Published in final edited form as:

Proteins. 2015 November ; 83(11): 1929–1939. doi:10.1002/prot.24737.

Modulation of HIV Protease Flexibility by the T80N Mutation

Hao Zhou¹, Shangyang Li¹, John Badger², Ellen Nalivaika³, Yufeng Cai³, Jennifer Foulkes-Murzycki³, Celia Schiffer³, and Lee Makowski^{4,5}

¹Department of Electrical and Computer Engineering, Northeastern University, Boston, MA

²DeltaG Technologies, San Diego, CA

³Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA

⁴Department of Bioengineering, Northeastern University, Boston, MA

⁵Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA

Abstract

The flexibility of HIV protease plays a critical role in enabling enzymatic activity and is required for substrate access to the active site. While the importance of flexibility in the flaps that cover the active site is well known, flexibility in other parts of the enzyme is also critical for function. One key region is a loop containing Thr 80 which forms the walls of the active site. Although not situated within the active site, amino acid Thr80 is absolutely conserved. The mutation T80N preserves the structure of the enzyme but catalytic activity is completely lost. To investigate the potential influence of the T80N mutation on HIVp flexibility, wide-angle scattering (WAXS) data was measured for a series of HIV protease variants. Starting with a calculated WAXS pattern from a rigid atomic model, the modulations in the intensity distribution caused by structural fluctuations in the protein were predicted by simple analytic methods and compared to the experimental data. An analysis of T80N WAXS data shows that this variant is significantly more rigid than the WT across all length scales. The effects of this single point mutation extend throughout the protein, so as to alter the mobility of amino acids in the enzymatic core. These results support the contentions that significant protein flexibility extends throughout HIV protease and is critical to catalytic function.

Keywords

x-ray solution scattering; molecular dynamics simulation

Introduction

HIV protease (HIVp) is a homodimer of 99 amino acid polypeptides that carries out sequence-specific cleavages of the viral polyprotein needed to create the protein components

Corresponding Author: John Badger, DeltaG Technologies, 4360 Benhurst Ave, San Diego, CA 92122. info1.dgtech@gmail.com.

Work site: Department of Electrical and Computer Engineering, Northeastern University, Boston, MA

that are essential for viral replication. HIVp is a principal drug target for AIDS therapies and has become one of the most thoroughly studied of all proteins. The emergence of multi-drug resistant mutations in HIVp has motivated studies that characterize the functional roles of the most highly conserved amino acids in the HIVp sequence. The identification of conserved amino acids that are remote from the active site has suggested the potential importance of specific amino acids in allowing the protein flexibility necessary for enzymatic activity^{1,2}. The mutation site that is the subject of this work, threonine 80, is highly conserved in all patient populations, both treated and untreated, yet makes no direct contact with substrate or inhibitors³. Replacement of threonine 80 with any other amino acid greatly diminishes or abolishes enzymatic activity. CD and tryptophan fluorescence spectroscopy data³ indicate that there is very little difference between the apo structures of the WT and T80N mutant forms.

Crystal structures of HIVp both with and without bound ligands have been solved and compared⁴. Excluding the flap regions (amino acids 35–57) the conformations of apo and ligand-bound structures (PDB IDs 3HVP, 4HVP) are rather similar, with an rms deviation between aligned pairs of C α atoms of 0.7 Å. Large differences between the apo and ligand-bound structures are limited to the conformations of a small number of residues at the tips of the flaps (Fig 1). Furthermore, it seems likely that in solution the flaps explore a variety of conformations that might include a sub-population of the more closed conformation characteristic of the ligand-bound structures⁵. The ligand-bound crystal structure of HIVp incorporating the T80N mutation is also essentially unchanged from the WT structure with an rms deviation between all pairs of C α atoms of 0.6 Å but HIVp containing the T80N mutation is devoid of enzymatic activity and cannot support viral replication³. In the earliest MD simulations of the T80N mutant it was noted that there was decreased conformational mobility in the flap regions that open to allow substrate access to the enzymatic site and close over bound compounds³. Later analysis suggested that the greater global flexibility of WT, extending through the protein core, might be necessary for function⁶. Replacement of amino acid pairs in the hydrophobic core of HIVp with cysteines positioned so that they may be connected by disulphide bonds has demonstrated the essential role that the mobility of core residues must play in protease activity. The formation of disulfide bonds between cysteines has the effect of greatly muting the enzymatic activity and their reduction restores a higher level of enzymatic function². Experimental measurements using NMR relaxation and pulse-EPR techniques may be interpreted as indicating that the rate limiting step in catalysis is *not* the opening and closing of flaps but, rather, the degree of pre-organization of the flaps correlated with a structural reorganization of active site residues⁷.

