Journal of Medicinal Chemistry Article

J. Med. Chem. **2009,** *52,* 6413–6420 **6413**
DOI: 10.1021/jm900854h

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[a]) 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 subdomain[1−3] 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.[8] 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 enzyme[10,11] (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 $\pi-\pi$ stacking interactions with aromatic residues, as well as making hydrophobic contacts with other nonpolar pocket residues, and (usually) hydrogen bond (H-bond) donors capable

*To whom correspondence should be addressed. For E.A.: (address) Center for Advanced Biotechnology and Medicine, Rutgers University, 679 Hoes Lane, Piscataway, NJ 08854; (phone) 732-235-5323; (fax) 732-235-5788; (e-mail) arnold@cabm.rutgers.edu. For R.M.L.: (address) Department of Chemistry and Chemical Biology, Rutgers University, 610 Taylor Road, Piscataway, NJ 08854; (phone) 732-445-3947; (fax) 732-445-5312; (e-mail) ronlevy@lutece.rutgers.edu.

[a] Abbreviations: HIV-1, human immunodeficiency virus type 1; RT, reverse transcriptase; NNIBP, non-nucleoside reverse transcriptase inhibitor binding pocket; NNRTI, non-nucleoside reverse transcriptase inhibitor; PDB, Protein Data Bank; REMD, replica exchange molecular dynamics.

Published on Web 09/08/2009

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)[14] 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 elasticity 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.

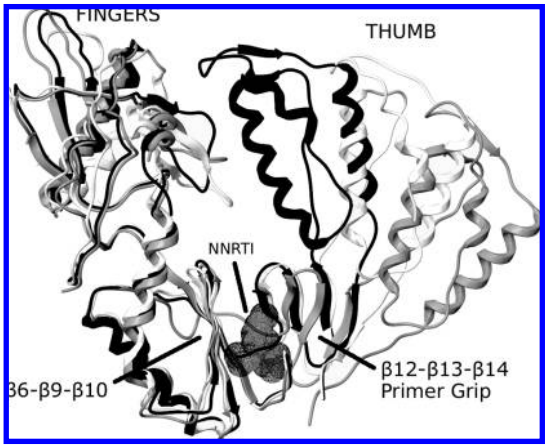

Figure 1. Rearrangement of HIV-1 RT upon binding. Upon binding substrates such as DNA or NNRTIs such as nevirapine, RT undergoes a global conformational change in the thumb of the enzyme. Upon binding an NNRTI, there is also a shift in the primer grip region shown in the movement from black/light-gray to dark-gray: black, unbound RT (PDB ID 1DLO); light-gray, RT bound to DNA (PDB ID 1RTD); dark-gray, RT bound to the NNRTI nevirapine (PDB ID 1VRT).

The goal of characterizing a conformational landscape and its corresponding energy landscape has come to occupy a central role in biophysics.[15−19] 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.[20,21]

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.[15−19] 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 Cα 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 Cα atoms. This increases to 1.58 ± 0.59 Å for all Cα atoms within 10 Å of any NNRTI. An analysis of the radii of gyration for each of the Cα atoms within 10 Å of any NNRTI across the ensemble of 82 RT/NNRTI complex conformations shows large variation in

Table 1. The 99 Crystal Structures of HIV-1 RT Used in This Analysis[a]

	PDB code
WT/NNRTI	1BQM, 1C0T, 1C0U, 1C1B, 1C1C, 1DTQ, 1DTT, 1EET, 1EP4, 1FK9, 1HNI, 1HNV, 1IKW, 1JLQ, 1KLM, 1LW0, 1LW2, 1LWE, 1REV, 1RT1, 1RT2, 1RT3, 1RT4, 1RT5, 1RT6, 1RT7, 1RTH, 1RTI, 1S6P, 1S6Q, 1S9E, 1S9G, 1SUQ, 1TKT, 1TKX, 1TKZ, 1TL1, 1TL3, 1TV6, 1TVR, 1 VRT, 1 VRU, 2B5J, 2B6A, 2BAN, 2BE2, 2OPP, 2 VG5, 2 VG6, 2 VG7, 2ZD1, unpublished
Mut/NNRTI	1FKO, 1FKP, 1HPZ, 1HQU, 1IKV, 1IKX, 1IKY, 1JKH, 1JLA, 1JLB, 1JLC, 1JLF, 1JLG, 1LWC, 1LWF, 1S1T, 1S1U, 1S1V,1S1W, 1S1X, 1SV5, 2HND, 2HNY, 2HNZ, 2IC3, 2OPQ, 2OPR, 2OPS, 2ZE2, 3BGR
WT unbound	1DLO, 1HMV, 1RTJ
Mut unbound	1HQE, 1JLE, 1QE1
DNA/RNA/ATP bound	1HYS, 1J5O, 1N5Y, 1N6Q, 1R0A, 1RTD, 1T03, 1T05, 2HMI, 2IAJ
RT/RNase H I	2I5J

