Conformational Landscape of the Human Immunodeficiency Virus Type 1 Reverse Transcriptase Non-Nucleoside Inhibitor Binding Pocket: Lessons for Inhibitor Design from a Cluster Analysis of Many Crystal Structures

Kristina A. Paris,†,‡ Omar Haq,†,‡ Anthony K. Felts,†,‡ Kalyan Das,†,§ Eddy Arnold,*,†,§ and Ronald M. Levy*,†,‡

†Department of Chemistry and Chemical Biology, Rutgers University, 610 Taylor Road, Piscataway, New Jersey 08854, ‡BioMaPS Institute for Quantitative Biology, Rutgers University, Piscataway, New Jersey 08854, and §Center for Advanced Biotechnology and Medicine, Rutgers University, 679 Hoes Lane, Piscataway, New Jersey 08854

Received June 11, 2009

Clustering of 99 available X-ray crystal structures of HIV-1 reverse transcriptase (RT) at the flexible non-nucleoside inhibitor binding pocket (NNIBP) provides information about features of the conformational landscape for binding non-nucleoside inhibitors (NNRTIs), including effects of mutation and crystal forms. The ensemble of NNIBP conformations is separated into eight discrete clusters based primarily on the position of the functionally important primer grip, the displacement of which is believed to be one of the mechanisms of inhibition of RT. Two of these clusters are populated by structures in which the primer grip exhibits novel conformations that differ from the predominant cluster by over 4 Å and are induced by the unique inhibitors capravirine and rilpivirine/TMC278. This work identifies a new conformation of the NNIBP that may be used to design NNRTIs. It can also be used to guide more complete exploration of the NNIBP free energy landscape using advanced sampling techniques.

Introduction

Current strategies for treatment of HIV involve hindering different steps in the retrovirus’ life cycle. This study focuses on the inhibition of the viral enzyme reverse transcriptase (RT) by non-nucleoside reverse transcriptase inhibitors (NNRTIs). NNRTIs are noncompetitive inhibitors that bind to a pocket called the non-nucleoside reverse transcriptase binding pocket (NNIBP) which lies ∼10 Å from the enzyme’s polymerase active site.1

Analysis of crystal structures has suggested three possible mechanisms of inhibition (which are not mutually exclusive) of HIV-1 RT by NNRTIs: (1) restriction of the p66 thumb flexibility; (2) distortion of catalytically essential residues at the polymerase active site; (3) displacement of the primer grip. In each of these mechanisms, the binding of NNRTIs is proposed to lead to conformational perturbations and to limit conformational flexibility required for efficient DNA synthesis by RT. In the “molecular arthritis” mechanism, conformational restriction of the p66 thumb subdomain was suggested to limit flexibility of the enzyme required for catalysis.1 NNRTI binding may restrict the mobility of the thumb subdomain1–3 or may change the direction of movement of the thumb subdomain,4 thus slowing down or preventing template-primer translocation and inhibiting facile elongation of nascent viral DNA. NNRTI binding perturbs the configuration of the RT polymerase active site region, including the catalytically essential D110, D185, and D186 residues,5 and limits conformational changes of the “YMDD loop” containing M184 and D185.6 The primer grip is a structural element in HIV-1 RT that has been proposed to be essential for positioning the primer 3’ terminus at the active site,7 and movements of the primer grip and the associated thumb subdomain are thought to be critical for the translocation of nucleic acid following incorporation of nucleotides during polymerization.2 NNRTI binding causes a significant displacement (∼4 Å) of the primer grip, leading to possible inappropriate positioning of the primer terminus at the active site; this conformational alteration and possible restriction of primer grip mobility may be a major contributor to inhibition by NNRTIs.4 Movement in the primer grip is also thought to affect allosteric hinge-bending movements in the position of the thumb subdomain (the tip of which lies ∼30 Å from the NNIBP).1,9,10 If structures are superimposed on the basis of the β6-β9-β10 strands, the thumb subdomain of NNRTI-bound RT is rotated by ∼40° relative to that in the unbound apo enzyme10,11 (see Figure 1).