Here, we use a combination of wide-angle x-ray solution scattering (WAXS) and molecular dynamics simulations (MD) to further characterize the relationship between flexibility and function in HIVp. WAXS is a natural extension of small-angle x-ray solution scattering (SAXS) involving the collection and analysis of x-ray scattering data to wider angles. Development of WAXS has been supported by the availability of high-brilliance synchrotron x-ray sources that allow precise measurement of scattered intensities to much wider scattering angles than were possible with laboratory-based sources⁸. The sensitivity

of WAXS data to protein flexibility was not anticipated, but became apparent in studies of proteins over a wide range of protein concentrations⁹. Changes in flexibility due to changes in protein concentration (macromolecular crowding effects), unfolding, and ligand binding¹⁰ have now been characterized using WAXS. Increased flexibility leads to a broader structural ensemble that expresses itself in solution scattering patterns by filling in troughs in the scattered intensity and muting the intensity of peaks. The range of motion of interatomic vectors can be estimated by comparison of the scattering pattern expected for a rigid protein with the observed scattering pattern¹⁰. Here we use WAXS to estimate the range of motion that WT HIVp undergoes in solution, and the degree to which that range of motion changes in response to amino acid replacements and the binding of the inhibitor, pepstatin. These results are compared with MD simulations of the WT protein and the T80N variant.

Materials and Methods

Protein production

Synthesis of protein genes, expression and purification of HIVp was as previously described³. WT included a substitution of Q7K to prevent autoproteolysis¹¹. Mutagenesis was performed³ using the Stratagene QuikChange site-directed mutagenesis kit and confirmed by sequencing. Expression and purification of protease were carried out as previously described^{12,13}. Briefly, the HIV protease gene was cloned into plasmid pXC-35 (American Type Culture Collection, Manassas, Va.) which was transfected into *E. coli* TAP106. Transfected cells were grown in a 12-liter fermenter and, after protein expression, lysed to release inclusion bodies containing the protease¹⁴. Inclusion bodies were isolated by centrifugation, after which the pellet was dissolved in 50% acetic acid to extract protease. High-molecular-weight proteins were separated from the desired protease by size exclusion chromatography on a 2.1-liter Sephadex G-75 superfine (Sigma Chemical) column equilibrated with 50% acetic acid. Refolding was accomplished by rapidly diluting the protease solution into a 100-fold excess of refolding buffer. Excess acetic acid was removed through dialysis. Protein used for crystallization was further purified with a Pharmacia Superdex 75 fast-performance liquid chromatography column equilibrated with refolding buffer. Protein was concentrated to 2–4 mg/ml for WAXS experiments.

WAXS

All data were collected at the BioCAT undulator beamline (18ID) at the Advanced Photon Source¹⁵ using methods previously described in detail⁹. Data were collected using a sample cell consisting of a thin-walled quartz capillary held at an ambient temperature of 4°C. To minimize radiation damage, protein samples were made to flow through the X-ray beam at a rate that limited X-ray exposure of any one protein to no more than 100 ms. At these exposure levels, the effect of radiation damage on radiosensitive test proteins is undetectable. Typically, a data-set consisted of a series of 2 s exposures with five from buffer, 15 to 20 from protein solution, and five from the empty capillary. The two-dimensional (2D) scattering patterns were circularly averaged and the resulting one-dimensional intensity distribution was plotted as a function of spacing, 1/d. Standard

deviations of the observed data were calculated, with error propagation formulae used to calculate their effect on the final estimate of scattering from protein.

The program XS was used to predict scattering patterns from sets of atomic coordinates¹⁶. This program uses an explicit atom representation of water to overcome limitations in the use of continuum models of the hydration layer and the modeling of excluded volume¹⁷. NAMD¹⁸ was used to generate a 20 ps equilibration followed by a 100 ps MD simulation during which 100 snapshots of the water of hydration were captured. During the MD simulation the protein was held rigid. The scattering due to the solvated protein was then calculated and the corresponding scatter from a comparable “droplet” of water was subtracted.

Calculation of three-dimensional molecular envelopes

Three-dimensional molecular envelopes for HIVp were determined from observed data from HIVp (WT) and HIVp (T80N) and from the WAXS profile calculated on the basis of atomic coordinates (1F7A) using programs from the ATSAS suite¹⁹ for the analysis of solution x-ray scattering data. The GNOM program²⁰ was used to calculate the particle distance distribution function, $p(r)$, and the GASBOR program²¹ was used to generate three-dimensional models to fit this data. The x-ray solution scattering data used in reconstruction calculations spanned the resolution range $0.0086 < 1/d < 0.1252 \text{ \AA}^{-1}$ and, therefore, included experimental information to a length scale that exceeds twice the maximum dimension of the HIVp dimer. Reconstruction averages were obtained by taking ten independently calculated reconstructions and aligning them with the DAMSEL and SUPCOMB programs²², using the most representative model within each subset as reference. These averaged reconstructions were transformed into grid objects with volumes equal to the partial specific volume calculated from the HIVp amino acid sequence. Finally, the molecular envelopes represented by these grid objects were aligned and averaged to create an improved low-resolution image of the protein.

Vector length Convolution

The intensity calculated using XS corresponds to that expected for a population of completely rigid proteins. We used the vector-length convolution to predict the effect of structural fluctuations on the computed intensities. Starting with a reference scattering pattern an indirect Fourier transform²⁰ was used to compute the pair correlation function - a histogram of the interatomic vector lengths within a protein.

Predicted patterns for model ensembles were computed by replacing each interatomic vector in the pair correlation function of the reference structure by a Gaussian distribution of vector lengths. The calculations were carried out as described previously¹⁰. Model ensembles with distinctly different properties can be generated by varying the way in which the Gaussian distribution varies with interatomic vector length. The pair correlation function corresponding to the model structural ensemble, $p_m(r)$, is computed from the convolution of the pair correlation function of the reference structure, $p_r(r)$, and a Gaussian of half width $\sigma(r)$ which may be a function of the interatomic vector length, r , according to

$$p_m(r) = p_r(r) * \exp(-\sigma(r)^2/2r^2)$$

The '*' in the equation denotes convolution. Using $\sigma(r) = cr^e$ where c and e are free parameters, has, empirically, provided models adequate to fit the scattering from a broad range of proteins in solution.

