

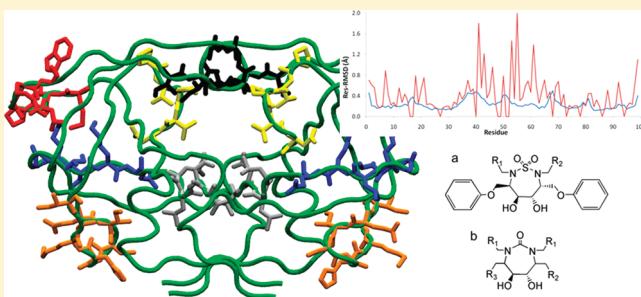
# Effect of Structural Stress on the Flexibility and Adaptability of HIV-1 Protease

Daniel P. Oehme, David J. D. Wilson, and Robert T. C. Brownlee\*

Department of Chemistry and La Trobe Institute for Molecular Sciences (LIMS), La Trobe University, Bundoora, Victoria, 3086, Australia

Supporting Information

**ABSTRACT:** Resistance remains a major issue with regards to HIV-1 protease, despite the availability of numerous HIV-1 protease inhibitors and copious amounts of structural and binding data. In an effort to improve our understanding of how HIV-1 protease is able to “outsmart” new drugs, we have investigated the flexibility of HIV-1 protease and in particular how it adapts to different structural stresses. Our analysis has highlighted the effects of space group on the variability between structures of HIV-1 protease and suggests that consideration of multiple structures and appropriate consideration of different conformations of the Ile50 residue is necessary in any structural analysis. Calculation of the root-mean-square deviation on a per-residue basis has been used to identify ‘natural variation’, while mutational and ligand analyses have been carried out to identify the effect on structure as a result of specific stresses. It was observed that mutations readily cause changes to occur at sites both close to and distant from a mutation site, with changes more likely to occur at residues that are sites of other major mutations. It is also revealed that HIV-1 protease adaption is dependent on the type and the structure of any bound ligand. Identification of the specific changes that occur due to these stresses will aid in the understanding of resistance and also aid in the design of new drugs.



## 1. INTRODUCTION

The link between molecular structure and activity has long been recognized;<sup>1–3</sup> it is a primary motivation for structural X-ray and nuclear magnetic resonance (NMR) studies of protein, DNA, and small-ligand complexes. While it has regularly been demonstrated that proteins with very different sequences can have similar functions when their three-dimensional (3D) structures are comparable,<sup>4</sup> mutational studies indicate that small changes in the sequence of a protein can have a significant effect on the activity of that protein.

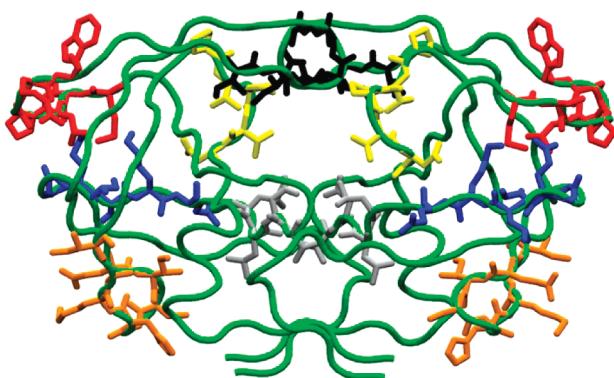
HIV-1 protease, a  $C_2$  symmetric aspartic acid homodimer with 99 amino acids in each subunit, is a system for which the link between structure and activity has been actively investigated. Hundreds of structures generated from both X-ray and NMR are available, encompassing a large array of ligands bound to HIV-1 protease structures with a variety of amino acid sequences. It has been identified that although there is variation in the protein structure due to the different bound ligands and amino acid sequences, the overall deviation between the structures is quite minimal.<sup>5–15</sup> However, these small variations in structure have been demonstrated to have a significant effect on the ligand binding strength, as reflected in the resistance profiles that develop when administering inhibitors to HIV positive patients. Modern structure-based drug design (SBDD) relies heavily on the ability to predict binding affinities, which is tested by the occurrence of mutations.<sup>16</sup> It is therefore imperative to

understand the effect that protein mutations and different ligands have on protein structure, since it will not only help in identifying better drugs but also has the potential to improve theoretical models of protein systems.

There are two previous studies that have analyzed structural changes in X-ray structures of HIV-1 protease, although in both cases the analysis was limited to the backbone structure. Zoete and co-workers have carried out a flexibility study of HIV-1 protease using a database of 73 X-ray structures.<sup>17</sup> Consistent backbone variation was identified at residues 15–21, 36–42, 47–52, 65–72, and 79–83 of HIV-1 protease (Figure 1). Their analysis was used to support the hypothesis that X-ray structures of different sequences of proteins correspond to different points on the potential energy surface of that protein. In contrast, Kumar and Hosur have studied the adaptability of the residues of HIV-1 protease.<sup>18</sup> They defined adaptability to be the ability of a residue to adjust to a stress or a change in its environment as opposed to flexibility, which they defined as the conformational variability of a residue. It is important to note that by definition, these stresses or changes in environment include the effect of a protein mutation or change in the bound ligand. Surprisingly it was found that the region containing residues 23–26 was the most adaptable, even though this region was the least flexible.

**Received:** February 9, 2011

**Published:** April 18, 2011



**Figure 1.** HIV-1 protease represented in licorice form with regions identified as being flexible by Zoete et al.<sup>17</sup> colored blue (residues 15–21), red (36–42), black (47–52), orange (65–72), and yellow (79–83). The adaptable region as noted by Kumar and Hosur<sup>18</sup> (residues 23–26) is colored silver.

In the present work, we have utilized the definition of adaptability and flexibility as outlined by Kumar and Hosur.<sup>18</sup>

The work presented here includes a detailed analysis of the flexibility and adaptability in HIV-1 protease, which has been carried out in order to gain insight into the structural changes that take place due to (i) mutations and (ii) binding of different ligand types. Our aim was two-fold: first to increase our understanding of resistance (as a result of mutation) and second to understand the effect that different ligand types have on the HIV-1 protease structure. An enhanced understanding of mutation could assist in the development of strategies to overcome resistance more effectively, while understanding the effect of ligands will guide the development of new drugs and, moreover, will indicate how these new drugs may be effected by resistance.

To this end, we have analyzed the root-mean-square deviations (RMSDs) in the backbone and side chain of available X-ray structures of HIV-1 protease with bound ligands as a function of residue number (Res-RMSD). Only bound structures were considered since only these structures are relevant for the present study of the effect of mutations and bound ligand. The natural flexibility of each residue was quantified and subsequently compared to the flexibility of residues after stress was applied by (i) mutation or (ii) when the ligand type was modified. Statistical analysis was then carried out to indicate which residues adapted most readily to the specific stress. The analysis of the flexibility of side chain atoms as well as backbone atoms presented here significantly extends the work of Zoete et al.<sup>17</sup> and Kumar and Hosur,<sup>18</sup> who only considered backbone atoms. In addition, we report the first analysis of the effect of space group on the Res-RMSD.

## 2. MATERIALS AND METHODS

A total of 235 HIV-1 protease complex structures were selected to be included in this study (see Supporting Information). A selection criteria was employed, which specified that structures (i) were generated using X-ray crystallography, (ii) had both chains present, and (iii) had only a single inhibitor bound in the active site (alternate conformations of inhibitors were allowed). Unbound structures were excluded since these structures are not relevant for the present study of the effect of mutations and bound ligand. Moreover, there are significant structural differences in the flap region between the bound and unbound structures, from

1	25
P Q I T L W Q R P L V T I K I G G Q L K E A L L D	
26	50
T G A D D T V L E E M S L P G R W K P K M I G G I	
51	75
G G F I K V R Q Y D Q I L I E I C G H K A I G T V	
76	99
L V G P T P V N I I G R N L L T Q I G C T L N F	

**Figure 2.** Consensus sequence of HIV-1 protease.

which inclusion of unbound structures would result in substantial Res-RMSD, masking the subtle changes that occur as a result of mutation and bound ligand. Additionally, there are insufficient unbound structures available to carry out a comparable study.

Residue sequences were then extracted for each of the 235 structures, and a consensus sequence was defined as the sequence that occurred most frequently (Figure 2). The mutant type of each structure could then be identified by comparison with this consensus sequence. Structures with an identical sequence to the consensus sequence were termed wild-type (WT). Each structure was then edited to remove all hydrogen atoms, solvent water molecules, counterions, and ligands such that only the HIV-1 protease protein remained.

