BIOINFORMATICS ORIGINAL PAPER Vol. 28 no. 3 2012, pages 350–357
doi:10.1093/bioinformatics/btr683

*Structural bioinformatics*

# Inhibition of HIV-1 protease: the rigidity perspective

J. W. Heal[1,*], J. E. Jimenez-Roldan[2,3], S. A. Wells[2], R. B. Freedman[3] and R. A. Römer[2,*]

[1]MOAC Doctoral Training Centre, [2]Department of Physics and Centre for Scientific Computing and [3]School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK

Associate Editor: Anna Tramontano

**ABSTRACT**

**Motivation:** HIV-1 protease is a key drug target due to its role in the life cycle of the HIV-1 virus. Rigidity analysis using the software FIRST is a computationally inexpensive method for inferring functional information from protein crystal structures. We evaluate the rigidity of 206 high-resolution (2 Å or better) X-ray crystal structures of HIV-1 protease and compare the effects of different inhibitors binding to the enzyme.

**Results:** Inhibitor binding has little effect on the overall rigidity of the protein homodimer, including the rigidity of the active site. The principal effect of inhibitor binding on rigidity is to constrain the flexibility of the $\beta$-hairpin flaps, which move to allow access to the active site of the enzyme. We show that commercially available antiviral drugs which target HIV-1 protease can be divided into two classes, those which significantly affect flap rigidity and those which do not. The non-peptidic inhibitor tipranavir is distinctive in its consistently strong effect on flap rigidity.

**Contact:** jack.heal@warwick.ac.uk; r.roemer@warwick.ac.uk

**Supplementary information:** Supplementary data are available at *Bioinformatics* online.

Received on August 15, 2011; revised on October 28, 2011; accepted on December 3, 2011

## 1 INTRODUCTION

HIV-1 protease is a key drug target due to its role in the life cycle of the HIV virus (Erickson, 1995), and many protease inhibitors have been designed to block its function and to prevent virus replication (Flexner, 1998). HIV-1 is a retrovirus; after it enters a host cell, its viral RNA is reverse transcribed into double-stranded DNA and then integrated into host DNA. The host cell subsequently transcribes this integrated virus-specific DNA and translates the transcribed message into a single extended translation product containing the sequences of all the viral-encoded proteins required to make further copies of the virus. The role of HIV-1 protease is to cleave this product at specific cleavage sites to generate the individual functional viral proteins. Successful inhibition of the protease prevents this maturation step and hence blocks proliferation of the virus.

Many X-ray crystal structures of HIV-1 protease have been resolved and deposited in the RSCB Protein Data Bank (PDB) (Berman, 2000). The wild-type protease is a symmetric homodimer with a largely $\beta$-sheet secondary structure (Wlodawer *et al.*, 1989). The active site of the protease lies in a cleft where the two chains meet; this cleft is covered by $\beta$-hairpin 'flaps' (Hornak and Simmerling, 2007). Experimentally, nuclear magnetic resonance (NMR) (Freedberg *et al.*, 2001; Ishima *et al.*, 1999; Katoh *et al.*, 2003; Nicholson *et al.*, 1995) and fluorescence spectroscopy (Rodríguez *et al.*, 1993) have been used to study the flexibility of the protein. The flaps have been found to be highly flexible in the unbound enzyme, with internal flap motion occurring on a sub-ns timescale (Freedberg *et al.*, 2001; Katoh *et al.*, 2003). On a timescale of 100 $\mu$s, they undergo a different large amplitude motion in slow dynamic equilibrium between semi-open and less ordered open conformations (Ishima *et al.*, 1999). The conformational dynamics of the protease have also been investigated using molecular dynamics simulations (Harte and Beveridge, 1993; Swaminathan *et al.*, 1991; York *et al.*, 1993; Zhu *et al.*, 2003) and coarse-grained models (Chang *et al.*, 2006; Tozzini and McCammon, 2005). Again, the flaps have been found to be highly flexible with distances between their tips ranging from 7 Å to 25 Å (Collins *et al.*, 1995). The conversion between closed, semi-open and fully open flaps has been simulated in qualitative agreement with experimental data (Hornak and Simmerling, 2007; Hornak *et al.*, 2006).

Here we examine HIV-1 protease by performing rigidity analysis on a set of 206 X-ray crystal structures that have been solved to a resolution of 2 Å or better and made available in the PDB. In our analysis, we determine the locations of flexible regions within the protein and hence study the potential for motion. FIRST identifies rigid clusters and flexible regions by matching atomic degrees of freedom with the constraints present due to bonding (Jacobs and Thorpe, 1995; Jacobs *et al.*, 2001; Thorpe *et al.*, 2001). We examine the effect of inhibitor binding on flap rigidity using a comparative approach; for each structure crystallized with an inhibitor bound, we perform rigidity analysis on the structure with the inhibitor present and on the same structure with the inhibitor deleted (Hornak *et al.*, 2006).

## 2 METHODS

### 2.1 Structure of HIV-1 protease

Wild-type HIV-1 protease is a symmetrical homodimer comprising polypeptide chains each containing 99 amino acid residues. We refer to the two monomers as chain A and chain B, with residues numbered from 1 to 99 in chain A and from 101 to 199 in chain B.

Figure 1 shows the structure of HIV-1 protease. We show the structure 3LZU because its inhibitor darunavir is one which we study in detail later in this article. The secondary structure motifs are clearly shown as arrows ($\beta$-sheet) and helices ($\alpha$-helix). The terminal region is situated at the base of the enzyme and consists of the N- and C-termini of both monomers, the turn located immediately above this and the two $\alpha$-helices towards the

*To whom correspondence should be addressed.

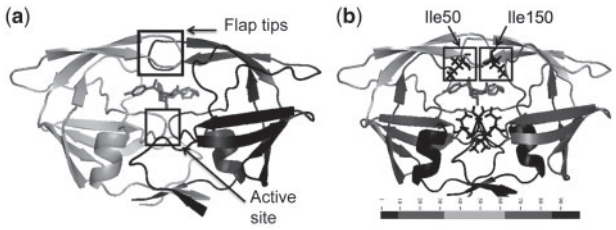


**Fig. 1.** Structure of HIV-1 protease. (**a**) The structure 3LZU presented using the 'cartoon' option in PYMOL (http://www.pymol.org), with chain A shown in grey and chain B in black. This highlights the secondary structure of the protein, which is largely $\beta$-sheet. The heavy atoms of the inhibitor darunavir are shown as grey sticks, between the active site and the $\beta$-hairpin flap tips. (**b**) The same structure is shown, this time shaded by region. The primary structure of chain A, labelled with residue numbers, is given beneath with the same shading. The terminal region is shown in black, the core in dark grey and the flaps in light grey; the inhibitor is shown in the same way as in (a). The active site and the two isoleucine residues Ile50 and Ile150 at the very tips of the flaps are shown as black sticks.