The NNIBP is very flexible, changing conformation when different NNRTIs are bound.12 This has been described as a “shrink-wrap” effect where the binding pocket residues change conformation to form stabilizing interactions with a ligand.13 Although available inhibitors have different shapes, sizes, functional groups, and binding modes, they display a number of common features in their interactions with the NNIBP residues: aromatic ring(s) capable of forming π–π stacking interactions with aromatic residues, as well as making hydrophobic contacts with other nonpolar pocket residues, and (usually) hydrogen bond (H-bond) donors capable

© XXXX American Chemical Society
of forming an H-bond with the backbone carbonyl oxygen of K101. Current design of inhibitors focuses on developing the “required” interactions in such a way to ensure high activity, reasonable solubility, and broad potency against drug-resistant variants.

As there are many different NNRTIs, it is reasonable to assume that there are many different conformations of the NNIBP. Here, 99 X-ray structures of HIV-1 RT from the Protein Data Bank (PDB) are examined. Of these, 52 are wild-type (WT) RT bound to NNRTIs, 30 are mutant forms of RT bound to NNRTIs, 3 are unliganded WT RT, 3 are unliganded mutant RT, 10 contain DNA or RNA substrates or ATP, and 1 is bound to the RNase H inhibitor DHBNH. (Note: In this instance, WT refers to an enzyme with no mutations within 15 Å of the NNIBP.) These 99 structures represent an ensemble of observed conformations of the NNIBP with perturbations created by mutation, binding of different ligands (induced fit effects), and different crystal forms and are listed in Table 1.

The conformational plasticity of the binding pocket plays an important role in drug design. Here we report the analysis of conformational states of the NNIBP in 99 available crystal structures of HIV-1 RT. Each X-ray structure represents a point on the conformational landscape for binding NNRTIs.

The goal of characterizing a conformational landscape and its corresponding energy landscape has come to occupy a central role in biophysics. This study infers information about the landscape for ligand binding to HIV-1 RT by performing a cluster analysis of this large data set of X-ray crystal structures. Our cluster analysis focuses on the conformational states of the binding pocket, whereas previously published studies have used clustering primarily to characterize the flexibility, chemical class, and binding mode of the ligand.

The availability of a large data set of HIV-1 RT crystal structures in the PDB and their clustering provides information about the locations of free energy basins and their shapes. Ideally, the populations of the different X-ray resolved conformations of the NNIBP of HIV-1 RT could be transformed through Boltzmann statistics into a free energy landscape of the receptor in the spirit of free energy folding funnels proposed for proteins in general. Folding funnels are rugged in the vicinity of the native fold of the protein, suggesting energetically competitive and similar conformations that provide an enhanced means of interaction between the protein and either ligands or other proteins. The landscape provides useful information about both the different means for inhibitors to bind to HIV-1 RT and the strain free energy required to adopt a particular conformation for binding.

Highly populated clusters may suggest that the deformations within the NNIBP are locally elastic with small free energy penalties. In contrast, sparsely populated clusters are suggestive of more steeply sloped free energy basins. However, as the 99 X-ray structure data set does not represent a systematic sampling of the landscape, the populations may reflect the bias found in the drug design process where inhibitors are often designed on the basis of earlier inhibitors or are designed for previously determined structures of the NNIBP. Even so, the NNIBP conformations representative of the sparsely populated basins provide opportunities for exploiting new interactions and ligand conformational freedom in developing new more potent NNRTIs.

### Results

The average root-mean-square deviation (rmsd) of all Ca atoms within 15 Å of any NNRTI across the set of 99 X-ray structures of RT is only 1.23 ± 0.48 Å when the superposition is performed on the same set of Ca atoms. This increases to 1.58 ± 0.59 Å for all Ca atoms within 10 Å of any NNRTI. An analysis of the radii of gyration for each of the Ca atoms in 82 RT/NNRTI complexes shows large variation in