[a] 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.

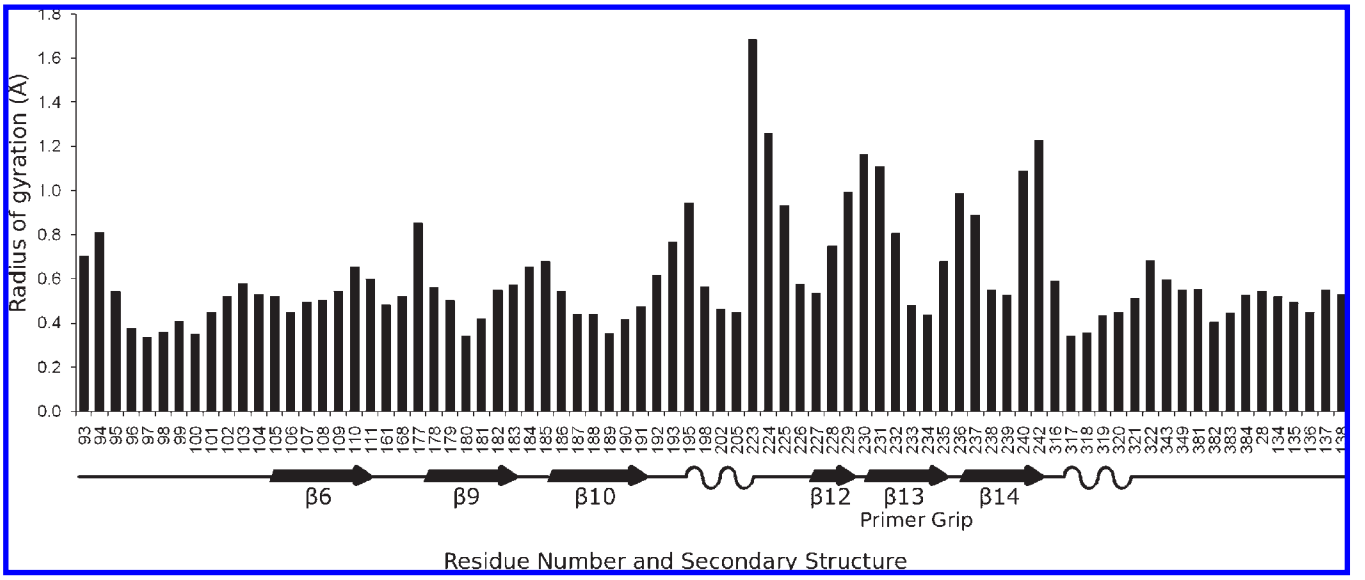

Figure 2. Radius of gyration (R_g) for each Cα atom of the residues within 10 Å of any NNRTI over the set of RT/NNRTI complexes (superposition was based on all Cα atoms in the set). The radius of gyration is an estimate for the spread of an atom's position across the ensemble of structures. A higher R_g signifies a large spread, whereas a smaller R_g signifies little fluctuation of position of the atom across different structures. Secondary structure is labeled with arrows for β strands and with curves for α helices. A straight line designates regions with no evident secondary structure using DSSP.[39]

