



RESEARCH ARTICLE

The denatured state of HIV-1 protease under native conditions

Heike I. Rösner^{1,2} | Martina Caldarini³ | Gregory Potel⁴ | Daniel Malmodin² |
 Maria A. Vanoni⁵ | Alessandro Aliverti⁵ | Ricardo A. Broglio^{3,6} |
 Birthe B. Kragelund²  | Guido Tiana^{3,7} 

¹BRIC, University of Copenhagen, Copenhagen N, Denmark

²Structural Biology and NMR Laboratory (SBiNlab), Department of Biology, University of Copenhagen, Copenhagen N, Denmark

³Department of Physics, Università degli Studi di Milano and INFN, Milan, Italy

⁴Lawrence Livermore National Laboratory, Livermore, California, USA

⁵Dipartimento di Bioscienze, Università degli Studi di Milano, Milan, Italy

⁶Niels Bohr Institutet, University of Copenhagen, Copenhagen Ø, Denmark

⁷Center for Complexity and Biosystems, Università degli Studi di Milano, Milan, Italy

Correspondence

Birthe B. Kragelund, Structural Biology and NMR Laboratory (SBiNlab), Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, DK-2200 Copenhagen N, Denmark.
 Email: bbk@bio.ku.dk

Guido Tiana, Department of Physics and Center for Complexity and Biosystems, Università degli Studi di Milano, via Celoria 16, 20133 Milan, Italy.
 Email: guido.tiana@unimi.it

Funding information

Novo Nordisk UK Research Foundation, Grant/Award Number: NNF18OC0033926; U. S. Department of Energy, Grant/Award Number: DE-AC52-07NA27344; Università degli Studi di Milano, Grant/Award Number: RV_PSR_SEED_2019_GTIAN; Open Access Funding provided by Università degli Studi di Milano within the CRUI-CARE Agreement.

Abstract

The denatured state of several proteins has been shown to display transient structures that are relevant for folding, stability, and aggregation. To detect them by nuclear magnetic resonance (NMR) spectroscopy, the denatured state must be stabilized by chemical agents or changes in temperature. This makes the environment different from that experienced in biologically relevant processes. Using high-resolution heteronuclear NMR spectroscopy, we have characterized several denatured states of a monomeric variant of HIV-1 protease, which is natively structured in water, induced by different concentrations of urea, guanidinium chloride, and acetic acid. We have extrapolated the chemical shifts and the relaxation parameters to the denaturant-free denatured state at native conditions, showing that they converge to the same values. Subsequently, we characterized the conformational properties of this biologically relevant denatured state under native conditions by advanced molecular dynamics simulations and validated the results by comparison to experimental data. We show that the denatured state of HIV-1 protease under native conditions displays rich patterns of transient native and non-native structures, which could be of relevance to its guidance through a complex folding process.

KEYWORDS

advanced molecular dynamics, denatured state, NMR

1 | INTRODUCTION

The denatured state D_0 that proteins populate transiently under native conditions¹ is important to determine their folding,² stability,³ aggregation,⁴ and misfolding;⁵ properties that can have direct

implication for disease states. Except for a few specific proteins,⁶⁻⁸ D_0 is so poorly populated that it escapes experimental observation. To overcome this problem, induced denatured states can be stabilized by chemical agents like urea, guanidine hydrochloride (GdmCl) or acids, populating the states D_{urea} , D_{GdmCl} , and D_{acid} , respectively; states that

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Proteins: Structure, Function, and Bioinformatics* published by Wiley Periodicals LLC.

are not necessarily similar to D_0 and which show variation among themselves. However, from a thermodynamic point of view, calorimetry experiments⁹ showed that the unfolding enthalpy of lysozyme, denatured by acid, GdmCl, and temperature, is identical once the energy associated with the denaturant mean (e.g., the ionization energy in the case of pH) was subtracted. From these data, it was concluded that the states denatured by different means are thermodynamically indistinguishable.⁹

One could then ask whether the conformational properties of the different denatured states D_{urea} , D_{GdmCl} , D_{acid} , and D_0 are similar as well. Although these states were originally believed to be randomly disordered,¹⁰ recent studies have revealed them to contain transient secondary^{11–15} and even tertiary structures.^{16,17} Such results were made possible mainly thanks to the development of NMR techniques and in particular of secondary chemical shift analysis.

In the present work, we studied the denatured states of a monomeric variant of human immunodeficiency virus (HIV)-1 protease* (mHIV-1-PR_{1–95}), a protein necessary for HIV-1 to replicate in infected cells.¹⁸ The denatured state of HIV-1 protease under native conditions is particularly important because it was suggested as a possible target of antiretroviral drugs that prevent the correct folding of the protein and thus of its enzymatic activity.^{19–21} Moreover, the native conformation of mHIV-1-PR_{1–95} displays a topology, which is more complex than that of typical proteins of comparable size, a feature possibly encoded also in its denatured state. In fact, its native conformation displays two pseudo-knots and the associated Plaxco's contact order,²² quantifying the nonlocality of native contacts, is

15, much larger than the values 8–10 of typical proteins of comparable length.

HIV-1 protease is an aspartic acid protease, which in its active form exists as a homodimer²³ (Figure 1A). Analysis of its folding kinetics identified a monomeric intermediate that associates to form the native dimer structure.²⁴ Deletion of the last four C-terminal residues stabilizes a monomeric, fully folded form.²⁵ In fact, the native structure of this mHIV-1-PR_{1–95}, which predominantly contains β -sheet structure and a C-terminal α -helix,¹⁸ is highly similar to the structure in the dimer (cf. Figure 1B). Both the unfolding and refolding kinetics of mHIV-1-PR studied in urea by fluorescence display two time scales, suggesting the presence of at least one kinetic intermediate and the typical refolding time of mHIV-1-PR_{1–95} is of the order of a minute.²⁴ Also, mechanical unfolding experiments suggest the presence of folding and unfolding intermediates.²⁶ Interestingly, mHIV-1-PR was shown to display cold denaturation well above zero degrees Celsius,²⁷ a feature that allowed us to compare the denatured states D_{urea} , D_{GdmCl} , and D_{acid} to a further state D_{cold} .

The native and non-native states of the wild-type and of several variants of HIV-1-PR have also been characterized both *in silico* and *in vivo*.^{28–30} In spite of its central role as a target for anti-retroviral therapies, biochemical and biophysical data on HIV-1 protease are still limited. A tethered dimer in GdmCl,^{31,32} a wild-type dimer in acetic acid³³ and HIV-1-protease embedded in its viral precursor protein in urea³⁴ constitute some of these states. However, none of these studies were performed on the same variant of the protein, prohibiting a direct comparison of the results.

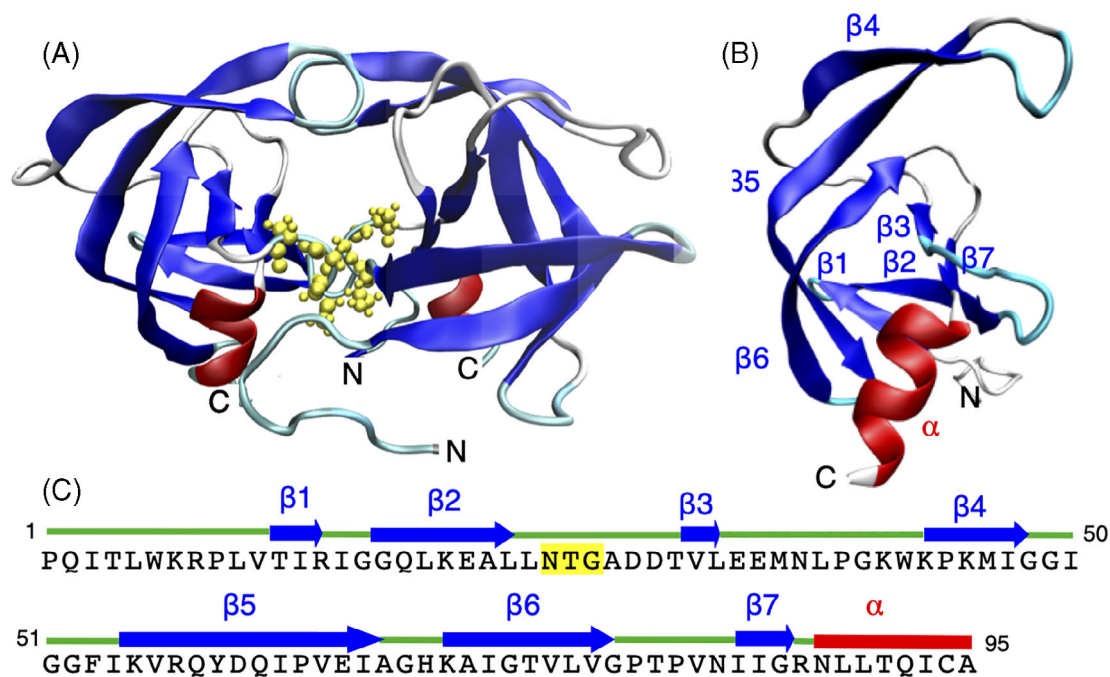


FIGURE 1 The native conformation of HIV-1-PR and mHIV-1-PR_{1–95}. (A) Structure of the HIV-1-PR homodimer (PDB code 1BVG). The active site is highlighted by yellow spheres. (B) Structure of the 1–95 variant mHIV-1-PR_{1–95} (PDB code 1Q9P). (C) The sequence of mHIV-1-PR; the active site is highlighted in yellow

In the present work, we have performed titration experiments in different chemical denaturants using the exact same mHIV-1-PR₁₋₉₅ variant and followed the changes by spectroscopy. This allowed us to monitor how the conformational properties of the denatured state depend on the kind and concentration of the denaturing agent, eventually extrapolating the properties of D_0 . The main quantity we investigated was the secondary chemical shifts, measured by heteronuclear NMR experiments. A nontrivial problem one has to face is then to interpret these data in terms of conformational properties of the protein. To assist us in this goal, we performed advanced molecular dynamics simulations of mHIV-1-PR₁₋₉₅ in water, building an ensemble of conformations that describes the denatured state D_0 . The correctness of the simulated D_0 was validated by back-calculating the secondary chemical shifts from the simulation and comparing them with those obtained from the extrapolation to zero denaturant of the NMR results.

