevirapine.

The optimal charge distribution of rilpivirine (fig. 15e) complexed with wild-type of HIV-1 RT is remarkable for its nearly identical appearance to the actual calculated charges for the drug. The adjacent nitrogen and carbon atoms forming dipoles on the pyrimidine ring are similarly charged to those of the actual charges. Interestingly, the N2 nitrogen forming the hydrogen bond with the amide on K101 appears highly optimized for this interaction. Figure 15f dramatically illustrates the very small differences between the optimal and actual charges of TMC278. Only the cyano and cyanovinyl groups and carbon C12 on the pyrimidine ring require a small amount of modification of charge changes ranging from a magnitude of 0.33 to 0.41.

4.2 Variants of HIV-1 reverse transcriptase

4.2.1 Nevirapine

The optimal charge distribution of nevirapine complexed with the K103N mutant (fig. 15g) appears less hydrophobic than the optimal charge distribution toward wild type. The optimal charge on the C9 carbon is positive (0.46) and the adjacent carbon C8 is negative (−0.37), forming a dipole. There is a very strong dipole between C2 and C4 on the cyclopropyl ring (0.86 and −0.79). As with optimal charges for wild type, the carbonyl group is neutralized. Figure (15h) shows the charge difference between the actual and optimal charge distributions of nevirapine complexed with the K103N mutant of HIV-1 RT. The oxygen and the nitrogens are too negative and the carbons on the diazepine ring are too positive. A difference with respect to the wild type is that the carbon in the methyl group and the dipole on the cyclopropyl ring are nearly optimal for this mutation. The dipole was absent in the optimal charges for the wild-type structure.

In order to understand why there is a strong dipole on the cyclopropyl ring, we looked at the neighboring residues and observed that there was a negative residue E138 5.76 Å away from C2 atom on the ring. This distance is closer than the 6.24 Å in wild-type. In order to understand the effect that E138 had on the optimal charge distribution, we changed the charges on the E138 to be neutral, and reran charge optimization. Figure 15l displays the values for charge differences between actual and optimal on nevirapine after the charges on the E138 were neutralized. The dipole on the ring has shifted—the C2 is slightly too positive, and the C4 is strongly too negative. In the optimal charge figure below (fig. 15k) note that the charges on the dipole are similar, but slightly weaker than the optimal charges before the E138 was neutralized. One next step is to see whether

these differences are a result of minor conformational changes in the surrounding residues or whether they are due to the actual mutation itself.

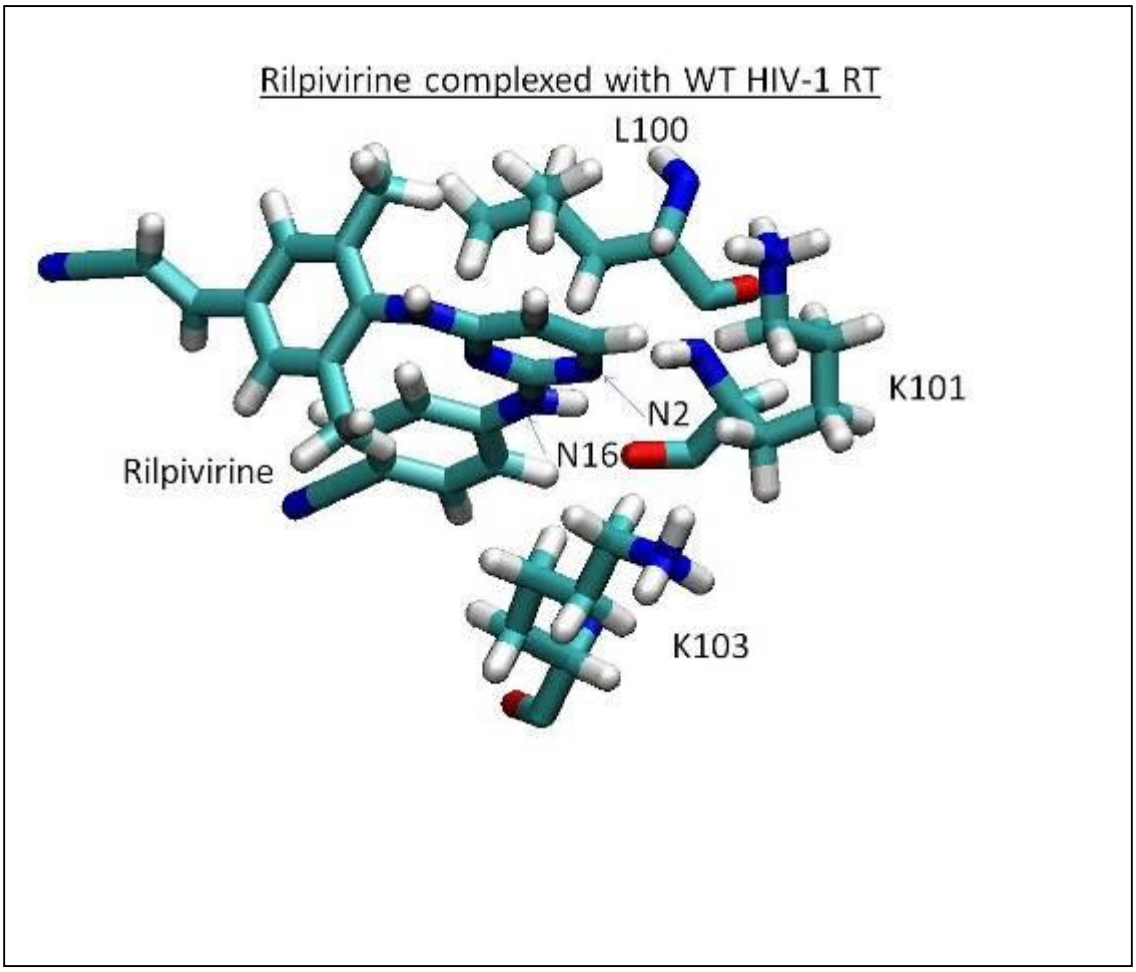
The optimal charge distribution of nevirapine complexed with the Y188C mutant (fig 15i) shows changes from the actual charge configuration including a dipole on the cyclopropyl ring and a pair of dipoles on the left pyridine ring involving C8, C11, C10, and N5. Carbon C19 is slightly negative. The other atoms appear white, meaning their ideal charge for this mutant is neutral. Figure 15j demonstrates the many changes that would be necessary for NVP to achieve optimal electrostatic binding. As compared to the mutant K103N, the dipole on the cyclopropyl ring has shifted. Nitrogen atoms N5, N12 and N20, and oxygen O14 and the neighboring carbons require positive or negative shifts to construct a series of contiguous dipoles. Carbons C2 and C4 on the cyclopropyl ring are too negative. Additional studies are needed in order to understand these results.