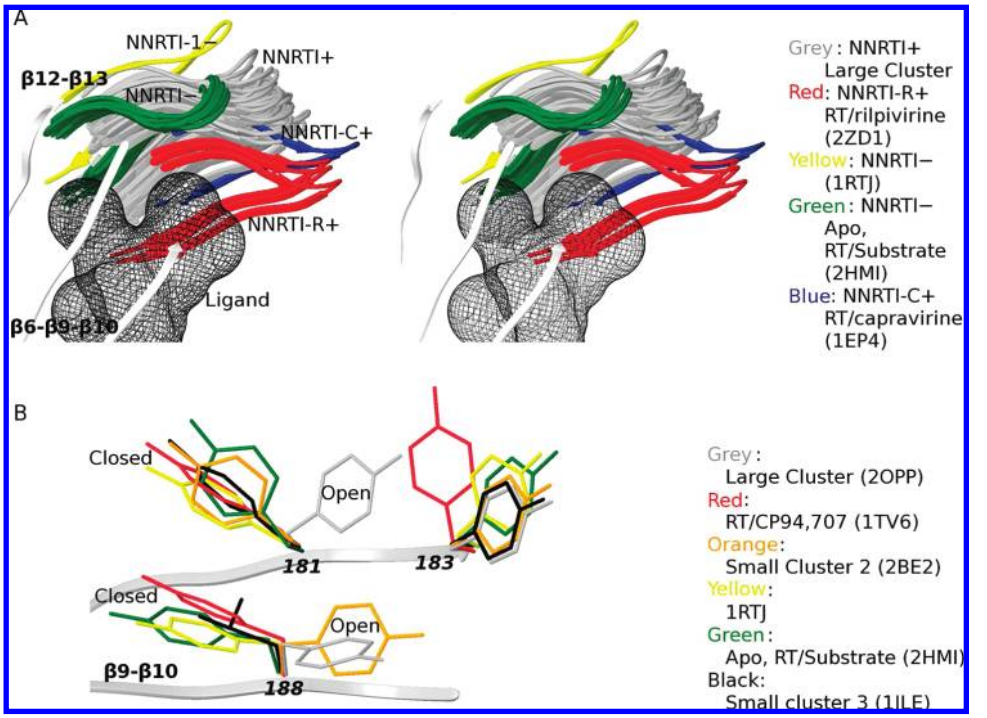

Figure 3. Separation of 99 X-ray structures. Structure representatives from each cluster are listed in parentheses. A further description can be found in Table 2. (A) Stereoview of the primer grip regions of all 99 X-ray structures (only β12-β13 shown) with alignment on β6-β9-β10. (B) Cluster representatives showing separation in residues 181, 183, and 188.

some regions in the vicinity of the NNIBP such as the primer grip β12-β13-β14 and the loop around P95 through L100 while other regions such as the β6-β9-β10 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 least variable) and on which to focus clustering experiments (the most 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

Table 2. Clustering Results: Eight Basins with Different Features[a]

cluster	cluster members, PDB IDs	residue 181, 188	$\beta12$-$\beta13$	rmsd of $\beta12$-$\beta13$-$\beta14$ (Å) to large Clust Rep (2OPP)	rmsd of $\beta12$-$\beta13$-$\beta14$ (Å) to Apo (1DLO)	NNRTI bound
large	**2OPP**, etc. (73 structures)	open, open	NNRTI+	0.0	4.0	y
small 1	2IAJ,1HMV,1RTD, 1HQE,1T05, 1N6Q, **2HMI**,1R0A,1T03, 1N5Y, 1J5O,1QE1, 1HYS,1DLO	closed, closed	NNRTI−	3.4	1.8	n
small 2	1FKO,1RTI,**2BE2**, 2B5J,1RT3	closed, open	NNRTI+	1.2	3.3	y
small 3	**1JLE**,2I5J	closed, closed	NNRTI+	2.2	3.8	n
small 4	**2ZD1**,2ZE2,3BGR	open, open	NNRTI-R+	3.9	6.6	y
1EP4	1EP4	open, open	NNRTI-C+	4.5	6.2	y
1TV6[b]	1TV6	closed, closed	NNRTI+	2.6	2.8	y
1RTJ	1RTJ	closed, closed	NNRTI-1−	4.1	3.6	n

[a] 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.[38] [b] 1TV6 represents the only case in which Y183 is in a largely differing position in the ensemble of structures.