### Table 1. The 99 Crystal Structures of HIV-1 RT Used in This Analysis

<table>
<thead>
<tr>
<th>PDB code</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT/NNRTI</td>
</tr>
<tr>
<td>1BQM, 1C0T, 1C0U, 1C1B, 1C1C, 1D7Q, 1D1T, 1EET, 1EP4, 1FK9, 1HNI, 1HNV, 1IKW, 1JLQ, 1KLM, 1LW0, 1LW2, 1LWE, 1REV, 1RT1, 1RT2, 1RT3, 1RT4, 1RT5, 1RT6, 1RT7, 1RTH, 1RTI, 1SEP, 1S6Q, 1S9E, 1S9G, 1SUQ, 1TFT, 1TXX, 1TKZ, 1TL1, 1TL3, 1TV6, 1VT1, 1WT, 1VRU, 2B5J, 2B6A, 2BAN, 2BE2, 2OPP, 2VG5, 2VG6, 2VG7, 2ZD1, unpublished</td>
</tr>
<tr>
<td>Mut/NNRTI</td>
</tr>
<tr>
<td>1FKO, 1FKP, 1HPZ, 1HQQ, 1IKV, 1IKX, 1IKY, 1IKH, 1JLA, 1JLB, 1JLC, 1JLF, 1JLG, 1LWC, 1LWF, 1S1T, 1S1U, 1S1V, 1S1W, 1S1X, 1S5V, 2HND, 2HNY, 2HNZ, 2IC3, 2OPQ, 2OPR, 2OPS, 2ZEE, 2ZGR</td>
</tr>
<tr>
<td>WT unbound</td>
</tr>
<tr>
<td>1D1O, 1HMV, 1RTJ</td>
</tr>
<tr>
<td>Mut unbound</td>
</tr>
<tr>
<td>1HQE, 1JLE, 1QEE</td>
</tr>
<tr>
<td>DNA/RNA/ATP bound</td>
</tr>
<tr>
<td>1HY5, 1J5O, 1N5Y, 1N6Q, 1ROA, 1RTD, 1T03, 1T05, 2HML, 2IAJ</td>
</tr>
<tr>
<td>RT/RNase H I</td>
</tr>
<tr>
<td>2B5J</td>
</tr>
</tbody>
</table>

**WT/NNRTI**: WT RT complexed with an NNRTI. Mut/NNRTI: mutant RT complexed with an NNRTI. WT unbound: Apo WT RT. Mut unbound: Apo mutant RT. DNA/RNA/ATP bound: RT bound to substrates DNA, RNA or ATP. RT/RNase H I: RT complexed with a RNase H inhibitor.
some regions in the vicinity of the NNIBP such as the primer grip $\beta_{12}$-$\beta_{13}$-$\beta_{14}$ and the loop around P95 through L100 while other regions such as the $\beta_{6}$-$\beta_{9}$-$\beta_{10}$ sheet remain more static (see Figure 2). This analysis serves to indicate regions on which to align the ensemble of 99 X-ray structures (the most variable) and on which to focus clustering experiments (the least variable).

Analysis of the side chains that point into the NNIBP of the 52 WT NNRTI-bound conformations of RT displays only three residues that give discrete clusters when clustered individually: Y181, Y183, and Y188. The remaining side chains (L100, K103, D186, F227, W229, and L234 from p66; N136 and E138 from p51) fluctuate across the ensemble of structures, but the distribution is quasi-continuous and therefore does not allow separation into meaningful clusters.

Further investigation of the NNIBP backbone (primarily in the primer grip region) and side chains (residues 181, 183, 188) utilizing hierarchical clustering techniques elucidates eight
basins that depict varying conformations of the flexible binding pocket. The eight basins include one large cluster of 73 structures, four small clusters composed of 2–14 structures, and three singletons (Figure 3 and Table 2). The large cluster is composed solely of NNRTI-bound structures with Y181 and Y188 side chains both occupying bound “open” conformations where the two side chains have moved to open a pocket that accommodates the ligand. The small cluster of 14 (small cluster 1) includes unliganded structures and those containing dsDNA, RNA/DNA, and/or dNTP. Since no NNRTI is bound in these 14 conformations, W229, F227, and Y188 fill the space where the ligand would be found. This difference in positioning of W229 and its connected primer grip as well as the positioning of the side chains of residues 181 and 188 in the unbound “closed” position allows for separation from structures that are bound to NNRTIs. The primer grip conformation seen in this small cluster will be referred to as NNRTI−, while the primer grip conformation seen in the large cluster, where there is a shift in the primer grip of −3.4 Å (see Table 2), will be referred to as NNRTI+.