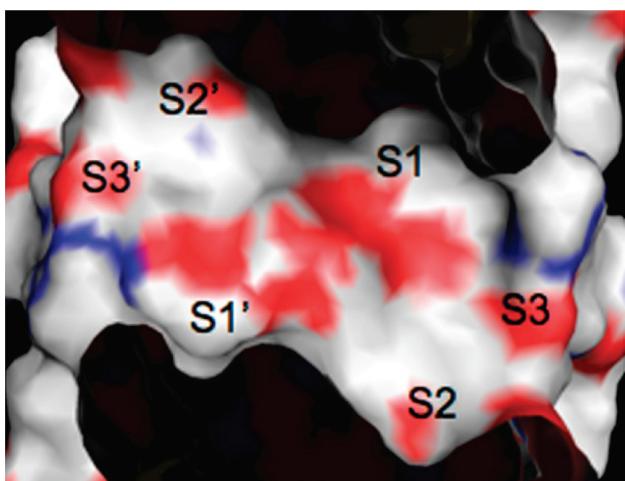
**Chain Switching for Accurate Superimposition.** To ensure that the Res-RMSDs were calculated accurately, chain A of the structure to be analyzed had to be superimposed onto chain A of the template structure. For a number of structures, this required a switch of the labels of the two chains of HIV-1 protease. Structures that needed to be switched were identified by first aligning all structures onto 1hpv and then identifying the position of the Ile50 carbonyl group. Structures that did not need to be switched were those where the carbonyl groups of Ile50 (i) pointed toward each other, (ii) pointed away from each other, or (iii) aligned perfectly onto the 1hpv structure. All other structures were tagged as needing to be switched. For these structures, chains A and B were switched.

**Refined RMSD Calculation Algorithm.** The final step in the preparation of pdb files was to convert equivalent atoms, such as the two  $C_{\gamma}$  atoms in valine, to dummy atoms with a position halfway between the equivalent atoms. This was required as the RMSD calculation algorithm used was unable to distinguish these chemically equivalent atoms that have differing atom names. This step of “unconfusing” the atoms decreases the possibility of obtaining erroneously high side chain RMSDs. Other residues with atoms that are chemically equivalent yet named differently, and for whom this procedure was applied, included arginine, aspartic acid, glutamic acid, and leucine. Res-RMSDs were calculated using tcl scripts and visual molecular dynamics (VMD).<sup>19</sup>

**Symmetry Analysis.** Each representative structure of the three main space groups (1hpv, 1hsg, and 1ohr) were split into two files: one for chain A and one for chain B. The two chains of each structure were then aligned onto each other, and backbone Res-RMSDs were calculated.

**Res-RMSD Analysis.** Space group analysis was conducted by calculating the Res-RMSD between the representative structure of the three main space groups and (i) all structures, (ii) the  $P6_1$  representative structure 1hpv, (iii) the  $P2_12_12$  representative structure 1hsg, and (iv) the  $P2_12_12$  representative structure 1ohr, after alignment of backbone atoms.

Same structure analysis was conducted within VMD for structures that had the same sequence, ligand, and space group.



**Figure 3.** View of the HIV-1 protease active site, detailing the binding subsites S3–S1 and S1’–S3’, colored by electrostatic potential. All peptidic ligands in this study fill the S1/S1’ and S2/S2’ subsites, with the 11223 ligands also filling one of the S3 subsites. The 112233 fill both S3 subsites. Note that the cyclic structures are classified as being 1122 binding ligands.

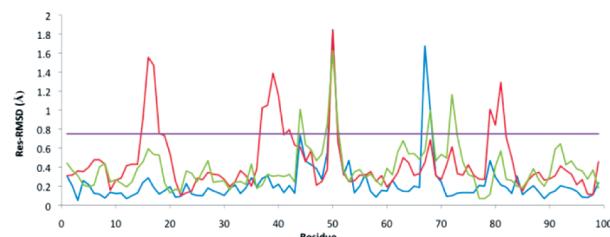
For the five pairs of structures studied, backbone and side chain Res-RMSD were calculated, along with averages and standard deviations of each Res-RMSD. These data define the natural variation of HIV-1 protease and were used as the control data for all subsequent analysis.

Mutant analysis was undertaken by first selecting structures that had the same space group and bound ligand but differed in sequence by at most two residues. The groups of structures selected were aligned onto a template structure for each group, after which backbone and side chain Res-RMSD were calculated. Each Res-RMSD was then averaged over the two chains and compared to the control data from the same structure analysis. Statistical analysis to identify where significant variation had taken place in comparison to the control data was undertaken in the form of a Z- or t-test.

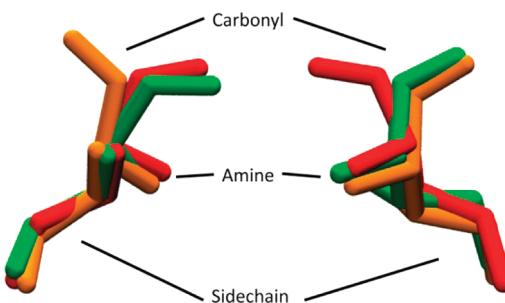
Ligand analysis was undertaken by selecting structures from the P2<sub>1</sub>2<sub>1</sub>2 space group with the WT sequence but with different bound ligands. The ligands were grouped as cyclic sulfamides, cyclic ureas, and peptidic ligands. The peptidic ligands were further grouped by the specific subsites that they fill (Figure 3). Average structures for each ligand group were determined, from which backbone and side chain Res-RMSDs were calculated between the average structures. As there was only one data set for each sample, a standard deviation could not be calculated, and thus statistical analysis to identify significant variation was in this case performed using Z-tests, which use the standard deviation of the population.

### 3. RESULTS AND ANALYSIS

Initial analysis of the available bound X-ray structures uncovered two important structural issues which, without being identified, could have led to incorrect conclusions regarding the effect that mutations and ligand type have on the structure of HIV-1 protease. The first is the similarity (or lack thereof) of the two chains that form the homodimer, and the second is the consequence of the X-ray space group on the structure of the homodimer as a whole.



**Figure 4.** Plot of backbone Res-RMSD as a function of residue between chains of the template structures 1hpv (blue), 1hsg (green), and 1ohr (red). The cutoff at 0.75 Å is labeled with a purple line.

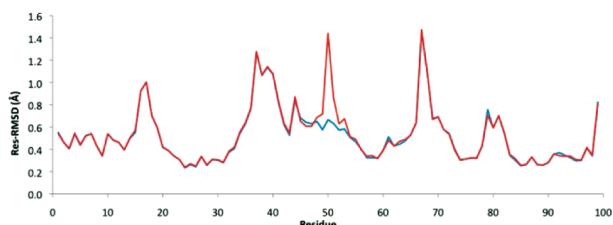


**Figure 5.** Residue Ile50/50' of structures 1hpv (green), 1a9m (orange), and 1aaq (red), illustrating the three possible conformations of the carbonyl groups: pointing in the same direction (green), directed away from each other (orange), or toward each other (red).

**3.1. Symmetry of HIV-1 protease.** HIV-1 protease has regularly been defined as being a C<sub>2</sub>-symmetric homodimer.<sup>20,21</sup> However, RMSD analysis has shown that both the backbone and side chains of the two chains are not symmetric.<sup>22–24</sup> This is reflected in results from several studies of flexibility analysis, which have demonstrated that residues from opposing chains have differing B-factors and differing interactions with bound ligands.<sup>12,13,25</sup> It has been suggested by some researchers that crystal-packing effects could be the cause of the observed asymmetry.<sup>23,26,27</sup>

To investigate symmetry, the pdb structures 1hpv, 1hsg, and 1ohr were selected as template structures from the most populated space groups P6<sub>1</sub>, P2<sub>1</sub>2<sub>1</sub>, and P2<sub>1</sub>2<sub>1</sub>2, respectively. These template structures were chosen as they all had resolutions under 2.5 Å, possessed the WT sequence, and were bound by well-characterized FDA-approved inhibitors.