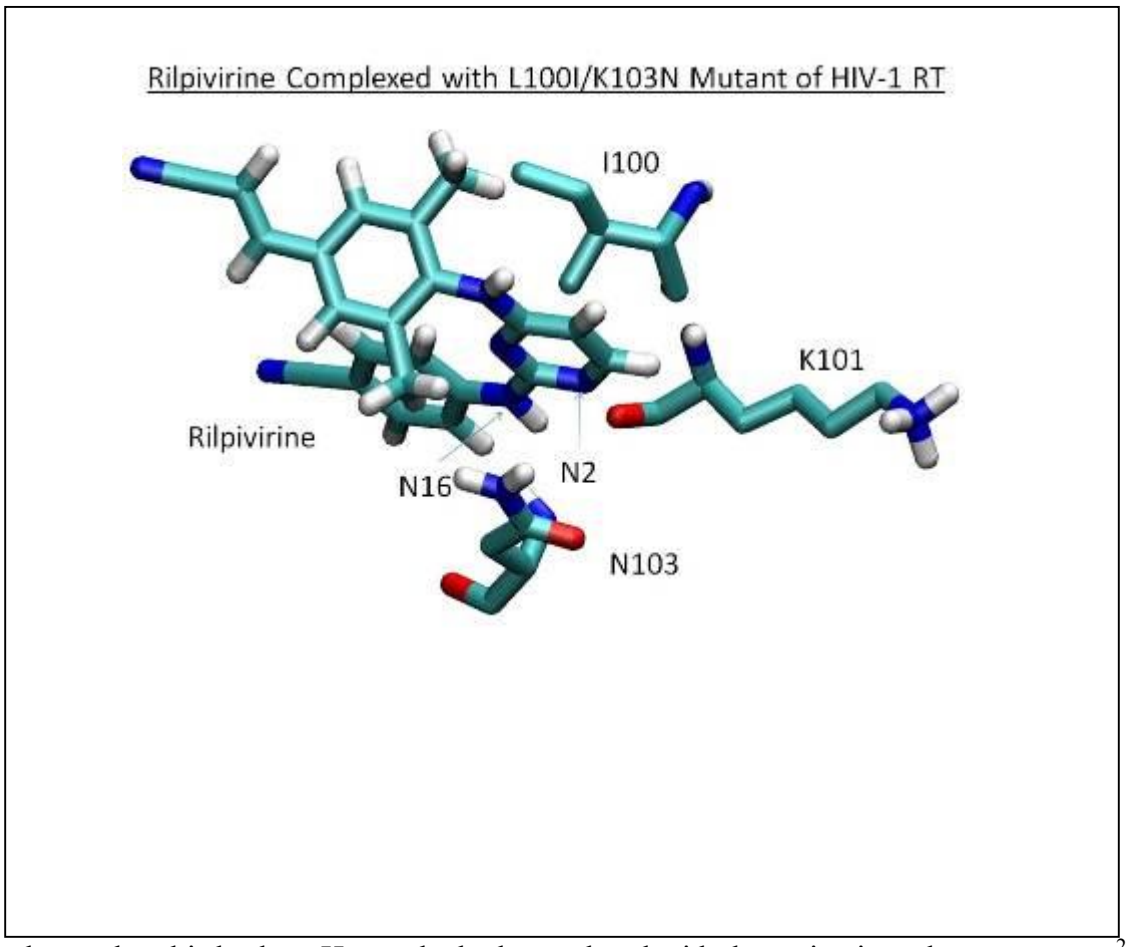
Across all three structures, the carbonyl group was not optimally charged. It needs to be neutralized to achieve optimal ΔG. The most contrasting differences across all three structures are the optimal charges on the cyclopropyl ring. For the wild type, optimal charges are neutral. For 1FKP, the K103N mutation, there is a strong dipole on C2 and C4, on Y188C mutant, there is a dipole on C2 and C3. It would be interesting to see what role the ring plays in binding.

4.2.2 Rilpivirine

The optimal charges (fig. 15m) for TMC278 for the L100I/K103N mutant shows a strong dipole at atoms C19 and C20 with charges of 0.55 and −0.55. There are two other weaker dipoles at atoms N16 and C15 and N2 and C11. The rest of the structure is neutral which is very similar to the optimal charges in the wild type. A significant

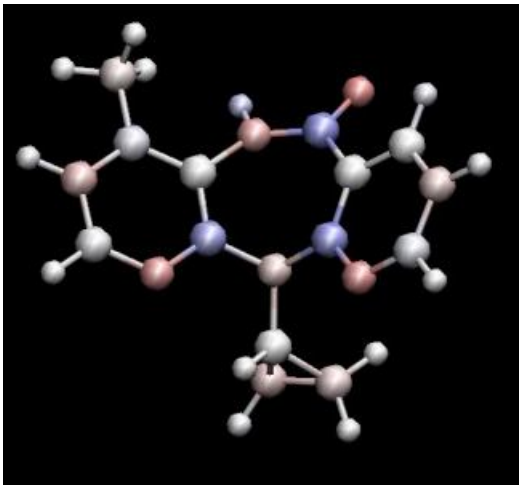
difference with wild type is that the nitrogen N2 achieves optimal binding when strongly negative, but the same atom is optimally slightly positive on the L100I/K103N mutant. Figure 15n shows the changes necessary to achieve optimal charge configuration. The charges on the adjacent carbons and nitrogens including C12, C11, N2, C15, N16, C20 and C21 require significant charge shifts in order to achieve the set of dipoles required for optimal binding. The N2 nitrogen on the central pyrimidine ring is slightly positive for optimal charge. Due to the L100I mutation, if TMC278 were to bind in the same conformation as in wild type, steric hindrance would occur. After the conformational change that the inhibitor undergoes in order to bind with this mutant, the distance between the drug and the K101 residue is larger and the geometry that is conducive to a