C-terminal of the monomers. The core region is composed primarily of $\beta$-strands. The active site is situated within the core and consists of three residues from each of the monomers. This pair of Asp–Thr–Gly triads is a form of active site common to aspartic proteases (Toh *et al.*, 1985). The catalytic triads are located at residues 25–27 and, due to symmetry, residues 125–127. The active site lies beneath the flaps, the tips of which are indicated in Figure 1. The flap region extends out to the flap 'elbows' which, it has been suggested, act as a cantilever for flap motion (Harte *et al.*, 1992; Swaminathan *et al.*, 1991). These are located at the widest section of the protein. The flap tips are a glycine-rich area of the protein and thus have enhanced conformational flexibility (Hong *et al.*, 1997).

## 2.2 Selection of structures and inhibitors

A large number of structures of HIV-1 protease, with various mutations in the sequence and with various inhibitors bound, are available in the PDB. We have selected a set of 206 structures (Supplementary Fig. S4 and Table S2) with a resolution of 2 Å or better so as to ensure that the hydrogen bond geometry assigned and used by FIRST is accurate.

In order to prepare the crystal structures for rigidity analysis, all the crystal water molecules are removed, leaving the dimer unit of HIV-1 protease along with the inhibitor with which it was crystallized. It has previously been suggested that the presence of buried water is important as it coordinates the protease flap residues (Wlodawer and Vondrasek, 1998). However, we note that this should not affect the rigidity analysis presented by FIRST, as molecular dynamics simulations (Mamonova *et al.*, 2005) have shown that the hydrogen bonds formed between the protein and the solvent have a short duty cycle of typically <1 ps, comparable to those in liquid water (Matsumoto *et al.*, 2002), and should be neglected when considering rigidity. The REDUCE software (Word *et al.*, 1999) is then used to add the hydrogen atoms that are not present in the X-ray crystal structure and to flip side chains of residues such as asparagine, glutamine and histidine to avoid clashes and improve the hydrogen-bonding network. For a set of eight test cases, we have examined the effect of running REDUCE before deletion of the crystal waters, and no significant difference in our results was observed. The 206 structures have a range of crystallization conditions, inhibitors and mutations. For each structure that was crystallized with a ligand, a copy is made with its inhibitor manually deleted, as in the molecular dynamics study by Hornak *et al.* (2006). We take it that this inhibitor-free structure is part of the ensemble of conformations explored by the inhibitor-free protein during its natural flexible motion (Hornak *et al.*, 2006).

Within the full set of 206 structures, we select a subset of 'main inhibitors' based on some of the commercially available protease inhibitors approved for use in antiretroviral therapies. In accordance with Pokorná *et al.* (2009), we refer to the US Food and Drug Administration (FDA) approved inhibitors by their three-letter abbreviations: APV (amprenavir, available as the pro-drug fosampranavir), ATZ (atazanavir), DRV (darunavir), IDV (indinavir), NFV (nelfinavir), SQV (saquinavir) and TPV (tipranavir). We also include DMP-323 and the polypeptidic inhibitor (ACE)TI(NLE)(NLE)QR due to the relatively large number of high-resolution structures crystallized with these inhibitors. We label DMP-323 as DMP and (ACE)TI(NLE)(NLE)QR as PEP. There are a different number of structures for each of the main inhibitors, which are as follows: APV – 8, ATZ – 4, DRV – 16, IDV – 5, NFV – 4, SQV – 6, TPV – 4, DMP – 5 and PEP – 5, making a total of 57 structures.

## 2.3 Rigidity analysis

A protein can be considered as a 'molecular framework' consisting of atoms and bonding constraints. The software FIRST carries out rigidity analysis on such a framework using the 'pebble game' algorithm (Jacobs and Thorpe, 1995; Jacobs *et al.*, 2001), which matches degrees of freedom to constraints, dividing the structure into overconstrained, isostatic or underconstrained regions. An underconstrained region is flexible, in the sense that dihedral angles can vary and so atoms can move without violating distance constraints (Thorpe *et al.*, 2001). An isostatic region has an exact balance of constraints and degrees of freedom; the removal of any constraint would make the region flexible. In overconstrained regions, it is possible to remove a constraint without altering the rigidity of the region—there are constraints to spare. A division such as this is called a rigid cluster decomposition (RCD).

The results of an RCD obviously depend on the set of bonds that are included to rigidify the structure. In FIRST, the strong bonding forces such as covalent bonds and hydrophobic interactions are always included. Long-range electrostatic forces and van der Waals forces do not contribute to the distance constraints between atoms because they are generally weak at this level and not highly directional (Jacobs and Thorpe, 1995). Hydrogen bonds and salt bridges are included in the constraint network on a selective basis. In FIRST, the energy of each potential hydrogen bond in the structure is calculated using the Mayo potential (Dahiyat *et al.*, 1997) as a function of the geometry of the donor, hydrogen and acceptor atoms. Bonds are included as constraints if their energy lies below an energy cutoff $E_{\mathrm{cut}}$, which is usually given in kcal/mol. Thus, $E_{\mathrm{cut}}$ determines the bond network, which in turn determines the RCD. This dependence on hydrogen bond geometry means that the atomic coordinates need to be precise, hence our selection of high-resolution structures. An RCD can be visualized by appropriate labelling of the 3D structure of a protein. For easier comparison between different structures, it is also useful to consider the *mainchain rigidity* of the protein. On a 1D representation of the protein sequence, residues are labelled according to the rigid cluster membership of the $C_\alpha$ atom. Large rigid clusters are represented by coloured blocks, while flexible regions are represented by a thin black line.

Supplementary Figure S1 gives examples of the rigidity of the structure 3LZU, in both 3D and 1D representations. Some key features of the RCD are clearly visible; a single rigid cluster can include residues from several non-contiguous sections of the primary sequence, and indeed in this dimeric protein a single large rigid cluster can include residues from both chains A and B. The 1D representation shows clearly the change from a largely rigid structure to a largely flexible structure with a small shift in $E_{\mathrm{cut}}$. A series of 1D representations makes up a 'rigidity dilution plot', examples of which are shown in Figure 2.