2 | MATERIALS AND METHODS

2.1 | Protein expression, purification, and sample preparation

A synthetic gene encoding the HIV-1 protease monomer lacking the last 4 residues, mHIV-1-PR₁₋₉₅ was a kind gift from Dr. Celia Schiffer, University of Massachusetts Medical School, and was cloned into a pET11a vector. The protein was expressed in *Escherichia coli* Rosetta (DE3) cells upon induction with 0.2 mM IPTG. For the synthesis of isotope labeled protein, Spectra9 LB media (Euriso-top, France—cod. CGM-3030-CN-1, 1 L) enriched with ¹⁵N and ¹³C was used. Cells obtained from 0.4 L of culture were lysed by sonication at 4°C in extraction buffer: 20 mM Tris/HCl, 1 mM ethylenediaminetetraacetic acid (EDTA), and 10 mM 1,4-dithiothreitol (DTT), pH 8. The protein was refolded as described previously.²⁶ For spectroscopic measurements, the protein was dialyzed against 20 mM sodium phosphate, pH 6.0.

2.2 | Fluorescence and CD experiments

Fluorescence experiments were performed with a Varian Eclipse fluorimeter on 4 μM protein in 20 mM sodium phosphate at pH 6.0 and 25°C by adding different concentrations of denaturant. Circular dichroism (CD) measurements were conducted at 230 nm and a protein concentration of 15 μM in 20 mM sodium phosphate, pH 6, and containing different amounts of denaturant at 25°C using a JASCO J810 spectropolarimeter and a 1 mm path length. A total of 120 data points were recorded over 1 min and averaged. The actual urea and GdmCl concentrations were confirmed by refractive index measurements. For the temperature transition, CD measurements were conducted at 205 nm and a protein concentration of 10 μM in 20 mM sodium phosphate, pH 6. The temperature was increased in 1°C steps from 3°C to 20°C and in 2°C steps from 20°C to 90°C using a Peltier control unit. To account for the slow refolding kinetics, each point was allowed to equilibrate 5 min prior to detection.

2.3 | NMR experiments

2.3.1 | Backbone assignment and R_1 , R_2 , and hetNOE relaxation experiments

All NMR spectra were recorded either on an Agilent DD2 800 MHz or a Varian INOVA 750 MHz spectrometer using a room temperature probe, and standard pulse programs from the Vnmrj BioPack. For assignment, we prepared 11 different aliquots of ¹⁵N-¹³C-labeled ~200 μM protein solution in 20 mM sodium phosphate, pH 6.0, and 10% D₂O (v/v), 125 μM 2,2-dimethyl-2-silane-pentane-5-sulfonic acid (DSS) containing 4, 6, and 8 M urea, 0.75, 1, 2, and 4 M GdmCl, or 9% (v/v), 25% (v/v), and 45% (v/v) acetic acid, respectively, and one containing no extra additives. For relaxation experiments, identical samples were prepared containing a ¹⁵N-labeled ~200 μM protein solution. The backbone nuclei were assigned using HSQC,³⁵ HNCA, HNCO,³⁶ HN(CA)CO,³⁷ HNCOCA,³⁸ HNCACB,³⁹ CBCACONH,⁴⁰ HNN,⁴¹ and ¹⁵N-edited NOESY-HSQC⁴² spectra recorded at 25°C for the samples containing 4 and 8 M urea, and using only HSQC, HNCA, HNCO, HNCOCA, HNN, and ¹⁵N-edited NOESY-HSQC spectra for 1 M GdmCl and 25% (v/v) acetic acid. For the remaining samples, only the HSQC, HNCA, HNCOCA, and HNCO spectra were used for backbone assignment. The assignment was completed for 95% of all non-proline residues for samples containing acetic acid, 96% for samples containing GdmCl, 97% for samples containing urea, and 97% for cold denatured protein.

To analyze the T_1 and T_2 relaxation times and heteronuclear nuclear Overhauser effects (hetNOEs), five series of spectra were recorded on ¹⁵N-labeled protein in 20 mM sodium phosphate, pH 6.0, and 10% D₂O (v/v), 125 μM DSS, also containing 4 or 8 M urea, 1 M GdmCl or 25% (v/v) acetic acid, at 25°C.⁴³ We chose eight different delay times: 0, 100, 200, 300, 500, 700, 900, and 1200 ms for recording T_1 and nine different delay times: 10, 50, 90, 130, 170, 190, 210, 230, and 250 ms for recording T_2 relaxation times. For the hetNOE a relaxation delay of 8 s was used.

2.3.2 | Pulsed-field-gradient NMR diffusion experiments

The above-described protein samples were used to record sets of 60 bipolar pulse-pair stimulated echo experiments using a watergate scheme for water suppression with varying gradient strength.⁴⁴ As internal reference, 0.5% (v/v) dioxane was added to all samples to correct for viscosity effects by the solvent. All spectra were obtained at 25°C using 32 transients on a 750 MHz Varian INOVA spectrometer.

2.3.3 | 2-D and 3-D NMR spectra processing

The X-carrier frequency was determined by referencing to internal DSS. The DSS frequency was obtained from a 1D ¹H spectra recorded immediately before the remaining experiments. Indirect referencing

was used in the ^{15}N and ^{13}C dimensions by use of conversion factors.⁴⁵ The spectra were processed using nmrPipe⁴⁶ and qMDD.⁴⁷ Spectrometer frequencies and carrier frequencies in ppm were inserted with four decimals. Zero-filling to nearest power of 2 was used. The processed spectra were assigned and analyzed in CcpNmr Analysis.⁴⁸ The assigned HSQC spectra were further used to extract the relaxation decays from the series of spectra recorded to determine the T_1 and T_2 relaxation times. Relaxation decay curves were fitted to single exponentials and relaxation times determined using the *relax* software.^{49,50} The values of R_1 , R_2 , and the hetNOE recorded at 17.6 Tesla were used to derive the spectral density function at three frequencies (0, ω_H , and ω_N) analyzed by reduced spectral density mapping using *relax*.^{49,50}

2.3.4 | DOSY processing

Each set of 60 1D- ^1H spectra was separately processed and analyzed using The DOSY Toolbox⁵¹ and MATLAB.⁵² Spectra were phased in zero order and smoothed using a 10 Hz Lorentzian efficiently removing most visible noise. The function *msbackadj* was used rather than the internal DOSY Toolbox baseline correction routine.

2.3.5 | Analysis of the chemical shifts

Secondary chemical shifts associated with different atoms were systematized using the formula $(\Delta(\delta\text{C}^\alpha) + \Delta(\delta\text{C}^\beta) - 0.5 \cdot \Delta(\delta\text{N}))$.⁷

2.3.6 | Fit of dynamics parameters

The R_2 parameters were fitted with the function described in equation (3) in the supplementary materials of Reference 53. The fit was done with a nonlinear least-square algorithm employing a Levenberg-Marquardt algorithm. To avoid overfitting, we performed fits with different number of exponentials, eventually choosing the minimum number of exponentials which gave a χ^2 lower than 5.

NMR data have been deposited at the BioMagResBank with the accession number: 25255.

2.4 | Molecular dynamics simulations

The mHIV-1-PR₁₋₉₅ system was described with the Amber 99SBdisp force field⁵⁴ in Tip4/pd water and simulated with Gromacs 2020.4.⁵⁵ The protein was prepared in a dodecahedral box of 571 nm³ with 19160 water molecules and 4 Cl⁻ ions to neutralize the charge. A preliminary simulation of 50 ns at 700 K and constant volume was carried out, followed by 100 ns at 300 K and 1 atm. From the latter simulation, 110 conformations were extracted to act as starting conformations of the production run. A replica-exchange simulation was then performed with 110 replicas whose temperature range from 300 to 500 K for a total of 68 μs .

Once the first 30 ns were removed, the replica at 300 K was analyzed to validate the simulation against the NMR data. Secondary chemical shifts were calculated for each conformation with Sparta+⁵⁶ and averaged over all of them. To calculate secondary chemical shifts, we used Bax's reference values.⁵⁶

To predict the R_1 relaxation parameters qualitatively, we extracted 50 conformations from the 300 K trajectory, using each of them as starting point of a 1 ns simulation at fixed temperature. The root mean square fluctuations (RMSF) around each of the 50 average conformations were calculated and then averaged together. The experimental R_2 values were compared to the solvent-accessible surface area (SASA) of each residue, averaged over the full 300 K trajectory.

The clustering of the 300 K trajectory was performed with a tailor-made Python code that uses the fraction q of common contacts as underlying metric, normalized to the maximum between the numbers of contacts of the two structures. A contact is defined if the center of mass of two residues is closer than 0.65 nm. In the calculation of q , only pairs of residues which were further apart by at least three other residues along the chain were considered.