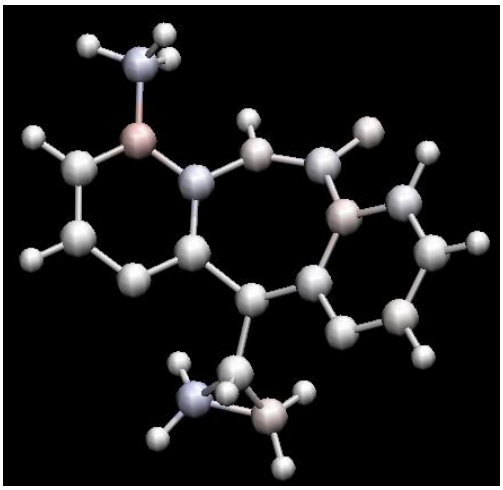


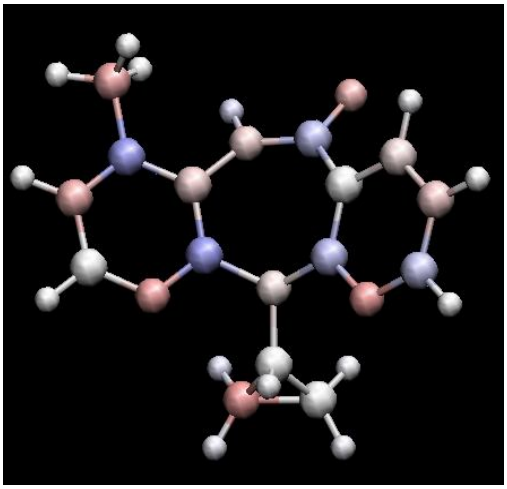


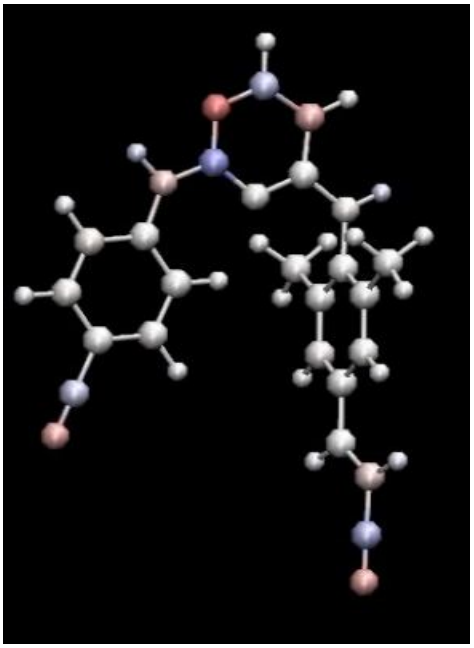
hydrogen bond is broken. Hence the hydrogen bond with the amine is no longer present.[21] In addition, the nitrogens on the cyano and cyanovinyl groups are too negative and the carbons are too positive.

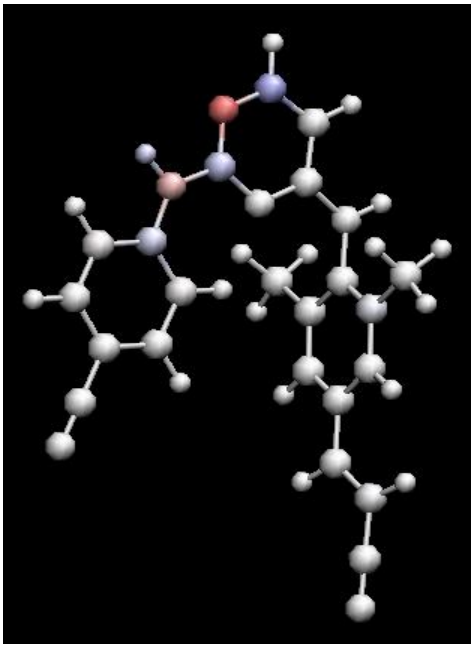
The optimal charge distribution of rilpivirine complexed with the K103N/Y181C mutant shows a strongly charged pair of dipoles (from −1.04 to 0.69) including atoms N16, C15, N2, and C11 (fig. 15o). Another dipole is optimal at C3 and C4. The remaining atoms are neutral for optimal binding configuration. Charge difference figure 15p reveals that the nitrogen atom N16 requires a strong negative shift to achieve optimal charge. The nitrogen atoms on the cyano and cyanovinyl groups are too negative and the carbons are too positive. Both require moderate shifts.