Molecular Dynamics

MD simulations extending to 100 ns were run on both WT and T80N HIVp structures. The initial coordinates were taken from the 9-ns time point of a previous simulation where the protease flap tips were already folded in toward the P1 loop²³. Asparagine was modeled by MIDAS²⁴ into residue 80 in both monomers, and MD simulations were started with the AMBER8 software package²⁵. Two hundred cycles of restrained steepest-descent energy minimization were performed on each structure. The protein was restrained with a harmonic force constant of 9.55 kcal mol⁻¹ Å⁻². Each structure was solvated with a TIP3P water box to allow for at least 8 Å of water on each face of protease. The initial periodic box dimensions were 82.2 by 82.6 by 80.4 Å³. Approximately 14,000 water molecules were added to each system, and 1,000 cycles of steepest-descent energy minimization were performed, keeping the protein fixed.

Equilibration of the system continued with 5,000 steps of restrained MD with a 9.55 kcal mol⁻¹ Å⁻² force constant and constant temperature (300 K) and pressure (1 atm). This equilibration was followed by 25,000 steps of unrestrained MD. The data-collecting portion of the simulation was performed at constant temperature and pressure for 10 or 100 ns with 1 fs time steps.

Results

Solution Scattering Data

Solution scattering data were collected from a series of HIVp variants which included: WT, T80N, and the flap variant G48V that included the secondary mutation L63P. For each protein form, WAXS data was collected both in the presence and absence of the inhibitor, pepstatin. All data sets were collected at beam line 18ID at the Advanced Photon Source using methods essentially identical to those described previously⁹ from protein samples at 2–4 mg/ml concentration.

The main focus of this work is elucidating the mechanism by which the T80N variant is inactive compared to WT. A comparison WAXS data collected from T80N and WT variants of HIVp reveals some significant differences in the distribution of intensities between the two samples (Fig 2). The scattering pattern from the T80N variant shows sharper features than the WT, indicating a more rigid protein structure. For comparison with these data, an accurate calculation of the scattering pattern from a rigid protein model was performed using the program XS¹⁶ and atomic coordinate set PDB ID:1F7A²⁶. The pattern predicted from the rigid HIVp model is much more sharply modulated than either set of experimental

measurements and, in particular, shows a much more prominent intensity trough at $1/d \sim 0.05 \text{ \AA}^{-1}$.

Calculations of the radii of gyration (R_g) from the x-ray scattering data give 18.47 $\text{\AA}$ for WT and 18.10 $\text{\AA}$ for the T80N mutant. The estimated standard deviation for the difference between these values for R_g was 0.36 $\text{\AA}$, which is almost equal to the observed difference. These values compare well with the exact value obtained from an atomic model (PDB ID: 1F7A) of 18.40 $\text{\AA}$. As characterized by R_g and within the error estimates for the solution scattering data, the shapes of the HIVp dimer obtained from the experimental data for WT, T80N and the protein crystal structure are indistinguishable. In order to identify any large conformational differences between HIVp structures in solution, low-resolution reconstructions of the molecular envelopes were calculated from the solution scattering data from both the WT and T80N samples using *ab initio* (model independent) methods²¹. As a control, the molecular envelope for the rigid protein in the conformation of a crystal structure (PDB ID: 1F7A) was calculated by the same reconstruction methodology using simulated WAXS data (Fig 3). Despite the observed differences in intensities (Fig 2), the three-dimensional reconstructions resulting from these three intensity distributions are all very similar. Although the resolution of molecular envelopes reconstructed from solution scattering data is limited these reconstructions are sufficient to eliminate the possibility that the T80N mutation alters HIVp dimerization or causes a conformational change in the protein outside the range expected from known crystal structures. These structural results are consistent with other experimental data, including CD measurements and tryptophan fluorescence spectra, which also indicate a very high degree of similarity between the T80N and WT structures either in the presence or in the absence of ligand³.

When the calculated solution scattering pattern from an apo crystal structure (PDB ID: 3HVP) is compared with the calculated pattern from ligand-bound structures there is only a small difference in the relative emphasis of the intensity minima at $1/d \sim 0.05 \text{ \AA}^{-1}$ and the secondary maxima at $1/d \sim 0.075 \text{ \AA}^{-1}$. This calculation shows that structural differences in the flap conformation that characterize the apo and ligand-bound conformations are not large enough to alter the form of the x-ray scattering curve. In contrast, in the experimentally observed scattering differences between WT and the T80N mutant the secondary maxima is almost completely eliminated from the WT data and the entire scattering pattern appears smoothed out (Fig 2). Our determinations of R_g and the close similarity of the shapes and dimensions of the three-dimensional reconstructions with an atomic model demonstrate that the intensity curves that we have analyzed correspond to known structures of HIVp and are not corrupted by protein aggregation. Furthermore, aggregation would typically lead to a sharp intensity peak at extremely low resolution but would not significantly alter the relative sizes of the modulations in the intensity curve that we observe at higher resolution.