Several interesting features regarding symmetry are identified by plotting the Res-RMSD of the backbones between chains 1 and 2 as a function of residue number for all three template structures (Figure 4). First it is clear that the two chains of HIV-1 protease are not symmetry related. If the two chains were symmetry related, then the Res-RMSD should all be zero. Second, Ile50 (a residue in the flexible flap sitting above the active site) has high RMSD values in all three template structures. Finally, by using a cutoff of 0.75 Å, residues that vary significantly between chains for each template structure could be identified. For the 1hpv structure there were significant differences for residues 67–68; for 1ohr this included residues 15–18, 37–42, and 79–81; for 1hsg, residues 44, 68, and 72. Interestingly, all these residues except for 44 were flexible residues, as predicted by Zoete et al.<sup>17</sup> This symmetry analysis confirms that HIV-1 protease is asymmetric upon binding of a ligand. It also indicates that the degree of asymmetry differs from residue to residue and structure to structure.



**Figure 6.** Plot of the average backbone Res-RMSD of all structures against 1hpv before (red) and after (blue) ‘switching’ of chains to highlight the dramatic changes about Ile50 after ‘switching’ was performed.

The high RMSD values for Ile50 arise from different conformations of the two Ile50 between the two chains. This observation was investigated further by calculating distances between the carbonyl oxygen atoms of Ile50 from opposing chains (see Tables S1 and S2 of Supporting Information for data). On the basis of this measure, all structures could be grouped into three distinct groups relating to three different conformations of Ile50 (Figure 5). The first conformation corresponds to oxygen–oxygen distances greater than 7.3 Å (orange structure in Figure 5) and has the carbonyl groups directed away from each other. The second conformation exhibits oxygen–oxygen distances of approximately 5.8 Å and has both carbonyl groups pointing in the same direction (green structure in Figure 5). The third conformation, with distances less than 3.5 Å, has the carbonyl groups pointing toward each other (red structure in Figure 5).

The conformation with the carbonyl groups pointing in the same direction would appear to be most electrostatically favored, which is reflected in this conformation being the most highly populated (82%). Interestingly, the conformation with the carbonyl groups pointing toward each other, which would seem to be least electrostatically favored, is actually more prevalent than the conformation with the carbonyl groups pointing away from each other. It appears unusual that this conformation is favored, given that it places two electronegative oxygen atoms in close proximity. However, 13% of structures from the population are present in this conformation, compared to just 5% of structures having the carbonyls pointing away from each other. Both of these lesser favored symmetric conformations are energetically hindered as a result of the inability to form a stabilizing hydrogen bond between the Ile50 carbonyl of one chain and the backbone Gly51 amine of the other chain,<sup>26,28</sup> which is common to many structures. The question arises as to whether these structures have been assigned correctly, and if so, why would these conformations be preferred? One possible explanation comes from the observation that the conformation with the carbonyl groups pointing toward each other is favored in open flap structures.<sup>28</sup> It could also indicate that the conformation of the Ile50 carbonyl plays a role in flap tip handedness and thus plays a role in the dynamics of the flap. A comparable study of Res-RMSD in unbound structures may shed light on this.

Initial backbone Res-RMSDs generated after aligning all structures onto 1hpv, exhibited an unexpectedly high Res-RMSD value for Ile50, which results from the asymmetry about Ile50. This is illustrated by the red line in Figure 6. Visual inspection of the structures indicated that the aligning algorithm was aligning the backbone of HIV-1 protease without taking into consideration the conformation around residue Ile50. To account for this, chains were ‘switched’ for structures with high Ile50 Res-RMSD values with reference to 1hpv. Subsequently, all structures could

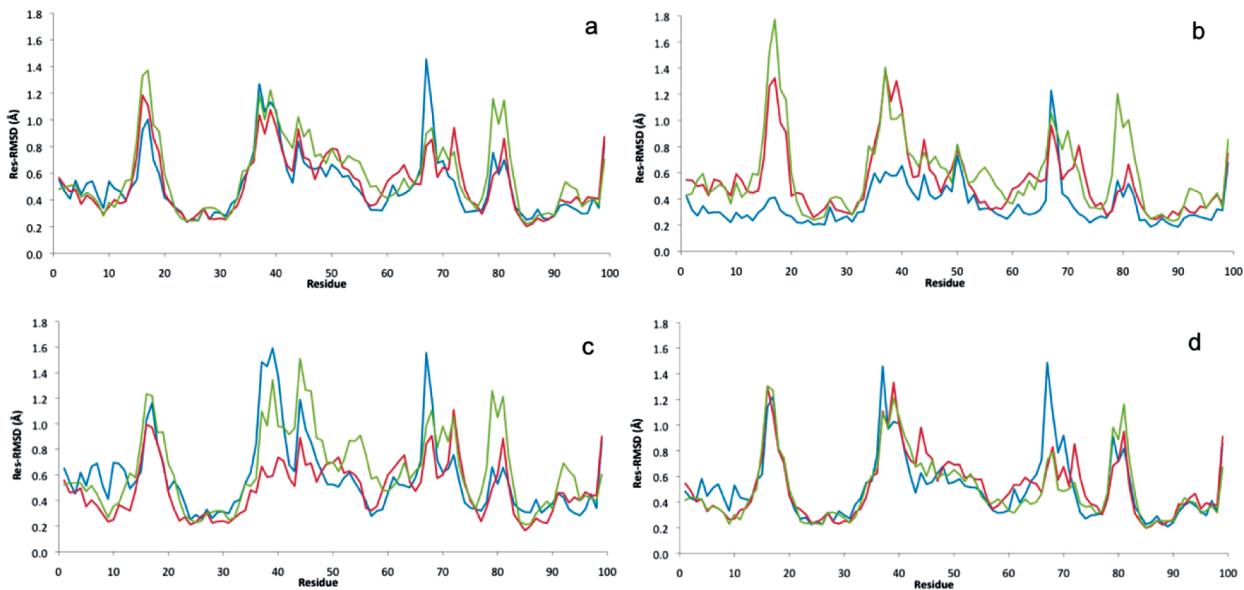
be aligned satisfactorily with the high Res-RMSD values for residue Ile50 no longer observed, as illustrated by the blue line in Figure 6. These switched structures were used for all further analysis. Since chains A and B can be differentiated for most structures, we have employed a consistent naming scheme for discriminating the two chains. The chain with the carbonyl group pointing toward the other chain is labeled chain A, while the other chain whose carbonyl group points out toward the solvent is labeled chain B.

**3.2. Space Group.** It has previously been noted that different lattice contacts can have an effect on the X-ray structure in cases where structures have been produced in different space groups.<sup>12,27,29</sup> The X-ray structures used in this study span nine different space groups, however the vast majority (93%) were formed in the P6<sub>1</sub>, P2<sub>1</sub>2<sub>1</sub>2, and P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> space groups. Backbone Res-RMSDs were calculated for each template structure against (i) all structures, (ii) those from space groups P6<sub>1</sub>, (iii) P2<sub>1</sub>2<sub>1</sub>2, and (iv) P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> (Figure 7).

The first thing to note is that the variability pattern reported by Zoete et al.<sup>17</sup> is reproduced. As is illustrated in Figure 7a, the template structures give very similar plots for the Res-RMSD against all structures. However, there are increased Res-RMSD values for 1hpv at residues 67–68 and for 1ohr at residues 15–18 and 79–81. The increased variance at these residues correlates with the data from the symmetry analysis, suggesting that the variance is due to differences at one of the chains of these regions in the template structures. With regards to 1hpv, the carbonyl group of Cys67 from chain A points upward into the center of mass of the structure, instead of downward and out toward the solvent as observed in the other structures. For 1ohr, residues 15–18 are found at a region called the cantilever, while it appears that chain B of 1ohr sits a bit further from the flap elbow (residues 38–42) than in the other template structures. This is consistent with the symmetry data, which indicates that these flap elbow residues have high Res-RMSDs. For residues 79–81 (a loop of residues which form the walls of the binding site), the loop is bent toward the solvent causing a slight increase in the size of the active site in 1ohr. Interestingly, high Res-RMSD values are not observed at residues 37–42 of 1ohr or for any of the residues in the 1hsq analysis. This seems to suggest that these symmetry differences are not particularly significant but blend in with the remainder of the population. The changes identified at the residues in 1hpv and 1ohr are unique to these structures and highlight the limitations of employing only a single template structure in any structural analysis.

