Measuring rapid hydrogen exchange in the homodimeric 36 kDa HIV-1 integrase catalytic core domain

Nicholas C. Fitzkee, Dennis A. Torchia, and Ad Bax*

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520

Received 3 December 2010; Accepted 7 October 2010
DOI: 10.1002/pro.582
Published online 6 January 2011 proteinscience.org

Abstract: Measurements of rapid hydrogen exchange (HX) of water with protein amide sites contain valuable information on protein structure and function, but current NMR methods for measuring HX rates are limited in their applicability to large protein systems. An alternate method for measuring rapid HX is presented that is well-suited for larger proteins, and we apply the method to the deuterated, homodimeric 36 kDa HIV-1 integrase catalytic core domain (CCD). Using long mixing times for water-amide magnetization exchange at multiple pH values, HX rates spanning more than four orders of magnitude were measured, as well as NOE cross-relaxation rates to nearby exchangeable protons. HX protection factors for the CCD are found to be large (>10⁴) for residues along the dimer interface, but much smaller in many other regions. Notably, the catalytic helix (residues 152-167) exhibits low HX protection at both ends, indicative of fraying at both termini as opposed to just the N-terminal end, as originally thought. Residues in the LEDGF/p75 binding pocket also show marginal stability, with protection factors in the 10–100 range (~1.4–2.7 kcal/mol). Additionally, elevated NOE cross-relaxation rates are identified and, as expected, correspond to proximity of the amide proton to a rapidly exchanging proton, typically from an OH side chain. Indirect NOE transfer between H₂O and the amide proton of I141, a residue in the partially disordered active site of the enzyme, suggests its proximity to the side chain of S147, an interaction seen in the DNA-bound form for a homologous integrase.

Keywords: NMR; protein stability; dynamics; HX; retroviral integration; active site dynamics

Introduction
First proposed by Kaj Ulrik Linderstrøm-Lang, amide hydrogen exchange (HX) with water has become an exceedingly useful tool for understanding protein structure and function.¹ Modern techniques have evolved to study the full range of HX rates, using NMR to measure the rates of rapidly exchanging protons,⁵⁴ or NMR/mass spectrometry to measure the slower rates.⁵⁻⁷ In the simplest case, the HX rate $k_{\text{H₂O}}$ contains information on both the local folding/unfolding kinetics of the amide group and the energetic stability. As local unfolding (or solvent penetration) occurs, an amide group becomes accessible to exchange:

$$
\text{Closed} \xrightarrow{k_{\text{op}}} \text{Open} \xrightarrow{k_{\text{cl}}} \text{Exchanged}
$$

(Scheme 1)

where $k_{\text{op}}$ and $k_{\text{cl}}$ correspond to the opening and closing rates for the local folding event, respectively, and $k_{\text{re}}$ corresponds to the intrinsic “random coil” exchange rate. Exchange is highly sensitive to

Abbreviations: ASV, avian sarcoma virus; CCD, catalytic core domain; CTD, C-terminal DNA-binding domain; HIV, human immunodeficiency virus; HX, hydrogen exchange; IN, integrase; LEDGF, lens epithelium-derived growth factor; NTD, N-terminal zinc-binding domain; PFV, prototype foamy virus; RDC, residual dipolar coupling; SM, supplementary material; SSR, sum of squared residuals; Vpr, viral protein R.

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Ad Bax, Building 5, Room 126, NIH, Bethesda, MD 20892-0520, E-mail: bax@nih.gov
[OH\(^-\)] concentration, since above pH \(\sim 5\), HX is catalyzed by hydroxyl ions.\(^7\) Exchange rates can span many orders of magnitude, with amide hydrogen life times \((k_{\text{H2O}}^{-1})\) ranging from as low as a few milliseconds at physiological pH for amide protons exposed to solvent and not stabilized by H-bonding, to weeks or longer for H-bonded amides buried in the core of a highly stable protein like myoglobin.\(^8\) Rapidly exchanging protons reveal information about H-bonding, surface dynamics, allostery, and binding\(^1,9\) while slowly exchanging protons report on the stability of the protein and mechanisms of folding.\(^10,11\)

NMR is the primary tool for measuring fast HX, because the HX-mediated transfer of magnetization from water to amide sites can be measured with millisecond resolution. However, since proton-proton cross-relaxation can also transfer magnetization through multiple pathways, care must be taken to distinguish genuine HX from cross-relaxation contributions to NMR \(k_{\text{H2O}}\) measurements.\(^12\) To this end, a number of pulse methods have been developed to measure fast exchange.\(^13\) The MEXICO,\(^14\) WEX,\(^15\) WEX-II,\(^16\) spin-echo filtered WEX-II,\(^17\) and CLEANEX-PM\(^18,19\) sequences measure HX by inverting water spins and examining transfer to amide protons. Amide HX can also be measured using diffusion-ordered spectroscopy (DOSY)\(^20\) or by saturation transfer.\(^2,21\) Of these techniques, CLEANEX-PM has become the method of choice for folded proteins, which typically experience rotational diffusion in the slow tumbling limit. CLEANEX-PM effectively removes contributions from cross relaxation using a spin-lock sequence during the HX mixing time. The recent SOLEXSY experiment\(^22\) solves the cross-relaxation problem using a different approach and is particularly applicable to disordered proteins, where the slow tumbling limit does not apply.