# 3 RESULTS

## 3.1 Rigidity dilution

In a 'rigidity dilution' (Wells *et al.*, 2009), hydrogen bonds are removed in order from weakest to strongest by a gradual lowering of $E_{\mathrm{cut}}$, as discussed in the previous section. The resulting loss in rigidity is visualized by plotting a new 1D representation of the RCD whenever main chain rigidity changes. This highlights where

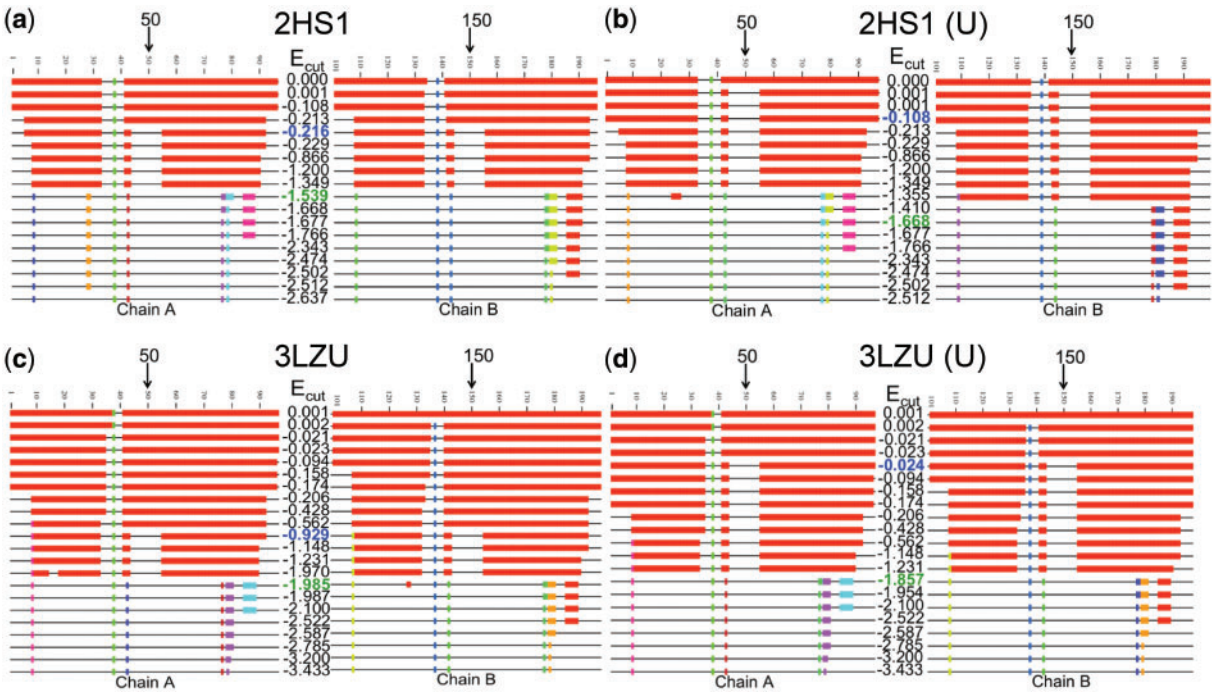


**Fig. 2.** Rigidity dilution plots for the two crystal structures with PDB codes 2HS1 and 3LZU. Rigidity dilution was carried out on the structures as crystallized (**a**), (**c**) and after deletion of the inhibitor (**b**), (**d**), indicated by the suffix (U). The areas surrounding the 50th and the 150th residue of each structure, which correspond to the flap tips are indicated by labelled arrows. The highest $E_{cut}$ at which residue 50 is flexible is $E_{flap}$ and is highlighted in blue on the vertical axis. As explained in the text, in (b) we take $E_{flap}$ to be the highest $E_{cut} < 0$. The energy cutoff at which 70% of the backbone becomes flexible, $E_{body}$, is highlighted in green on the vertical axis of each plot.

in the structure most of the rigidity resides, and where rigidity is most readily lost. We note that the manner in which these plots are constructed means that the vertical axis is non-linear with $E_{cut}$ (Wells *et al.*, 2009).

Figure 2 consists of four dilution plots describing the rigidity of two crystal structures, 3LZU and 2HS1, before and after the deletion of an inhibitor. The structures 2HS1 and 3LZU were both selected because they were crystallized with the inhibitor darunavir, an antiviral drug that was approved by the FDA in June 2006 and was designed to form strong interactions with many different mutated structures of the protease (Ghosh *et al.*, 2007). The dilution plots of the structures as they were crystallized are given in Figure 2a and c. In both cases, the protein loses the majority of its rigidity abruptly, as the 1D profiles change from being mostly rigid to being mostly flexible with a single step in $E_{cut}$. This is first-order rigidity loss, as expected for a protein with a largely $\beta$-sheet secondary structure (Wells *et al.*, 2009). The flap regions around residues 50 and 150 are visible in the rigidity dilution as they become flexible at higher $E_{cut}$ than the main body of the protein. The helical regions near the terminal residues 86–94 are also visible as their rigidity persists after the main body of the protein has become flexible. Previous studies (Tan and Rader, 2009; Wells *et al.*, 2009) have included structures with resolution worse than 2 Å, and so we also examined a small selection of such structures and verified that rigidity analysis was qualitatively consistent: each showed first-order rigidity loss with the flaps becoming flexible before the main body of the protein. (Supplementary Table S1 and Fig. S2).

Plots for the same structures with the inhibitors deleted are given in Figure 2b and d. Deletion of the inhibitor does not appear to alter the basic pattern of rigidity loss during the dilution. However, the flap regions become flexible at even higher $E_{cut}$. The active site of the enzyme is not distinctive in the rigidity analysis of either the inhibited or uninhibited structures, forming part of the largest rigid cluster until the protein becomes mostly flexible. The principal effect of the inhibitor upon the rigidity profile of the protein is to rigidify the flap tip region rather than the active site.

For quantitative analysis, it is typical to extract significant energy cutoff values from the RCD plots. Previous studies such as those on rigidity in the context of thermostability (Rader, 2010; Radestocke and Gohlke, 2008, 2011) have considered measures such as the folding core energy. This is defined to be the lowest line in the dilution plot where at least three consecutive residues are mutually rigid with at least three other consecutive residues of another secondary structural element (Hespenheide *et al.*, 2002; Tastan *et al.*, 2007). In this instance we are interested in the function rather than the folding or melting point of the protein and so, in order to quantify the influence of inhibitor binding on the overall rigidity of the protein and on the flexibility of the flap regions, we define two significant values of $E_{cut}$ for our analysis. We define $E_{body}$ to be the highest $E_{cut}$ at which at least 70% of the protein is flexible. Due to the first-order nature of the rigidity loss in HIV-1 protease, this value would not change much if we made our definition of $E_{body}$ based on any flexibility in the 50–90% range (data not shown). We also note that for HIV-1 protease the folding core energy and $E_{body}$ are similar, typically separated in the dilution plots by just one line.