The results plotted in Figure 7b–d highlight a significant difference from the results reported by Zoete et al.<sup>17</sup> In their work, they inferred that the different crystallization conditions, and therefore the space group of each structure, did not affect the flexibility of HIV-1 protease. On a qualitative level, the data plotted in Figure 7b–d appear to support their conclusion, however quantitatively, this is not the case. In each of Figure 7b–d, it is the template structure whose space group is being investigated that has the lowest Res-RMSD. For the P2<sub>1</sub>2<sub>1</sub>2 analysis (Figure 7c), there is a higher degree of variability than for all other data, while for the P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> data (Figure 7d), very little variance is identified between the data from the three different template structures. The average RMSDs (Table 1) confirm that the comparison of 1hpv against P6<sub>1</sub> has the smallest average Res-RMSD, and the P2<sub>1</sub>2<sub>1</sub>2 data has the greatest overall average Res-RMSD for all template structure, while P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> has the least.

It can be concluded that space group does indeed have a quantitative effect on the variability of the backbone Res-RMSD



**Figure 7.** Plot of backbone Res-RMSD for 1hpv (blue), 1hsg (red), 1ohr (green) against (a) all structures, (b)  $P_6_1$  structures, (c)  $P_{2,1,2,2}$  structures, and (d)  $P_{2,1,2,1,1}$  structures; 1hpv, 1hsg, and 1ohr belong to space groups  $P_6_1$ ,  $P_{2,1,2,2}$ , and  $P_{2,1,2,1,1}$ , respectively.

**Table 1. Average Backbone Res-RMSD (Å) between Template Structures and Space Group Subsets of the Population**

PDB ID	all	$P_6_1$	$P_{2,1,2,2}$	$P_{2,1,2,1,1}$
1hpv	0.516	0.356	0.583	0.520
1hsg	0.527	0.534	0.495	0.521
1ohr	0.586	0.590	0.642	0.502

values, and thus to identify the effects that mutations and changing ligands have on the structure of HIV-1 protease, space group considerations must be taken into account.

We did not carry out the analogous side chain analysis, since the backbone analysis had already validated the hypothesis that space group has a quantitative effect on the structure of HIV-1 protease. Any side chain analysis would be expected to provide the same conclusion.

**3.3. Natural Variation.** In order to identify how specific residues adapt to particular stresses, such as mutations or changes of ligand, it is first necessary to investigate the natural flexibility of residues. Natural variation of both the backbone and side chains of residues was identified by analyzing closely related structures. The structures selected for this analysis were identical in sequence, ligand, and space group, which reflects our earlier conclusion that space group has an effect on Res-RMSD (see Table S3 of Supporting Information). Of the 236 structures in the population, five pairs of structures met these criteria. Average Res-RMSDs were calculated over the five pairs to produce natural variation Res-RMSDs for HIV-1 protease (Table 2 and Figure 8). This enabled the natural variation to be quantified. In the subsequent analysis of the effect of mutation and bound ligand on the structure, these data are used to perform statistical tests in order to identify structural differences that are statistically beyond natural variation.

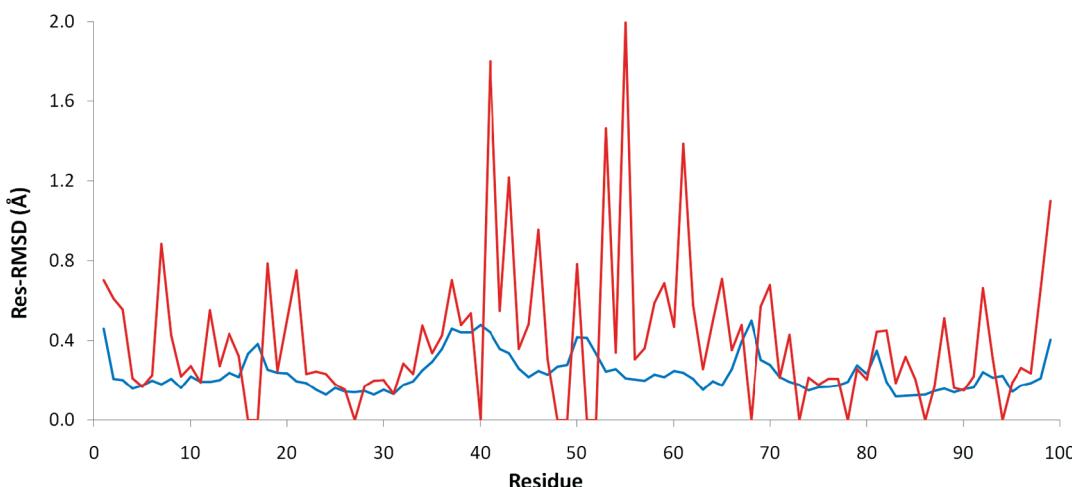
Residues are labeled as having significant natural side chain variation if their side chain Res-RMSD is greater than 0.4 Å (Table 2). It should be noted that as expected, the majority of residues identified as showing significant natural variation are

**Table 2. Natural Variation Data for the Backbone (Back) and Side Chain (Side) of Each Residue of HIV-1 Protease<sup>a</sup>**

PRO 1	GLN 2	ILE 3	THR 4	LEU 5	TRP 6	GLN 7	ARG 8	PRO 9	LEU 10
Back	0.461	0.208	0.202	0.160	0.173	0.200	0.179	0.209	0.161
Side	0.703	0.612	0.556	0.210	0.168	0.227	0.884	0.427	0.222
VAL 11	<b>THR 12</b>	ILE 13	<b>LYS 14</b>	ILE 15	GLY 16	GLY 17	<b>GLN 18</b>	LEU 19	<b>LYS 20</b>
Back	0.194	0.192	0.203	0.239	0.217	0.333	0.383	0.253	0.240
Side	0.189	0.554	0.274	0.435	0.321	0.000	0.000	0.787	0.249
GLU 21	ALA 22	LEU 23	LEU 24	ASP 25	THR 26	GLY 27	ALA 28	ASP 29	<b>ASP 30</b>
Back	0.196	0.187	0.152	0.128	0.162	0.145	0.140	0.148	0.130
Side	0.753	0.233	0.245	0.232	0.180	0.152	0.000	0.170	0.198
THR 31	<b>VAL 32</b>	<b>LEU 33</b>	<b>GLU 34</b>	GLU 35	MET 36	SER 37	LEU 38	PRO 39	GLY 40
Back	0.132	0.177	0.195	0.252	0.295	0.358	0.462	0.442	0.441
Side	0.130	0.285	0.233	0.477	0.337	0.277	0.702	0.480	0.536
ARG 41	TRP 42	LYS 43	PRO 44	LYS 45	<b>MET 46</b>	<b>ILE 47</b>	<b>GLY 48</b>	GLY 49	<b>ILE 50</b>
Back	0.443	0.358	0.337	0.261	0.217	0.249	0.231	0.270	0.279
Side	1.800	0.549	1.216	0.358	0.483	0.957	0.306	0.000	0.784
GLY 51	GLY 52	<b>PHE 53</b>	<b>ILE 54</b>	<b>LYS 55</b>	VAL 56	ARG 57	<b>GLN 58</b>	TYR 59	ASP 60
Back	0.414	0.335	0.246	0.257	0.211	0.204	0.198	0.228	0.219
Side	0.000	0.000	1.464	0.341	1.995	0.307	0.363	0.590	0.687
GLN 61	ILE 62	LEU 63	ILE 64	GLU 65	ILE 66	CYS 67	GLY 68	HIS 69	LYS 70
Back	0.238	0.208	0.154	0.194	0.173	0.257	0.394	0.500	0.304
Side	1.386	0.577	0.258	0.497	0.711	0.353	0.480	0.000	0.572
ALA 71	<b>ILE 72</b>	GLY 73	THR 74	VAL 75	<b>LEU 76</b>	VAL 77	GLY 78	PRO 79	THR 80
Back	0.216	0.192	0.176	0.148	0.165	0.167	0.174	0.194	0.277
Side	0.215	0.430	0.000	0.215	0.176	0.208	0.207	0.000	0.258
PRO 81	<b>VAL 82</b>	ASN 83	<b>ILE 84</b>	ILE 85	GLY 86	ARG 87	<b>ASN 88</b>	LEU 89	<b>LEU 90</b>
Back	0.351	0.193	0.119	0.122	0.125	0.129	0.147	0.158	0.141
Side	0.445	0.450	0.188	0.320	0.204	0.000	0.168	0.511	0.162
THR 91	<b>GLN 92</b>	ILE 93	GLY 94	CYS 95	THR 96	LEU 97	<b>ASN 98</b>	PHE 99	
Back	0.165	0.240	0.213	0.223	0.143	0.173	0.187	0.211	0.406
Side	0.219	0.663	0.311	0.000	0.191	0.265	0.236	0.661	1.101