Although well-suited for smaller proteins, practical considerations limit the application of CLEANEX-PM to larger systems. First, larger proteins typically require higher magnetic field strengths to yield the required spectral resolution. Higher field strengths require stronger radiofrequency fields during the HX mixing time and limit the duration of the mixing time such as to avoid sample heating or probe damage. Second, both longitudinal and transverse relaxation processes are operative during the CLEANEX spin lock HX mixing period, because protons are locked in the rotating frame to eliminate the NOE effects. The transverse relaxation rate \(R_2\) in proteins increases approximately linearly with size, also limiting the maximum durations of the mixing period that can be fruitfully used. The additional \(R_2\) relaxation process complicates existing models for HX, which typically only account for \(R_1\) relaxation during the mixing period. Where \(R_2\) relaxation is significant, an additional experiment must be performed to compensate for the relaxation that occurs during the mixing time.\(^23\) More importantly, however, the limited mixing time required to compensate for fast \(R_2\) relaxation has a strong adverse impact on the sensitivity at which HX rates themselves can be measured for large proteins. For example, for the 36 kDa homodimeric catalytic core domain (CCD, residues 50-212) of the HIV-1 integrase enzyme, where the backbone amides have high \(R_2\) rates and attainable spectral quality is limited by both its solubility and poor spectral properties,\(^24\) measuring HX rates by the CLEANEX-PM method only yields adequate data for the fastest exchanging amide protons.

Full-length integrase is tetrameric in solution, and each 32 kDa monomer contains three domains an N-terminal zinc-binding domain (NTD), the CCD that contains the active site, and a C-terminal DNA-binding domain (CTD).\(^25\) Integrase is responsible for insertion of the viral cDNA into the host genome, and it is therefore necessary for viral replication. Since integrase catalyzes two different reactions, 3’ processing and strand transfer, dynamics at the active site are likely to be important.\(^26\) Although the structure of the homologous prototype foamy viral (PFV) integrase recently has been solved in complex with DNA,\(^27\) with reasonably well-defined electron density for its active site, very little is known about the active site structure during 3’ processing or when DNA is absent. The integrase CCD (IN\(^{50-215}\)) in solution exhibits poor NMR spectral properties, and measurement of HX rates is therefore of interest as it may reveal regions of the protein that are prone to unfolding and aggregation. Additionally, rapid HX promises to be a useful probe of the stability of various proposed binding sites of the CCD.

In this paper, we present an alternate method for measuring HX processes in larger proteins and apply the technique to IN\(^{50-212}\). Perdeuteration in combination with TROSY spectroscopy\(^28\) is used to maximize attainable spectral resolution at higher fields, and the \(^2\)H isotopic labeling as well as generation of suitable difference spectra minimize the NOE contributions of non-exchangeable protons. By measuring HX at multiple pH values, accurate neutral pH \(k_{\text{H2O}}\) values can be extracted over a wide range of rates, from \(\sim 0.001\) to 35 \(s^{-1}\), thus spanning more than four orders of magnitude. The method also yields an estimate for the NOE contribution to magnetization transfer from solvent to protein and shows a correspondence between high NOE rates and proximity of amide protons to other rapidly exchanging protons. The dynamic nature of IN\(^{50-212}\) is highlighted by our results, and combining our HX data with previously determined RDC measurements,\(^29\) we find evidence for millisecond timescale fluctuations throughout the entire catalytic helix as well as in many of the putative binding pockets of the CCD.
**Results and Discussion**

**Experimental approach**

When water protons are selectively inverted in a solution of protein molecules, magnetization transfer will occur between the water and protein protons, perturbing the protein nuclear spin magnetization. This transfer can occur by means of several processes. First, water protons can physically exchange with amide protons, the HX process of key interest in this study. However, bound water molecules can also transfer magnetization to $\text{H}^\text{N}$ protons via cross-relaxation (i.e., the homonuclear NOE). Moreover, water magnetization can also be transferred to protein by an indirect NOE process, for example, in a situation where a water proton exchanges with an OH proton, which in turn exchanges magnetization with a nearby amide proton via cross-relaxation.

Since hydroxyl protons often exchange very rapidly with solvent (relative to $k_{\text{H}^\text{O}}$ and NOE rates), a solvent-exposed OH group effectively acts as a slow magnetization donor. Protons resonating very close to the water signal, such as nonexchangeable $\text{H}^\text{a}$ protons, will also contribute to magnetization transfer; however, they will occur between the water and protein protons, the HX process of key interest, will be much weaker and non-linear. For the time being ignoring process 4, the signal transferred from water to an amide proton, $S(T)$, is initially proportional to the mixing time $T$, but subsequently competes with decay caused by relaxation. To a first approximation (see also Supplementary Material) the amide signal is described by:

$$S(T) = \frac{M_z^o(T) - M_z^e(T)}{2M^0} = \frac{\Delta f (k_{\text{H}^\text{O}} + k_{\text{obs}})}{k_{\text{H}^\text{O}} + k_{\text{H}^\text{N}}} \left[ 1 - e^{-[(R_1 + k_{\text{H}^\text{O}} + R_{\text{H}^\text{N}})T]} \right]$$