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, 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+.

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 structural determination), which have different primer grip conformations from those seen in the large cluster. The small cluster of three RT/rilpivirine complexes will be called NNRTI-R+, the singleton 1EP4 will be called NNRTI-C+, and the singleton 1RTJ will be called NNRTI-1−. The unexpected effect of rilpivirine and capravirine on the distortion of the primer grip was discovered through the cluster analysis performed in this work. These conformations will be discussed at length in the Discussion.

Inclusion of the side chains of residues 181, 183, and 188 in the clustering analysis provides separation of several structures that have the large cluster primer grip NNRTI+ conformation but have differing conformations of these side chains. The small cluster of 5 (small cluster 2 in Table 2) is composed of RT/NNRTI complexes where Y181 is in the "closed" position, while Y188 is in the "open" position. The small cluster of 2 (small cluster 3 in Table 2) displays both Y181 and Y188 in the "closed" position. Both structures in this cluster represent interesting cases in which an NNRTI is not bound but the primer grip is in a bound conformation: an NNRTI was removed prior to structure determination of 1JLE, and 2I5J is bound to the RNase H inhibitor DHBNH. DHBNH does not bind in, but instead binds adjacent to, the NNIBP[22] 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[7] (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.

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.[10,24]

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 β-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 β12-β13-β14 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 β12-β13-β14 sheet: the large NNRTI+ cluster, the small NNRTI-R+ cluster of three structures bound to 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 β12-β13 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-1−, 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 interrogation 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, 3BRG), 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 cyanovinyl group extends into a hydrophobic tunnel formed by the side chains of Y188, F227, W229, and L234. The extensive interaction of this cyanovinyl 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 β12-β13-β14 strands over the binding pocket as the positions of F227, W229, and L234 are reconfigured to make room for the cyanovinyl group. One other crystallized NNRTI, seen in 2B5J, acts similarly because its cyanovinyl group also extends into the hydrophobic tunnel.[27] However, instead of causing a displacement of the β12-β13-β14 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 cyanovinyl 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 β12-β13-β14 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 EC_{50} 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 techniques[31] 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).[19,32,33] 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.

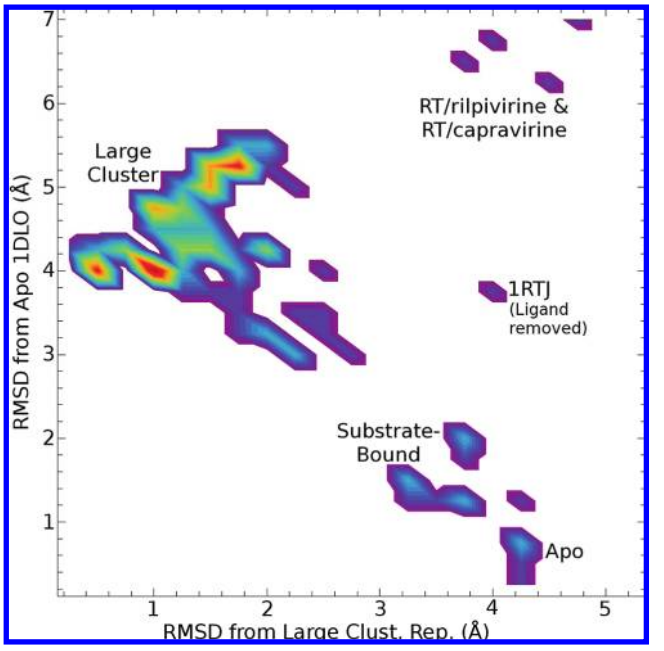

Figure 4. Populations based on rmsd (to apo 1DLO and large cluster representative 2OPP) of the primer grip with alignment on β6-β9-β10. Red designates a higher population, with tones progressing to dark blue being regions with lesser occupancy. The 99 experimental X-ray structures are as follows: locations of apo RT, substrate bound RT, the large NNRTI+ cluster, the NNRTI-1− singleton (1RTJ) and the RT/rilpivirine and RT/capravirine complexes (NNRTI-R+ and NNRTI-C+, respectively). The large cluster (NNRTI+) is descibed as broad and rugged.