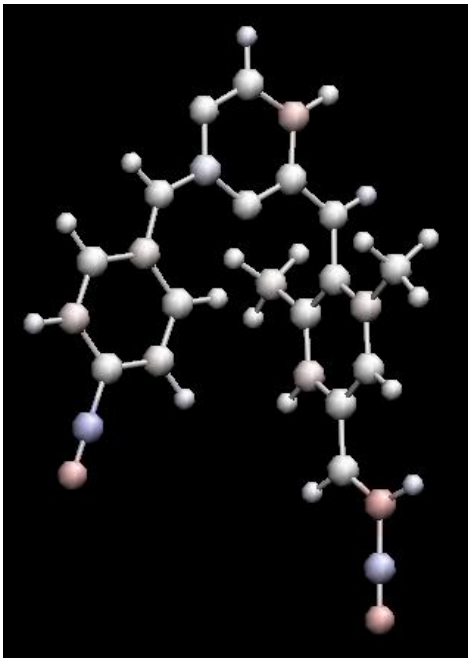


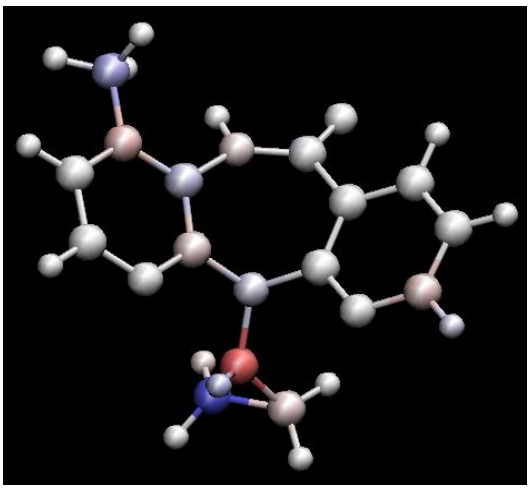


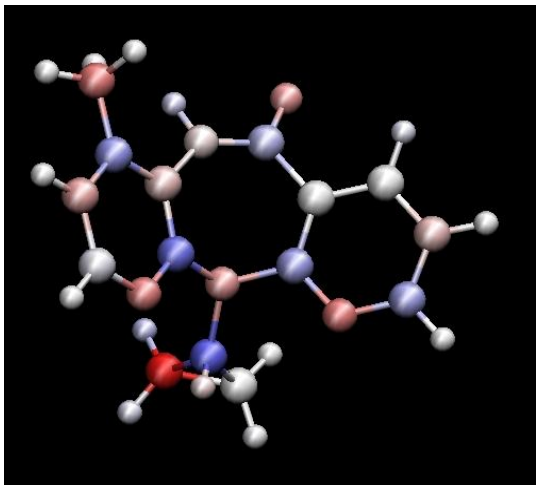


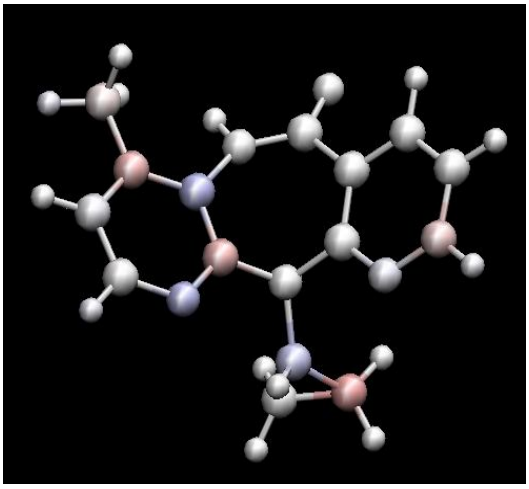


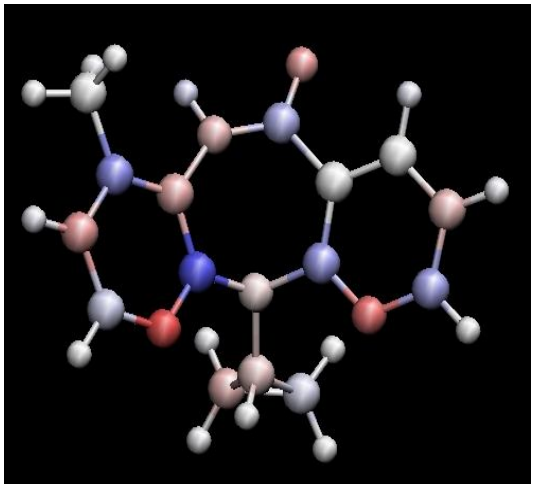


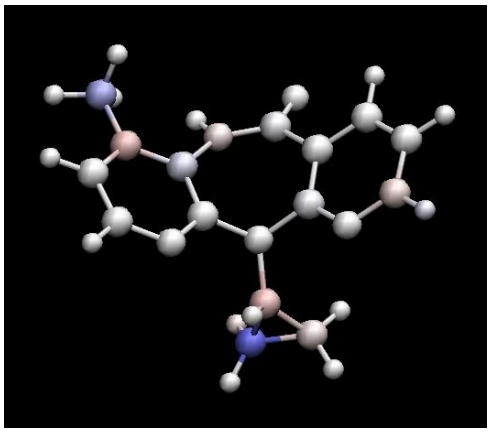


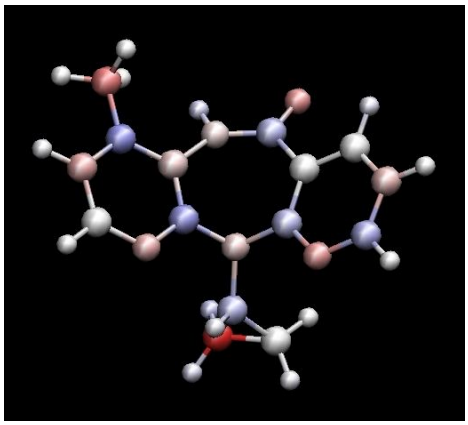


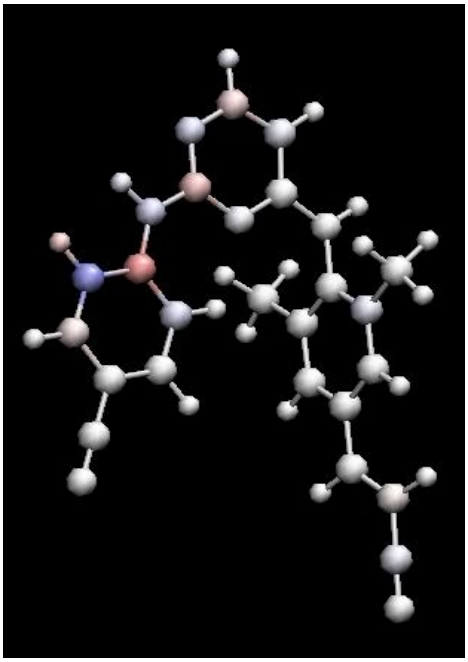


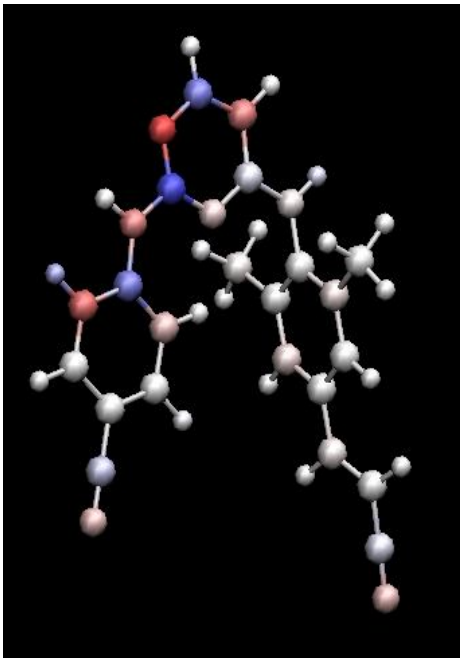


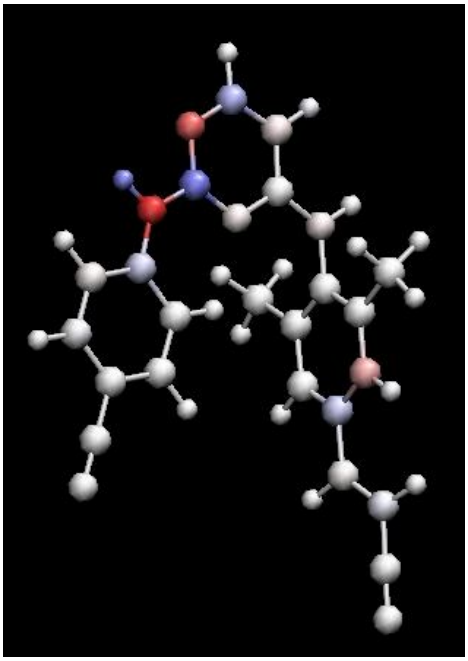


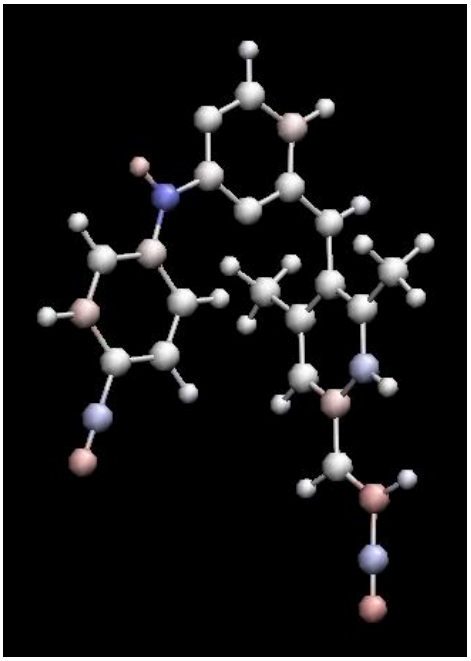












4.3 Sensitivity analysis

4.3.1 Sensitivity Figures

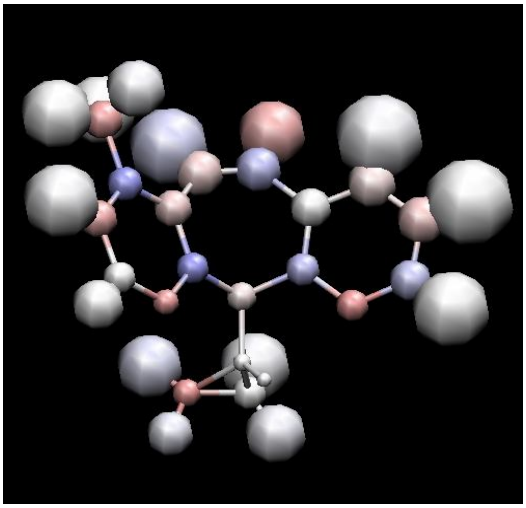
Sensitivity analysis offers guidelines for evaluating which atoms are most important as contributors to binding. In order to determine sensitivity values, we consider how much changing the charge on a given atom will affect Gibbs free energy. In the figures below, the radius of each atom is proportional to its corresponding diagonal desolvation element in the L matrix. Each element represents the difference in energy upon binding per charge per charge for a given atom center, thus capturing the change in an atom center's interaction with its own reaction field. Reality is much more complicated. To accurately measure how a change in charge will affect Gibbs free energy, it is necessary to measure the change in interaction between a point charge and the whole reaction field produced from all other point charges between the bound and the unbound states. We approximate this value when we measure sensitivity by measuring only the interaction between the point charge and its reaction to its own reaction field. Generally, ΔG will be highly sensitive to an atom's charge if the charge is highly desolvated upon binding, and far less so if the charge remains exposed to solvent upon binding.