<sup>a</sup> Residues that show significant structural side chain variation are shaded in blue. Residues that are sites of major mutation according to the Stanford HIV Drug Resistance Database are bold and italicized (ref 30).

polar and sit on the surface of HIV-1 protease, however some nonpolar surface residues and buried residues also show significant variation. It is expected that mutations at these sites of significant variation (shaded blue in Table 2), where the protein is flexible, will not cause significant changes in the rest of the structure of HIV protease and thus not have a significant effect on the function. On the other hand, mutations at residues where the protein is relatively rigid (the unshaded regions in Table 2) are more likely to cause the rest of the protein to adapt to this change,



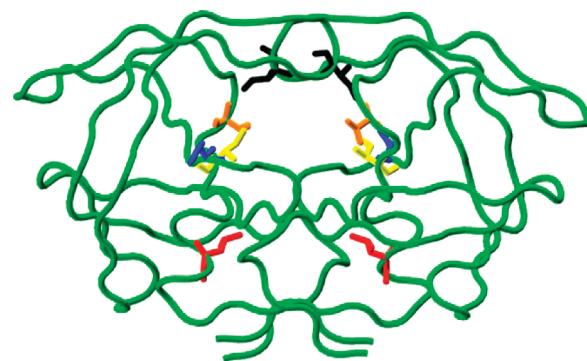
**Figure 8.** Plot of natural variation Res-RMSD with backbone data colored blue and side chain data colored red.

resulting in a change in structure and by implication function. Interestingly, the majority of major mutations (as identified by the Stanford HIV Database) occur at the unshaded sites in Table 2, which agrees with our discussion above.

It is instructive to consider the effect of the cutoff threshold of 0.4 Å on this analysis. If a lower threshold of 0.3 Å was employed, then an additional 11 residues would be categorized as showing significant side chain variation. However, these additional residues all neighbor residues identified as showing side chain variation at the 0.4 Å cutoff. The exception is Ile84, which is nevertheless identified as a site of major mutation. Increasing the cutoff from 0.4 to 0.45 Å would reduce the number of variable side chain residues by 4, while a cutoff of 0.5 Å reduces the number of variable side chain residues by 9. In each case, the same conclusions regarding variability remain: the majority of residues identified as having significant side chain variation are polar, surface residues. If we couple this with the fact that the largest side chain Res-RMSD was 1.995 Å, it can be readily argued that the 0.4 Å threshold is reasonable.

**3.4. Mutational Analysis of HIV-1 protease.** The majority of structures deposited in the PDB are the result of investigations into the effect of mutations on ligand binding strength. One could make the simple assumption that when a mutation occurs, the effect on the structure is the same no matter what ligand is bound. That is, changes in the protein structure that arise from a mutation are independent of the complexed ligand. Here we explore the validity of this assumption by considering if consistent structural changes exist for particular mutations. Structures were selected that had the same space group and the same bound ligand but whose sequence differed by at most two residues (see Table S5 of the Supporting Information for a list of structures used). To ensure a sufficiently large sample, structures were chosen containing the major mutations D30N, I50V, V82S, V82A, I84V, V82F/I84V, and L90M. The sites of these mutations are detailed in Figure 9. Statistical analysis of the backbone and side chain Res-RMSD (in comparison with natural variation) enabled identification of residues where statistically significant changes occurred beyond natural variation. Results are detailed in Table 3.

In general, backbone and side chain variation is observed for residues both close to and far away from the mutation site. Despite this, the data for the I50V mutation indicate that no specific changes occur with this mutation beyond natural variation. This



**Figure 9.** HIV-1 protease illustrating major mutation sites: Asp30 (blue), Ile50 (black), Val82 (orange), Ile84 (yellow), and Leu90 (red).

is understandable given that this mutation only changes isoleucine to valine: two relatively similar residues that are both hydrophobic and differ only by a methyl group.

For the V82A mutation, only the side chain of Leu33 was identified as undergoing significant change. The Leu33 residue is of particular interest, since significant changes were noted at this residue for all but the I50V and V82F/I84V mutations. The Leu33 residue resides downstream from Asp30 in a large hydrophobic pocket, in very close proximity to residues 82 and 84, all of which are sites of major mutations. The fact that significant variation occurs for Leu33 in a number of mutations suggests that this residue may play a significant role in resistance.

The D30N mutation, in which an aspartic acid is mutated to an asparagine, causes significant and complex changes to the hydrogen-bonding interactions around the S2 subsite (Figure 3). Significant changes were noted at residues 50, 84, and 88, all of which are sites of major mutation. For Ile84, this change corresponds to 120° rotations about the C<sub>α</sub>—C<sub>β</sub> and C<sub>γ</sub>—C<sub>β</sub> bonds. For Asn88 rotameric switching was observed, which is considered to play a key role in resistance arising from D30N mutations.<sup>15</sup> In the WT, Asn88 is usually hydrogen bonded to Thr74, which allows Asp30 to form a hydrogen bond to the bound ligand. However, Asn88 can alternatively interact with Asp30, which forces a change in the conformation of Asp30 and results in the breaking of the hydrogen bond to the ligand. In the case of the D30N mutation, where Asp30 is replaced by an Asn, the side chain of Asn sits in a

**Table 3.** Residues That Exhibit Significant Structural Variation beyond Natural Variation Due to a Mutational Stress

mutation	backbone	side chain
D30N	50	33, 35, 84, 88, 97
I50V		
V82A		33
V82S	33, 34, 38, 46, 63, 65, 75, 78, 83	6, 33, 57, 83
I84V		33, 65, 97
V82F, I84V	2, 10, 27, 30, 32, 69, 76, 86, 88	5, 9, 25, 28, 31, 65, 85
L90M	12, 90	11, 19, 20, 24, 31, 33, 35, 50, 63, 65, 82, 84, 97

slightly different conformation to Asp due to an 180° rotation about the C<sub>β</sub>–C<sub>γ</sub> bond of Asn88. This allows a significant increase in the interaction between Asn88 and Asn30, with a subsequent decrease in the interaction between Asn88 and Thr74 and between Asn30 and the ligand. The only other significant change is with Leu97, which results from a 120° rotation about the C<sub>α</sub>–C<sub>β</sub> bond causing the C<sub>δ</sub> to sit in a slightly different environment.

Changes to the backbone as a result of the I84V mutation are largely insignificant; significant side chain changes were identified only at residues 33, 65, and 97. The changes to residues Leu33 and Leu97 are also present in the D30N mutation, while for Glu65, rotation about the C<sub>β</sub> is observed. The Glu65 residue actually borders on being a surface residue with a solvent accessible surface area (SASA) of ~30 Å<sup>2</sup>. Moreover, since it has a long and polar side chain, it is not unexpected to see a change at this residue without a tangible link to a mutation. Significant changes at this residue are also observed as a result of several other mutations (V82F/I84V and L90M).

The V82S mutation is a substantial mutation, and as a result, a number of significant changes were identified. Changes in the 80s loop, which forms the outer edge of the binding site at the front/back of the S1/S1' subsite, cause a shift of the nearby 33–35 loop, which initiates correlated motions in Leu38. Changes to the side chain of Glu35 cause significant changes in the side chain of Arg57, such that both side chains are able to occupy the small void they fill without sterically clashing. Changes at the nonpolar surface residues Trp6 and Ile63 are unexplained, although it is postulated that these changes may be due to unidentified natural variation.

For the V82F/I84V mutant a number of residues show significant variation at backbone atoms, however none of these variations are transferred to their side chains. A few important changes also occur at sites of major mutation, with direct variation noted at residues 30 and 88, while variation at residues 27, 31, 32, and 76 occurs in regions adjacent to residues 30 and 84. The most compelling change involves Asp25, the noted catalytic site and one of the most stable residues. Although initially unexpected, this variation does correlate with the work by Kumar and Hosur, who concluded that Asp25 is an adaptable residue.<sup>18</sup> This feature of the 23–26 stretch of residues is considered to be important in the ability of HIV-1 protease to cleave eight different natural substrates.