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[14] 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 Cβ and Cγ 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 backbones 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, 134−138 from p51) was performed by aligning the ensemble of structures based on the Cα of each of the 81 residues and calculating the radius of gyration (R_g) of the point cloud of all the positions for each Cα atom in the entire ensemble of structures. A low R_g coincides with little movement 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.[37,38] 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 Cα 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 Cα 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.[37] 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 Cα 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 Cα 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 χ_1 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

(1) Kohlstaedt, L. A.; Wang, J.; Friedman, J. M.; Rice, P. A.; Steitz, T. A. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **1992**, *256*, 1783–1790.
(2) Tantillo, C.; Ding, J.; Jacobo-Molina, A.; Nanni, R. G.; Boyer, P. L.; Hughes, S. H.; Pauwels, R.; Andries, K.; Janssen, P. A.; Arnold, E. Locations of anti-AIDS drug binding sites and resistance mutations in the three-dimensional structure of HIV-1 reverse transcriptase. Implications for mechanisms of drug inhibition and resistance. *J. Mol. Biol.* **1994**, *243*, 369–387.
(3) Shen, L.; Shen, J.; Luo, X.; Cheng, F.; Xu, Y.; Chen, K.; Arnold, E.; Ding, J.; Jiang, H. Steered molecular dynamics simulation on the binding of NNRTI to HIV-1 RT. *Biophys. J.* **2003**, *84*, 3547–3563.