These results suggest that the cause of the observed differences in WAXS patterns is largely the difference in the relative flexibilities of these structures rather than a difference in conformation between structures. Furthermore, the types of differences observed in the WAXS data and the intensities predicted from the crystal structure are suggestive of changes in flexibility – the characteristic pattern of peaks and troughs in the calculated intensities is

present but muted in the observed intensities. This flattening out of WAXS features is a hallmark of intensity changes caused by increased protein flexibility¹⁰.

Vector-length convolution analysis of WAXS data

Vector-length convolution¹⁰ was used to quantify the magnitudes and ranges of motions in the HIVp samples. Using this approach the average range of motion of atoms in the protein is determined as a function of inter-atomic vector length. Protein mobility is characterized by two free parameters and their values are chosen to be those providing the best prediction of the observed intensity. Beginning with a reference intensity distribution calculated for a rigid protein (using an atomic coordinate set), the method predicts the scattering that would be observed if every inter-atomic vector of a particular length were replaced by a distribution of vector lengths. Here, the distribution of lengths is assumed to be Gaussian with standard deviation, $\sigma(r)$, that varies with vector length, r , such that $\sigma(r) = cr^e$. The parameters c and e are chosen to minimize the difference between observed and calculated intensities. A positive exponent, e , implies that atom pairs that have the greatest separations also have correspondingly greater fluctuations in separation. Using the intensity distribution calculated from PDB ID: 1F7A (as plotted in Fig 2), the values for c and e that result in the best fit to the observed data were determined using a least-squares criterion. In detailed examinations of the discrepancy between observed data and calculated scattering plotted over a wide range of different pairs of values, $\{c, e\}$, the discrepancy function was found to be characterized by a single minimum, indicating that this parameterization leads to a unique solution. Single-parameter fits, equivalent to setting $e = 0$, misfit the data, showing that models that posit only local, uncorrelated atomic fluctuations are not sufficient to represent the type of conformational variation observed in HIVp.

Vector-length convolution calculations were carried out to fit WAXS patterns from WT, T80N, and G48V/L63P HIVp variants using data from protein both in the presence and absence of inhibitor. The scattering patterns calculated using this two parameter model (Fig 4) agree reasonably well with the observed intensities out to a spacing of $1/d = 0.1 \text{ \AA}^{-1}$ (10 $\text{\AA}$ spacing) but diverge somewhat at higher resolution. Prediction of the impact of flexibility on higher resolution scattering would require a more elaborate model for protein flexibility. The current formulation of vector-length convolution is in terms of global, average descriptors; some structural regions (for example, the flaps) may be more flexible than average but their contribution is balanced by less flexible regions. Nevertheless, the vector-length convolution approach provides a simple and direct way of characterizing and quantifying the average mobility of atoms in these samples using WAXS data.

The flexibility parameters, c and e , show significant variation in values for the six HIVp samples (Table 1). In addition to these two parameters, Table 1 also provides the atomic fluctuation between atom pairs at the benchmark separation of 10 $\text{\AA}$. Alpha helices in proteins tend to pack with separations of $\sim 10 \text{ \AA}$, as do adjacent β -sheets. The range of motion at $r = 10 \text{ \AA}$ provides a measure of how much movement might be occurring between adjacent secondary structures in the protein. A value of $\sigma(r = 10 \text{ \AA}) = 2.01 \text{ \AA}$ means that, on average, interatomic vectors of 10 $\text{\AA}$ length will be varying by $\pm 2.01 \text{ \AA}$ in the protein.

As forecast by the initial visual inspection of WAXS data, the WT protein appears to be much more flexible than the T80N mutant over length scales involving atomic separations up to 30 Å (Fig 6). The quantification of protein flexibility provided in Table 1 shows that, on average, the distances between atom pairs separated by 10 Å fluctuate by ~2 Å in the T80N mutant and the magnitude of this fluctuation is ~0.8 Å greater in WT. The addition of pepstatin inhibitor to either WT or the G48V/L63P mutant results in a large reduction in the flexibility of these proteins such that they approach the reduced levels of motion observed for the T80N mutant. The addition of inhibitor to T80N has very little effect on the parameters representing protein flexibility. For HIVp, the loss of function in the inhibitor-bound sample is due to the competitive binding of the inhibitor and these data sets provide controls for the accuracy of the derived parameters, c and e . For nearly equivalent protein forms the differences in these parameters is small, indicating that the values obtained by the vector-length convolution method are well-defined by the WAXS data. Classifying the HIVp samples by the parameters c and e , the T80N mutant clusters with the inhibitor-bound forms rather than the functional apo forms of WT and G48V L63P (Fig 5). The impact of the T80N point mutation is to reduce the protease flexibility so that it resembles the inhibitor-bound states. Another view of the variation in flexibility for the different samples is the plot of $\sigma(r)$ versus r (Fig 6A) or, more strikingly, $\sigma(r)/r$ versus r (Fig 6B) for the six data sets. The latter plot measures the variation in interatomic vector length as a percentage of interatomic vector length. Here the apo WT and apo G48 L63P samples also appear distinct from the other samples in that they exhibit relatively large-scale fluctuations.