3 | RESULTS

3.1 | Denaturation of mHIV-1

Following the far-UV CD spectra of folded mHIV-1-PR₁₋₉₅ and cold-denatured mHIV-1-PR₁₋₉₅ (Figure S1 in the SI), we observed nearly identical spectra over a wide range of wavelength spanning from 208 to 250 nm. This is due to the presence of dominating aromatic contributions in the far-UV region,⁵⁷ which result in an atypical CD spectrum of a β -sheet protein. To monitor the unfolding temperature of mHIV-1-PR₁₋₉₅, we therefore chose to record the mean residue ellipticity at 205 nm as a function of increasing temperature from 3 °C to 90 °C (Figure S2).

Besides cold denaturation occurring at 10 °C, already described in Reference 27, we observed heat denaturation with an apparent midpoint temperature T_m^{app} of approximately 50 °C and a third transition at ~ 80 °C, corresponding to the irreversible aggregation of the protein. Due to aggregation, the heat-denatured state was not considered for high-resolution NMR studies. Under all conditions explored, the native state was never fully populated and hence all equilibrium unfolding transitions could not be satisfactorily fitted to a standard equilibrium transition curve.

In the presence of increasing amounts of urea, mHIV-1-PR₁₋₉₅ showed a very broad transition indicative of a noncooperative unfolding (Figure 2). Interestingly, close to 2 M urea, the unfolding transition was more than 95% complete as judged from CD measurements, but not according to fluorescence emission. Thus, the data did not seem to agree with the expected behavior of a two-state unfolding mechanism. At protein concentrations as high as those used for the NMR experiments, mHIV-1-PR₁₋₉₅ showed strong visible aggregation making reliable measurements below 4 M urea

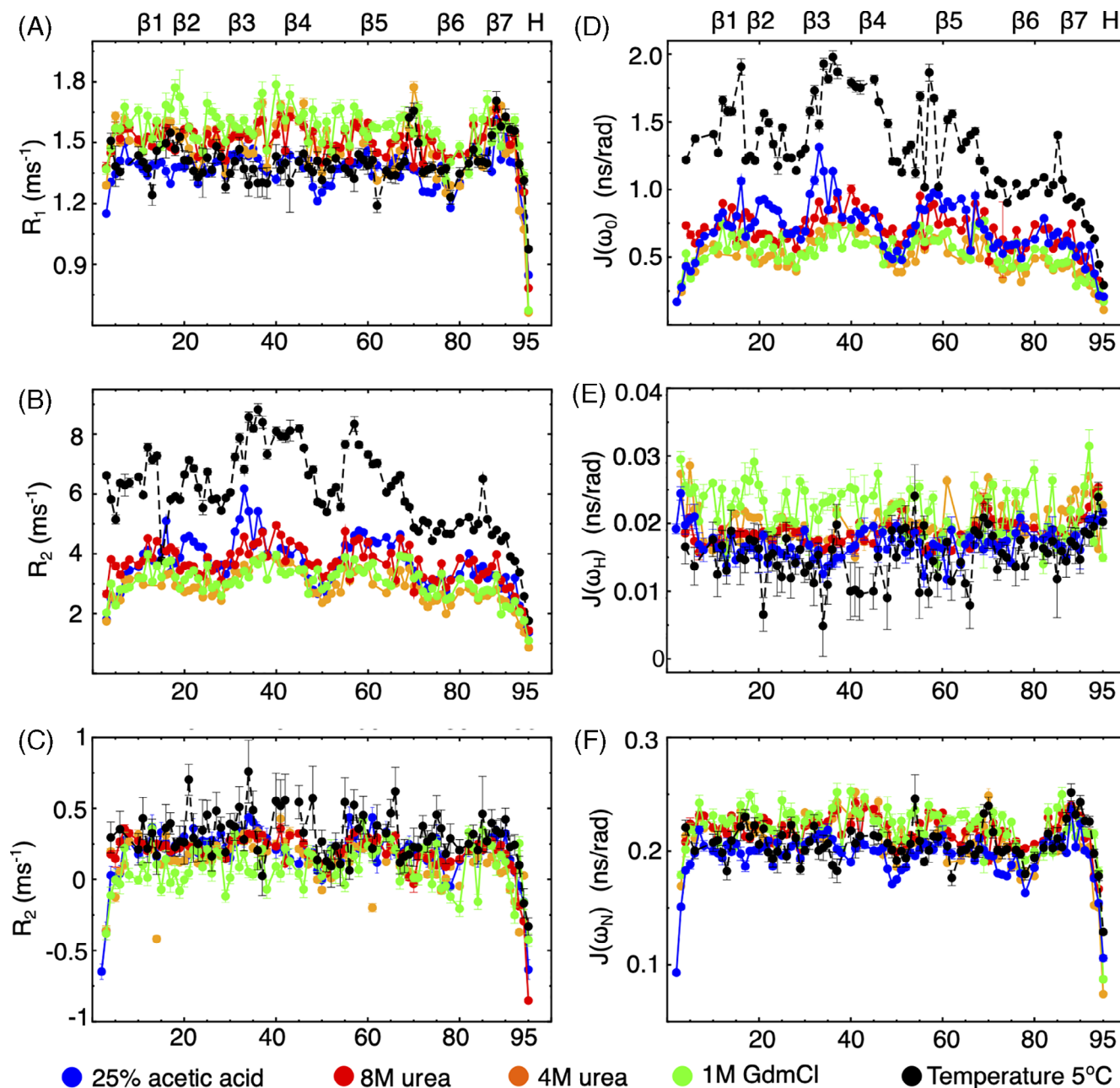


FIGURE 2 Equilibrium unfolding of HIV-1-PR₁₋₉₅. Top: mean residue ellipticity at 230 nm of a 15 μ M HIV1-PR₁₋₉₅ in 20 mM sodium phosphate, pH 6, measured in the presence of increasing concentrations of urea (left), GdmCl (middle) and acetic acid (right) at 25°C. Bottom: wavelength of maximum fluorescence emission of 4 μ M mHIV-1-PR₁₋₉₅ sample 20 mM sodium phosphate buffer, pH 6, measured in the presence of increasing concentrations of urea, GdmCl or acetic acid, at 25°C (excitation wavelength: 295 nm)

impossible. In all NMR experiments, the protein was >95% unfolded as judged from the CD signal. Monitoring the hydrodynamic radius R_h by pulsed-field-gradient (PFG) NMR experiments at 4 M urea showed that the hydrodynamic radius, $R_h = 27.2 \pm 0.5 \text{ \AA}$, was comparable to data in Reference 58. However, when increasing the urea concentration from 4 to 8 M urea, mHIV-1-PR₁₋₉₅ underwent further expansion from 27.2 ± 0.5 to $28.0 \pm 0.6 \text{ \AA}$ (Table 1).

Compared to urea denaturation, the equilibrium transition curve was steeper and appeared more cooperative using GdmCl. The secondary structure of mHIV-1-PR₁₋₉₅ had already fully disappeared in the

presence of less than 0.5 M GdmCl as monitored by CD (Figure 2). Again, fluorescence emission indicated mHIV-1-PR₁₋₉₅ to be >95% unfolded at a much higher concentration of denaturant than for CD, indicating that the monomer did not follow a two-state unfolding mechanism. At a denaturant concentration below 0.75 M GdmCl, protein aggregation was observed and NMR experiments were only recorded when more than 95% of the protein was denatured. Similar to the case in urea, the R_h increased with increasing concentration of GdmCl. For three selected samples, the R_h increased from $24 \pm 0.5 \text{ \AA}$ at 0.75 M GdmCl to $26.2 \pm 0.5 \text{ \AA}$ at 2 M GdmCl (Table 1).

TABLE 1 R_h of mHIV-1-PR₁₋₉₅ measured in different denaturants by PFG-NMR

| Urea | | GdmCl | | Acetic acid | |
|----------|-----------|-----------|-----------|---------------|----------------|
| c (urea) | R_h (Å) | c (GdmCl) | R_h (Å) | % acetic acid | R_h (Å) |
| 4 M | 27.2 | 0.75 M | 24.0 | 9 | 27.2 |
| 6 M | 27.7 | 1 M | 25.8 | 25 | 29.8 |
| 8 M | 28.0 | 2 M | 26.4 | 45 | — ^a |

Notes: All measurements were performed in 90% H₂O/10% D₂O at 25°C. The experimental uncertainty was estimated to $\pm 2\%$ based on triplicate measurements.

^aThe fitted data to this measurement were not available due to overlap with signals arising from acetic acid.

The acid denatured state appears crucial for successful refolding of the dimeric protein⁵⁹ and changes in protonation states can result in small but distinct differences in the preferences for local structure. The addition of just 0.1% acetic acid to 20 mM sodium phosphate, pH 6, caused the pH of the sample to drop to 4, and in an identical buffer containing 0.75% acetic acid, pH was 3.4. Addition of 5% acetic acid or more decreased pH below 3, where dimeric HIV-1-PR is reported to be largely unfolded.⁵⁹ We observed a midpoint of denaturation at about 0.5% acetic acid, which corresponded to a measured pH of 3.6. From CD experiments, further addition of acetic acid caused additional structural changes even when full acid denaturation was complete, when judged from fluorescence emission spectra (Figure 2). In addition, we observed an increase of R_h from 27.2 Å \pm 0.5 at 9% acetic acid to 29.8 \pm 0.6 Å at 25% acetic acid (Table 1). This increase was significantly larger than for the other two denaturants.