(4) Temiz, N. A.; Bahar, I. Inhibitor binding alters the directions of domain motions in HIV-1 reverse transcriptase. *Proteins* **2002**, *49*, 61–70.
(5) Esnouf, R.; Ren, J.; Ross, C.; Jones, Y.; Stammers, D.; Stuart, D. Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside inhibitors. *Nat. Struct. Biol.* **1995**, *2*, 303–308.
(6) Ding, J.; Das, K.; Hsiou, Y.; Sarafianos, S. G.; Clark, A. D., Jr.; Jacobo-Molina, A.; Tantillo, C.; Hughes, S. H.; Arnold, E. Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 Å resolution. *J. Mol. Biol.* **1998**, *284*, 1095–1111.
(7) Jacobo-Molina, A.; Ding, J.; Nanni, R. G.; Clark, A. D., Jr.; Lu, X.; Tantillo, C.; Williams, R. L.; Kamer, G.; Ferris, A. L.; Clark, P.; Hizi, A.; Hughes, S. H.; Arnold, E. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6320–6324.
(8) Das, K.; Ding, J.; Hsiou, Y.; Clark, A. D., Jr.; Moereels, H.; Koymans, L.; Andries, K.; Pauwels, R.; Janssen, P. A. J.; Boyer, P. L.; Clark, P.; Smith, R. H., Jr.; Kroeger Smith, M. B.; Michejda, C. J.; Hughes, S. H.; Arnold, E. Crystal structures of 8-Cl and 9-Cl TIBO complexed with wild-type HIV-1 RT and 8-Cl TIBO complexed with the Tyr181Cys HIV-1 RT drug-resistant mutant. *J. Mol. Biol.* **1996**, *264*, 1085–1100.
(9) Ding, J.; Das, K.; Tantillo, C.; Zhang, W.; Clark, A. D., Jr.; Jessen, S.; Lu, X.; Hsiou, Y.; Jacobo-Molina, A.; Andries, K.; Pauwels, R.; Moereels, H.; Koymans, L.; Janssen, P. A. J.; Smith, R. H., Jr.; Koepke, M. K.; Michejda, C. J.; Hughes, S. H.; Arnold, E. Structure of HIV-1 reverse transcriptase in a complex with the non-nucleoside inhibitor alpha-APA R 95845 at 2.8 A resolution. *Structure* **1995**, *3*, 365–379.
(10) Hsiou, Y.; Ding, J.; Das, K.; Clark, A. D., Jr.; Hughes, S. H.; Arnold, E. Structure of unliganded HIV-1 reverse transcriptase at 2.7 Å resolution: implications of conformational changes for polymerization and inhibition mechanisms. *Structure* **1996**, *4*, 853–860.
(11) Rodgers, D. W.; Harrison, S. C. The structure of unliganded reverse transcriptase from the human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1222–1226.
(12) Das, K.; Clark, A. D., Jr.; Lewi, P. J.; Heeres, J.; De Jonge, M. R.; Koymans, L. M.; Vinkers, H. M.; Daeyaert, F.; Ludovici, D. W.; Kukla, M. J.; De Corte, B.; Kavash, R. W.; Ho, C. Y.; Ye, H.; Lichtenstein, M. A.; Andries, K.; Pauwels, R.; De Béthune, M. P.; Boyer, P. L.; Clark, P.; Hughes, S. H.; Janssen, P. A.; Arnold, E. Roles of conformational and positional adaptability in structure-based design of TMC125-R165335 (etravirine) and related non-nucleoside reverse transcriptase inhibitors that are highly potent and effective against wild-type and drug-resistant HIV-1 variants. *J. Med. Chem.* **2004**, *47*, 2550–2560.
(13) Das, K.; Bauman, J. D.; Clark, A. D., Jr.; Frenkel, Y. V.; Lewi, P. J.; Shatkin, A. J.; Hughes, S. H.; Arnold, E. High-resolution structures of HIV-1 reverse transcriptase/TMC278 complexes: strategic flexibility explains potency against resistance mutations. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 1466–1471.
(14) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
(15) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545–600.
(16) Becker, O. M.; Karplus, M. The topology of multidimensional potential energy surfaces: theory and application to peptide structure and kinetics. *J. Chem. Phys.* **1997**, *106*, 1495–1517.
(17) Dill, K. A.; Chan, H. S. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **1997**, *4*, 10–19.
(18) Tsai, C. J.; Kumar, S.; Ma, B.; Nussinov, R. Folding funnels, binding funnels, and protein function. *Protein Sci.* **1999**, *8*, 1181–1190.
(19) Ravindranathan, K. P.; Gallicchio, E.; Levy, R. M. Conformational equilibria and free energy profiles for the allosteric transition of the ribose-binding protein. *J. Mol. Biol.* **2005**, *353*, 196–210.
(20) Xu, Q. S.; Daeyaert, F.; Lewi, P. J.; Massart, D. L. Studies of relationship between biological activities and HIV reverse transcriptase inhibitors by multivariate adaptive regression splines with curds and whey. *Chemom. Intell. Lab. Syst.* **2006**, *82*, 24–30.
(21) O'Brien, S. E.; Brown, D. G.; Mills, J. E.; Phillips, C.; Morris, G. Computational tools for the analysis and visualization of multiple protein−ligand complexes. *J. Mol. Graphics Modell.* **2005**, *24*, 186–194.