For the L90M mutants (a nonactive site mutation that occurs at the interface between the two monomers of HIV-1 protease), significant side chain variation is observed at three of the other major mutation sites, residues 50, 82, and 84. Previous analysis of

the L90M mutant suggests that changes at the 80s loop are to be expected,<sup>31</sup> however the changes at Ile50 have not been previously identified. Variation was also observed at Leu24 and Leu97, both of which sit adjacent to residue 90. This correlates with previous studies, where significant changes have been identified at the catalytic triad (24–26) as a result of the L90M mutation.<sup>9,29,31</sup> Changes at Glu35 correlate with the V82S mutation, whereby modification of the 80s loop causes changes at the 33–35 loop. Interestingly, this analysis does not show Asp30 to vary significantly, even though previous studies have observed significant change.<sup>32</sup> However, variation is exhibited by residues 31 and 33, which sit close to, or are part of, the 33–35 loop and also sit close to Asp30. Finally, significant changes were observed at Lys20, which is a buried polar residue (see Table 2). Despite being defined as buried, Lys20 has moderate SASA (~30 Å<sup>2</sup>), and since lysine is a polar residue with a long side chain, it is highly likely that the change observed is due to natural variation. Changes at Leu19 and Leu63 are expected to be due to the unidentified natural variation of nonpolar surface residues.

The data presented for the L90M mutation support the hypothesized resistance mechanism, which suggests that mutation causes a change in interactions with the catalytic triad, in turn affecting the stability between the chains of HIV-1 protease, and in the position of the ligand. The catalytic triad forms what is called the ‘fireman’s grip’, which is a key point for stability of the homodimer. This destabilization of the two chains is thought to favor the natural substrate over the ligands, while the change of interactions between the catalytic triad and L90M can also cause a shift in the ligand position in the active site. This is highlighted in this and previous work<sup>31</sup> where changes are observed in the position of residues 30, 50, 82, and 84, which are key residues in the active site.

Mutation analysis has identified that specific mutations force specific residues to adapt to the stress, both close to and far from the mutation site. In general it was observed that when a mutation occurred at the active site, residues around the active site adapted, while for mutations away from the active site, residues at both the site of mutation and at the active site adapted. The data presented here is in agreement with the hypothesized resistance mechanisms and highlights the importance of key residues around the active site that are also sites of major mutations.

**3.5. Ligand Analysis of HIV-1 Protease.** In the same manner as the mutation analysis, we investigated whether HIV-1 protease residues adapt in a consistent manner when different types of ligands are bound. Confirmation of this hypothesis would suggest that the different scaffolds are causing changes in the structure of HIV-1 protease. If the changes that take place when different ligands bind can be understood, then this could lead to the design of better scaffolds and therefore better ligands.

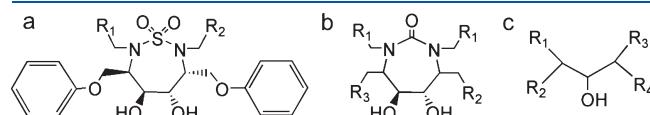
A large number of ligands and ligand-types have been cocrystallized with HIV-1 protease. Analysis of the WT structures from the P2<sub>1</sub>2<sub>1</sub> space group (the most populated sequence–space group combination) identified three major types of ligands: cyclic sulfamides, cyclic ureas, and peptidic analogues (Figure 10). The peptidic ligands can be further grouped by the subsite of the HIV-1 protease active site (Figure 3) that they occupy: 1122, 11223, 112233. A list of the structures used in this analysis is provided in Table S6 of the Supporting Information.

Unlike the mutant analysis, it was not possible to assign a template structure for which all ligands could be compared. The assumption was made that it was only the ligand scaffold type that

was affecting the structure of HIV-1 protease beyond natural variation. As a result, analysis of the effect that ligands had on the structure of HIV-1 protease was carried out by first calculating an average structure of HIV-1 protease for each of the five ligand scaffold types: cyclic sulfamides, cyclic ureas, and the three peptidic ligands identified above. We compared these five averaged structures by calculating the backbone and the side chain RMSD between pairs of the averaged structures. This approach enabled us to identify specific residues that show significant variation between each pair of averaged structures. Residues were classified as varying significantly (Table 4) if they failed a Z-test at the 1% level in comparison to the natural variation data (Table 2).

Minimal variation was observed between the average peptidic 1122 structure and the remaining structures, as highlighted in Table 4 (first data column). The minimal variation observed against the cyclic ligands is thought to be due to the peptidic 1122 ligands being of a similar size to the cyclic inhibitors and thus binding in the same subsites. The limited variation against the larger peptidic inhibitors could arise from all the ligands being peptidic inhibitors. Thus, the average peptidic 1122 structure may almost be considered a template for the other average structures. A change observed in most of the peptidic 1122 analyses is the variation at Leu63. Given that Leu63 sits at the 'cheek' (as named in Perryman et al.),<sup>13</sup> where it is near the surface yet far from the active site, this variation is unexpected. The only conclusion that can be made is that the data for the natural variation must be too small (average RMSD 0.258 Å, standard deviation 0.119 Å, and critical value 0.396 Å). Another residue of interest from the peptidic 1122 analysis is Asp30. This residue, as noted above, is a site of major mutation and is demonstrated to be variable when compared to the other peptidic ligands, suggesting that the extra length of the peptidic 1122/3 ligands may be causing this variation. We suggest that the larger peptidic ligands are able to interact with Asp30 more favorably and thus cause a change in the position of Asp30. This provides a possible explanation why resistance is observed for D30N mutations, as the mutation may cause unfavorable interactions; smaller ligands will be able to adapt to this change, however the larger ligands will not.

Despite there being little variation from all structures to the peptidic 1122 average structure, there is significant variation



**Figure 10.** Three types of ligands used in the ligand analysis: (a) cyclic sulfamide, (b) cyclic urea, and (c) peptidic analogues.

**Table 4. Residues That Exhibit Significant Variation between Pairs of Average Structures for the Five Different Ligand Scaffold Types**

ligand type	peptidic 1122	cyclic sulfamide	cyclic urea	peptidic 11223
cyclic sulfamide	backbone	24, 27, 29, 48, 84		
	side chain	63		
cyclic urea	backbone	63	27, 29, 30, 48	
	side chain		29, 30	
peptidic 11223	backbone	63, 72	4, 24, 27, 29, 47, 48, 63, 72, 73, 74, 80, 83, 84	4, 18, 46, 47, 48, 49, 63, 72
	side chain	30, 63	6, 19, 29, 30, 63, 72	6, 19, 63, 72
peptidic 112233	backbone		24, 27, 29, 30, 48, 49, 83	24, 63, 72, 83
	side chain	30	30	30, 31

between the peptidic 1122/3 and the cyclic structures. The cyclic sulfamides cause the most change, with variation consistently observed at Leu24, Gly27, Asp29, Asp30, Gly48, and Asn83. The majority of these residues are found around the S2 subsite, suggesting that the cyclic sulfamides cause changes to the S2 subsite. If we compare the peptidic 11223 and peptidic 112233 data, then it is noted that variation at residues Leu63 and Ile72 is consistent for all peptidic 11223 analyses. As described above, Leu63 sits at the 'cheek' region of HIV-1 protease, while Ile72 also sits in this space and is a noted site of variation from Zoete et al.<sup>17</sup> In addition to this unexpected variation, the changes observed at residues Leu24 and Asn83 for both the peptidic 11223 and the peptidic 112233 data cannot readily be explained, although it does suggest that the longer peptidic ligands may be causing some of this variation.

Finally, we have analyzed data from Table 4 to identify which residues show variation when comparing the cyclic and the peptidic ligands. If we focus first on the peptidic 11223 data, then variation was observed at the backbone of residue Gly47 and at the side chain of Trp6 and Leu19, in comparison to both cyclic urea and sulfamide ligands. In addition, changes occurred between the cyclic sulfamide and the peptidic 11223 ligands at Gly73, Thr74, Thr80, and Ile84. For the cyclic ureas, the only additional variation was noted at the backbone of Gly49. Analysis of the peptidic 112233 data indicates that the residues with common variance for both cyclic ligand types were Asp29, Asp30, Gly48, Gly49 (backbone), and Asp30 (side chain). Additional variation at residue Gly27 was observed for the sulfamides and at Thr31 and Thr74 for the ureas.