Interestingly, in the absence of denaturant, mHIV-1-PR₁₋₉₅ is folded except for the N-terminal region.¹⁸ In addition, the wild-type protein folds through a monomeric phase before dimerization.^{24,30,60,61} Inspection of the HSQC spectrum of mHIV-1-PR₁₋₉₅ recorded in 20 mM sodium phosphate (pH 6.0) at 25°C revealed a small but nondisputable second population. Under these experimental conditions, the folding rate of the monomer²⁴ is about 1 min⁻¹, the equilibrium thus being in the slower regime of chemical exchange for NMR experiments. Hence, the second set of peaks most likely originated from the denatured state D_0 .

3.2 | Chemical shift analysis

For each type of denaturant, the heteronuclear backbone resonances were assigned at three different denaturant concentrations (Figure S3 in the SI). Moreover, the cold denatured state described in Reference 27 was taken into account. The secondary chemical shifts of the protein in different denaturation conditions were rather similar to each other (Figure S4).

To describe the transient structures in the denatured state under nondenaturing conditions, C α , C', N, and H^N chemical shifts from individual titration series were extrapolated to the low intensity peaks

observed at zero denaturant, as described in Figure 3. As a result, in 16 cases, the weak cross peaks observed at the position defined by the extrapolated values could be unambiguously assigned in the set of spectra recorded at physiological conditions at 25°C in the absence of any denaturant. For these 16 cases, the assignment by extrapolation was cross-checked and confirmed by 3D backbone spectra. The same extrapolation procedure was applied to all residues and the remaining plots are shown in Figure S3 together with nine of the identified cross peaks in the HSQCs.

In Figure 4, we report the extrapolated secondary chemical shifts for the C α , C', N, and H^N backbone atoms averaged over the chemical shifts obtained from the four different extrapolations under different denaturing conditions. In these plots, we make use of intrinsic reference (i.e., the chemical shifts at highest denaturant concentration), although other choices gave similar results (see Figure S5 in the SI). The error bars indicate the associated SE and quantify the precision of the assignment under native conditions. Most residues displayed small errors compared to the average. A few discrepancies were observed for charged residues. The largest deviations were associated with the titration with acetic acid and were observed for three aspartic acids, D29, D30, and D60, and the single histidine, H69. Weaker effects were seen for four glutamates, E21, E34, E35, and E65. All the fits are displayed in Figure S6 in the SI.

3.3 | Polypeptide chain dynamics

We next measured ¹⁵N spin-lattice/spin-spin relaxation rates as well as heteronuclear NOEs for the mHIV-1-PR₁₋₉₅ at 25°C, at a field strength of 17.6 T (750 MHz). These relaxation parameters are sensitive towards motions on the subnanosecond timescale. In addition, the R_2 relaxation rate provides insights into motions on the millisecond to microsecond timescale. Full sets of relaxation data could be extracted for a total of 79 (5°C), 82 (4 M urea), 85 (8 M urea), 87 (1 M GdmCl), and 88 (25% acetic acid) residues, respectively.

For all five denatured states, the R_1 values remained more or less constant throughout the sequence, with average values of 1.44 \pm 0.04 (5°C), 1.48 \pm 0.01 (4 M urea), 1.50 \pm 0.01 (8 M urea), 1.56 \pm 0.03 (1 M GdmCl), and 1.36 \pm 0.02 ms⁻¹ (25% acetic acid), respectively (Figure 5). The N- and the C-termini showed lower R_1 compared to the rest of the protein, consistent with faster timescale movements usually experienced for chain termini. In all five profiles, we observed a stretch (V77–V82) of significantly lower values followed by a stretch (I84–L89) of significantly increased values. The average R_1 rates for the acetic acid and the cold denatured states were clearly reduced compared to those associated with the other two denatured states.

Measurements of the heteronuclear steady-state NOEs showed mostly positive values apart from those associated with the N- and C-terminal regions. The profile of the heteronuclear NOE did not agree with a fully unfolded state, but rather followed a profile of four arcs for all four denatured states.

The R_2 value is usually the most informative parameter for denatured proteins as it can reveal regions that undergo chemical

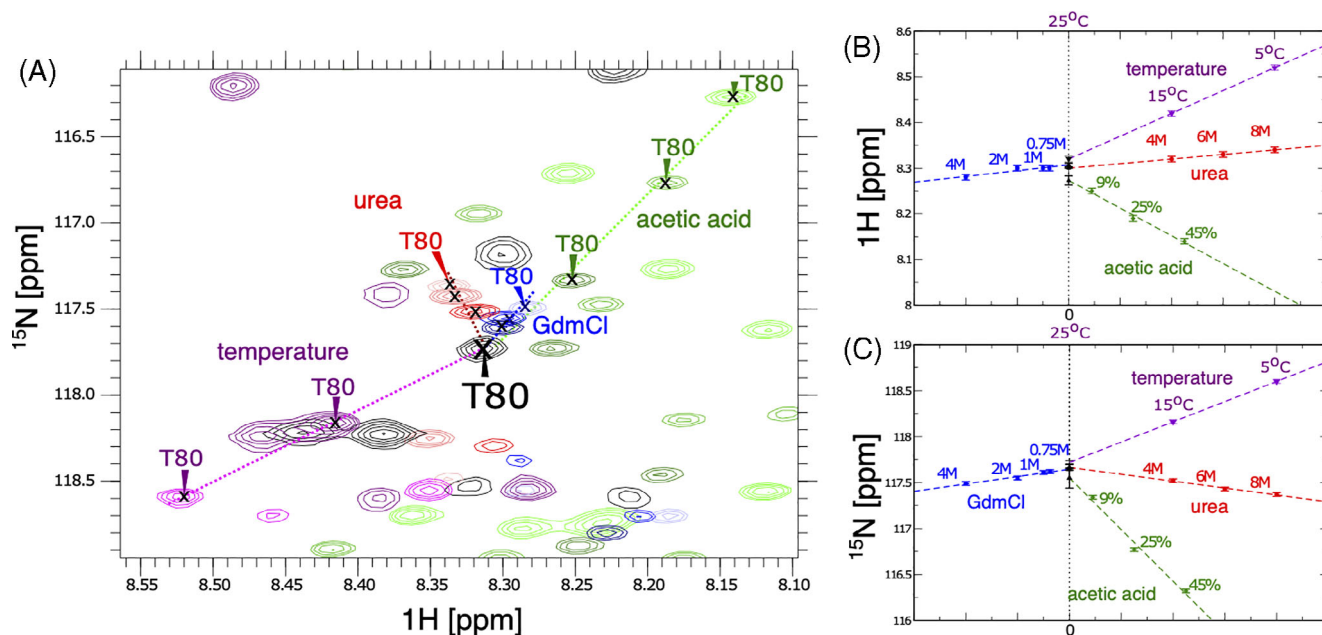


FIGURE 3 Convergence to the denatured state D_0 . (A) ^{15}N -HSQC spectra (zoom) of mHIV-1-PR $_{1-95}$ showing T80 in 12 different ^{15}N -HSQC spectra recorded under various denaturing conditions. All peaks converge to the same position for T80 in D_0 . Red peaks: urea 8 M (light red), urea 6 M (medium red) and urea 4 M (dark red). Green peaks: acetic acid 45% (light green), acetic acid 25% (medium green) and acetic acid 9% (dark green). Blue peaks: GdmCl 4 M (light blue), GdmCl 2 M (medium blue), GdmCl 1 M (dark blue). Purple peaks: sodium phosphate 20 mM, pH 6 at 5°C (light violet), 15°C (dark violet). Black peaks: sodium phosphate 20 mM, pH 6 at 25°C (denatured state D_0). (B) Schematic of T80 ^1H chemical shifts of mHIV-1-PR as a function of denaturant concentrations; the x-values are annotated in the graph (red: urea; blue: GdmCl; green: acetic acid; purple: different temperatures; black: physiological condition at 25°C [D^{phys} state]). All ^1H chemical shifts converge to the D_0 state value. (C) Same for ^{15}N chemical shifts

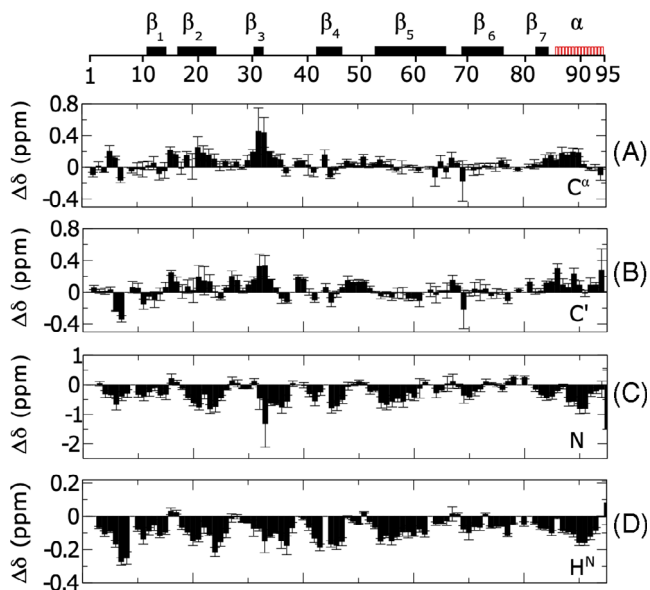


FIGURE 4 Secondary chemical shift analysis extrapolated for the D_0 state of mHIV-1-PR $_{1-95}$. For each residue, the chemical shifts under different experimental conditions were extrapolated to zero denaturant. Here we report the average of these extrapolations. The secondary chemical shifts were calculated by the use of the intrinsic random coil reference. The error bars indicate the total error of the procedure. Different nuclei were monitored (A) $\text{C}\alpha$, (B) C' , (C) N , and (D) H^{N} . The secondary structure of mHIV-1-PR $_{1-95}$ is shown at the top

exchange. For a fully extended protein where chain dynamics is dominated by unrestrained segmental motion, this profile usually adopts the shape of an inverted U, with a plateau along the chain and steep drops at the N- and C-terminal ends.⁶² For all five denatured states, the R_2 profiles deviated from an inverted U-shape. Instead, they displayed a four-arcs-like pattern distributed almost evenly over the sequence, and covering R8-L24, V32-G48, V56-G68, and G78-A95 (cf. Figure S7 in the SI), respectively. This unusual pattern of R_2 rates persisted at 8 M urea where the unfolded mHIV-1-PR $_{1-95}$ showed a more elongated conformation, as testified by the corresponding R_1 value (Table 1).