Here, $M_z^o(t)$ and $M_z^e(t)$ are the signal intensities observed in the WEX-III TROSY spectrum with and without selective water inversion, respectively; $M^0$ is a reference signal, taken in the absence of water inversion at a very long recycle delay in order to eliminate the effects of incomplete $R_1$ relaxation; $\Delta f$ is given as $f^\text{r} - f^\text{w}$, where $f^\text{r}$ ($f^\text{w}$) is the fractional steady-state water signal with (without) inversion; the observed build-up rate is a sum of contributions from processes 2–3 ($k_{\text{obs}}$ or $-\sigma$ in the original Solomon notation$^{33}$), and the HX rate of interest, $k_{\text{H}^\text{O}}$. $R_1$, and $R_{\text{H}^\text{N}}$ are the longitudinal spin relaxation rates of the amide and water signal, respectively.

$$S(T) = \frac{M_z^o(T) - M_z^e(T)}{2M^0} = \frac{\Delta f k_{\text{obs}} e^{-[R_1 + k_{\text{obs}} + R_{\text{H}^\text{N}}]T]}{R_1 + k_{\text{obs}} + R_{\text{H}^\text{N}} \left[ 1 - e^{-[(R_1 + k_{\text{obs}} + R_{\text{H}^\text{N}})T]} \right]}$$

Separating the $k_{\text{H}^\text{O}}$ and the $k_{\text{H}^\text{N}}$ contributions to $k_{\text{obs}}$ is straightforward if the $\text{H}^\text{N}$ is exchanging under EX2 conditions. In that case, the H-bond closing rate, $k_{\text{cl}}$, is much faster than the rate of exchange, $k_{\text{rec}}$, and the overall exchange rate $k_{\text{H}^\text{O}}$ should depend exponentially on pH.$^1$ At the same time, if
where pH 0 is an arbitrary reference pH, chosen in a weighted-average measure to combine 15N and 13C. In the region where HX is dominantly base catalyzed, the 1H-15N correlation map provides a convenient probe to assess whether significant structural changes occur in IN50-212 as a function of pH. Using a weighted-average measure to combine 15N and 1H resonances, we find the chemical shifts in IN50-212 to be nearly unchanged from pH 6.4 to 9.2. Even for NOE transfer from water to H N mediated by the scheme is run without the water inversion pulse at time point a. These two FIDs are subtracted from one another to create the difference spectrum as described in the text. The measurement is performed as a single interleaved experiment, iterating through all values of pH, on 503 ms range. For each value of T, two experiments are recorded, with and without the hatched water inversion pulse at a. These two FIDs are subtracted from one another to create the difference spectrum as described in the text. The measurement is performed as a single interleaved experiment, iterating through all values of T, with and without water inversion. The reference experiment begins with the 1H pulse at b and lacks the initial water inversion module.

the protein structure remains constant, the NOE contribution to exchange will not depend on pH. Even for NOE transfer from water to H N mediated by side-chain OH groups, the OH exchange will generally be much faster than NOE transfer, and therefore such processes will be rate limited by the pH-independent NOE transfer rate. Thus, by performing experiments at multiple pH values, it is possible to determine the relative contributions to kobs using the following model:

\[ k_{\text{obs}}(pH) = k_{\text{noe}} + k_{\text{h2o}} 10^{(\delta pH - \Delta pH)} \]  (2)

where pH 0 is an arbitrary reference pH, chosen in the region where HX is dominantly base catalyzed (i.e., pH 0 > ~5).

**pH independence of IN50-212 structure**

The 1H-15N correlation map provides a convenient probe to assess whether significant structural changes occur in IN50-212 as a function of pH. Using a weighted-average measure to combine 15N and 1H resonances, we find the chemical shifts in IN50-212 to be nearly unchanged from pH 6.4 to 9.2. This is consistent with NOE transfer rates being pH independent, as one would expect for a process that is not rate limited by the pH-dependent chemical exchange between HX and H N. This is also consistent with the assumption that the NOE contribution to exchange will not depend on pH.

**1H R1’ and kobs values**

At any given pH, measurement of the apparent exchange rates kobs and the R1’ values is straightforward. In a 15N-1H TROSY spectrum at 800 MHz, most peaks in IN50-212 are well-resolved, permitting the observation of separate signals for approximately 90% of the nonprolyl residues. Signals undergo an initial buildup of intensity as a function of the mixing time, T, and if the observed rate is fast enough to reach equilibrium with the water magnetization within the T period, the signals will begin to decay.
as the water signal itself decays [Fig. 2(A)]. Although most recent studies have largely focused on short mixing times, where $S(T)$ is approximately linear,$^{16,19}$ Eq. (1) remains a good approximation for much longer times too.$^{15}$ Sampling long mixing times is important for reliable measurement of slow $k_{\text{obs}}$ rates that correspond to intermediate protection factors, many of which are observed in the CCD [Fig. 3(A)]. Even with only six mixing times, it is possible to obtain accurate values of $k_{\text{obs}}$ over a broad range, and comparing measurements at multiple pH values immediately reveals which residues have a significant HX component [Figs. 3(A–C)]. Accurate $^1H$ R1' values are harder to obtain, because this parameter is dependent on the maximum of $S(T)$. However, for all but the slowest exchanging resonances it was possible to find at least one pH value where R1' could be determined with an uncertainty of $\pm 50\%$ or better. For cases where R1' could be determined under multiple experimental conditions, it was generally found to be consistent within the experimental uncertainty (SM, Table S1).