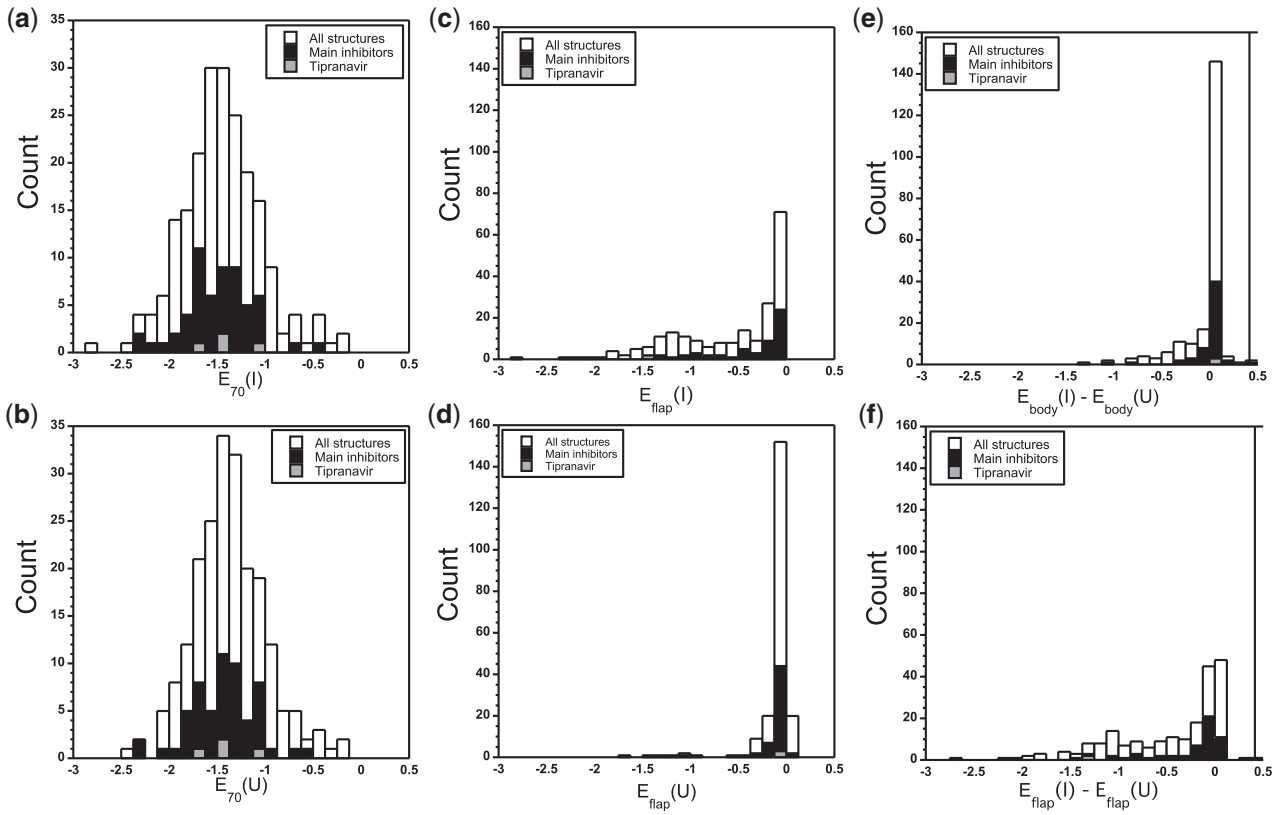


**Fig. 3.** Distributions of $E_{body}$ and $E_{flap}$ values for 206 structures. Structures complexed with TPV are coloured grey, the other 'main inhibitors' are shown in black, and the rest of the population is shown in white. Population distributions of $E_{body}$ and $E_{flap}$ are given for structures with inhibitors present (**a**) and (**c**), and after the inhibitors have been removed (**b**) and (**d**). The distributions of differences $E_{body}(I) - E_{body}(U)$ in (**e**) and $E_{flap}(I) - E_{flap}(U)$ in (**f**) give a clearer idea of what changes occur within individual structures.

Residues Ile50 and Ile150 are residues at the tip of the flaps and have been tracked previously in molecular dynamics so as to assess the motion of the flaps as the protein samples an ensemble of conformations in its native state (Zhu *et al.*, 2003). Due to the symmetry of the crystal structures used here, these residues become flexible at the same $E_{cut}$ in almost all the structures. We select $E_{flap}$ as our second important value of $E_{cut}$ to be the $E_{cut}$ at which Ile50 first becomes flexible in the rigidity dilution process. For numerical reasons, we sometimes observe that $E_{flap}$ has a small positive value, as observed in Figure 2b, in such a case, we take $E_{flap}$ to be the highest negative value of $E_{cut}$. This means that in Figure 2a for example, $E_{body} = -1.539$ kcal/mol and $E_{flap} = -0.216$ kcal/mol.

### 3.2 Inhibition and overall rigidity

The loss of rigidity occurs at different values of $E_{body}$ for different crystal structures—for example, −1.539 kcal/mol in Figure 2a and −1.985 kcal/mol in Figure 2c. Deletion of the inhibitor can shift $E_{body}$ to slightly higher values: −1.355 kcal/mol in Figure 2b and −1.857 kcal/mol in Figure 2d. We now consider whether the presence of the inhibitor significantly affects the *overall* rigidity of the protease.

In Figure 3a, we show the distribution of $E_{body}$ values for structures with the inhibitor present, which we call $E_{body}(I)$. The distribution is approximately normal with a peak around −1.5 kcal/mol; almost all values lie between −0.5 kcal/mol and −2.5 kcal/mol (Wells *et al.*, 2009). Figure 3b shows the distribution of $E_{body}$ values for 'uninhibited' structures, that is, with the inhibitor deleted. This distribution of $E_{body}(U)$ is very similar to that of $E_{body}(I)$; deletion of the inhibitor does not appear to have a strong effect on overall rigidity. Figure 3c and d show the corresponding plots for the distribution of $E_{flap}$ values. In contrast to the situation with $E_{body}$, there is a shift in the population when the inhibitors are removed, with the $E_{flap}$ values in general shifting towards zero after the removal. We note that these distributions do not tell us about the effect of inhibitor deletion on each structure individually. The distribution of individual differences $E_{body}(I) - E_{body}(U)$ and $E_{flap}(I) - E_{flap}(U)$ are shown in Figure 3e and f, respectively. Figure 3e shows a distribution strongly peaked around a modal value close to zero, with few structures having a difference of less than −0.5 kcal/mol. This confirms that deletion of the inhibitor does not typically have a strong effect on overall rigidity. However, Figure 3f shows a bimodal distribution, with many structures exhibiting a large change in $E_{flap}$ following the removal of an inhibitor. Throughout Figure 3, we also show the distribution for our set of 'main inhibitors'. The distributions for this set closely resemble the distributions for all structures, illustrating that in terms of $E_{body}$ and $E_{flap}$, the selection of 'main inhibitors' is representative of the entire population.

### 3.3 Flap rigidity and flexibility

We have seen that the flap region becomes flexible at a higher $E_{\text{cut}}$ than the main body of the dimer. We can quantify this by comparing the $E_{\text{cut}}$ value at which the flap tips become flexible, $E_{\text{flap}}$, to the value at which the main body becomes flexible, $E_{\text{body}}$. We define the *flexibility fraction* $\Phi = \frac{E_{\text{flap}}}{E_{\text{body}}}$. This measure runs from zero, when the flaps are flexible from the beginning of the dilution (in which case $E_{\text{flap}} = 0$), to 1, if the flaps become flexible only when the main body becomes flexible (i.e. $E_{\text{flap}} = E_{\text{body}}$); this normalization aids comparison between structures with different values of $E_{\text{body}}$.