The probability function of finding motions at a given angular frequency ω can be described by the spectral density function $J(\omega)$. As unfolded states cannot be described in terms of an overall rotational correlation time, we instead chose to describe the relaxation data by reduced spectral density mapping.⁶³ We used the values of R_1 , R_2 , and the heteronuclear NOE recorded at 17.6 T to derive the spectral density function at three frequencies (0, ω_{H} , and ω_{N} , cf. Figure 5). Neither $J(\omega_{\text{H}})$ nor $J(\omega_{\text{N}})$ showed large variation in their profiles when plotted against the sequence, in agreement with the related profiles for the hetNOE and the R_1 values, respectively. Instead, the $J(0)$ values displayed the same pattern of four arcs as described for R_2 . Of importance, we note that the arches mostly revolve around prolines.

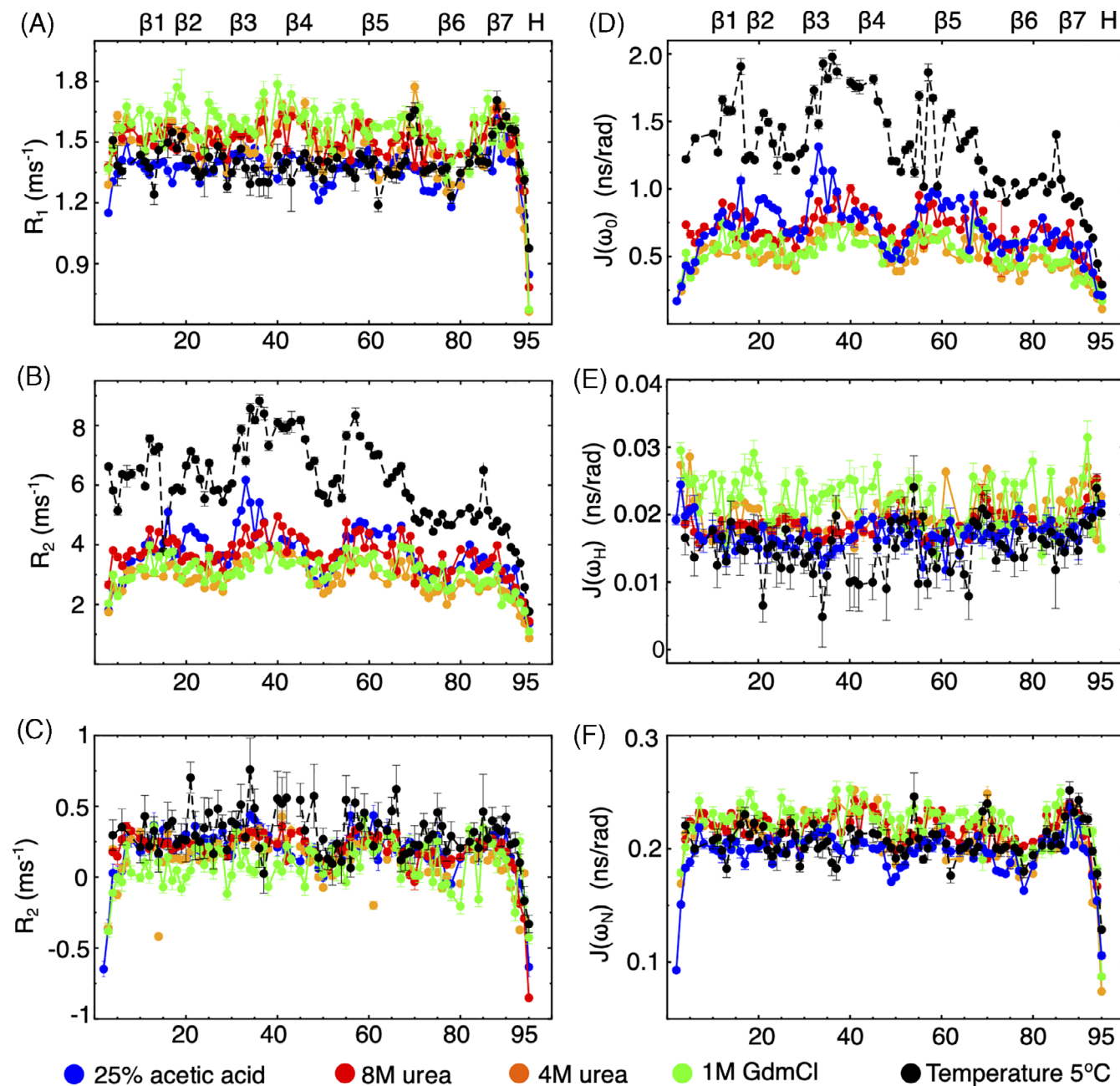


FIGURE 5 Chain dynamics in HIV-1-PR₁₋₉₅ at different denaturant conditions. NMR relaxation rates (A–C) and spectral densities (D–F) in 25% acetic acid (blue dots), in 8 M urea (red dots), in 4 M urea (orange dots), in 0.9 M GdmCl (green dots), and in buffer at 5°C (black dots), respectively. Secondary structures of folded HIV-1-PR₁₋₉₅ are shown in the top

3.4 | Experimental validation of molecular dynamics simulation of D_0

A replica-exchange simulation of 68 μ s of mHIV-1-PR₁₋₉₅ in water is performed with 110 temperatures in the range from 300 to 500 K, as described in the Materials and Methods. The degree of equilibration of the simulation seems acceptable, as testified by the good exchange between replicas (cf. Figure S8) and by the convergence of the average contact map (cf. Figure S9).

To validate the simulation, we calculated the average secondary chemical shifts from the simulated trajectory using Sparta⁵⁶ and compared them with the experimental values extrapolated for D_0 (Figure 6). The Pearson's correlation coefficients are $r = 0.68$ for CA, $r = 0.63$ for C', $r = 0.67$ for N, and $r = 0.54$ for HN (also cf. Figure S10). Thus, the simulated data are in good agreement with the experimental values ($p < 10^{-5}$, as calculated from a random bootstrap of the data).

The average radius of gyration, R_g , calculated from the simulated conformations, was 2.19 ± 0.48 nm. The corresponding R_h can be

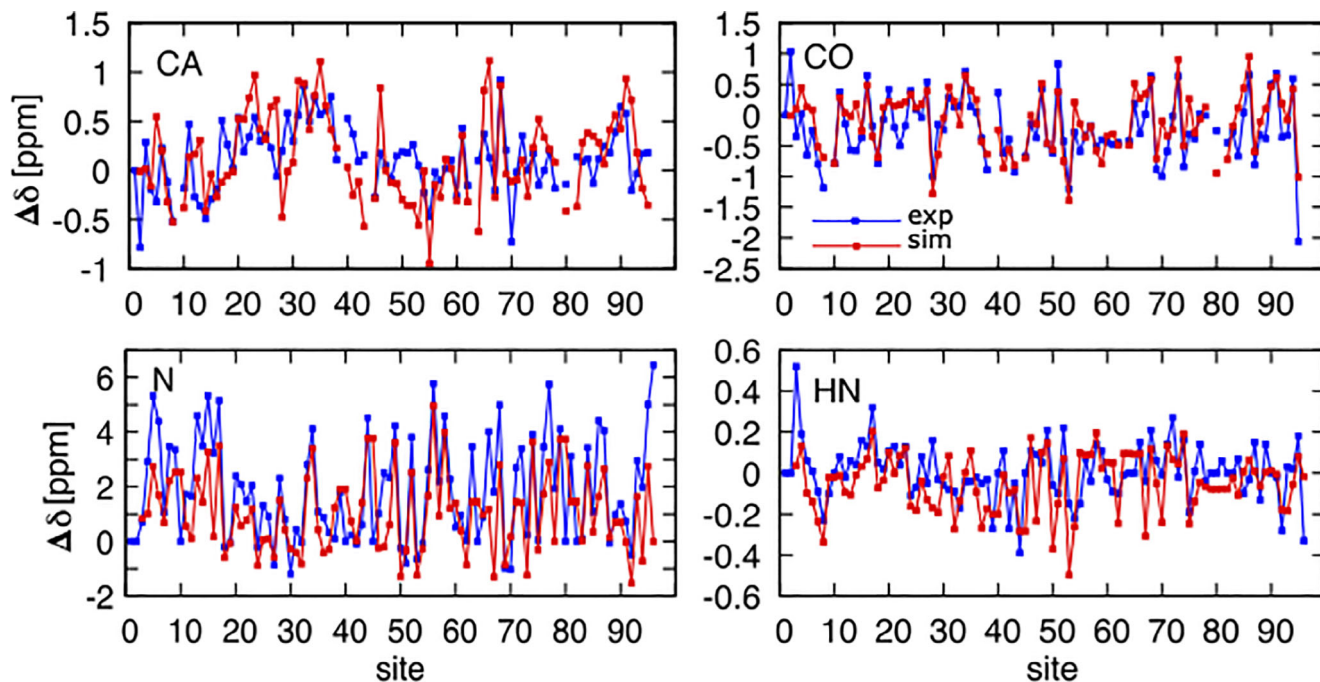


FIGURE 6 Secondary chemical shift comparison between experiment and simulation. A comparison between the average secondary chemical shifts of D_0 predicted by the MD simulation (red line) and those extrapolated from the experiment to zero denaturant (blue line)

estimated⁶⁴ to be 2.45 ± 0.61 nm. This is equal, within the error bars, to the hydrodynamic radius 2.51 ± 0.19 nm obtained as extrapolation to zero denaturant from the data of Table 1. Another, more qualitative comparison was done between the experimental and simulated relaxation parameters R_1 and R_2 . The reason why a direct comparison cannot be done is that a replica-exchange simulation is efficient in sampling the equilibrium conformations of the protein at the price of generating an unphysical time-dependent trajectory that would be necessary for calculating the NMR relaxation parameters.