When $k_{\text{obs}}$ is determined for multiple pH values, it is possible to separate the contributions from NOE cross-relaxation and HX for protons exchanging in the EX2 limit. Because IN$^{50-212}$ is deuterated and the pulse scheme (Fig. 1) is implemented as a difference experiment, NOEs with nonexchangeable protons are already largely eliminated, except for a very small fraction ($= \sim 5\%$) of residual random protonation. As discussed above, however, isotopic labeling and pulse sequence design will not eliminate the contributions of rapidly exchanging protons, such as those from side-chain OH groups. Using Eq. (2) to model $k_{\text{obs}}$ at multiple pH values, we determined the aggregate NOE exchange rate $k_{\text{noe}}$ and the actual HX rate $k_{\text{h}_2\text{o}}$ (SM, Table S1). The fitting procedure yielded well-defined values of the parameters, with uncertainties of typically $\pm 10\%$ or lower (Fig. 4). For several residues, it was possible to measure exchange only at pH 6.4—these residues exchange too rapidly to be detected at higher pH. For these residues, we assumed that the $k_{\text{noe}}$ term was negligible compared to the $k_{\text{h}_2\text{o}}$ rate. Thus, at pH 6.4, the upper bound for $k_{\text{h}_2\text{o}}$ values is approximately $35\, s^{-1}$, as observed for S119. By examining the uncertainties in our data over the pH range from 6.4 to 9.2, the lowest unambiguously detectable intrinsic HX rates correspond to approximately $10^{-3}\, s^{-1}$ at pH 6.4. For rates slower than this, changes in $k_{\text{obs}}$ with increasing pH only become visible at the final pH (9.2) and would not significantly exceed the noise threshold established by $k_{\text{obs}}$ values at lower pH. Thus, the approach described here is able to determine accurate $k_{\text{h}_2\text{o}}$ rates in IN$^{50-212}$ over a range spanning more than four orders of magnitude. Even for very slowly exchanging amides, however, it is frequently possible to determine $k_{\text{noe}}$ quite accurately, even if $k_{\text{h}_2\text{o}}$ cannot be determined (Fig. 4, open circles).

Figure 2. Signal transferred from water to amide sites during the mixing period, $T$. Each panel shows a $^1H$ one-dimensional cross-section taken through the 2D ($^1H$, $^{15}N$) TROSY-HSQC difference spectra at various recovery time points, $T$. Insets show the 2D peaks at identical contour levels. G189 (A) and A105 (B) represent rapidly and slowly exchanging $^1H$ atoms, respectively. Signals of fast exchanging amides rapidly attain a maximum and then decay at the $H_2O$ R1 rate, whereas the decay phase is not yet reached at $T = 480\, ms$ for slowly exchanging amides. The dotted line indicates the maximum observed intensity for each residue.
For the vast majority of residues in IN50-212, Eq. (2) represents an excellent model for obtaining both $k_{\text{h}_2\text{O}}$ and $k_{\text{noe}}$ from a series of HX measurements. However, several residues were observed which did not fit well to the model (SM, Table S1). An example is I204 (Fig. 4), where the $k_{\text{obs}}$ values are precisely defined but nevertheless do not fit to Eq. (2). A poor fit will result if a nearby amide proton itself shows pH-dependent HX. In this case, the $k_{\text{obs}}$ will also be pH-dependent, and Eq. (2) will no longer hold. This situation arises for I204, located at the N-terminus of helix $\alpha_6$, close to the disordered $\alpha_5$--$\alpha_6$ loop, where HX values are strongly pH-dependent. As a result, its $k_{\text{h}_2\text{O}}$ value (≈ 0.002 s$^{-1}$) is an upper bound for HX, rather than a true measurement.

The situation of pH-dependent $k_{\text{noe}}$ values is mitigated by several factors: (1) It is straightforward to identify questionable resonances by their poor fits to Eq. (2). A pH-dependent $k_{\text{noe}}$ will cause an increase in $k_{\text{obs}}$ with increasing pH, but it will be far from the exponential dependence expected for an exchanging residue. These residues are easily identified by careful examination of the data. (2) It is possible to predict pH-dependent $k_{\text{noe}}$ values if the structure is known. Likely candidates will be well-packed residues within 4 Å of other rapidly exchanging amide protons in the protein structure. This is a fairly short distance, so helical residues will be more affected by interactions with sequentially neighboring amide groups than strand residues. (3) Finally, the overall effect on $k_{\text{h}_2\text{O}}$ will be small. In the IN50-212 monomer, we find evidence for pH-dependent $k_{\text{noe}}$ values in 15 residues. In five of these, the $k_{\text{h}_2\text{O}}$ remains below the detection threshold of 0.001 s$^{-1}$. In the remaining 10, eight of which are in the carboxy-terminal helix (residues 197–204), $k_{\text{h}_2\text{O}}$ values (back-calculated for pH 6.4) are elevated on average by only 0.002 s$^{-1}$. Thus, even when neighboring residues have strongly pH-dependent HX rates, they are rarely close enough to create a significant impact on $k_{\text{h}_2\text{O}}$. 