Molecular Dynamics Simulations

MD simulations were carried out for both WT and T80N HIVp as previously described^{1,3}. To facilitate comparisons of motions observed in these simulations with the vector-length convolution analysis from the WAXS data, we calculated the function $\sigma(r)$ for the C α atoms for these two proteins using coordinates from the MD trajectories. Atomic coordinate snapshots taken every 10 ps over the 100 ns simulations were used to construct $\sigma(r)$. According to NMR relaxation data^{5,27}, structural dynamics involving flap movements in the apo samples of HIVp may be accommodated within nanosecond simulations but other dynamical transitions may occur on much slower time scales. Since both the WT and T80N simulations were started from the same well-equilibrated structure, this pair of 100 ns simulations compares the ambient dynamics of these proteins in the conformational region of this starting point. A comparison of $\sigma(r)$ for WT and T80N for the 100 ns simulations shows that at all inter-atomic vector lengths, $\sigma(r)$ is smaller for T80N than WT (Fig 7A). Since the flaps are not involved in interatomic vectors greater than about 40 Å, it is clear that this amino acid replacement has a broader impact than simply locking the flaps in place. Analysis of the $s(r)$ from MD simulations excluding atoms in the flap regions gives a very similar $\sigma(r)$ to that calculated using all the C α atoms. Rather than simply providing a local effect on the mobility of the flaps, the effect of the T80N mutation is to cause a general increase in the rigidity of the protein through the entire structure. Additional simulations are required to know if the types of motion observed in these MD runs are sufficiently decoupled from the effects of much slower transitions to other structural states for the MD to reliably predict that the WT HIVp is more flexible than the T80N mutant. A preliminary analysis of replicated 100 ns MD runs for these two structural forms does show a greater

flexibility in the simulated dynamics of WT than T80N in the majority of runs (work in progress).

A comparison of $\sigma(r)$ for WT HIVp calculated from MD trajectories of varying length exposes the effect of time scale on the magnitude of observed structural fluctuations (Fig 7B). The standard deviations of interatomic vector lengths increase as a function of MD simulation time but, even when MD simulations are extended to 100 ns, do not approach the magnitudes of the fluctuations estimated from the analysis of the WAXS data. Convergence of $\sigma(r)$ over 100 ns simulations has been investigated by dividing the simulation into consecutive 10 ns segments and comparing $\sigma(r)$ for each segment. Some of these consecutive simulation segments give similar curves for $\sigma(r)$, which would appear to indicate convergence of conformational sampling around particular a substate. However, occasional abrupt conformational transitions also sometimes result in much larger values of $\sigma(r)$ and the overall value of $\sigma(r)$ over 100 ns is greater than over 10 ns (Fig 7B). These observations suggests that a 100 ns simulation is still insufficiently long to sample across all of the relatively rare conformational transitions into different conformational substates for this protein. The inter-atomic standard deviations estimated from the WAXS data are approximately 3.6x greater than those estimated from the 100 ns MD trajectory. Although the magnitude of the fluctuations are substantially different, their variation as a function of interatomic vector length is similar up to an interatomic separation of $r \sim 35$ Å. The difference may be rationalized in terms of the extent to which each data set reflects the structural ensemble of the protein. The $\sigma(r)$ curve derived from the WAXS data is a reflection of the entire structural ensemble of the protein as it is calculated from the scattering of $\sim 10^{17}$ molecules in the scattering volume over a period of ~ 20 s (ten 2 s exposures in a flow cell where no individual molecule is exposed to more than 100 ms of radiation). Conversely, the 100 ns MD simulation may not have explored the full range of protein motion. Exploration of the full structural ensemble may require micro- to milliseconds or more in order to allow the protein to sample a range of distinct conformations²⁸.

Discussion

The leading contention of this work is that the scattering differences between WT and T80N forms of HIVp are accounted for by differences in protein flexibility rather than by differences in conformation. CD and fluorescence spectroscopy measurements³ indicate that the secondary structure composition and tryptophan environments are very similar in both forms and our reconstruction calculations (Fig 3) show that the overall shapes of the two proteins are not distinguishable. Furthermore, it is the experimentally observed scattering pattern from WT HIVp rather than the T80N mutant that is most dissimilar to predicted scattering patterns from (rigid) crystal structure models. Although experimental^{5, 27} and simulation studies^{29–31} of WT HIVp typically show variability in the flap regions these studies have not suggested that there is any large global structural difference between the conformation of HIVp in solution and the structures obtained in crystal environments. Our example calculations of scattering curves from apo and ligand-bound crystal structures show that the intensity changes that result from this scale of conformational variation (~ 1 Å overall but larger in limited regions of the flaps) does not

match the type of intensity difference that we observe between WT and T80N. Specifically, the significant loss of intensity at $1/d \sim 0.075 \text{ \AA}^{-1}$ in the WT data relative to expected scattering from relatively rigid models is not reproduced by these conformational variations. For these reasons we do not consider that it is very likely that there are relatively large structural differences between WT and T80N forms of HIVp in solution or that these structures deviate from the crystal structures to a large enough extent to undermine WAXS analysis. In contrast, in our analysis and modeling protein flexibility using the vector length convolution formalism we are able to explain the scattering differences between WT and T80N HIVp based on standard crystallographic models and differences in the overall flexibility of the two molecules.