(22) Himmel, D. M.; Sarafianos, S. G.; Dharmasena, S.; Hossain, M. M.; McCoy-Simandle, K.; Ilina, T.; Clark, A. D., Jr.; Knight, J. L.; Julias, J. G.; Clark, P. K.; Krogh-Jespersen, K.; Levy, R. M.; Hughes, S. H.; Parniak, M. A.; Arnold, E. HIV-1 reverse transcriptase structure with RNase H inhibitor dihydroxy benzoyl naphthyl hydrazone bound at a novel site. *ACS Chem. Biol.* **2006**, *1*, 702–712.
(23) Das, K.; Lewi, P. J.; Hughes, S. H.; Arnold, E. Crystallography and the design of anti-AIDS drugs: conformational flexibility and positional adaptability are important in the design of non-nucleoside HIV-1 reverse transcriptase inhibitors. *Prog. Biophys. Mol. Biol.* **2005**, *88*, 209–231.
(24) Das, K.; Sarafianos, S. G.; Clark, A. D., Jr.; Boyer, P. L.; Hughes, S. H.; Arnold, E. Crystal structures of clinically relevant Lys103Asn/Tyr181Cys double mutant HIV-1 reverse transcriptase in complexes with ATP and non-nucleoside inhibitor HBY 097. *J. Mol. Biol.* **2007**, *365*, 77–89.
(25) Ren, J.; Nichols, C.; Bird, L.; Fujiwara, T.; Sugimoto, H.; Stuart, D. I.; Stammers, D. K. Binding of the second generation non-nucleoside inhibitor S-1153 to HIV-1 reverse transcriptase involves extensive main chain hydrogen bonding. *J. Biol. Chem.* **2000**, *275*, 14316–14320.
(26) Lewi, P. J.; de Jonge, M.; Daeyaert, F.; Koymans, L.; Vinkers, M.; Heeres, J.; Janssen, P. A.; Arnold, E.; Das, K.; Clark, A. D., Jr.; Hughes, S. H.; Boyer, P. L.; de Béthune, M. P.; Pauwels, R.; Andries, K.; Kukla, M.; Ludovici, D.; De Corte, B.; Kavash, R.; Ho, C. On the detection of multiple-binding modes of ligands to proteins, from biological, structural, and modeling data. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 129–134.
(27) Himmel, D. M.; Das, K.; Clark, A. D., Jr.; Hughes, S. H.; Benjahad, A.; Oumouch, S.; Guillemont, J.; Coupa, S.; Poncelet, A.; Csoka, I.; Meyer, C.; Andries, K.; Nguyen, C. H.; Grierson, D. S.; Arnold, E. Crystal structures for HIV-1 reverse transcriptase in complexes with three pyridinone derivatives: a new class of non-nucleoside inhibitors effective against a broad range of drug-resistant strains. *J. Med. Chem.* **2005**, *48*, 7582–7591.
(28) Das, K.; Bauman, J. D.; Arnold, E. Unpublished results.
(29) Fujiwara, T.; Sato, A.; el-Farrash, M.; Miki, S.; Abe, K.; Isaka, Y.; Kodama, M.; Wu, Y.; Chen, L. B.; Harada, H.; Sugimoto, H.; Hatanaka, M.; Hinuma, Y. S-1153 inhibits replication of known drug-resistant strains of human immunodeficiency virus type 1. *Antimicrob. Agents Chemother.* **1998**, *42*, 1340–1345.
(30) Janssen, P. A.; et al. In search of a novel anti-HIV drug: multidisciplinary coordination in the discovery of 4-[[4-[[4-[(1*E*)-2-cyanoethenyl]-2,6-dimethylphenyl]amino]-2-pyrimidinyl]amino]benzonitrile (R278474, rilpivirine). *J. Med. Chem.* **2005**, *48*, 1901–1909.
(31) Debnath, A. K. Application of 3D-QSAR techniques in anti-HIV-1 drug design—an overview. *Curr. Pharm. Des.* **2005**, *11*, 3091–3110.
(32) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
(33) Ravindranathan, K. P.; Gallicchio, E.; Friesner, R. A.; McDermott, A. E.; Levy, R. M. Conformational equilibrium of cytochrome P450 BM-3 complexed with *N*-palmitoylglycine: a replica exchange molecular dynamics study. *J. Am. Chem. Soc.* **2006**, *128*, 5786–5791.
(34) Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Ross, C.; Kirby, I.; Keeling, J.; Darby, G.; Jones, Y.; Stuart, D.; Stammers, D. High resolution structures of HIV-1 RT from four RT-inhibitor complexes. *Nat. Struct. Biol.* **1995**, *2*, 293–302.
(35) Pata, J. D.; Stirtan, W. G.; Goldstein, S. W.; Steitz, T. A. Structure of HIV-1 reverse transcriptase bound to an inhibitor active against mutant reverse transcriptases resistant to other nonnucleoside inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 10548–10553.
(36) Spallarossa, A.; Cesarini, S.; Ranise, A.; Ponassi, M.; Unge, T.; Bolognesi, M. Crystal structures of HIV-1 reverse transcriptase complexes with thiocarbamate non-nucleoside inhibitors. *Biochem. Biophys. Res. Commun.* **2008**, *365*, 764–770.
(37) Johnson, S. C. Hierarchical clustering schemes. *Psychometrika* **1967**, *32*, 241–254.
(38) Shenkin, P. S.; McDonald, D. Q. Cluster analysis of molecular conformations. *J. Comput. Chem.* **1994**, *15*, 899–916.
(39) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–637.