Figure 3. Determining $k_{\text{obs}}$ and $R_1$ values in IN50-212. Each panel displays the signal intensity versus mixing time $T$ for pH 6.4 (A), 7.35 (B), and 8.3 (C). The dotted gray line at the top of each panel shows the calculated intensity for an isolated spin exchanging infinitely fast with water, decaying at the water $R_1$ rate. Residues G189 (●, —); S147 (○, — e); V79 (▲, — —); A105 (●, —); Q137 (▲, — —) each exhibit different behavior as a function of pH. Signals of fast-exchanging spins (G189, S147) are not detected at higher pH, whereas those of slowly exchanging spins persist. Lines are the result of curve fitting signal intensities for each residue to Eq. (1). $R_1$ values cannot be determined for slowly-exchanging residues, but reasonable values are obtained for G189 and S147.

Figure 4. Identifying contributions of $k_{\text{noe}}$ and $k_{\text{h}_2\text{O}}$ to $k_{\text{obs}}$. Measurements of $k_{\text{obs}}$ versus pH are shown for M154 (▲, — —); G140 (X, ...); S123 (■, — —); G106 (○, — —); and I204 (▲, —). In the EX2 regime, residues will exchange predictably faster at higher pH, according to Eq. (2). Residues whose $k_{\text{obs}}$ values are independent of pH (e.g., G106) relax primarily via homonuclear NOE pathways ($k_{\text{noe}}$). Poor agreement or a plateau in $k_{\text{obs}}$ may result from a pH-dependent $k_{\text{noe}}$ value. An example of this is seen for I204, where the best fit curve falls well outside the error bars for the observed $k_{\text{obs}}$. 

Fitzkee et al.
Protection factors and catalytic significance in integrase

The HX protection factor \( P \) of a backbone amide is a useful measure of the relative water-accessibility and state of H-bonding of the exchangeable proton, where \( P = k_{rc}/k_{h2o} \). Here, \( k_{rc} \) is the intrinsic exchange rate expected for the proton if the polypeptide chain is in a disordered, water-accessible random coil conformation. The log of the protection factor is proportional to a free energy and is often interpreted as the local unfolding free energy for a given exchangeable proton. Although the maximum protection factors probed in our study are lower than obtainable via direct H-D exchange measurements, they were measured over a wide range, up to fairly high values (\( \text{log} \, P \approx 4 \)), because of the long mixing times and multiple pH values probed in our experiments. While not as pertinent to global protein unfolding, the protection factors observed in our study cover the range relevant to protein dynamics occurring on time scales shorter than global unfolding, such as allostery and transient sampling of local unfolding.

Log protection factors in IN\(^{50-212}\) reveal a strong segregation between the solvent-accessible face of the protein and the dimer interface (Figs. 5 and 6). Those elements of secondary structure that are most stable, including helices \( \alpha_1 \), \( \alpha_5 \), and \( \alpha_6 \), as well as strands \( \beta_2 \) and \( \beta_3 \), all line the dimer interface. The protection factors of these residues, on average, are more than two orders of magnitude higher than for the rest of the protein. Therefore, it seems likely that dimer formation is a key to the stability of IN\(^{50-212}\). After expression in deuterated growth media, several days are required for L74 and A76 to back-exchange (data not shown), so the exchange rates for these residues are orders of magnitude slower than the lowest rates detectable in our experiment. Because L74 and A76 are involved in hydrogen bonding between strands \( \beta_2 \) and \( \beta_3 \), these strands probably form the most stable region of the core domain dimer.

Of the backbone amide protons in the active site that are predicted to be H-bonded on the basis of the PFV integrase structure, only N144 shows significant protection \( ( \text{log} \, P = 4) \) from solvent exchange. Other CCD active site backbone amides protons that are H-bonded in the PFV structure (Q146, S147, and Q148) show very little protection \( ( \text{log} \, P = 1.4 \) to 3.9), indicating these H-bonds are not highly populated in the HIV-1 CCD.

The HX data also provide new insights into the stability of the C-terminal \( \alpha_6 \) helix. In crystal structures of the CCD alone, this helix is well packed across the dimer interface and generally extends at least to residue 1208.\(^{39,43-45}\) Additionally, Chen \textit{et al.}\(^{46}\) observed a much longer CT helix in an integrase construct containing both the CCD and the CTD, suggestive of increased helical propensity for the linker between the two domains. On the other hand, the orientation observed by Chen \textit{et al.} may be impacted by crystal packing interactions: In a