To give an approximate estimate of R_1 from the simulation, we performed 20 plain-MD simulations at fixed temperature (300 K) starting from 20 conformations extracted from the replica-exchange trajectory. Each simulation lasted for 1 ns, that is the time scale described by the R_1 parameter. From each simulation, we calculated the RMSF around the average conformation. We expected that R_1 is anticorrelated with the RMSF. The comparison between the experimental R_1 and the (rescaled and shifted) RMSF is displayed in Figure 7. Although the linear correlation is not high ($r = 0.21$), 74% of points stay on the same side with respect to the median ($p = 10^{-8}$), suggesting the two curves indicate similar regions of rigid and flexible residues (black bars above the curves).

The values of R_2 that reflect the conformational freedom of residues on the μ s–ms timescale, were compared with the total SASA of each amino acid, calculated on the replica-exchange simulation. Again, the linear correlation is low ($r = 0.16$) but 68% of the points stay on the same side with respect to the median ($p = 10^{-4}$), indicating that residues that are experimentally more flexible are those less constrained by other parts of the polymer in the simulation.

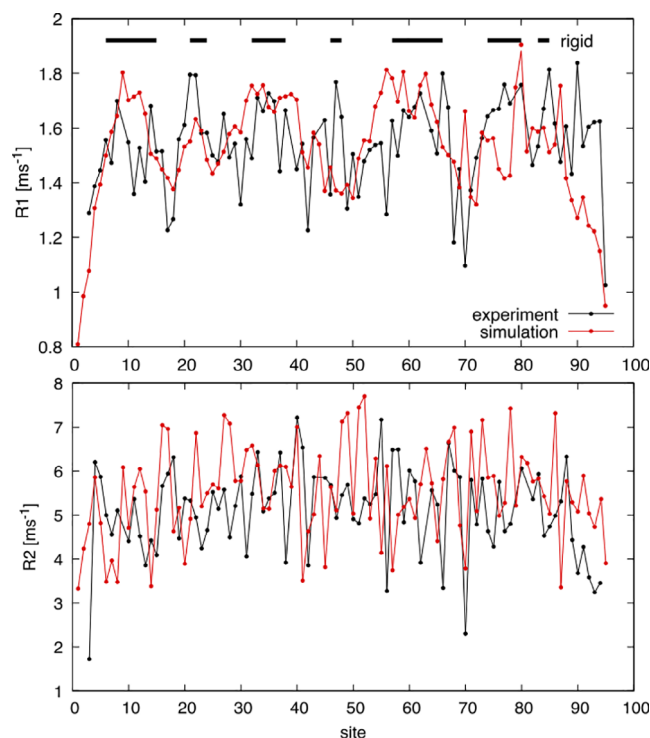


FIGURE 7 The experimental R_1 and R_2 relaxation parameters compared with proxies of the same quantities calculated from the MD simulations. In the case of R_1 , we plotted (in red) the function $(2.7 \cdot 2 \cdot \text{RMSF}/\text{nm})/\text{ms}$, where the RMSF is calculated from 1 ns MD simulations starting from the sampled conformations. In the case of R_2 , we used as proxy the function $(8.7 \cdot 3 \cdot \text{SASA}/\text{nm}^2)/\text{ms}$

3.5 | Analysis of the ensemble of conformations of D_0

The ensemble of conformations generated by the replica-exchange algorithm at 300 K was further analyzed to characterize D_0 . The average R_g of value 2.19 ± 0.48 nm (cf. Figure S11) was consistently larger than the value 1.28 nm of the native conformation. The fact that the contact probability between pairs of residues as a function of their distance along the chain is a power law with an exponent ≈ -1.8 (cf. Figure S12) suggests that the chain is, on average, in a coil state.

The average number of contacts is 60.6 ± 15.3 (cf. Figure S13) and the fraction q_N of native contacts is low ($\approx 0.018 \pm 0.011$). To be noted the fraction q_N of native contacts in the denatured state is poorly correlated with the commonly employed root mean square deviation (RMSD) (cf. Figure S14). In fact, the RMSD is a highly nonlinear function of the diversity between conformations in that it is very sensitive to conformational changes between similar

conformations and quite insensitive to large conformational changes between dissimilar conformations. Since the denatured state is expected to be conformationally very heterogeneous, we compared pairs of conformations using the fraction q of common contacts (and we compared a conformation to the native one using the fraction q_N of native contacts).

In Figure 8A, the distribution of common contacts q between denatured conformations is plotted. Its average is 0.09 ± 0.08 but it displays a tail up to 0.7. Not surprisingly, the denatured state D_0 thus appears conformationally very heterogeneous. However, its average contact map (cf. Figure 8B) displays well-defined secondary structures that can reach a probability of 0.4 and also tertiary structures populated with probabilities up to 0.15. Some of these structures are native-like and include the hairpin $\beta 1-\beta 2$, the hairpin $\beta 4-\beta 5$, the hairpin $\beta 5-\beta 6$, and the terminal α -helix (cf. Figure 8C). Non-native contacts (cf. Figure 8D) include a set of alternative structures in the region of the hairpin $\beta 4-\beta 5$, some fluctuating structure around P63 and a small amount of tertiary contacts.

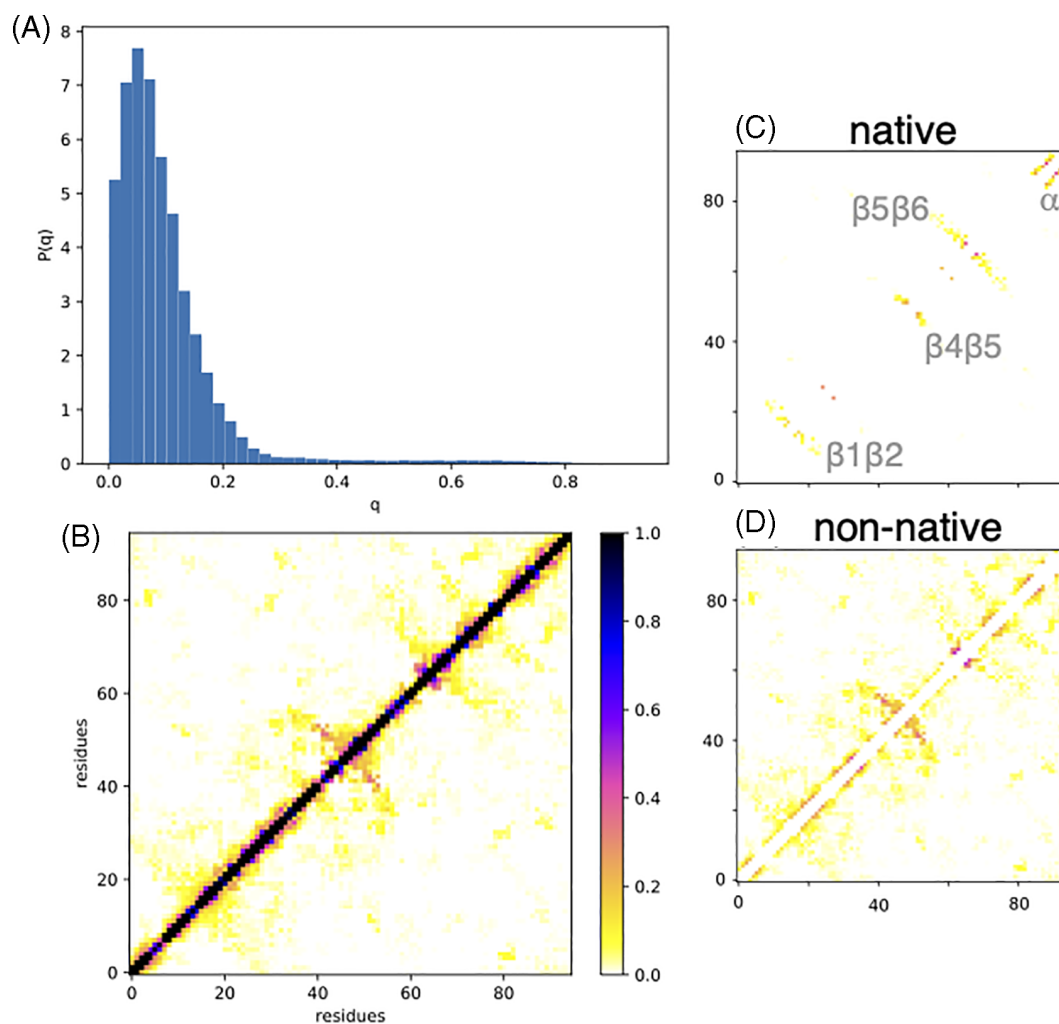


FIGURE 8 Equilibrium properties of the simulated protein (A) the distribution of similarity q between the conformations of D_0 , (B) the average contact map of mHIV-1-PR₁₋₉₅, (C) the average contact map limited to native contacts, (D) the average contact map limited to non-native contacts

A cluster analysis was performed for the simulated conformations of D_0 at 300 K with the Ward algorithm. The fraction q of common contacts was used as underlying metrics for the clustering instead of the more common RMSD because of the reasons described above. We could identify 17 clusters. In Figure 9, we displayed the three most populated clusters (others can be found in Figure S15). The most populated cluster (labeled A) has a population of 21% and is poorly structured; it contains most conformations with a low number of contacts. The only stable structure is a turn involving P63. Clusters B and C have a population of 8% each. Cluster B displays a non-native β -hairpin involving residues 39–45 and the native, C-terminal α -helix. Cluster C displays a β -hairpin involving residues 56–63, a β -turn 80–83, and tertiary contacts between this and the N-terminal region 4–6.