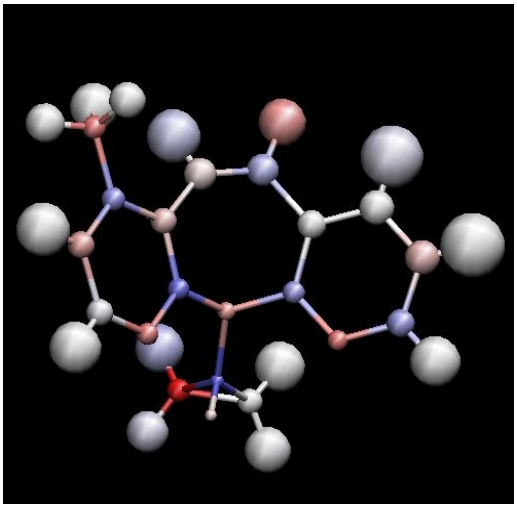
In the charge difference figures below, atoms that are larger in diameter and white are therefore "ideal" from a charge optimization perspective because they are optimally charged and their charge will have a more significant impact on ΔG of binding than atoms that are smaller in diameter. Atoms that are larger in diameter and colored are more important to "correct" to the optimal charge by reformulation because their charge will have a larger unfavorable impact on ΔG of binding when their charge is not optimal.

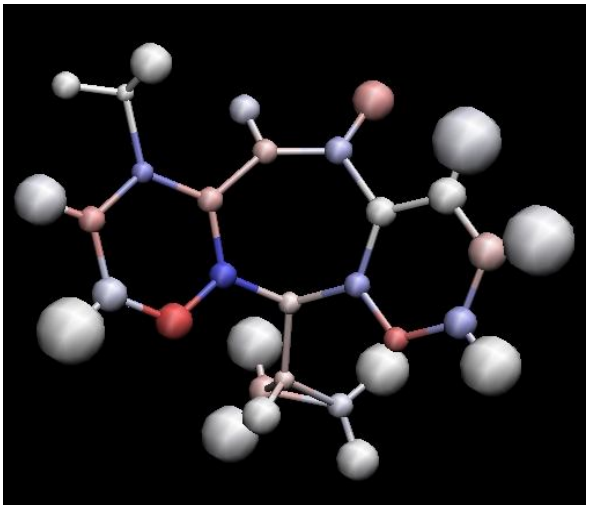
Charges on atoms with smaller diameters have a smaller impact on ΔG. For example, consider the dipole in the cyclopropyl group in figure 16b. Although the carbons are strongly charged, their small size indicates that they have a relatively small impact on ΔG of binding, even though their charges are far from optimal.

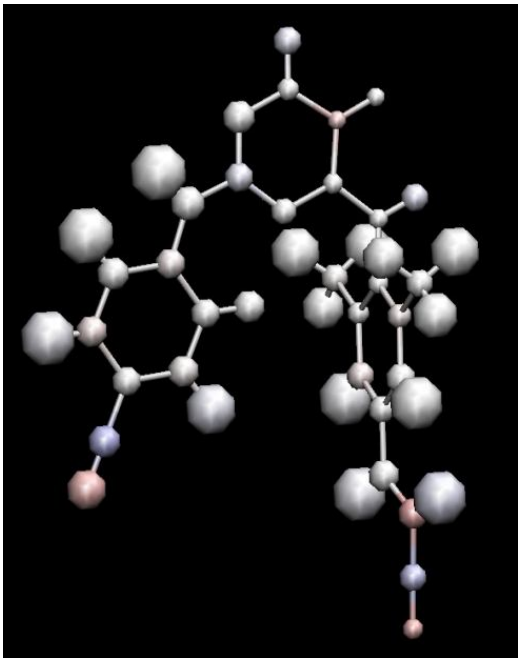
It is important to remember that sensitivity analysis give us only approximations. It would be interesting to carry out a more rigorous analysis that takes into account not only a charge's reaction to its own reaction field, but also interactions with other charges' reaction fields. This is a simple extension of our current analysis. It would also be useful to test the robustness of the current results to exact positions of each atom. If the conformation of the structure were slightly different, how much would our sensitivity analysis be affected?