This analysis of WAXS data indicates that the T80N variant exhibits less flexibility than WT across all length scales within the protein. This result is consistent with studies suggesting that a ‘hydrophobic sliding’ of core residues provides the necessary flexibility for HIVp function⁶ and that when the protein is engineered to restrict mobility, function is lost². WAXS patterns from WT and the functional, flap variant G48V/L63P indicate that these two forms have more global flexibility, with WT only somewhat more flexible than the double mutant. Conversely, inhibitor-binding rigidifies the protein and the similarity of flexibility parameters obtained for all three inhibitor-containing examples and the apo T80N mutant form support the consistency and reliability of the WAXS analysis of protein flexibility.

HIVp, in WT and various mutated forms, has been subject to numerous MD simulation studies and one focus of the analysis has often been the mobility and conformations of the flaps as determinants of substrate binding and hence, activity^{29–31}. Nevertheless, some of these studies also report more global differences in the dynamics of HIVp mutants from WT³². The WAXS analysis supports the view that global flexibility is required for catalytic activity and that protein motions throughout the entire structure can be impacted by a point mutation at a critical site. More generally, this demonstration of the non-local impact of a point mutation on protein flexibility in conjunction with a functional requirement for atomic mobility provides an example of a critical amino acid site with a role that is not readily apparent.

The method described in this paper for deriving $\sigma(r)$ from WAXS patterns provides a novel experimental approach for characterizing the global dynamics of a protein in solution. The simple functional form where $\sigma(r) = cr^e$ that was used to fit the scattering data enabled the quantitative exploration of models of motion in which the fluctuations of interatomic distances between pairs of atoms can vary depending on the atomic separation. Using this functional form, the best fits to the scattering data were obtained when the distances between more widely separated atoms, on average, fluctuated more than closely-spaced atoms. However, in this example the largest population of interatomic vectors is around the maxima in $p(r)$ at $\sim 20 \text{ \AA}$ (*i.e.* on a length scale in the protein comparable to R_g). The fit of this simple two-parameter function is expected to breakdown at the extreme ends of the range of interatomic vectors within the protein. Specifically, protein flexibility is over-estimated by this simple function for extremely short vectors ($r < 5 \text{ \AA}$) where many atomic pairs are related through covalent bonds that allow little variation in interatomic separations

(including, for example, atoms within peptide planes and aromatic ring systems). For very long interatomic vectors ($r > 30 \text{ \AA}$) it is also unclear whether the monotonic increase in flexibility with separation that is necessitated by the two-parameter function is justified since there are also relatively few vectors in this range and this distance exceeds two of the three dimensions of the HIVp dimer. Furthermore, MD simulation studies suggest a flattening out of $\sigma(r)$ for $r > 30 \text{ \AA}$.

Since the same function can also be calculated from analysis of MD simulations the reliability of the atomic-level detail that is provided by these simulations is supported if the predicted $\sigma(r)$ resembles the function that fits experiment. While the 3.6-fold discrepancy in the magnitude of the flexibility between MD and WAXS in the longest simulations presented here (100 ns) may be plausibly rationalized as being primarily due to incomplete conformational sampling in the MD simulation, other causes for this discrepancy might also be considered. While not as overt as the constraints of a crystal lattice, protein flexibility might still be influenced by the protein environment in solution. For example, the exact buffer constituents, protein concentration and temperature are experimental factors that are not typically replicated in detail in MD simulations. Nevertheless, the similarity in functional form of $\sigma(r)$ as derived from both WAXS data and MD simulation, as well as the finding that the T80N mutant is *globally less flexible* than WT in both cases is reassuring and supports extending the global characterization of protein mobility obtained from WAXS data with the detailed atomic-level observations of protein behavior obtained from MD simulation.

Acknowledgments

This work was supported by grants RO1 GM65347, RO1 GM85648, RO1 GM085648 and P01 GM66524 from the NIH. We thank Dr. Nese Kurt Yilmaz for insightful comments. BioCAT is a National Institutes of Health-supported Research Center RR-08630. Use of the Advanced Photon Source, an Office of Science User Facility operated for the U.S. Department of Energy (DOE) Office of Science by Argonne National Laboratory, was supported by the U.S. DOE under Contract No. DE-AC02-06CH11357.

References

1. Cai Y, Schiffer C. (2012) Decomposing the energetic impact of drug-resistant mutations: the example of HIV-1 protease-DRV binding. *Methods Mol Biol.* 2012; 819:551–560. [PubMed: 22183557]
2. Mittal S, Cai Y, Nalam MNL, Bolon DNA, Schiffer CA. Hydrophobic core flexibility modulates enzyme activity in HIV-1 protease. *J Am Chem Soc.* 2012; 134:4163–4168. [PubMed: 22295904]
3. Foulkes JE, Prabu-Jeyabalan M, Cooper D, Henderson GJ, Harris J, Swanstrom R, Schiffer CA. Role of Invariant Thr80 in Human Immunodeficiency Virus Type 1 Protease Structure, Function, and Viral Infectivity. *J Virology.* 2006; 80:6906–6916. [PubMed: 16809296]
4. Wlodawer A, Miller M, Jaskolski M, Sathyanarayana BK, Baldwin E, Weber IT, Selk LM, Clawson L, Schneider J, Kent SB. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science.* 1989; 245:616–621. [PubMed: 2548279]
5. Freedberg DI, Ishma R, Jacob J, Wang X-Y, Kustanovich I, Louis JM, Torchia DA. Rapid structural fluctuations of the free HIV protease flaps in solution: Relationship to crystal structures and comparison with predictions of dynamics calculations. *Protein Science.* 2002; 11:221–232. [PubMed: 11790832]
6. Foulkes-Murzycki JE, Scott WR, Schiffer CA. Hydrophobic sliding: a possible mechanism for drug resistance in human immunodeficiency virus type 1 protease. *Structure.* 2007; 15:225–233. [PubMed: 17292840]