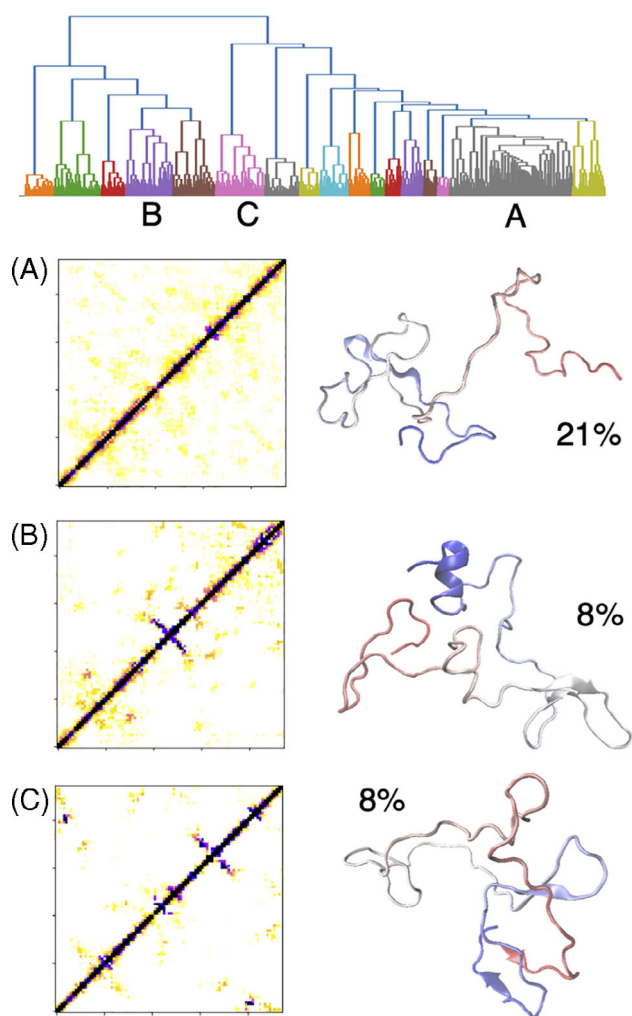


FIGURE 9 Cluster analysis of the conformations sampled at 300 K. In the dendrogram of structural similarity some clusters are indicated with Latin letters. For the three most populated clusters (labeled A, B, and C) the average contact map (normalized to the number of conformations of each cluster, same color code as Figure 8) and the central conformation (the N terminus in red, the C terminus in blue) are shown. The percental equilibrium population is also indicated for each cluster

In the other clusters (cf. Figure S15), particular recurrent non-native contacts in the region 40–50 and the native α -helix are seen.

4 | DISCUSSION

The denatured state D_0 of a protein under native solvent conditions is important to determine its behavior in the cell, but it is usually hard to characterize because of its intrinsic instability and low population. By studying the dependence of NMR observables as secondary chemical shifts and relaxation parameters under different denaturing conditions and extrapolating their values to native conditions, we could provide a conformational characterization of D_0 of the HIV-1-PR₁₋₉₅.

A remarkable result was that the extrapolations of these quantities to native conditions were rather independent on the denaturant and matched the minor population of D_0 present at native conditions.

In 1976, Pfeil and Privalov showed in a series of experiments⁹ that the unfolding enthalpy of lysozyme, denatured by pH, GdmCl, and temperature was identical, once the mean energy associated with the denaturant (e.g., the ionization energy in the case of pH) was subtracted. From this, they concluded that the states denatured by different means are thermodynamically indistinguishable. Ever since it has been discussed whether the denatured states generated by different means of denaturation were structurally different or not. In the present structural study, the extrapolation of chemical shifts to non-denaturing conditions plays a similar role to that of subtracting the denaturing energy in Privalov's experiment and all the extrapolations seem to agree very well with the presence of a single denatured state.

The interpretation of the raw data produced by NMR experiments in terms of conformational properties of D_0 is particularly difficult for a state composed of a plethora of heterogeneous conformations. In this case, MD simulations can be a valuable complement to the experimental data because of their ability to probe the system at atomic scale. A critical issue in this respect is whether MD simulations can provide a realistic picture of the state of interest of the protein. To address this concern, we compared the secondary chemical shifts, the hydrodynamic radius and the relaxation parameters predicted by the simulation with the experimental values. The good agreement we found is a consequence of two factors. First, we used advanced sampling techniques of simulation that favor the diffusion in the conformational space of the system, allowing it to sample a heterogeneous conformational space. Second, we employed a force field⁵⁴ that was particularly adjusted to simulate intrinsically-disordered proteins,⁵⁴ namely systems with conformational properties that are analogous to those of the denatured state of a structured protein. It is important to stress that the tools to analyze a simulation of the denatured state are different than those typically used for native-like states. For example, while the commonly used RMSD is a poor quantifier of the similarity between pairs of conformations with subtle common features, the fraction q of common contacts is a more sensitive tool.

The detection of transient native and non-native structures in the denatured state of proteins is important to understand their fast

folding^{65,66} and their aggregation.^{4,67} In the case of a viral protein as HIV-1-PR, such structures can also be relevant as targets of antiviral molecules.¹⁹⁻²¹ We found in D_0 specific secondary structures, both native and non-native, displaying an equilibrium probability of up to $\approx 30\%$ and also specific tertiary structures with equilibrium probability of up to $\approx 10\%$. Among them, the most stable elements seem to be the native C-terminal α -helix and a non-native β -like structure at the center of the protein. In particular, we observed transient populations of the hairpin $\beta 1$ – $\beta 2$, the hairpin $\beta 4$ – $\beta 5$, the hairpin $\beta 5$ – $\beta 6$, and the terminal α -helix. Except for the C-terminal helix, the remaining structures correlate with the arches described by elevated R_2 values (Figure 4) and further suggest that the simulations are capturing the details of the ensemble. Furthermore, several substates of D_0 were identified by cluster analysis, each with peculiar conformational features, both native and non-native. Interestingly, we observed positive secondary chemical shift values for C^α and C' at few places along the chain (Figure 4), suggesting the presence of transient helicity in regions that in the native structure form β -strands. Thus, non-native interactions appear relevant to the unfolded state of HIV-1-PR and may play roles in guidance through the folding process.

5 | CONCLUSIONS

The combination of experimental NMR techniques and of advanced MD simulation allowed us to characterize the denatured state of a complex protein as the HIV-1 protease under native conditions. This state displays transient native and non-native secondary and tertiary structures which could be of key relevance for guidance through the complex folding process. The strategy we used for HIV-1 protease can be easily applied to other proteins of comparable length.

ACKNOWLEDGMENTS

This work was performed under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344. Heike I. Rösner and Birthe B. Kragelund were supported by the Carlsberg Foundation, Birthe B. Kragelund by the Novo Nordisk Foundation Challenge Program REPIN (#NNF18OC0033926), and Guido Tiana by the SEED program of the University of Milano (RV_PSR_SEED_2019_GTIAN). Open Access Funding provided by Università degli Studi di Milano within the CRUI-CARE Agreement.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ENDNOTE

This is exactly the same protein used in Reference 27, in spite of the unfortunate notation used in that reference.

PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1002/prot.26189>.

DATA AVAILABILITY STATEMENT

The NMR data are deposited to the BMRB database with access number 25255.