In general, we denote the flexibility fraction of an inhibited structure as $\Phi(\text{I})$ and the flexibility fraction of a structure with the inhibitor deleted as $\Phi(\text{U})$. The influence of the inhibitor on the flap rigidity is measured as the difference between them and is denoted $\Delta\Phi$, so that $\Delta\Phi = \Phi(\text{I}) - \Phi(\text{U})$. Since $\Phi(\text{I}), \Phi(\text{U}) \in [0, 1]$, we have that $\Delta\Phi \in [-1, 1]$ but note that $\Delta\Phi < 0$ corresponds to the physically improbable situation where the flaps show enhanced rigidity after the deletion of the inhibitor. In the majority of cases, we observe that $\Delta\Phi \in [0, 1]$. Low positive values of $\Delta\Phi$ indicate that the presence of the inhibitor makes little difference to the flap rigidity, and high positive values indicate that the flaps are rigidified by the presence of the inhibitor. It has been previously noted that stabilizing mutations can affect the rigidity analysis computed using FIRST (Tan and Rader, 2009). The value of $\Phi$ is therefore influenced not only by the presence and identity of the inhibitor but also by the presence of any stabilizing mutations in the enzyme. However $\Phi(\text{I})$ and $\Phi(\text{U})$ are calculated each time from *the same structure* and so the effects of mutations are accounted for by considering the difference $\Delta\Phi$. We also note that our choice of $E_{\text{body}}$ over the similar folding core energy does not substantially change the value of $\Delta\Phi$ or the conclusions drawn from these values (Supplementary Fig. S2).

Initially, we examine structures crystallized with the inhibitor darunavir, as in Figure 2. Darunavir is chosen as it is a recently FDA-approved protease inhibitor (Ghosh *et al.*, 2007), and there are a relatively large number (16) of high-resolution structures of HIV-1 protease crystallized with darunavir in the PDB. The $\Phi$ values for all 16 of these structures are presented in Table 1.

We find that the $\Phi(\text{I})$ values range from 0.031 to 0.468. Two of the values, for 3D1Z and 3LZU, are noticeably higher than the others. Discounting these, there is little variation and the remaining values have a mean of 0.074. There is similar variation in the $\Phi(\text{U})$ values, although as they are smaller than the $\Phi(\text{I})$, the variation itself is of a smaller magnitude. In almost every case, we have that $0 \leq \Delta\Phi \leq 0.075$, suggesting that darunavir has in general only a small rigidifying effect on the flap tips. 2HS2 is different in that $\Delta\Phi < 0$, but the magnitude of the negative value is so small that the effect of the inhibitor on this structure is minimal, similar to the structure 3JW2. Contrastingly in 3LZU, $\Delta\Phi = 0.455$ and the inhibitor appears to have a sizeable rigidifying effect, as can be seen by comparing Figure 2c and 2d. The differences in values of $\Delta\Phi$ underline the fact that different crystallization conditions can affect the rigidity of the enzyme–inhibitor complex.

### 3.4 Effect of inhibitors on flap flexibility

The principal results of our article are shown in Figure 4. We review the distributions of $\Phi(\text{I})$, $\Phi(\text{U})$ and $\Delta\Phi$ for the full dataset of 206 structures and for our subset of 'main inhibitors'.

**Table 1.** The values of $\Phi(\text{I})$, $\Phi(\text{U})$ and $\Delta\Phi$ for 16 high-resolution structures of HIV-1 protease crystallized under various conditions with the antiviral drug darunavir, which we have chosen as a test case for investigation. We note that there is no apparent correlation between $\Delta\Phi$ and the location of mutations or the resolution (Res) of the structures

| PDB code | Mutation | Res (Å) | $\Phi(\text{I})$ | $\Phi(\text{U})$ | $\Delta\Phi$ |
|---|---|---|---|---|---|
| 1T3R | No mutation | 1.20 | 0.053 | 0.037 | 0.016 |
| 1T7I | No mutation | 1.35 | 0.102 | 0.053 | 0.049 |
| 2F80 | Core | 1.45 | 0.083 | 0.014 | 0.069 |
| 2F81 | Terminal | 1.25 | 0.076 | 0.008 | 0.068 |
| 2F8G | Flaps | 1.22 | 0.071 | 0.009 | 0.062 |
| 2HS1 | Core | 0.84 | 0.140 | 0.065 | 0.075 |
| 2HS2 | Flaps | 1.22 | 0.064 | 0.070 | −0.006 |
| 3BVB | Active-site | 1.30 | 0.083 | 0.011 | 0.072 |
| 3CYW | Flaps | 1.40 | 0.068 | 0.013 | 0.055 |
| 3D1Z | Flaps | 1.30 | 0.250 | 0.232 | 0.018 |
| 3D20 | Flaps | 1.05 | 0.053 | 0.011 | 0.042 |
| 3EKT | Multiple sites | 1.97 | 0.093 | 0.020 | 0.073 |
| 3JVY | Terminal | 1.60 | 0.079 | 0.013 | 0.066 |
| 3JW2 | Terminal | 1.80 | 0.034 | 0.034 | 0.000 |
| 3LZS | Terminal | 1.95 | 0.031 | 0.015 | 0.016 |
| 3LZU | Terminal | 1.76 | 0.468 | 0.013 | 0.455 |

Figure 4a, c and e show the distributions of $\Phi(\text{I})$, $\Phi(\text{U})$ and $\Delta\Phi$, respectively. Again, we find that the distributions for the set of 'main inhibitors' are similar to the distributions for all structures. Figure 4a shows that, unlike the distribution of $E_{\text{body}}$ values, the distribution of $\Phi(\text{I})$ values is bimodal; there are peaks near values of zero, indicating that the flaps are flexible from near the beginning of the rigidity dilution, and also near 1, indicating that the flaps form an integral part of the main rigid body of the dimer. In contrast, the distribution of $\Phi(\text{U})$ has a single peak at near-zero values, seen in Figure 4c. After deletion of the inhibitor from the structure, it is much more common for the flaps to be flexible relative to the rest of the protein at high $E_{\text{cut}}$ values. The distribution of $\Delta\Phi$ values in 4e confirms this impression. We see a broad and bimodal distribution of values with peaks around zero (minimal effect on flap rigidity) and around values near 1, indicating a strong effect on flap rigidity. The bimodality of the distribution suggests that inhibitors might fall into different classes—those which typically affect flap rigidity and those which do not.