In summary, ligand analysis has identified that HIV-1 protease does adapt to different ligand types, with the ligand size and scaffold being determining factors. Our analysis reveals that the peptidic 1122 structures sit in a conformational space between that of the larger peptidic structures and the cyclic structures. The cyclic ureas bind in a exceptionally similar manner to the peptidic 1122 ligands, while for the cyclic sulfamides adaption was observed in residues above and below the S2/S2' subsite. It is thought that this similarity is due to the peptidic 1122 and the cyclic ligands being of a similar size. Moreover, comparison of the larger peptidic and the peptidic 1122 ligands consistently identified that adaption was only observed at Asp30. It is suggested that this variation is due to the increased size of these ligands and that the overall similarities between the two groups of ligands are due to them both having the same scaffold type. When comparing the cyclic ligand structures to the larger peptidics, significant variation is observed suggesting that the different scaffolds and ligand sizes have a significant effect on the structure of HIV-1 protease. Interestingly, the majority of the significant changes observed

occur at the backbone and not at the side chain, as might have been expected.

**3.6. Comments on the Structural Analysis.** Despite all that has been discovered from this structural analysis, there are a number of limitations in the data set that may influence the conclusions that have been reached. First, the fact that space group has a quantitative effect on Res-RMSD meant the sample sizes were limited, even though there is an extremely large population of HIV-1 protease structures to select from. With small sample sizes, it is difficult to be definitive since the data could possibly be biased by increased or decreased Res-RMSD values in the natural variation analysis. We have investigated the effect of space group, mutation, and bound ligand independently of each other, however we cannot rule out the interdependence of these effects due to the limited data available.

Another limitation is that the paired structures in the natural variation analysis have different resolutions. It is possible that a lot of the variation between ‘same structures’ is actually due to this resolution difference and not to the natural variation in HIV-1 protease. As an example, the 2bxp and 1hsg structures have resolutions of 2.8 and 2.0 Å, respectively. These two structures also highlight a further limitation in that the derivation of structures can be fraught with human error. In the 2bxp structure, the carbonyl group of Asp60 points toward the oxygen of Gln58, which is electrostatically unfavored. This contrasts with 1hsg, which has the Asp60 carbonyl group pointing toward the nitrogen atom of Gln58. This is an example of a well-known rotamer issue in crystal structures for glutamine and asparagines residues, resulting from the fact that they both have amide functionality, and it is difficult to distinguish oxygen and nitrogen electron densities.

These issues will likely be resolved when more high-resolution structures become available and when the use of validation tools such as NQ flipper<sup>33</sup> are a standard part of the protocol for the production of crystal structures. This would allow greater confidence in the Res-RMSD data and would also diminish the effect that any outlying data has on the analysis.

X-ray structures are static representations of dynamic molecular systems and are predominantly the basis for modeling studies. However it is appreciated that these X-ray structures are influenced by crystal packing, crystal contact, solvent effects, and even difficulties in the crystallographic refinement procedures.

## 4. CONCLUSIONS

We have investigated the flexibility and adaptability of HIV-1 protease arising from the stresses of key mutations and binding of different ligands. The analysis has focused on both the backbone and side chains of residues, which significantly extends previously reported analysis. The flexibility analysis agrees qualitatively with previous work but shows significant quantitative differences, in particular that the space group of the complex does have an effect on the structural analysis. By calculating control Res-RMSD data from identical structures, residues that vary naturally have been identified. It has also been demonstrated that when HIV-1 protease has a bound ligand, it no longer retains its  $C_2$ -symmetry. Additionally, in the majority of structures the conformation of Ile50 can be used to distinguish the two chains of the homodimer, for which we have introduced nomenclature to uniquely identify the two chains.

It has been shown that mutation results in structural changes at the site of mutation and at residues far from the mutation site.

Of particular significance for the study of resistance, we have observed that it is common for residues at major mutation sites to adapt to mutations made at other sites. It was also identified that major mutations generally occur at sites of limited natural variation; that is, rigid regions of the structure. These results suggest that the standard methods employed in most visualization programs for including mutations in structures are incorrect, as they generally do not consider changes distant from the mutation site.

Analysis of the effect that different ligands had on the structure of HIV-1 protease indicated that ligands of similar scaffold type and size produced similar structural effects. For example, the cyclic urea and the cyclic sulfamide inhibitors, with similar scaffolds, produced a similar effect on the protease structure. In the same manner, the peptidic ligands could additionally be classified by size, with the larger 11223 and 112233 ligands grouped together, separate from the 1122 ligands. Interestingly, the peptidic 1122 and the cyclic ligands, which are of similar size but possess different scaffolds, produced similar effects on the structure of HIV-1 protease. Moreover, the smaller 1122 peptidic ligands and the larger peptidic ligands (11223, 112233), which are of different sizes but possess a similar scaffold, also produced similar effects. However, the structural changes produced by the larger peptidic ligands were significantly different than those of the cyclic ligands, which results from the different size and scaffold of these ligands. This highlights the fine balance between ligand size and scaffold type and indicates why the peptidic 1122 ligands have been grouped separately from the larger peptidic ligands.

Conclusions from this analysis should be considered in the design of new ligands for HIV-1 protease. That is, it is suggested that active site residues identified as being rigid should be the focus of binding in ligand design, while flexible active site residues should not be targeted. It is hoped that with a greater understanding of the changes that occur for a particular mutation or type of ligand, new ligands can be developed and adapted such that they will not be affected by the changes that take place upon mutation, and therefore the resistance can eventuate. Conclusions from this analysis could also be used for the generation of hypothetical models of complexes by virtual MD docking process where no X-ray or NMR structures are available. It would be possible to place a drug in the active site based on the structure of another similar complex, mutate the HIV-1 protease to the particular sequence required, and then allow the variable regions as identified from this work to move while keeping the rest of the complex restrained. After sufficient equilibration, it would be expected that a reliable and an accurate model of the new complex would be produced, from which further investigations could result.

## ■ ASSOCIATED CONTENT

**S Supporting Information.** Tables identifying structures used in the overall, natural variation, mutational, and ligand analysis; tables of carbonyl Ile50 distances; and tables of the full natural variation data. This information is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: r.brownlee@latrobe.edu.au.

## ■ ACKNOWLEDGMENT

D.O. was supported by an Australian Postgraduate Award (APA) scholarship. The authors acknowledge support from the National Computational Infrastructure National Facility (NCI-NF), Victorian Partnership for Advanced Computing (VPAC), Victorian Life Science Computing Initiative (VLSCI), and the high-performance computing facility of La Trobe University. We also thank the reviewers for insightful comments.