\[ \text{Figure 5. HX protection factors, } P, \text{ and } k_{h2o} \text{ values for IN}^{50-212}. \] (A) Log protection factors and (B) \( k_{h2o} \) values are plotted for each residue. HX is generally lowest for regular secondary structure and highest in loops. Depressed protection factors are observed for the entire catalytic loop (residues 140-152), but low protection persists well beyond residue 152 in to helix \( \alpha_4 \). In (A), the gray background highlights the upper observation limit for HX protection of each residue (using \( k_{h2o} = 0.001 \text{ s}^{-1} \)). Residues colored blue are H-bonded to the main chain in the structure of Maignan \textit{et al.}\(^{39}\) (PDB entry 1BL3, chain C), residues colored violet are H-bonded to side chain acceptors, and residues in red are not H-bonded to protein acceptor atoms. The secondary structure is plotted along the top of panel A. In both panels, dotted lines connect continuous residues.
recent crystal structure of the homologous PFV integrase,27 the corresponding helix terminates at F270 (residue V201 in HIV-1 integrase) and the helix does not participate in the dimer interface. This is also the case for a HIV-1 model based on the PFV crystal structure.47 Our data show strong HX protection up to residue T206 (Figs. 5 and 6). Beyond this residue, protection falls off sharply. Complementary data from residual dipolar couplings (RDCs) indicate that the orientation of the CT helix is consistent with the crystal structure up to Q209.29 Therefore, it is likely that the CT helices of IN\textsuperscript{50-212} dimerize in solution, as seen in crystal structures of the CCD alone, and that they are well packed up to T206. The agreement between RDCs and the crystal structures for residues 207–209 suggests these remain helical but that transient opening of their amide H-bonds occurs on a fast time scale, in agreement with below average $^{15}$N-{$^1$H} NOE and $^{15}$N R\textsubscript{2} values, as well as increased $^{15}$N R\textsubscript{1} values.24 At this point, it remains unclear whether the CT helices dimerize in the full-length HIV-1 integrase.

Unlike the dimer interface, the solvent exposed face of IN\textsuperscript{50-212} shows strikingly low HX protection factors [Fig. 6(A)]. This is most likely related to a large portion of the solvent exposed face being involved in catalysis or DNA binding.39 It is, therefore, not surprising that both the catalytic loop (residues 140-152) as well as the $\alpha_5$-$\alpha_6$ linker (residues 187-194) experience low HX protection, in particular considering that these regions also show relatively high levels of dynamic disorder as assessed by $^{15}$N relaxation data and smaller RDCs.24,29 In PFV integrase, both of these regions interact directly with the DNA, and they may require some degree of pliability to establish optimal interaction with the DNA. Indeed, flexibility may also be required for rearrangement of the CCD active site, which functions to catalyze the two distinct reactions of 3'-processing and strand transfer.26,48 The degree to which nearly the entire catalytic $\alpha_4$ helix is found to be unstable is surprising, however. It is perhaps not unexpected that the N-terminal end of $\alpha_4$ is unstable, considering that many crystal structures show this helix to start as late as residue N155,43,46,49 as assessed by the DSSP program,50 and our data indicate that this helix experiences diminished protection up to residue L158. Remarkably, however, our data also indicate fraying at the C-terminal end of this helix, suggesting that the observed H-bonding in $\alpha_4$ is dynamic on the microseconds time scale [Fig. 6(A); H-bonding is indicated by the data color]. In fact, only residues 159-162 exhibit protection factors above the upper detection limit (Fig. 5).

In addition to the active site, several other areas on the CCD have been proposed as drug binding targets. The integrase binding domain (IBD) of lens epithelium-derived growth factor (LEDGF, also known as transcriptional co-activator p75) binds near the dimer interface, interacting with helices $\alpha_1$, $\alpha_3$, and $\alpha_5$, as well as the linker between helices $\alpha_4/5$ (residue 166–171).51 Binding of LEDGF is important (though not essential) for allowing efficient integration of the DNA in the nucleus.52 Because integration is reduced in LEDGF-knockout cells,53 it has been proposed that the CCD-LEDGF binding site may be a potential drug target. Our data show that this region (particularly residues 167-173) has depressed HX protection, possibly making structure-based drug design more challenging. From a thermodynamics perspective, tight binding to this location would have to overcome an entropic penalty for stabilizing the protein. Indeed, current inhibitors

**Figure 6.** Log protection factors plotted on the CCD structure. Color-coded log protection factors are displayed on the structure of IN\textsuperscript{50-212} by Maig\textit{n} et al.39 (PDB entry 1BL3, chain C). Protection factors are calculated relative to predicted HX rates40,37 for random coils of the same amino acid sequence. The water-accessible face of the protein (A) shows relatively low protection, especially near the active site (highlighted by the spherical crystallographically observed Mg\textsuperscript{2+} ion). The dimer interface (B) shows higher protection, except at the LEDGF binding pocket (arrowhead, residues 166-171). Panels A and B are related by a 90° rotation about the vertical axis. In the top panel, the rear-facing monomer has been rendered semitransparent for clarity. Rendering was performed using MOLSCRIPT\textsuperscript{41} and Raster3D.\textsuperscript{42}

Fitzkee et al.
targeting this region all have relatively low affinities, with IC\textsubscript{50} values in the low micron range.\textsuperscript{54,55}

In addition to LEDGF, HIV viral protein R (Vpr) has been found to interact with integrase and to inhibit integration.\textsuperscript{56} Peptides derived from Vpr have been found to be potent inhibitors of integrase, with IC\textsubscript{50} values as low as 4 nM.\textsuperscript{57,58} Molecular docking predictions localized the peptide to the binding pocket between the NTD and CCD. Many of the proposed interactions involve highly protected residues, including T112, S195, and D207.\textsuperscript{58} However, several other contacts show very low HX protection in our data, including those in the \(\alpha\)5/6 linker. Conceivably, the peptide-integrase interactions are able to sufficiently stabilize the more dynamic regions of integrase, much like the concerted folding-binding events observed for intrinsically disordered proteins.\textsuperscript{59–61} Alternatively, it is possible that DNA binding is required for tight peptide-integrase interaction. Litttle is known about the mechanism of Vpr peptide-based integrase inhibition, but attempts to bind Vpr peptides to the CCD alone have so far been unsuccessful (data not shown).