We now consider our set of 'main inhibitors' more carefully. The distributions of $\Phi(\text{I})$, $\Phi(\text{U})$ and $\Delta\Phi$ for these inhibitors are shown using box plots in Figure 4b, d and f, respectively. The box plots demonstrate the range of values observed for each inhibitor by indicating the total range of the distribution as well as the median and the inter-quartile range. In Figure 4b, we see that the inhibitors can be divided into three groups based upon the spread of the $\Phi(\text{I})$ values. APV, DRV, IDV and NFV all have consistently low $\Phi(\text{I})$ values, indicating that the flaps are typically flexible in structures crystallized with these inhibitors. ATZ, DMP, PEP and SQV have a wide spread of values, with some of the individual structures crystallised with these inhibitors having much more rigid flaps. Tipranavir (TPV) is distinctive in that the flaps are consistently highly rigid in structures crystallized with this inhibitor. The groups have been colour coded in Figure 4b according to their different

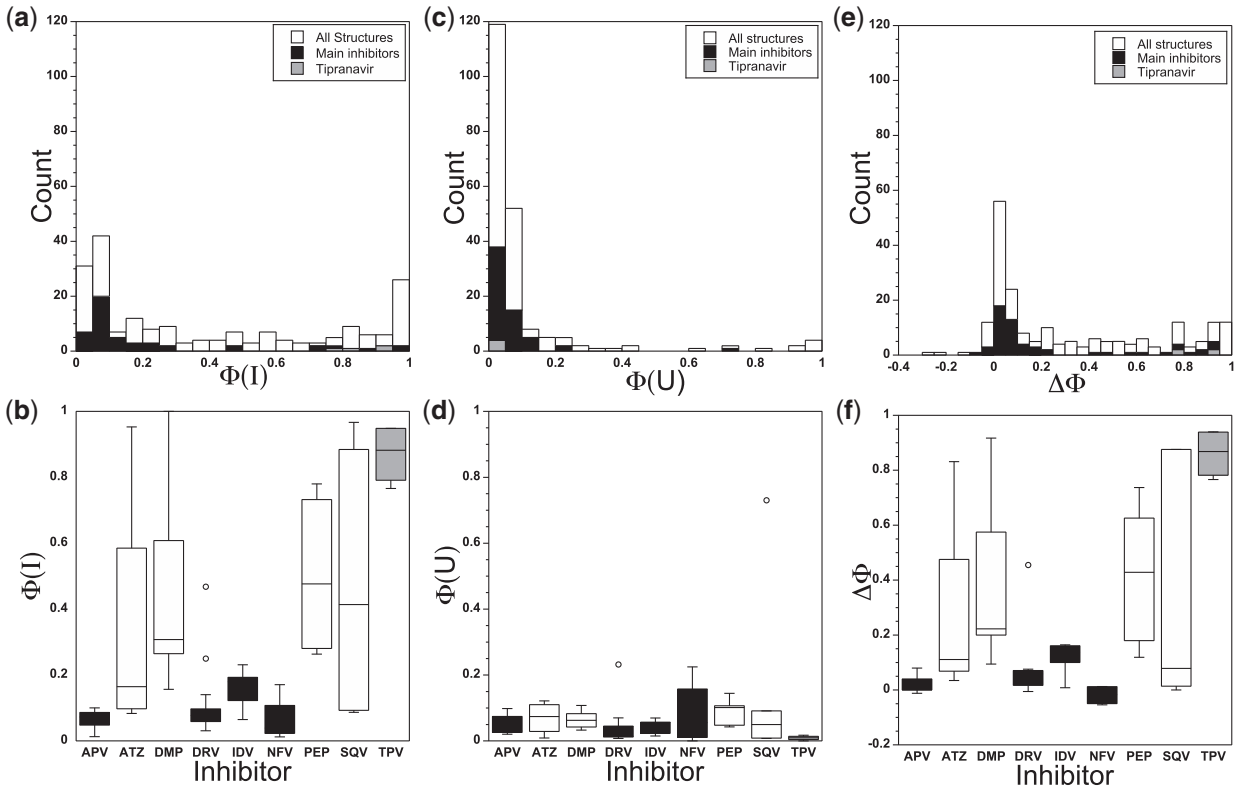


**Fig. 4.** Population distribution of flexibility fractions. Histograms for the entire population of structures are shown for Φ(I), Φ(U) and ΔΦ in (**a**), (**c**) and (**e**), respectively. Colours are as in Figure 3. Box plots of the distributions for the 'main inhibitors' are shown in (**b**) Φ(I), (**d**) Φ(U) and (**f**) ΔΦ. Inhibitors with a small effect on flap rigidity are shown in black and those with generally larger effect in white. Structures inhibited by tipranavir are shown in grey. Note that for each 'box', the horizontal line in the box corresponds to the median value and the vertical size of the box denotes the interquartile range. Error bars correspond to maximum and minimum values, with circles indicating extreme outliers.

rigidifying effects. The colour coding is maintained in Figure 4d and f. Figure 4d shows that Φ(U) values are low across all of the 'main inhibitors' including tipranavir. That is, with the inhibitors deleted, all the structures now have flexible flaps. This is evidence that the variations in flap rigidity observed in the Φ(I) values are attributable to the inhibitors, rather than being coincidental effects of the conditions of crystallization. The distribution of ΔΦ values, shown in Figure 4f, confirms this impression.

### 3.5 Inhibitor binding and inhibition mechanisms

Tipranavir is distinctive in that ΔΦ is consistently high for structures inhibited by tipranavir. Reviewing the distributions of $E_{body}$ values in Figure 3, where the tipranavir structures are highlighted, demonstrates that tipranavir is not distinctive in its effect on the overall rigidity of the protease. Thus, tipranavir appears to act specifically by *rigidifying the flaps above the active site*.

This effect is not due to tipranavir forming an unusually large number of hydrogen bonds, or indeed hydrophobic interactions, with the protease. In Figure 5, we show distributions of the number of bonds formed by inhibitors. Inhibitors with a large effect on flap rigidity have a slight tendency to form more hydrogen bonds than inhibitors with a small effect; the number of hydrophobic interactions shows little difference between the two classes. Tipranavir is not distinctive in either case, suggesting that the number of interactions is not the driving force for its large effect upon rigidifying the flaps. The value of rigidity analysis is demonstrated by its sensitivity not only to the number but also to the location of constraint formation.

Tipranavir has a unique structure among the nine main inhibitors. It is the first true non-peptidic HIV-1 protease inhibitor and has shown considerable activity against many drug-resistant strains of HIV (Pokorná *et al.*, 2009). Its flexibility is not hampered by the presence of a rigid peptide bond. Flexibility is a property which increases the chance of a drug being able to resist some mutations in the protease (Velazquez-Campoy *et al.*, 2001). Darunavir was also designed to be adaptable to drug-resistant strains (Ghosh *et al.*, 2007), but rigidity analysis indicates that tipranavir and darunavir have quite different modes of action.

Examining the interactions identified by FIRST between tipranavir and the protease using PYMOL shows that tipranavir forms hydrogen bonds with the isoleucine residues 50 and 150, and not the residues of the active site. In contrast, other inhibitors such as darunavir typically bond with the active site and not with the flap tip residues. The difference between these two types of interaction can be seen in Figure 6.