7. Torbeev VY, Raghuraman H, Hamelburg D, Tonelli M, Westler WM, Perozo E, Kent SBH. Protein conformational dynamics in the mechanism of HIV-1 protease catalysis. *PNAS*. 2011; 108:20982–20987. [PubMed: 22158985]
8. Makowski L.; Makarov, OA.; Xu, S.; Fischetti, RF. Wide-angle X-ray Solution Scattering. In: Tschesche, H., editor. *Methods in Protein Biochemistry*. Walter De Gruyter, GMBH; Berlin: 2011. p. 165-186.
9. Makowski L, Rodi DJ, Mandava S, Minh D, Gore D, Fischetti RF. Molecular crowding inhibits intramolecular breathing motions in proteins. *J Mol Biol*. 2008; 375:529–546. [PubMed: 18031757]
10. Makowski L, Gore D, Mandava S, Minh D, Park S, Rodi DJ, Fischetti RF. (2011) X-ray solution scattering studies of the structural diversity intrinsic to protein ensembles. *Biopolymers*. 2011; 95:531–542. [PubMed: 21462170]
11. Rose JR, Salto R, Craik CS. Regulation of autoproteolysis of the HIV-1 and HIV-2 proteases with engineered amino acid substitutions. *J Biol Chem*. 1993; 268:11939–11945. [PubMed: 8505318]
12. King NM, Prabu-Jeyabalan M, Wigerinck P, De Bethune M-P, Schiffer CA. Structural and thermodynamic basis for the binding of TMC114, a next-generation human immunodeficiency virus type 1 protease inhibitor. *J Virol*. 2004; 78:12012–12021. [PubMed: 15479840]
13. Prabu-Jeyabalan M, Nalivaika E, King NM, Schiffer CA. Structural basis for coevolution of the human immunodeficiency virus type 1 nucleocapsid-p1 cleavage site with a V82A drug-resistant mutation in viral protease. *J Virol*. 2004; 78:12446–12454. [PubMed: 15507631]
14. Hui JO, Tomasselli AG, Reardon IM, Lull JM, Brunner DP, Tomich C-SC, Heinrikson RL. Large scale purification and refolding of HIV-1 protease from *Escherichia coli* inclusion bodies. *J Protein Chem*. 1993; 12:323–327. [PubMed: 8397790]
15. Fischetti R, Stepanov S, Rosenbaum G, Barrea R, Black E, Gore D, Heurich R, Kondrashkina E, Kropf AJ, Wang S, Zhang K, Irving TC, Bunker GB. The BioCAT undulator beamline 18ID: a facility for biological non-crystalline diffraction and X-ray absorption spectroscopy at the Advanced Photon Source. *J Synchrotron Radiat*. 2004; 11:399–405. [PubMed: 15310956]
16. Park S, Bardhan JP, Roux B, Makowski L. Simulated X-Ray Scattering of Protein Solutions Using Explicit-Solvent Molecular Dynamics. *J Chem Phys*. 2009; 130:134114. [PubMed: 19355724]
17. Bardhan JP, Park S, Makowski L. SoftWAXS: A Computational Tool for Modeling Wide-Angle X-ray Solution Scattering from Biomolecules. *J Appl Cryst*. 2009; 42:932–943. [PubMed: 21339902]
18. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten KJ. Scalable molecular dynamics with NAMD. *Comput Chem*. 2005; 26:1781–1802.
19. Konarev PV, Petoukhov MV, Volkov VV, Svergun DI. ATSAS 2. 1, a program package for small-angle scattering data analysis. *J Appl Cryst*. 2006; 39:277–286.
20. Svergun DI. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr*. 1992; 25:495–503.
21. Svergun DI, Petoukhov MV, Koch MHJ. Determination of domain structure of proteins from X-ray solution scattering. *Biophys J*. 2001; 80:2946–2953. [PubMed: 11371467]
22. Kozin MB, Svergun DI. Automated matching of high- and low-resolution structural models. *J Appl Cryst*. 2001; 34:33–41.
23. Scott WRP, Schiffer CA. Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance. *Structure*. 2000; 8:1259–1265. [PubMed: 11188690]
24. Ferrin TE, Huang CC, Jarvis LE, Langridge R. The MIDAS display system. *J Mol Graph*. 1988; 6:13–27.
25. Case, DA.; Darden, TA.; Cheatham, TE., III; Simmerling, CL.; Wang, J.; Duke, RE.; Luo, R.; Mertz, KM.; Wang, B.; Pearlman, DA.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, JW.; Ross, WS.; Kollman, PA. *AMBER 8*. University of California—San Francisco; San Francisco, Calif: 2004.
26. Prabu-Jeyabalan M, Nalivaika E, Schiffer CA. How does a symmetric dimer recognize an asymmetric substrate? a substrate complex of HIV-1 protease. *J Mol Biol*. 2000; 301:1207–1220. [PubMed: 10966816]