**Figure 7.** Structural basis for elevated \(k_{\text{noe}}\) values. Large \(k_{\text{noe}}\) values result from close proximity to a rapidly exchanging proton. Residues with elevated H\textsuperscript{N} \(k_{\text{noe}}\) values are rendered as CPK spheres, and relevant hydrogen atoms have been added using REDUCE.\textsuperscript{62} The H\textsuperscript{N} may relax via NOE transfer to a rapidly exchanging OH group, as shown for H67 (A) and H171 (B), where transfer is mediated through surface-exposed Thr side chains. The side-chain OH hydrogens are not observed crystallographically, so distances are noted between H\textsuperscript{N} and O\textsuperscript{\textdagger} in the figure. Alternatively, nearby amide protons with rapid HX can also contribute to a large \(k_{\text{noe}}\) value, as observed for I60 (C), where for G59 H\textsuperscript{N} we find \(k_{\text{coo}} = 11.5 \pm 0.4 \text{ s}^{-1}\). (D) The large \(k_{\text{noe}}\) contribution for T122 arises from transfer to its own side chain OH.

**NOE transfer rates**

While eliminated in most other HX detection methods, the NOE transfer rate \(k_{\text{noe}}\) can also provide valuable structural information. Homonuclear NOEs of amide protons to nonexchangeable protons are largely eliminated by deuteration of the protein and the difference method employed, respectively. However, nearby, rapidly exchangeable protons can participate in transferring magnetization between water and the amide proton, and we observe elevated \(k_{\text{noe}}\) rates both for residues with high (e.g., H67) and low (S195) protection (Fig. 6). There is a strong correlation of \(k_{\text{noe}}\) values in IN\textsuperscript{50–212} with structure: amide protons with elevated \(k_{\text{noe}}\) rates tend to be close to rapidly exchanging OH groups, such as surface-exposed Ser or Thr side chains or surface-exposed amides with high \(k_{\text{hydro}}\) values (Fig. 7). At the same time, residues that are distant from other rapidly exchangeable protons tend to have very low \(k_{\text{noe}}\) values. An interesting case is I141, one of the residues in the catalytic loop. It has a low protection factor, but it also has an uncharacteristically high \(k_{\text{noe}}\) rate compared with its neighbors. To interpret this in structural terms, we note that in the DNA-bound structure of PFV integrase, the amide of the corresponding residue (T210) is H-bonded to the OH of S216 (S147 in HIV-1 integrase).\textsuperscript{27} Thus, the observation of high \(k_{\text{noe}}\) values in the CCD strongly suggests that the catalytic loop in the absence of DNA nevertheless resembles the conformation observed in the DNA-bound state for a significant fraction of time. This agrees with previous conclusions based on the intermediate order parameters observed in the catalytic loop.\textsuperscript{24}

**Conclusions**

HX rates contain a wealth of information about the dynamics and stability of proteins. Here, we have presented an approach suitable for measuring HX in larger proteins and have applied the method to the CCD of HIV-1 integrase, a 36 kDa dimeric protein of high biological and pharmaceutical interest. The measured exchange rates span more than four orders of magnitude and reveal a dramatic difference in solvent protection between the two sides of the monomer. Amides at the dimer interface are highly protected, but the catalytic face of the protein is characterized by fraying helices and dynamic loops. Although the catalytic loop was known to be dynamic, the low protection factors observed for the C-terminal end of the catalytic helix, as well as the LEDGF binding pocket, were unexpected. The dynamic behavior of these regions may complicate the design of integrase inhibitors, since the free energy of binding must also overcome the instability of the regions themselves. In contrast to most other methods for measuring HX, the WEX-III TROSY
method also allows for estimation of NOE transfer rates to nearby exchangeable protons. Elevated rates are strongly correlated with close proximity to nearby OH groups, and by analyzing these rates, we find evidence that the catalytic loop may sample the DNA-bound conformation. Given that many of the regions prone to exchange are also involved in DNA binding interactions, it is highly likely that HX rates will change upon DNA binding. Work is ongoing to use the methods described here to probe the interactions of integrase with its binding partners.

Materials and Methods

Sample preparation

Uniformly $^{13}$C, $^{15}$N, $^{2}$H enriched Q53E C56S W131E F185K Q209E IN$^{50-212}$ was expressed in E. coli and purified as described previously.$^{24}$ Freshly purified IN$^{50-212}$ was exchanged in to NMR buffer and concentrated by ultrafiltration to 500 µM. The sample buffer contained 150 mM NaCl, 40 mM MgCl$_{2}$, 6% (v/v) D$_{2}$O, 0.02% (w/v) NaN$_{3}$, and EDTA-free complete protease inhibitor cocktail (Roche). Samples at pH 6.4 and 7.35 were buffered with 20 mM PIPES, and samples at pH 8.3 and 9.2 were buffered with 20 mM HEPES. To ensure pH stability, the sample pH was measured both before and after NMR experiments.