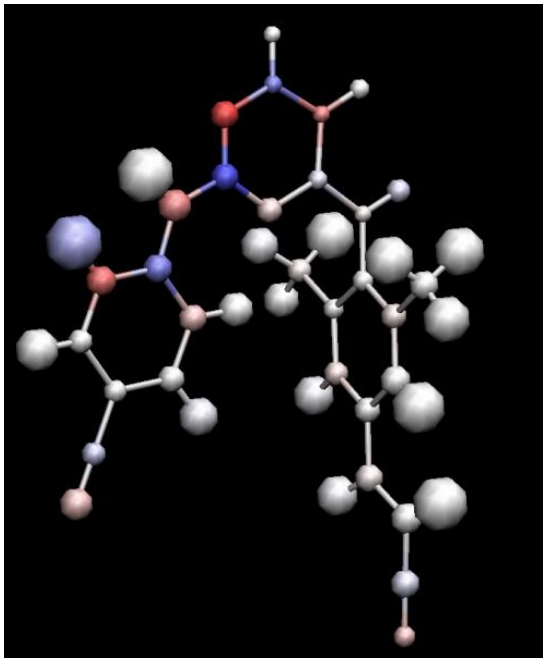
Δ Δ

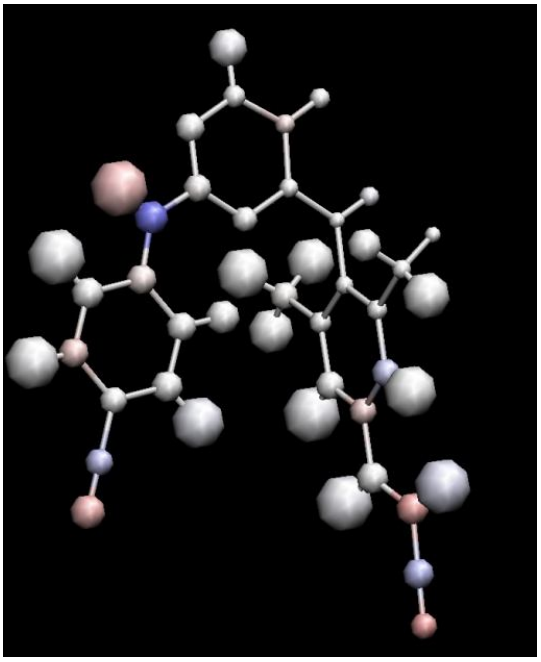












NVP complexed with wild-type HIV-1 RT (fig. 16a) shows that the binding free energy is most sensitive to the charges on the hydrogen atoms, though they are optimally charged except for H24 on the diazepine ring that forms a dipole with N12. The O14-C13 carbonyl group also highly influences ΔG, but it is not optimal. The ΔG is sensitive to all hydrogens on the cyclopropyl ring except for the H29 atom. The sensitivity values range from 27.27 kcal/mol/charge^2 for the H33 hydrogen to 4.67 for the H29 hydrogen on the cyclopropyl ring. Two of the smaller atoms, the N5 and H30, are too negative but their small radii suggests that their contribution to Gibbs free energy is relatively less significant.

Sensitivity analysis of NVP to the K103N mutant (fig. 16b) shows that ΔG is most sensitive to charge change on the hydrogen atoms, except for the H29 atom on the cyclopropyl ring. Many of the hydrogens are optimally charged, except for H24 on the diazepine ring which is too positive, and the H34 on the right pyridine ring. ΔG is not sensitive to charge change on the strong dipole on the 3-membered ring. Sensitivity values range from 27.13 for H33 to 4.94 for H29 on the cyclopropyl ring.

NVP complexed with the Y188C mutant of RT (fig. 16c) shows less sensitivity to charges than wild type or the K103N mutant. As with the other NVP variants, ΔG is most sensitive to charge change on the hydrogen atoms, and they are again optimally charged. The oxygen on the carbonyl group is too negative and binding would benefit from a charge modification because ΔG is moderately sensitive to this atom. There are several dipoles on the diazepine and pyridine rings but changing their charges will have little effect on ΔG. The lowest sensitivity value of 6.92 is on the carbon C9 on the methyl

group. As with the wild type and K103N variants, the highest sensitivity value (27.16) is on the H33 atom.

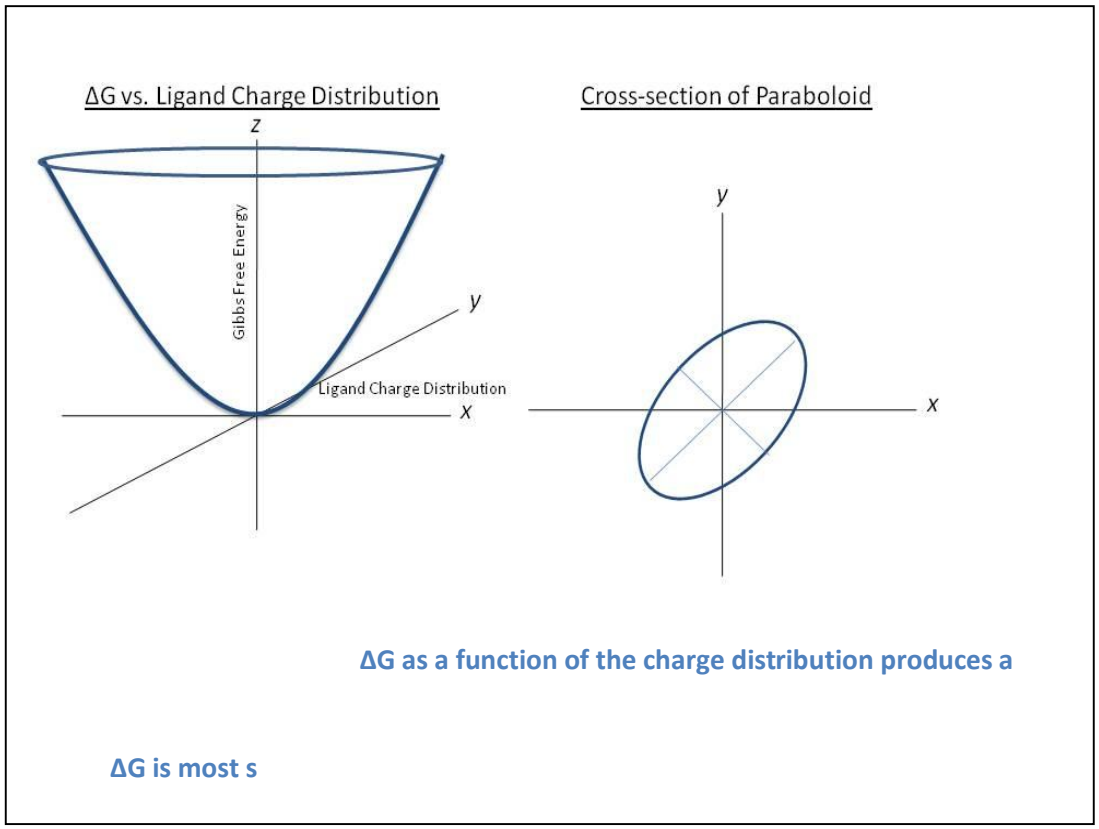
For TMC278 complexed with wild type of RT, ΔG is most sensitive to several of the hydrogen atoms (fig. 16d). It is notable that the self-contribution to the desolvation penalty is not sensitive for the cyanovinyl and cyano groups, which are the atoms that are furthest from optimal, so their lack of optimal charge may not have a measurable effect on the electrostatic binding free energy. The magnitude of sensitivity ranges from 8.64 for H35 to 26.23 for H33 and H29. Notable is the nearly optimal charge distribution across the molecule, as mentioned previously.

Charge sensitivity of TMC278 complexed with the mutant L100I/K103N (fig. 16e) shows that charge change on several of the hydrogen atoms would also have the greatest effect on ΔG. The charge on H29 is too positive, and correcting the charge would have a beneficial effect on ΔG. This mutant is interesting because, as compared to the wild type, several atoms are far from being optimally charged. Sensitivity analysis also reveals that while many atoms are far from optimally charged, their smaller radii indicate that ΔG would not be strongly sensitive to changing the charges on these atoms.

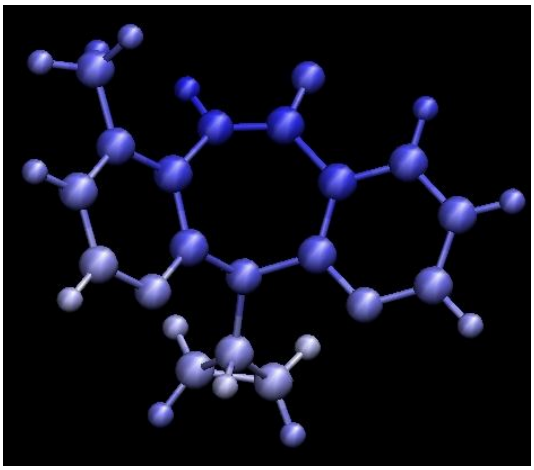
Sensitivity analysis of TMC278 complexed with mutant K103N/Y181C (fig. 16f) shows that ΔG is most sensitive to charge change on the hydrogens that are optimally charged except for H33, which is too negative. As with the other TMC278 complexes, the cyanovinyl and cyano groups are not optimally charged but they have a relatively small effect on ΔG.