Most interesting is the identification of a small cluster of three RT/rilpivirine complexes (small cluster 4; PDB IDs 2ZD1, 2ZE2, 3BGR) and two singletons (1EP4, bound to the NNRTI capravirine, and 1RTJ, where an NNRTI was washed out prior to structure determination). The primer grip conformation seen in these 14 structures, W229, F227, and Y181, and Y188 fill the space where the ligand would be found. This difference in positioning of W229 and its connected primer grip as well as the positioning of the side chains of residues 181 and 188 in the unbound “closed” position allows for separation from structures that are bound to NNRTIs. The primer grip conformation seen in this small cluster will be referred to as NNRTI−, while the primer grip conformation seen in the large cluster, where there is a shift in the primer grip of −3.4 Å (see Table 2), will be referred to as NNRTI+. Additionally, CP-94,707, is the only structure in which a different discrete NNIBP and therefore would not be expected to have a primer grip conformation similar to the NNRTI-bound large cluster. This small cluster therefore offers insight into possible unique interactions near the NNIBP that may be exploited for design of new NNRTIs that can stabilize the primer grip in a perturbed conformation that disrupts polymerase activity. The final singleton, 1TV6, bound to the large ligand CP-94,707, is the only structure in which a different discrete conformation of Y183 is seen. 1TV6 also is a case in which both Y181 and Y188 are in the “closed” position.

Additional Cluster Analysis of the RT “Thumb” Region. A large conformational change occurs upon binding of nucleic acid where the thumb and fingers of RT move to “clasp” the nucleic acid; a similar change in the thumb conformation is also apparent upon binding of an NNRTI (see Figure 1). As movement in the primer grip is thought to affect allosteric hinge-bending movements in the position of the thumb subdomain,1,9,10,23 an additional cluster analysis on three residues at the tip of the p66 thumb subdomain of the structures was performed in an attempt to give more information about the large cluster. The clustering level with the largest separation ratio yields nearly identical results to clustering on the primer grip. Further analysis using a smaller separation ratio for selection of the cluster level results in the separation of the 80 RT/NNRTI complexes that occupy the large cluster into one singleton (1JLE) and two subclusters of 28 and 51 structures between which a small shear or twist of the primer grip is seen. The separation of the two subclusters is due to a shift in position of the tip of the thumb corresponding to an average rmsd between clusters of 5.7 ± 1.7 Å. The shift in thumb position is most likely due to the different crystal forms used in structure determination, as the structures in the cluster of 51 utilize one crystal form while the cluster of 28 utilize one of two differing crystal forms. Influence of the crystal packing propagates to the NNIBP, causing a slight shear or twist seen in the primer grip. However, these subclusters overlap in the primer grip region and so are not discernible by clustering on the primer grip alone; the effect of crystal form on the primer grip conformation is minimal.

**Discussion**

The existence of several clusters indicates that structural variability is present, but since most of the structures are in one cluster, that variability is not evenly distributed across the NNIBP landscape. Most obvious, the NNRTI-bound structures are separated from those of RTs not bound to NNRTIs.

---

**Table 2. Clustering Results: Eight Basins with Different Features**

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Cluster Members, PDB IDs</th>
<th>Residue 181, 188</th>
<th>β12-β13</th>
<th>NNRTI Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large</td>
<td>2OOP, etc. (73 structures)</td>
<td>open, open</td>
<td>0.0</td>
<td>to large Clust Rep (2OOP)</td>
</tr>
<tr>
<td>Small 1</td>
<td>2I5L, 2I5D, 1Q0A, 1T05, 1N6Q, 2HML, 1R0A, 1T03, 1N5Y, 1J5O, 1Q0E1, 1HY5, 1DLO</td>
<td>closed, closed</td>
<td>3.4</td>
<td>to Apo (1DLO)</td>
</tr>
<tr>
<td>Small 2</td>
<td>1FKO, 1R0T, 2BE2, 2B5J, 1RT3</td>
<td>closed, open</td>
<td>1.2</td>
<td>NNRTI+</td>
</tr>
<tr>
<td>Small 3</td>
<td>1JLE, 2I5J</td>
<td>closed, closed</td>
<td>2.2</td>
<td>NNRTI+</td>
</tr>
<tr>
<td>Small 4</td>
<td>2D1L, 2JE2, 2BGR</td>
<td>open, open</td>
<td>3.9</td>
<td>NNRTI-R+</td>
</tr>
<tr>
<td></td>
<td>1EP4</td>
<td>open, open</td>
<td>4.5</td>
<td>NNRTI-C+</td>
</tr>
<tr>
<td></td>
<td>1TV6</td>
<td>closed, closed</td>
<td>2.6</td>
<td>NNRTI−</td>
</tr>
<tr>
<td></td>
<td>1RTJ</td>
<td>closed, closed</td>
<td>4.1</td>
<td>NNRTI−</td>
</tr>
</tbody>
</table>

*Structure representatives from each cluster are in bold. A cluster representative is a structure in the cluster that has the smallest rmsd when compared to the collection of centroids of each of the comparison atoms. 