27. Ishma R, Freedberg DI, Wang X-Y, Louis JM, Torchia DA. Flap opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease, and their implication for function. *Structure*. 1999; 7:1047–1055. [PubMed: 10508781]
28. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, Bank JA, Jumper JM, Salmon JK, Shan Y, Wriggers W. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science*. 2010; 330:341–346. [PubMed: 20947758]
29. Collins JR, Burt SK, Erickson JW. Flap opening in HIV-1 protease simulated by ‘activated’ molecular dynamics. *Nat Struct Biol*. 1995; 2:334–338. [PubMed: 7796268]
30. Perryman AL, Lin J-H, McCammon JA. 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 for drugs. *Protein Science*. 2004; 13:1108–1123. [PubMed: 15044738]
31. Hornak V, Okur A, Rizzo RC, Simmerling C. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *PNAS*. 2006; 103:915–920. [PubMed: 16418268]
32. Piana S, Carloni P, Rothlisberger U. Drug resistance in HIV-1 protease: Flexibility-assisted mechanism of compensatory mutations. *Protein Science*. 2002; 11:2393–2402. [PubMed: 12237461]

Figure 1.

Comparison of the crystal structures of HIVp WT and T80N. (A) The main chain traces for the two molecules in the HIVp dimer from the WT ligand-bound form (PDB ID: 1F7A) are shown (yellow and green) with the side chains of the two copies of amino acid Thr 80. (B) Main chain traces from the two molecules (yellow and green) for the WT apo form (PDB ID: 3HVP) showing the difference in the flap positions compared to the ligand-bound form. (C) Superposition of WT HIVp (PDB ID: 1F7A, green) and the T80N mutant (PDB ID: 2FGV, red).

Figure 2. Comparison of WAXS scattering patterns from HIVp for WT (short dashes), T80N mutant (long dashes) with the simulated pattern calculated from a rigid protein model (PDB ID: 1F7A) using the program XS (solid line).

Figure 3. Reconstructions of molecular envelopes for HIVp from WAXS data for WT (blue), T80N mutant (red) and data simulated from the rigid protein model (PDB ID 1F7A, green). The three reconstructions are superimposed in the lower image to demonstrate their similarities, with the purple surface contour indicating an average molecular envelope

Figure 4.

Comparison of the scattering pattern predicted from a rigid HIVp model (black) with WAXS data sets (red) and the predicted patterns obtained after applying vector-length convolution to best fit the WAXS data (blue) for (A) WT; (B) WT plus inhibitor; (C) T80N; (D) T80N plus inhibitor; (E) G48V/L63P; (F) G48V/L63P plus inhibitor.

Figure 5. Distribution of the flexibility parameters c and e obtained from vector-convolution analysis of WAXS data for WT, G48V/L63P mutant and T80N mutant samples of HIVp, in the presence and absence of inhibitor for each case.

Figure 6.

(A) Standard deviations in inter-atomic vector lengths, $\sigma(r)$, as estimated from the WAXS patterns from WT, WT plus inhibitor, G48V/L63P, G48V/L63P plus inhibitor, T80N and T80N plus inhibitor. (B) Standard deviations in inter-atomic vector lengths as a percentage of vector length, $\sigma(r)/r$, for the same samples as shown in (A). This plot emphasizes the greater flexibility of the apo WT and G48V/L63P protein at relatively short inter-atomic distances.

Figure 7.

Standard deviation of interatomic vector lengths as a function of their length as calculated from MD simulation trajectories. (A) Comparison of 100 ns MD simulations of both WT (black circles) and T80N mutant (white circles) proteins. (B) Comparison of a 10 ns simulation (black circles) with a 100 ns simulation (white circles) of the WT protein. For comparison to the standard deviation in inter-atomic distances, $\sigma(r)$ estimated for WT HIVp from WAXS data (solid line) the curve from the 100 ns simulation has been multiplied by a factor of 3.6 (white squares). For consistency, the functional form for $\sigma(r)$ is shown over the same range of vectors as mapped by the MD simulations but it is anticipated that this two parameter model will be inapplicable for very short ($r < 5\text{\AA}$) and very long ($r > 30\text{\AA}$) interatomic vectors.

Table I
Flexibility parameters, c and e, determined from vector-length convolution analysis of WAXS data for HIVp samples

The value of c is related to the magnitude of the fluctuation in inter-atomic separation between pairs of atoms and the value of e is related to increase in this fluctuation with increased inter-atomic separation. (See text for details). The value of $\sigma(r=10\text{\AA})$ is the average variation in inter-atomic separations for pairs of atoms separated by 10 Å. The values presented in this table were obtained by fitting WAXS data in the resolution range (1/d) 0.02–0.1 Å⁻¹.

Sample	c	e	$\sigma(r=10\text{\AA})$
WT	0.42	0.82	2.77
WT+inhibitor	0.20	1.06	2.29
T80N	0.16	1.10	2.01
T80N+inhibitor	0.17	1.08	2.04
G48V/L63P	0.32	0.89	2.48
G48V/L63P+inhibitor	0.18	1.08	2.16