NMR measurements

All NMR measurements were performed at 298K on a Bruker 800 MHz cryoprobe-equipped spectrometer. The pulse program for measuring proton exchange (Fig. 1; code included as SM) was run in three distinct modes. In the first mode, peak intensities for $M^{0}$ were recorded using a 2D TROSY reference experiment, beginning at time point $b$ of Fig. 1, and using a long recycle delay (8 s) to allow for full $T_{1}$ recovery of amide protons. The second mode is the actual exchange experiment. In this mode, the pulse program was run in its entirety with a shorter recycle delay of 2 s. For each measured value of $T$, the program records two transients, one with and one without the selective water-flip-back at time point $a$. The acquisition time for the $^{15}$N dimension was 87 ms over 250 complex points, and the acquisition for the $^{1}$H dimension was 120 ms over 1537 complex points. Using 16 scans per FID, the total measurement time for each pH value was approximately 70 h, including 18 h for the reference experiment.

Finally, for determining the values of $f^{-}$ and $f^{+}$, the program was run in a third (calibration) mode. In this mode, the program is configured as a 1D $^{1}$H experiment. To ensure steady state, the program was run in its entirety for eight scans (more scans did not affect the results). Then, on the ninth scan, the program applies a 1 µs, ∼5° flip angle “spy” pulse at various points in the sequence, where after it immediately jumps to acquisition. If the spy pulse is placed before the selective inversion at time point $a$, the volume of the water peak measures the bulk magnetization of water at steady state before the program begins. If the spy pulse is placed at $b$, with (or without) the water inversion, one measures the water magnetization after the inversion and delay $T$. Taking the ratio of these values to the reference volume, we determined the relative water magnetization with $(f^{-})$ and without $(f^{+})$ the selective inversion pulse. The value of $T$ used in the calibration mode was 10 ms, and the recycle delay was set to match the HX measurements (2 s).

The water spin-lattice relaxation rate, $R_{1W}$, was measured using a separate saturation-recovery experiment. Water was saturated for 2 s using a series of high power 150° flip angle pulses, separated by 50 ms delays. After saturation, a 1-ms gradient was applied at 40 G/cm, followed by a low power gradient (1 G/cm) to prevent radiation damping from incompletely dephased water magnetization, and the water recovery was monitored using a spy pulse as before. Water peak volumes were measured for recovery times of 50, 100, 250, 500, 1000, 2000, and 4000 ms. An exponential function was fit to the recovery data to determine $T_{1}$, which was found to be approximately 3 s at all of the pH values measured. This value falls within experimental error from the rate determined in a selective water inversion recovery experiment, where a 1 G/cm gradient is used during the delay between inversion and readout, indicating radiation damping is effectively suppressed by the 1 G/cm gradient.

Data analysis

Spectral data were processed using the NMRPipe package.$^{64}$ Peak picking was performed in NMRDraw, and assignments were transferred using internal scripts. All data sets at any given pH were merged in to a single pseudo-3D spectrum, with the reference spectrum as the first plane. Each additional plane corresponded to a difference spectrum ($M^{+}$-$M^{-}$) for each value of $T$. Intensities for each peak were determined using the seriesTab program in NMRPipe, which accurately measures peak intensities without shifting peak positions.$^{29}$

For each residue, Eq. (1) was fit to the resulting list of peak intensities using the $R$ statistics package,$^{65}$ optimizing the sum of squared residuals (SSRs). Uncertainties were estimated using 1000 Monte-Carlo simulations. For each simulation, the original data were resampled, using the residuals to the best fit as the standard deviation for each point. The resampled data were then used to calculate new fit parameters and a new SSR for each simulation. The mean SSR was then determined for the Monte-
Carlo simulations. If this value was significantly different from the original best fit SSR (by more than 10%), the standard deviations for each data point were scaled uniformly and the simulations were repeated. The result is a set of Monte-Carlo resampled data where the Monte-Carlo SSR was similar to the original best-fit SSR. The uncertainties for the parameters $\sigma_{\text{knoe}}$ and $\sigma_{\text{RI}}$ were then taken as the standard deviation of the set of parameters from Monte-Carlo simulations.

To extract $k_{\text{knoe}}$ and $k_{\text{h2o}}$ at a reference pH of 6.4, Eq. (2) was fit to the $k_{\text{obs}}$ values recorded for each pH. Uncertainties on $k_{\text{h2o}}$ and $k_{\text{knoe}}$ again were estimated from 1000 Monte-Carlo simulations of the fits. As before, the standard deviation of the simulated distribution was used to estimate the uncertainty. The following residues displayed a large distribution resulted from $k_{\text{h2o}}$ and that $k_{\text{knoe}}$ was zero. As described in the text, the lower detection limit for $k_{\text{h2o}}$ at pH 6.4 was taken as 0.001. Fits to high pH data that correspond to rates of less than 0.001 generally yielded higher fractional errors and no significant pH dependence.

Acknowledgments

The authors are indebted to Nick Fawzi and Justin Lorieau for helpful discussions. This work was supported by the Intramural AIDS-Targeted Antiviral Program of the Office of the Director, NIH. NCF is the recipient of an Intramural AIDS Research and Dynamics Program of the Office of the Director, NIH. NIDDK, NIH, and by the Intramural AIDS-Targeted Research Program of the Office of the Director, NIH. NCF is the recipient of an Intramural AIDS Research Postdoctoral Fellowship.

References


58. Suzuki S, Urano E, Hashimoto C, Tsutsumi H, Nakahara T, Tanaka T, Nakashima Y, Maddali K, Han Y,