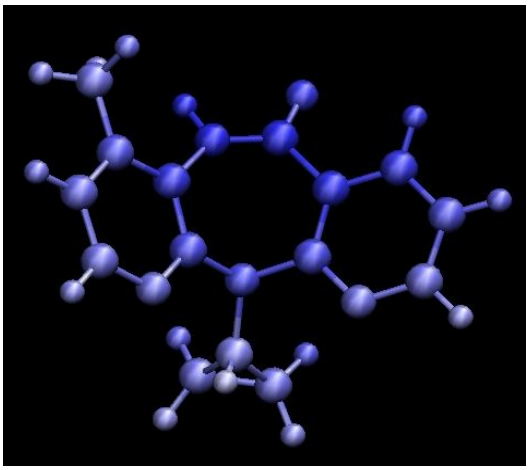
4.3.2 Eigenvalue and Eigenvector Sensitivity Analysis

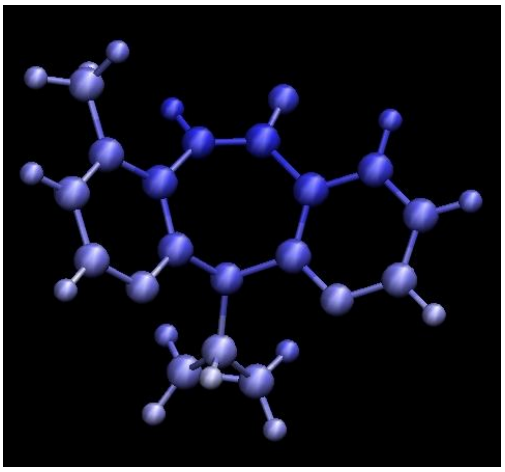
A more rigorous analysis of charge sensitivity is obtained by looking at the largest eigenvalue of the L matrix and its corresponding eigenvector. In the following figures, atoms are colored by their contribution to the eigenvector that corresponds to the largest eigenvalue. To understand what these values represent, consider the example of a simple diatomic molecule with atoms A and B with a single point charge on each atom. Graphing ΔG as a function of charge distribution on the molecule produces a paraboloid with an elliptical cross-section when viewed from above. The *x* and *y* axes are the distribution of charges on points A and B and the *z* axis is ΔG. As shown in the figure 17, the major and minor axes of the ellipse are simply the eigenvectors of the system and they generally will not be parallel with the *x* and *y* axes. The minor axis of the ellipse refers to the steepest part of the paraboloid with respect to the *z* axis. It is in this direction in charge space that ΔG is most sensitive to charge fluctuations.

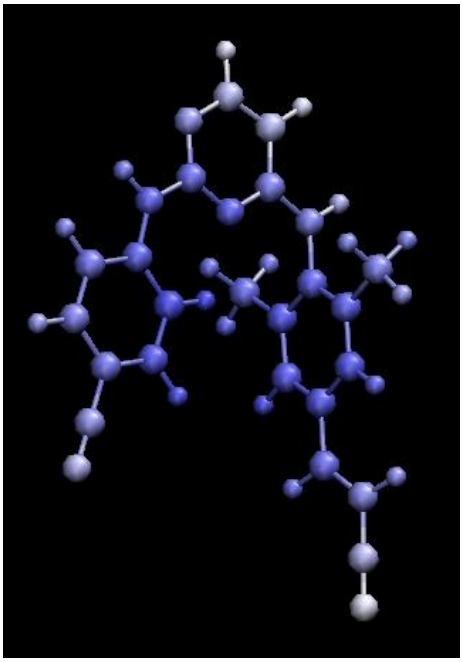


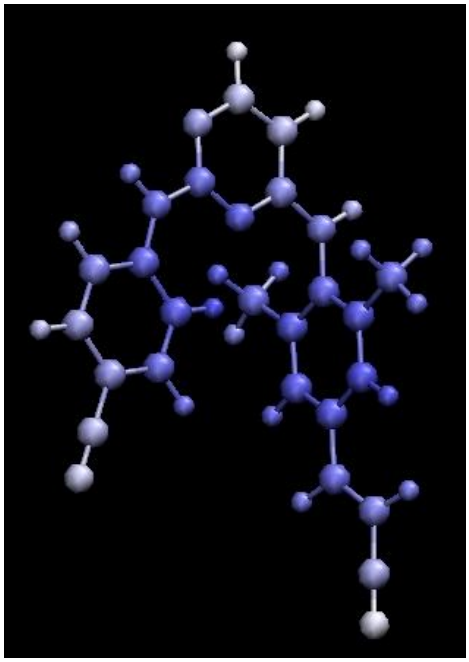
For a molecule made up of n atoms, we can take the n-dimensional vector that corresponds with the largest eigenvalue and color the atom by the value of the corresponding element of eigenvector. This will tell us the coupled direction in charge space to which the binding free energy will be most sensitive.

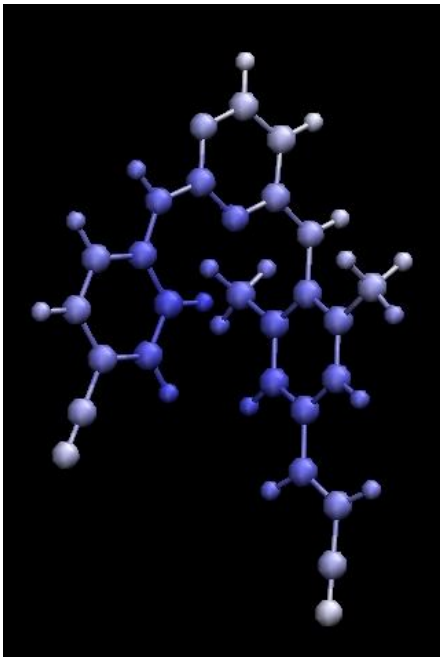












As illustrated in figure 18a, the binding free energy is most sensitive to the collective area on the NVP HIV-1 wild type complex that includes the group of atoms N12, the carbonyl group, and C15 on the diazepine ring, with values between −0.21 and −0.23. ΔG is least sensitive to charge changes on the hydrogen atoms on the cyclopropyl group with values as low as −0.09.

NVP shows ΔG sensitivity to charge changes on the K103 mutant of RT in figure 18b that are similar to the wild type but extending to C17 and H34 on the right pyridine ring. Values range between −0.20 and −0.23. The least influential group of atoms to ΔG includes the carbons and hydrogens on the cyclopropyl ring with H29 having the lowest value of −0.09.