---
This was anticipated, as the NNIBP undergoes large structural rearrangements upon binding of an NNRTI where the aromatic side chains of Y181 and Y188 swivel out of the binding pocket and the primer grip region moves to create space for the NNRTI; the non-nucleoside inhibitor binding pocket only exists in structures with an NNRTI present.\(^{10,25}\)

A cluster distribution like that observed for the 82 HIV-1 RT/NNRTI complexes, where the great majority of the structures are found in one large cluster, suggests the possibility that the receptor pocket has not been interrogated by as extensive a variety of ligands as may have been previously thought. However, a few other basins do emerge from this analysis of the large data set of (99) RT crystal structures. We can speculate that the sparsely populated basins are separated by relatively high free energy barriers from the largest basin representing 73 structures; otherwise we would expect to have observed structures populating these “barrier” regions of the landscape among the large number of RT complexes in the PDB. The large cluster, on the other hand, appears to represent a sampling from what is effectively a continuum of accessible states.

The large cluster can be pictured as a basin within which there are low energy barriers separating many minima. Several of the residues within 10 Å of the ligands sample a more or less continuous distribution of conformations; the flexible loop consisting of residues 95–100 is an example. The primer grip \(\beta\)-sheet displays a “shrink-wrap” effect involving small adjustments of position within the large basin to optimize interactions with different functional groups and chemically diverse NNRTIs. The average backbone rmsd of the \(\beta\)12−\(\beta\)13−\(\beta\)14 strands within the large cluster is 1.4 ± 0.5 Å; all NNIBP residues sample somewhat continuous distributions that span 1–3 Å of conformational space. The ligands bound to the complexes in the large cluster are diverse in their shapes, sizes, functional groups, and binding modes, creating a large basin within which many conformations of the binding pocket are explored.

The most significant new observation to arise from this cluster analysis of the RT data set is the description of four basins that are sampled by the functionally important primer grip \(\beta\)12−\(\beta\)13−\(\beta\)14 sheet: the large NNRTI+ cluster, the small NNRTI-R+ cluster of structures found to bind the NNRTI rilpivirine/TMC278,\(^{13}\) the NNRTI-C+ singleton bound to the NNRTI capravirine/S-1153,\(^{25}\) and the NNRTI-1- singleton 1RTJ in which a HEPT ligand was washed out prior to structure determination.\(^{5}\) Whereas the majority of the \(\beta\)12−\(\beta\)13 strands of the primer grip are repositioned upon binding a non-nucleoside inhibitor by 3.4 ± 0.5 Å, the NNRTI-R+ and NNRTI-C+ forms differ from the NNRTI− structure 1DLO by > 6.2 Å, setting them ~4 Å from the large NNRTI+ cluster representative (Table 2). All three clusters, NNRTI-R+, NNRTI-C+ and NNRTI−, are also separated via clustering on the thumb. The NNRTI-R+ and NNRTI-C+ forms, separated from each other by ~2.2 Å in the primer grip and ~6.6 Å in the thumb, can be rationalized as the interogation of the NNIBP by larger ligands that interact with residues in the NNIBP in distinct ways.

The ligand bound to the NNRTI-R+ form, rilpivirine/TMC278 (2ZD1, 2ZE2, 3BGR), is a diaryl pyrimidine (DAPY) analogue. DAPY compounds have been found to be effective against many mutant forms of RT by utilizing multiple binding modes.\(^{12,26}\) The binding of the DAPY rilpivirine differs from that seen in other DAPYs; its cyano-vinyl group extends into a hydrophobic tunnel formed by the side chains of Y188, F227, W229, and L234. The extensive interaction of this cyano-vinyl group with the hydrophobic tunnel is thought to explain why rilpivirine is the most potent of the DAPY analogues.\(^{13}\) The formation of the tunnel is apparently also responsible for the shift of the \(\beta\)12−\(\beta\)13−\(\beta\)14 strands over the binding pocket as the positions of F227, W229, and L234 are reconfigured to make room for the cyano-vinyl group. One other crystallized NNRTI, seen in 2B5J, acts similarly because its cyano-vinyl group also extends into the hydrophobic tunnel.\(^{27}\) However, instead of causing a displacement of the \(\beta\)12−\(\beta\)13−\(\beta\)14 sheet to form the tunnel, binding of this NNRTI is accompanied by a displacement of Y188. The three RT/rilpivirine complexes also correspond to a new crystal form of HIV-1 RT.\(^{13}\) To examine whether the crystal contacts in the NNRTI-R+ structures (2ZD1, 2ZE2, 3BGR) induce changes in the primer grip, a complex of RT with a non-DAPY ligand in the new crystal form was included in the clustering study.\(^{28}\) This structure clusters in the large cluster, implying that the NNRTI-R+ conformation is not due to the crystal contacts in the new crystal form but rather to the novel interactions of the inhibitor’s cyano-vinyl group with the hydrophobic tunnel of the enzyme.