## ■ REFERENCES

- (1) Karplus, M.; Kuriyan, J. Molecular Dynamics and protein function. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (19), 6679–6685.
- (2) Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002**, *9* (9), 646–652.
- (3) Karplus, M.; Petsko, G. A. Molecular dynamics simulations in biology. *Nature* **1990**, *347*, 631–639.
- (4) Lodish, H.; Berk, A.; Zipursky, S. L.; Matsudaira, P.; Baltimore, D.; Darnell, J. E. *Molecular Cell Biology*, 4th ed.; W. H. Freeman and Company: New York, 2000.
- (5) Amano, M.; Koh, Y.; Das, D.; Li, J.; Leschenko, S.; Wang, Y.-F.; Boross, P. I.; Weber, I. T.; Ghosh, A. K.; Mitsuya, H. A Novel Bis-Tetrahydrofurylurethane-Containing Nonpeptidic protease Inhibitor (PI), GRL-9806S, Is Potent against Multiple-PI-Resistant Human Immunodeficiency Virus In Vitro. *Antimicrob. Agents Chemother.* **2007**, *S1* (6), 2143–2155.
- (6) Kozisek, M.; Bray, J.; Rezacova, P.; Saskova, K.; Brynda, J.; Pokorna, J.; Mammano, F.; Rulisek, L.; Konvalinka, J. Molecular Analysis of the HIV-1 Resistance Development: Enzymatic Activities, Crystal Structures, and Thermodynamics of Nelfinavir-resistant HIV protease Mutants. *J. Mol. Biol.* **2007**, *374*, 1005–1016.
- (7) Mahalingam, B.; Louis, J. M.; Reed, C. C.; Adomat, J. M.; Krouse, J.; Wang, Y.-F.; Harrison, R. W.; Weber, I. T. Structural and kinetic analysis of drug resistant mutants of HIV-1 protease. *Eur. J. Biochem.* **1999**, *263*, 238–244.
- (8) Chen, X.; Weber, I. T.; Harrison, R. W. Molecular Dynamics simulations of 14 HIV protease mutants in complexes with indinavir. *J. Mol. Model.* **2004**, *10*, 373–381.
- (9) Hong, L.; Zhang, X. C.; Hartsuck, J. A.; Tang, J. Crystal structure of an *in vivo* HIV-1 protease mutant in complex with saquinavir: insights into the mechanisms of drug resistance. *Protein Sci.* **2000**, *9*, 1898–1904.
- (10) King, N. M.; Melnick, L.; Prabu-Jeyabalan, M.; Nalivaika, E. A.; Yang, S.-S.; Gao, Y.; Nie, X.; Zepp, C.; Heefner, D. L.; Schiffer, C. A. Lack of synergy for inhibitors targeting a multi-drug-resistant HIV-1 protease. *Protein Sci.* **2002**, *11*, 418–429.
- (11) Mahalingam, B.; Boross, P.; Wang, Y.-F.; Louis, J. M.; Fischer, C. C.; Tozser, J.; Harrison, R. W.; Weber, I. T. Combining Mutations in HIV-1 protease to Understand Mechanisms of Resistance. *Proteins* **2002**, *48*, 107–116.
- (12) Mahalingam, B.; Wang, Y.-F.; Boross, P. I.; Tozser, J.; Louis, J. M.; Harrison, R. W.; Weber, I. T. Crystal structures of HIV protease V82A and L90M mutants reveal changes in the indinavir-binding site. *Eur. J. Biochem.* **2004**, *271*, 1516–1524.
- (13) Perryman, A. L.; Lin, J.-H.; McCammon, J. A. HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: Possible contributions to drug resistance and a potential new target site for drugs. *Protein Sci.* **2004**, *13*, 1108–1123.
- (14) Tie, Y.; Boross, P. I.; Wang, Y.-F.; Gaddis, L.; Liu, F.; Chen, X.; Tozser, J.; Harrison, R. W.; Weber, I. T. Molecular basis for substrate recognition and drug resistance from 1.1 to 1.6 Å resolution crystal structures of HIV-1 protease mutants with substrate analogs. *FEBS J.* **2005**, *272*, 5265–5277.
- (15) Wartha, F.; Horn, A. H. C.; Meiselbach, H.; Sticht, H. Molecular Dynamics Simulations of HIV-1 protease Suggest Different Mechanisms Contributing to Drug Resistance. *J. Chem. Theory Comput.* **2005**, *1*, 315–324.
- (16) Bardi, J. S.; Luque, I.; Freire, E. Structure-Based Thermodynamic Analysis of HIV-1 protease Inhibitors. *Biochemistry* **1997**, *36*, 6588–6596.
- (17) Zoete, V.; Michelin, O.; Karplus, M. Relation between Sequence and Structure of HIV-1 protease Inhibitor Complexes: A Model System for the Analysis of Protein Flexibility. *J. Mol. Biol.* **2002**, *315*, 21–52.
- (18) Kumar, M.; Hosur, M. V. Adaptability and flexibility of HIV-1 protease. *Eur. J. Biochem.* **2003**, *270*, 1231–1239.
- (19) Humphrey, W.; Dalke, A.; Schulter, K. VMD - Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (20) Wlodawer, A.; Miller, M.; Jaskolski, M.; Sathyaranayana, B. K.; Baldwin, E.; Weber, I. T.; Selk, L. M.; Clawson, L.; Schneider, J.; Kent, S. B. H. Conserved Folding in Retroviral proteases: Crystal Structure of a Synthetic HIV-1 protease. *Science* **1989**, *245*, 616–621.
- (21) Navia, M. A.; Fitzgerald, P. M. D.; McKeever, B. M.; Leu, C.-T.; Heimbach, J. C.; Herber, W. K.; Sigal, I. S.; Darke, P. L.; Springer, J. P. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* **1989**, *337*, 615–620.
- (22) Clemente, J. C.; Hemrajani, R.; Blum, L. E.; Goodenow, M. M.; Dunn, B. M. Secondary Mutations M36I and A71V in the Human Immunodeficiency Virus Type 1 protease Can Provide an Advantage for the Emergence of the Primary Mutation D30N. *Biochemistry* **2003**, *42*, 15029–15035.
- (23) Andersson, H. O.; Fridborg, K.; Lowgren, S.; Alterman, M.; Muhllman, A.; Bjorsne, M.; Garg, N.; Kvarnstrom, I.; Schaaf, W.; Classon, B.; Karlen, A.; Danielsson, U. H.; Ahlsen, G.; Nillroth, U.; Vrang, L.; Oberg, B.; Samuelsson, B.; Hallberg, A.; Unge, T. Optimization of P1-P3 groups in symmetric and asymmetric HIV-1 protease inhibitors. *Eur. J. Biochem.* **2003**, *270*, 1746–1758.
- (24) Wittayanarakul, K.; Aruksakunwong, O.; Sompornpisut, P.; Sanghiran-Lee, V.; Parasuk, V.; Pinitglang, S.; Hannongbua, S. Structure, Dynamics and Solvation of HIV-1 protease/Saquinavir Complex in Aqueous Solution and Their Contributions to Drug Resistance: Molecular Dynamic Simulations. *J. Chem. Inf. Model.* **2005**, *45*, 300–308.
- (25) Tie, Y.; Boross, P. I.; Wang, Y.-F.; Gaddis, L.; Hussain, A. K.; Leshchenko, S.; Ghosh, A. K.; Louis, J. M.; Harrison, R. W.; Weber, I. T. High Resolution Crystal Structures of HIV-1 protease with a Potent Non-peptide Inhibitor (UIC-94017) Active Against Multi-drug-resistant Clinical Strains. *J. Mol. Biol.* **2004**, *338*, 341–352.
- (26) Fitzgerald, P. M. D.; McKeever, B. M.; VanMiddlesworth, J. F.; Springer, J. P.; Heimbach, J. C.; Leu, C.-T.; Herber, W. K.; Dixon, R. A. F.; Darke, P. L. Crystallographic Analysis of a Complex between Human Immunodeficiency Virus Type 1 protease and Acetyl-Pepstatin at 2.0-A Resolution. *J. Biol. Chem.* **1990**, *265* (24), 14209–14219.
- (27) Eigenbrot, C.; Randal, M.; Kossiakoff, A. A. Structural Effects Induced by Mutagenesis Affected by Crystal Packing Factors: The Structure of a 30–51 Disulfide Mutant of Basic Pancreatic Trypsin Inhibitor. *Proteins* **1992**, *14*, 75–87.
- (28) Prabu-Jeyabalan, M.; Nalivaika, E. A.; Romano, K.; Schiffer, C. A. Mechanism of Substrate Recognition by Drug-Resistant Human Immunodeficiency Virus Type 1 protease Variants Revealed by a Novel Structural Intermediate. *J. Virol.* **2006**, *80* (7), 3607–3616.
- (29) Kovalevsky, A. Y.; Tie, Y.; Liu, F.; Boross, P. I.; Wang, Y.-F.; Leshchenko, S.; Ghosh, A. K.; Harrison, R. W.; Weber, I. T. Effectiveness of Nonpeptide Clinical Inhibitor TMC-114 on HIV-1 protease with Highly Drug Resistant Mutations D30N, I50V, and L90M. *J. Med. Chem.* **2006**, *49*, 1379–1387.
- (30) Rhee, S.-Y.; Gonzales, M. J.; Kantor, R.; Betts, B. J.; Ravela, J.; Shafer, R. W. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res.* **2003**, *31*, 298–303.
- (31) Ode, H.; Neya, S.; Hata, M.; Sugiura, W.; Hoshino, T. Computational Simulations of HIV-1 proteases - Multi-drug Resistance Due to Nonactive Site Mutation L90M. *J. Am. Chem. Soc.* **2006**, *128*, 7887–7895.
- (32) Stoica, I.; Sadiq, S. K.; Coveney, P. V. Rapid and Accurate Prediction of Binding Free Energies for Saquinavir-Bound HIV-1 protease. *J. Am. Chem. Soc.* **2008**, *130*, 2639–2648.
- (33) Weichenberger, C. X.; Sippl, M. J. NQ-Flipper: validation and correction of asparagine/glutamine amide rotamers in protein crystal structures. *Bioinformatics* **2006**, *22*, 1397–1398.