For the Y188C mutant, ΔG is most sensitive to the area as shown in figure 18c that extends from N5 on the left pyridine ring to C10, N1, C16, C15, C17 and H34. Values range from −0.20 to −0.22. The area where ΔG is least sensitive to charge change includes the hydrogens on the methyl group with values from −0.09 to −0.11.

The most collectively sensitive area on TMC278 complexed with wild type HIV-1 RT, shown in figure 18d, where charge change would have the greatest influence on ΔG, includes C21, C22, H31 and H32 and all of the carbons and H44 on the right methyl phenyl ring. Values range as large as −0.19. The areas where ΔG would be least sensitive to charge change include the nitrogen atoms on the cyano and the cyanovinyl groups, and H34, H35, and N10. Values range as low as −0.06.

For the structure of TMC278 complexed with mutant L100I/K103N (fig. 18e), Gibbs free energy would show most sensitivity to charge changes in atoms including C20, C21 and C22 and H31 and H32 on the left ring, H37 on the left methyl group, N14

on the pyrimidine ring, and the carbons on the right methyl phenyl ring with values ranging from −0.16 to −0.19. Areas where ΔG would be least sensitive to charge change are the nitrogen atoms on the cyanovinyl and cyano groups and H34 and H35 on the pyrimidine ring with values ranging from −0.04 to −0.07.

The TMC278 molecule complexed with mutant K103N/Y181C of HIV1 RT in figure 18f shows that ΔG would show most sensitivity to charge change on the carbon-hydrogen pairs on the left methylphenyl ring, N14 on the pyrimidine ring, and the carbons and hydrogens on the inside of the right methyl phenyl ring with values as negative as −0.20. The areas least sensitive to influencing ΔG are the nitrogens on the cyanovinyl and cyano groups and the hydrogens H34, H35, H36, H40, H41, and H42. Their sensitivity values range from −0.07 to −0.11.

5. DISCUSSION

5.1 Nevirapine

The theoretical optimal charge distribution for nevirapine is strongly hydrophobic. This finding is in agreement with the known hydrophobic nature of the binding site.[27]

For wild-type and mutants, sensitivity analysis showed that the ΔG was especially sensitive to charge change on the carbonyl group but at the same time, the actual charges on the atoms were far from optimal. It is possible that the carbonyl group offers an ideal target for reformulation. Additional literature research is required in order to study the work done since the development of nevirapine for improved drug design. This finding may have already been addressed in developing phase two NNRTI's.

An additional area for research involves further exploring our finding that the theoretical optimal charges on the nitrogen atoms on nevirapine are positive. We are interested in studying whether drug improvement research after nevirapine addresses this apparent structural deficiency. From an electrostatics perspective, the drug may be improved if carbons replaced the nitrogens.

In order to understand the optimal charge distribution more fully, we neutralized charges on nearby residue GLU138 and examined the effects of this change on the drug's optimal charge distribution. This additional analysis illustrated the glutamic acid residue's contribution to ΔG of binding. More studies are required to understand the curious set of optimal charges on the 3-membered ring.

These considerations bring us back to our original problem—how to find a balance between specificity and affinity. If we redesign the drug to be more hydrophobic, it is possible that its specificity for reverse transcriptase will be lost. The carbonyl and

the nitrogen atoms may be necessary to bind to reverse transcriptase alone and not to other proteins. It may be necessary to give up complementarity of binding, with its possible toxic side effects, in order to achieve specificity.

5.2 Rilpivirine

According to our electrostatics studies, rilpivirine is nearly optimal across all three variants. This is likely reflective of the near decade of research and development that took place between nevirapine's introduction and that of rilpivirine which was marked by enormous progress in computational methods for improving drug design. Nevertheless there exist some interesting findings that warrant further study. The N2 nitrogen on the central pyrimidine ring is of particular interest. In the wild-type and K103N/Y181C mutant, the optimal charge on N2 is very negative. Upon visual exploration, we noted that the N2 forms a hydrogen bond with the amino terminus of the K101 residue. In the L100I/K103N mutant, however, the optimal charge for the N2 nitrogen is positive, requiring a very large shift to achieve optimal charge. According to Das[21], the K101 shifts in the L100I mutation due to steric hindrance when the leucine changes to the larger isoleucine. The hydrogen bonding between the N16 nitrogen, the "connector" nitrogen, and K101 breaks down due to conformational change. Based on our observations, a similar break occurs with the N2 nitrogen. In the L100I mutation, the N2 becomes closer to the oxygen on the carbonyl group of the K101, which would explain why the optimal charge for the N2 becomes positive.

An unusual finding that requires additional study is the actual charge distribution on the nitrogens on the pyrimidine ring that was obtained using the RESP fitting model. We obtained unexpected results with very different charges of -0.69 for N2 and -0.14 for

N14. The charges are unusual because we would have expected that the nitrogens would have similar charges. Additional studies required would be to run RESP fitting again with a structural variation of replacing one of the connector nitrogens, either N10 or N16, with a CH2 or C=CH2. This would help determine how much of an effect the area surrounding the nitrogens on the pyrimidine ring was having on the actual nitrogen charges, and to hopefully validate this unusual actual charge distribution.

5.3 Limitations and extensions of charge optimization

Charge optimization is a useful way to learn more about the electrostatic properties of the ligand and the receptor. However, this approach makes many assumptions that limit the application of its results. A major limitation is that it allows for comparisons of ΔG's only between identically-shaped ligand, receptor, and complex shapes. For instance, if we replace the carbonyl on nevirapine with a group whose electrostatic properties resembled the optimal distribution at this location, it would be necessary to replace it with a functional group that has the same shape in order for charge optimization results to rigorously hold. This limitation exists because we assume the shape of the ligand that we are studying remains fixed. However, recent work suggests that changes that change shape but alter the charge distribution to be closer to optimal may still be practically implemented with success.[64] A second limitation is that we assume rigid binding for both the ligand and the receptor. This means that we assume that the ligand and receptor do not internally change conformation upon binding, which is not true in most cases. Another limitation is the continuum electrostatics framework of our model. Charge optimization does not model the explicit dynamics of the water molecules, and so it simplifies reality. Of course, charge optimization overall only measures the

electrostatic interactions, ignoring van der Waals interactions and other components of ΔG. One must take into consideration these other components before implementing a change to the ligand if such a change alters the shape or other properties of the molecule. As a whole, charge optimization can provide interesting and valuable supplementary information about a system, and can inform design when used in conjunction with other modeling strategies.

To fully interpret the results of charge optimization, one should couple them with sensitivity analyses, that is, one should determine how sensitive ΔG of binding is to charge changes on individual atoms. One future goal is to develop sensitivity analyses that are more precise and can provide more easily-interpretable information. We can develop this analysis using eigenvalues and eigenvectors.

In this study, we analyzed the electrostatic determinants of the HIV-1 reverse transcriptase system using charge optimization by studying two non-nucleoside reverse transcriptase inhibitors, nevirapine and rilpivirine. This was done in order to understand the characteristics of the HIV-1 reverse transcriptase binding pocket so that it is possible to develop more effective drugs to treat HIV/AIDS.

6APPENDICES

7. REFERENCES