The ligand found in the NNRTI-C+ form, the imidazole capravirine (S-1153), is larger and more branched than others. Novel in this RT/capravirine complex (1EP4) is the formation of a main-chain hydrogen bond with P236.\(^{25}\) This H-bond causes the 3,5-dichlorophenyl ring to be in proximity with W229, which is shifted by ~4 Å over the NNIBP relative to the large cluster NNRTI+ representative.

The NNRTI-1− form shows a subtler shift in the binding pocket. Our clustering revealed that the different crystal forms of RT do not induce significant perturbations of the binding pocket structure except in the case of the NNRTI-1− form (PDB ID 1RTJ), where a weakly bound NNRTI was washed out to obtain an unliganded RT structure.\(^{5}\) Crystal contacts appear to stabilize the NNRTI-1− structure in the inhibited “open” form of the primer grip; this suggests that fluctuations of the binding pocket to the open form may occur even when no ligand is present.

The conformations of the primer grip \(\beta\)12−\(\beta\)13−\(\beta\)14 strands that are identified in the cluster analysis as NNRTI-R+ and NNRTI-C+ suggest routes for further exploration of new ligands that interrogate the NNIBP in ways that sample new and sparsely populated regions of the conformational landscape. Such conformations highlight receptor–ligand interactions such as additional H-bonds and formation of a stabilizing hydrophobic tunnel that appear resistant to several common mutations and may not be attainable in other conformations of the binding pocket. Design strategies based on the NNRTI-R+ and NNRTI-C+ basins can be utilized. These include further optimization of analogues of the highly active DAPY and imidazole compounds, focusing on interactions with the hydrophobic tunnel similar to rilpivirine and focusing on forming main-chain hydrogen bonds with P236 similar to capravirine.

**NNIBP Mutations: Conformational Effects.** Across the 82 RT/NNRTI complexes, mutations appear to have little effect on conformational change in the NNIBP but instead mainly affect the chemical signatures of the binding pocket, causing energetic penalties in binding of inhibitors. Minor changes in the NNIBP do occur in response to repositioning of the ligands, but these changes are minimal, causing mutants to be found in the conformational basins associated with their WT counterparts.
Interestingly, the ligands associated with both the NNRTI-R+ and NNRTI-C+ conformational basins, capravirine and rilpivirine, are highly active not just against the WT form of RT but also against many mutant forms. Both capravirine and rilpivirine have lower EC₅₀ values than many of the other NNRTIs, showing greater activity toward WT and commonly mutated forms of HIV-1 RT.²⁹,³⁰ The higher activities have been attributed to the interactions with the enzyme for these ligands as discussed above.

**Energy Landscape View of NNRTI Binding to HIV-1 RT.**

The overall affinity of a ligand for a receptor can be expressed as a balance between the strength of the interactions of a ligand for any particular binding-competent conformation of the receptor and the probability of occurrence of that conformation in the absence of the ligand. Another name for these receptor conformation probability distributions is the free energy landscape of the receptor from which the strain free energy required to move from one conformation to another in the absence of a ligand may be estimated.