Rigidity analysis indicates two different modes of action for HIV-1 protease inhibitors: direct interaction with the active site or interaction with the flaps to rigidify them in a closed conformation and thus indirectly prevent access to the active site. The HIV-1

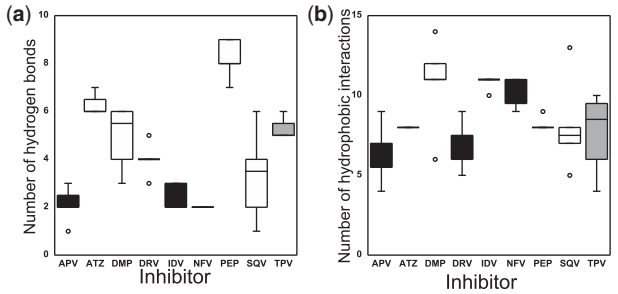


**Fig. 5.** Number of interactions between enzyme and inhibitor. The distributions of the number of hydrogen bonds (**a**) and hydrophobic interactions (**b**) between the different inhibitors and the protease are shown. The groups are coloured as in Figure 4.

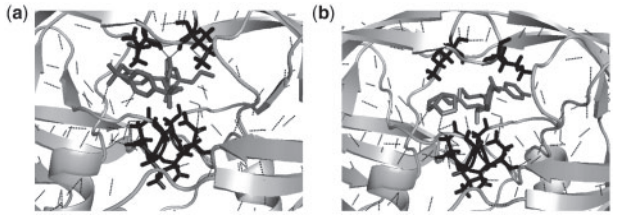


**Fig. 6.** Tipranavir binds directly to the flap tips. The hydrogen bond interactions in structures complexed with tipranavir (**a**) and darunavir (**b**). We show the secondary structure motif of the protein in grey, the active site and isoleucine flap tip residues as black sticks and the heavy atoms of the inhibitor as dark grey sticks. The hydrogen bond network is evaluated at an $E_{\text{cut}}$ of $-0.591$ kcal/mol, roughly corresponding to $kT$ at room temperature of 298 K. Each hydrogen bond is shown as a dashed black line.

protease inhibitor ritonavir became FDA-approved in 1996 and is often administered in combination with other protease inhibitors during antiretroviral therapy. There is a small selection of HIV-1 protease structures in the PDB which have been crystallized in complex with ritonavir. Although some of these structures have a worse resolution than 2 Å, we have performed a rigidity analysis and determined that they have consistently low $\Delta\Phi$ values, that is, ritonavir does not rigidify the flaps (Supplementary Fig. S3 and Table S1). Hence, the complementary approach of using one inhibitor that targets the flaps of the protease and one which does not may account for the efficacy of tipranavir–ritonavir in combination reported by Hicks *et al.* (2005).

## 4 CONCLUSIONS AND DISCUSSION

Our principal finding is that inhibitors of HIV-1 protease fall into two classes which differ in their effects on the rigidity of the $\beta$-hairpin 'flaps': those which restrict the flexibility and those which do not. Nine inhibitors, including seven FDA-approved retroviral drugs were studied in detail and of those which do restrict the flaps, tipranavir stands out as being consistently effective for this purpose. Rigidity analysis presents a way of studying the interactions an inhibitor creates with the enzyme, and hence can offer a new perspective during the process of drug design.

In summary, we have carried out an extensive comparative rigidity analysis on a set of 206 high-resolution X-ray crystal structures of HIV-1 protease in complex with a variety of inhibitors. We have performed rigidity dilution using FIRST on each structure with the inhibitor present and again with the inhibitor deleted. We describe the effect of inhibitor binding on the rigidity of the protein, using measures of overall protein rigidity and of the rigidity of the $\beta$-hairpin 'flaps' above the active site. HIV-1 protease in solution can have the flaps either open or closed. We predict that inhibitors that rigidify the flaps will bias the structure towards having the flaps closed, and that this bias will be weaker for inhibitors that do not rigidify the flaps. The prevalence of open-flap and closed-flap conformations could be probed using spectroscopic methods. We confirm the finding (Wells *et al.*, 2009) that HIV-1 protease shows an abrupt or first-order loss of rigidity. The presence or absence of an inhibitor does not typically have a significant effect on the energy cutoff at which rigidity is lost during the dilution. The active site of the protease is not distinctive in the rigidity analysis, typically forming part of the main rigid cluster until overall rigidity is lost. In contrast, the $\beta$-hairpin flaps are observable in the rigidity analysis as they typically become flexible at higher $E_{\text{cut}}$ than the main body of the protein. This effect is observed across the population of X-ray crystal structures. The flexible nature of the flaps is in qualitative agreement with molecular dynamics simulations as well as experimental work using NMR (Hornak *et al.*, 2006; Nicholson *et al.*, 1995; Zhu *et al.*, 2003).

The flaps are usually considered to have high conformational flexibility—due in part to their high glycine content (Freedberg *et al.*, 2001). In molecular dynamics simulations, the presence of glycine is conducive to flexibility as its minimal R group allows wider variation in backbone dihedral angles than for other amino acids. Here we find that glycine is also conducive to flexibility in *rigidity analysis*, but for a different reason: it is not capable of forming side chain interactions, either hydrogen bonds or hydrophobic tethers, and so tends to introduce fewer constraints into the system.

We construct the 'flexibility fraction' as a normalized measure of the rigidity of the flaps compared to the main body of the protein. We find that the effect of inhibitor binding on flap flexibility varies greatly, from zero/minimal to large rigidifying effect. Detailed examination of a set of inhibitors including FDA-approved antiretroviral medications shows that inhibitors can be divided into two classes: those that typically do not rigidify the flaps, and those which do. The rigidifying effect of the inhibitors is not a direct consequence of the number of hydrogen bonds or hydrophobic interactions that are formed with the protease, but rather the location of the interactions. Tipranavir, a non-peptidic inhibitor, is distinctive for the strength and consistency of its effect in rigidifying the flaps. The flexibility of the flaps could serve to permit the 'flexibility-assisted catalysis' mechanism introduced by Piana *et al.* (2002) and thus tipranavir may be acting to combat this mechanism. We observe that it eschews interaction with the active site while forming hydrogen bond interactions with the isoleucine residues at the tips of the flaps. This suggests that its mode of action is quite distinct from that of inhibitors such as darunavir, which interact with the active site rather than the flaps. An inhibitor capable of interacting with the active site and the flaps simultaneously may be particularly effective and/or particularly difficult for the virus to evade by mutation. The combination of one protease inhibitor from each class appears to be effective in current multi-drug therapies (Hicks *et al.*, 2005) and should be borne in mind in the selection of future multi-drug therapies.

## ACKNOWLEDGEMENTS

We thank V. Fulop and M.S. Turner for instructive discussions and H.G. Burgert for a critical reading of the manuscript.

*Funding*: EPSRC Life Sciences Interface programme (MOAC DTC EP/F500378/1 to J.W.H.); Leverhulme trust (Early career fellowship to S.A.W.).

*Conflict of Interest*: none declared.

## REFERENCES

Berman,H. *et al.* (2002) The protein data bank. *Nucleic. Acids Res.*, **28**, 235–242.