Clustering of the 99 available X-ray RT structures based on the functionally important primer grip residues has identified five clusters or strain free energy basins. These basins are depicted in Figure 4 where the rmsd of the primer grip β12-β13 strands, relative to the apo structure 1DLO and relative to the large cluster representative 2OPP, were chosen as order parameters. These coordinates best describe the degree to which the primer grip has moved due to binding of an inhibitor. The clusters described in the previous section are further illustrated by Figure 4: the main large cluster (80 structures), a cluster of 10 substrate-bound and 4 apo RTs, a cluster of 3 RT/rilpivirine complexes with 1 RT/capravirine complex (separated in the above section into a cluster of 3 and a singleton, respectively), and a singleton represented by 1RTJ (NNRTI-1−). The large cluster (NNRTI+) is described by a broad and rugged region of the landscape corresponding to fine-tuning of the NNIBP to fit various inhibitors. The region corresponding to the RT/capravirine (NNRTI-C+) and RT/rilpivirine (NNRTI-R+) complexes reflects inhibitors that have stretched the primer grip region, creating novel conformations of the NNIBP.

The populations of the different NNIBP conformational basins shown in Figure 4 cannot be directly inverted to estimate receptor strain free energies. The observed locations and populations of the basins depend not only on the receptor strain free energies but also on the averaged interaction energies of the ligands with the receptor. Additionally, the crystal structure database represents a nonsystematic sampling of the landscape, as many of the inhibitors have been designed on the basis of an earlier inhibitor through QSAR techniques³¹ or designed for previously determined structures of the receptor. Both design approaches limit the potential to discover novel conformations of the receptor and partially explain why the cluster analysis produces a large cluster of NNIBP structures with similar inhibitors and similar receptor conformations.

One possible route to construct the receptor strain free energy landscape for the binding of NNRTIs to the NNIBP is to integrate information from the cluster analysis with molecular simulations. We can use structures representative of the different basins as “landmarks” to guide and test physics-based simulations using modern effective potentials and advanced sampling techniques like replica exchange molecular dynamics (REMD).²⁹,³²,³³ As the enzyme is very large, performing simulations using the whole protein may not be the most effective way to carry out free energy simulations of the binding pocket. Information about flexibility acquired from the cluster analysis of the X-ray structures described here can be used to both create a suitable fragment of the enzyme and develop constraints on the system to limit the computational time needed while optimizing the sampling of the conformational landscape of the NNIBP. For example, the regions of the NNIBP that pertain to areas of little flexibility, e.g., the β6-β9-β10 strands, can be held fixed while the highly variable regions such as the β12-β13-β14 strands of the primer grip and neighboring residues can be allowed to move. The clustering results presented here also provide a benchmark for the performance of the conformational sampling of the landscape. Initial simulations appear promising, as all of the basins illustrated in Figure 4 are found to have substantial statistical weight in the physics-based exploration of the receptor free energy landscape using temperature replica exchange molecular dynamics. However, several of the basins are not fully explored. Incorporation of umbrella sampling and/or utilization of distance restraints will allow for a more complete picture.

**Conclusion**

Previous structural studies have compared and contrasted a limited number of HIV-1 RT NNIBP receptor conformations,¹³,²³,²⁵,³⁴–³⁶ while other studies have focused on the composition and conformations of the ligands (NNRTIs) without regard to the conformation of the NNIBP. This study, the first to take a comprehensive look at the conformational fluctuations of the NNRTI receptor pocket, fills in missing pieces by utilizing a clustering algorithm to compare...
and contrast 99 available conformations of the non-nucleoside inhibitor binding pocket. The cluster analysis reported here has identified the locations of several conformational basins of the receptor pocket. The separation found is very similar across multiple clustering algorithms and, as such, suggests that the results reported here are intrinsic to the data and represent a “natural” clustering of the experimentally determined NNIBP conformations. The different basins reflect the variation in the NNIBP; however, the basins are not evenly populated. The sparsely populated basins provide opportunities for the design and/or optimization of potent ligands that inhibit RT in conformations of the NNIBP that exhibit varied positions of the functionally important primer grip. Two of the sparsely populated basins highlight receptor–ligand interactions that may not be attainable in other conformations of the binding pocket and that may be exploited in drug design strategies. These include main-chain hydrogen bonds with P236 and interactions with the hydrophobic tunnel surrounded by Y188, F227, W229, and L234 and formed by repositioning the primer grip.

Information from this study also serves as an essential guide for theoretical studies to map the free energy landscape of the NNIBP using modern all atom effective potentials and advanced sampling techniques like replica exchange molecular dynamics (REMD). A free energy landscape for the NNIBP would allow calculation of the strain free energy of the receptor required to adopt the ligand-bound conformation. Simulations may also highlight previously unexplored conformations of the NNIBP that may be suitable for ligand design and lead to novel potent NNRTIs. The construction of a model for the free energy landscape of the NNIBP using REMD guided by the cluster landmarks described in this paper will be the subject of a future communication.

Experimental Section

Selection and Preparation of X-ray Structures. Careful preparation of the structural data set was essential, as simply clustering on the unedited set revealed mostly noise. X-ray crystal structures from the Protein Data Bank were first analyzed to determine an atom sequence common to all structures. Entries that were found to be missing a large amount of structural information were removed to leave the data set of 99 structures used in this study. The 99 structures were then renumbered and reordered to follow the common atom sequence discovered from the analysis of each entry. This allowed for a normalization of the entries. Residues that experienced mutations were stripped of their side chain atoms that were not shared by each residue type. For example, as residue 103 is found as either a lysine or an asparagine, only the backbone atoms and Ca and Cβ of the side chain were included. Regions where many structures were found to be missing atoms were also removed.

Analysis of Backbone and Side Chain Fluctuation. A fluctuation analysis of the backbone of residues within 10 Å of any NNRTI in RT/NNRTI complexes (residues 91, 93–111, 161, 168, 177–193, 195, 198, 202, 205, 223–240, 242, 316–322, 343, 381–384 from p66 and 28, 136–138 from p51) was performed by aligning the ensemble of structures based on the Ca atoms of each of the 81 residues and calculating the radius of gyration ($R_g$) of the point cloud of all the positions for each Ca atom in the entire ensemble of structures. A low $R_g$ value reflects the removal of the position of the atom across conformers, whereas a high $R_g$ coincides with an atom that takes on many different positions in the ensemble.

Fluctuation of the side chains within 10 Å of any NNRTI that point into the NNIBP was analyzed by clustering on each side chain individually using single-linkage hierarchical clustering. The best clustering was chosen as that which gave the highest minimum separation ratio (MSR), an empirical measure of the degree of separation.

Alignment of RT Structures. Clustering results are partially dependent on the alignment of conformers. In this study, the structures were aligned on the Ca atoms of residues 105–111, 178–183, and 186–191 that correspond to β6, β9, and β10, respectively. This alignment was chosen on the basis of the backbone analysis above, since their Ca atoms have low $R_g$ values across the 82 RT/NNRTI complexes (see Figure 2). Superposition on the β6-β9-β10 is also often used in the literature to show movement within the NNIBP as well as global changes in conformation due to binding different ligands and substrates. Results using alignment on β6-β9-β10 were compared to alternative alignments, including alignment on the backbone atoms of all residues within 15 Å of the NNIBP, and were shown to give similar results. However, alignment on β6-β9-β10 was found to give the best separation of primer grip conformations with respect to the minimum separation ratio and minimum distances between clusters.

Clustering of NNIBP Conformations. Clustering was performed using two different techniques: single-linkage hierarchical clustering and complete linkage hierarchical clustering. Single linkage forms clusters that are more connected, while complete linkage forms clusters that are optimally compact. However, in this case, both algorithms gave very similar results, which points to a clustering that is intrinsic to the data and not an artifact of the chosen method.

Several initial clustering experiments using different alignments, different atoms on which to perform the clustering analysis, and clustering of only torsion angles (which does not require alignment) were attempted. However, the results of these experiments were clouded by a large amount of noise.

Therefore, a more systematic approach to determine the alignment and clustering parameters was employed. The choice of atoms on which clustering was performed was based on the analysis of the backbone and side chain fluctuations above. The Ca atoms of residues associated with the primer grip region gave the highest radii of gyration across the 82 RT/NNRTI complexes (see Figure 2) and were therefore chosen for clustering. This corresponds to the Ca atoms of residues 224–242. Side chains were also chosen by reviewing their fluctuation analysis above. Side chains that gave clustering levels with both high minimum separation ratios and minimum distances between clusters were picked. This corresponds to the side chains of residues 181, 183, and 188. As all of these residues are tyrosines in the WT form of RT and either cysteines or leucines when mutated, the χ1 dihedral angle was chosen for clustering these side chains. Clustering on a dihedral angle also alleviates the need for alignment of the structures. The best clustering was chosen as that which gave the highest MSR and a minimum rmsd between clusters of greater than 1 Å.

Acknowledgment. This work was supported by NIH Grants AI27690 (MERIT award to E.A.) and GM30580 (to R.M.L.).

References