Chang,C.-E. *et al.* (2006) Gated binding of ligands to HIV-1 protease: Brownian dynamics simulations in a coarse- grained model. *Biophys. J.*, **90**, 3880–3385.

Collins,J.R. *et al.* (1995) Flap opening in HIV-1 protease simulated by 'activated' molecular dynamics. *Nat. Struct. Biol.*, **2**, 334–338.

Dahiyat,B.I. *et al.* (1997) Automated design of the surface positions of protein helices. *Protein Sci.*, **6**, 1333–1337.

Erickson,J.W. (1995) The not-so-great escape. *Nat. Struct. Biol.*, **2**, 523–529.

Flexner,C. (1998) HIV protease inhibitors. *N. Engl. J. Med.*, **338**, 1281–1293.

Freedberg,D.I. *et al.* (2001) Rapid structural fluctuations of the free HIV protease flaps in solution: relationship to crystal structures and comparison with predictions of dynamics calculations. *Protein Sci.*, **11**, 221–232.

Ghosh,A.K. *et al.* (2007) Darunavir, a conceptually new HIV-1 protease inhibitor for the treatment of drug-resistant HIV. *Bioorg. Med. Chem.*, **15**, 7576–7580.

Gohlke,H. *et al.* (2004) Change in protein flexibility upon complex formation: analysis of ras-raf using molecular dynamics and a molecular framework approach. *Proteins*, **56**, 322–337.

Harte,W.E. and Beveridge,D.L. (1993) Mechanism for the destabilization of the dimer interface in a mutant HIV-1 protease: a molecular dynamics study. *J. Am. Chem. Soc.*, **115**, 1231–1234.

Harte,W.E. *et al.* (1992) Molecular dynamics of HIV-1 protease. *Proteins*, **13**, 175–194.

Hespenheide,B.M. *et al.* (2002) Identifying protein folding cores: observing the evolution of rigid and flexible regions during unfolding. *J. Mol. Graph. Model.*, **21**, 195–207.

Hicks,C. B. *et al.* (2006) Durable efficacy of tipranavir-ritonavir in combination with an optimised background regimen of antiretroviral drugs for treatment-experienced HIV-1-infected patients at 48 weeks in the randomized evaluation of strategic intervention in multi-drug resistant patients with tipranavir (RESIST) studies: an analysis of combined data from two randomised open-label trials. *Lancet*, **368**, 466–475.

Hong,L. *et al.* (1997) Structure of a g48h mutant of HIV-1 protease explains how glycine-48 replacements produce mutants resistant to inhibitor drugs. *FEBS J.*, **420**, 11–16.

Hornak,V. and Simmerling,C. (2007) Targeting structural flexibility in HIV-1 protease inhibitor binding. *Drug Discov. Today*, **12**, 132–138.

Hornak,V. *et al.* (2006) HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proc. Natl Acad. Sci. USA*, **103**, 915–920.

Ishima,R. *et al.* (1999) Flap opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease, and their implications for function. *Structure*, **7**, 1047–1055, S1–S12.

Jacobs,D. and Thorpe,M. (1995) Generic rigidity percolation: the pebble game. *Phys. Rev. Lett.*, 1996, **75**, 4051–4054.

Jacobs,D. *et al.* (2001) Protein flexibility predictions using graph theory. *Proteins*, **44**, 150–165.

Katoh,E. *et al.* (2003) A solution NMR study of the binding kinetics and the internal dynamics of an HIV-1 protease-substrate complex. *Protein Sci.*, **12**, 1376–1385.

Mamonova,T. *et al.* (2005) Protein flexibility using constraints from molecular dynamics simulations. *Phys. Biol.*, **2**, S137-S147.

Matsumoto,M. *et al.* (2002) Molecular dynamics simulation of the ice nucleation and growth process leading to water freezing. *Nature*, **416**, 409-413.

Nicholson,L.K. *et al.* Flexibility and function in HIV-1 protease. *Nat. Struct. Biol.*, **2**, 274–280.

Piana,S. *et al.* (2002) Drug resistance in HIV-1 protease: flexibility-assisted mechanism of compensatory mutations. *Protein Sci.*, **11**, 2393–2402.

Pokorná,J., *et al.* (2009) Current and novel inhibitors of HIV protease. *Viruses*, **1**, 1209–1239.

Rader,A. J. (2010) Thermostability in rubredoxin and its relationship to mechanical rigidity. *Phys. Biol.*, **7**, 016002.

Radestocke,S. and Gohlke,H. (2008) Exploiting the link between protein rigidity and thermostability for data-driven protein engineering. *Eng. Life Sci.*, **8**, 507-522.

Radestocke,S. and Gohlke,H. (2011) Protein rigidity and thermophilic adaptation. *Proteins*, **79**, 1089-1108.

Rodríguez,E.J. *et al.* (1993) Inhibitor binding to the Phe53Trp mutant of HIV-1 protease promotes conformational changes detectable by spectrofluorometry. *Biochemistry*, **32**, 3557–3563.

Swaminathan,S. *et al.* (1991) Investigation of domain structure in proteins via molecular dynamics simulation: application to HIV-1 protease dimer. *J. Am. Chem. Soc.*, **113**, 2717–2721.

Tan,H. and Rader,A. J. (2009) Identification of putative, stable binding regions through flexibility analysis of HIV-1 gp120. *Proteins*, **74**, 881–894.

Tastan,O. *et al.* (2007) Comparison of stability predictions and simulated unfolding of rhodopsin structures. *Photchem. Photobiol.*, **83**, 351–362.

Thorpe,M.F. *et al.* (2001) Protein flexibility and dynamics using constraint theory. *J. Mol. Graph. Model*, **19**, 60–69.

Toh,H. *et al.* (1985) Synthetic non-peptide inhibitors of HIV protease. *Nature*, **315**, 691–692.

Tozzini,V. and McCammon,A. (2005) A coarse grained model for the dynamics of flap opening in HIV-1 protease. *Chem. Phys. Lett.*, **413**, 123–128.

Velazquez-Campoy,A. *et al.* (2001) The binding energetics of first- and second-generation HIV-1 protease inhibitors: implications for drug design. *Arch. Biochem. Biophys.*, 2001, **390**, 169–175.

Wlodawer,A. and Vondrasek,J. (1998) Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. *Science*, **27**, 249–284.

Wlodawer,A. *et al.* (1989) Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science*, **245**, 616–621.

Wells,S. *et al.* (2009) Comparative analysis of rigidity analysis across protein families. *Phys. Biol.*, **6**, 046005–11.

Word,J.M. *et al.* (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.*, **285**, 1735–1747.

York,D.M. *et al.* (1993) Molecular dynamics simulation of HIV-1 protease in a crystalline environment and in solution. *Biochemistry*, **32**, 1443–1453.

Zhu,Z. *et al.* (2003) Molecular dynamics study of the connection between flap closing and binding of fullerene-based inhibitors of the HIV-1 protease *Biochemistry*, **42**, 1326–1333.