ORCID

Birthe B. Kragelund  <https://orcid.org/0000-0002-7454-1761>

Guido Tiana  <https://orcid.org/0000-0001-9868-1809>

REFERENCES

- Dill KA, Shortle D. Denatured states of proteins. *Annu Rev Biochem.* 1991;60:795-825.
- Morrone A, McCully ME, Bryan PN, et al. The denatured state dictates the topology of two proteins with almost identical sequence but different native structure and function. *J Biol Chem.* 2011;286:3863-3872.
- Shortle D. The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J.* 1996;10:27-34.
- Fawzi NL, Chubukov V, Clark LA, Brown S, Head-Gordon T. Influence of denatured and intermediate states of folding on protein aggregation. *Protein Sci.* 2005;14:993-1003.
- Hetz C, Chevet E, Harding HP. Targeting the unfolded protein response in disease. *Nat Rev Drug Discov.* 2013;12:703-719.
- Alexandrescu AT, Abeygunawardana C, Shortle D. Structure and dynamics of a denatured 131-residue fragment of staphylococcal nuclease: a heteronuclear NMR study. *Biochemistry.* 1994;33:1063-1072.
- Reed MAC, Jelinska C, Syson K, et al. The denatured state under native conditions: a non-native-like collapsed state of N-PGK. *J Mol Biol.* 2006;357:365-372.
- Mok YK, Kay CM, Kay LE, Forman-Kay J. NOE data demonstrating a compact unfolded state for an SH3 domain under non-denaturing conditions. *J Mol Biol.* 1999;289:619-638.
- Pfeil W, Privalov PL. Thermodynamic investigations of proteins. *Biophys Chem.* 1976;4:23-32.
- Tanford C, Kawahara K, Lapanje S. Proteins in 6 M guanidine hydrochloride: demonstration of random coil behavior. *J Biol Chem.* 1966;241:1921-1923.
- Shortle D, Ackerman MS. Persistence of native-like topology in a denatured protein in 8 M urea. *Science.* 2001;293:487-489.
- Rösner HI, Poulsen FM. Residue-specific description of non-native transient structures in the ensemble of acid-denatured structures of the all-beta protein c-src SH3. *Biochemistry.* 2010;49:3246-3253.
- Sari N, Alexander P, Bryan PN, Orban J. Structure and dynamics of an acid-denatured protein G mutant †. *Biochemistry.* 2000;39(5):965-977.
- Teilum K, Thormann T, Caterer NR, et al. Different secondary structure elements as scaffolds for protein folding transition states of two homologous four-helix bundles. *Proteins Struct Funct Genet.* 2005;59:80-90.
- Wang Y, Shortle D. Residual helical and turn structure in the denatured state of staphylococcal nuclease: analysis of peptide fragments. *Fold des.* 1997;2:93-100.
- Ackerman MS, Shortle D. Robustness of the long-range structure in denatured staphylococcal nuclease to changes in amino acid sequence. *Biochemistry.* 2002;41:13791-13797.
- Bruun SW, Iesmantavicius V, Danielsson J, Poulsen FM. Cooperative formation of native-like tertiary contacts in the ensemble of unfolded states of a four-helix protein. *Proc Natl Acad Sci U S A.* 2010;107:13306-13311.
- Ishima R, Torchia DA, Lynch SM, Gronenborn AM, Louis JM. Solution structure of the mature HIV-1 protease monomer: insight into the tertiary fold and stability of a precursor. *J Biol Chem.* 2003;278:43311-43319.

19. Broglia RA, Provasi D, Vasile F, Ottolina G, Longhi R, Tiana G. A folding inhibitor of the HIV-1 protease. *Proteins Struct Funct Genet*. 2006; 62(4):928-933.
20. Broglia RA, Levy Y, Tiana G. HIV-1 protease folding and the design of drugs which do not create resistance. 2008;18:60-66.
21. Kimura S, Broglia RA, Tiana G. Thermodynamics of strongly allosteric inhibition: a model study of HIV-1 protease. *Eur Biophys J*. 2012 Oct; 8:1-13.
22. Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol*. 1998;277:985-994.
23. Kuzmic P. Kinetic assay for HIV proteinase subunit dissociation. *Biochem Biophys Res Commun*. 1993;191:998-1003.
24. Noel AF, Bilsel O, Kundu A, Wu Y, Zitewitz JA, Matthews CR. The folding free-energy surface of HIV-1 protease: insights into the thermodynamic basis for resistance to inhibitors. *J Mol Biol*. 2009;387: 1002-1016.
25. Ishima R, Ghirlando R, Tözsér J, Gronenborn AM, Torchia DA, Louis JM. Folded monomer of HIV-1 protease. *J Biol Chem*. 2001; 276:49110-49116.
26. Caldarini M, Sonar P, Valpapuram I, et al. The complex folding behavior of HIV-1-protease monomer revealed by optical-tweezer single-molecule experiments and molecular dynamics simulations. *Biophys Chem*. 2014;195C:32-42.
27. Rösner HI, Caldarini M, Prestel A, et al. Cold denaturation of the HIV-1 protease monomer. *Biochemistry*. 2017;56:1029-1032.
28. Tomasselli AG, Heinrikson RL. Targeting the HIV-protease in AIDS therapy: a current clinical perspective. *Biochim Biophys Acta*. 2000; 1477:189-214.
29. Cecconi F, Micheletti C, Carloni P, Maritan A. Molecular dynamics studies on HIV-1 protease drug resistance and folding pathways. *Proteins Struct Funct Genet*. 2001;43:365-372.
30. Kimura S, Caldarini M, Broglia RA, Dokholyan NV, Tiana G. The maturation of HIV-1 protease precursor studied by discrete molecular dynamics. *Proteins Struct. Funct Genet*. 2014;82:633-639.
31. Bhavesh NS, Panchal SC, Mittal R, Hosur RV. NMR identification of local structural preferences in HIV-1 protease tethered heterodimer in 6 M guanidine hydrochloride. *FEBS Lett*. 2001;509:218-224.
32. Bhavesh NS, Sinha R, Mohan PMK, Hosur RV. NMR elucidation of early folding hierarchy in HIV-1 protease. *J Biol Chem*. 2003;278: 19980-19985.
33. Rout MK, Hosur RV. Fluctuating partially native-like topologies in the acid denatured ensemble of autolysis resistant HIV-1 protease. *Arch Biochem Biophys*. 2009;482:33-41.
34. Chatterjee A, Mridula P, Mishra RK, Mittal R, Hosur RV. Folding regulates autoprocessing of HIV-1 protease precursor. *J Biol Chem*. 2005; 280:11369-11378.
35. Kay L, Keifer P, Saarinen T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J Am Chem Soc*. 1992;114:10663-10665.
36. Kay LE, Ikura M, Tschudin R, Bax A. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J Magn Reson*. 1990;89:496-514.
37. Clubb RT, Thanabal V, Wagner G. A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue ¹HN, ¹⁵N, and ¹³C chemical shifts in ¹⁵N–¹³C-labelled proteins. *J Magn Reson*. 1992;97:213-217.
38. Bax A, Ikura M. An efficient 3D NMR technique for correlating the proton and ¹⁵N backbone amide resonances with the alpha-carbon of the preceding residue in uniformly ¹⁵N/¹³C enriched proteins. *J Biomol NMR*. 1991;1:99-104.
39. Wittekind M, Mueller L. HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. *J Magn Reson Ser B*. 1993;101:201-205.
40. Grzesiek S, Bax A. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J Am Chem Soc*. 1992;114:6291-6293.
41. Panchal SC, Bhavesh NS, Hosur RV. Improved 3D triple resonance experiments, HNN and HN(C)N, for ¹HN and ¹⁵N sequential correlations in (¹³C, ¹⁵N) labeled proteins: application to unfolded proteins. *J Biol Chem*. 2001;20:135-147.
42. Zhang O, Kay LE, Olivier JP, Forman-Kay JD. Backbone ¹H and ¹⁵N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. *J Biomol NMR*. 1994;4:845-858.
43. Kay LE, Torchia DA, Bax A. Backbone dynamics of proteins studied by ¹⁵N inverse-detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*. 1989;28(23):8972-8979.
44. Balayssac S, Delsuc M-A, Gilard V, Prigent Y, Malet-Martino M. Two-dimensional DOSY experiment with excitation sculpting water suppression for the analysis of natural and biological media. *J Magn Reson*. 2009;196:78-83.
45. Wishart DS, Bigam CG, Yao J, et al. ¹H, ¹³C and ¹⁵N chemical shift referencing in biomolecular NMR. *J Biomol NMR*. 1995;6:135-140.
46. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR*. 1995;6:277-293.
47. Orekhov VY, Jaravine VA. Analysis of non-uniformly sampled spectra with multi-dimensional decomposition. *Prog Nucl Magn Reson Spectrosc*. 2011;59:271-292.
48. Vranken WF, Boucher W, Stevens TJ, et al. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*. 2005;59:687-696.
49. D'Auvergne EJ, Gooley PR. Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces. *J Biomol NMR*. 2009; 40(2):107-119.
50. D'Auvergne EJ, Gooley PR. Optimisation of NMR dynamic models II. A new methodology for the dual optimisation of the model-free parameters and the Brownian rotational diffusion tensor. *J Biomol NMR*. 2009;40(2):121-133.
51. Nilsson M. The DOSY toolbox: a new tool for processing PFG NMR diffusion data. *J Magn Reson*. 2009;200:296-302.
52. MATLAB, The MathWorks, Inc.
53. Klein-Seetharaman J, Oikawa M, Grimshaw SB, et al. Long-range interactions within a nonnative protein. *Science*. 2002;295:1719-1722.
54. Robustelli P, Piana S, Shaw DE. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc Natl Acad Sci U S A*. 2018;115:E4758-E4766.
55. Páll S, Zhmurov A, Bauer P, et al. Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS. *J Chem Phys*. 2020;153:134110.
56. Shen Y, Bax A. SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J Biomol NMR*. 2010;48:13-22.
57. Dagil R, Knudsen MJ, Olsen JG, et al. The WSXWS motif in cytokine receptors is a molecular switch involved in receptor activation: insight from structures of the prolactin receptor. 2012;20:270-282.
58. Kogo H, Takeuchi K, Inoue H, Kihara H, Kojima M, Takahashi K. Urea-dependent unfolding of HIV-1 protease studied by circular dichroism and small-angle X-ray scattering. *Biochim Biophys Acta*. 2009;1794(1): 70-74.
59. Todd MJ, Semo N, Freire E. The structural stability of the HIV-1 protease. *J Mol Biol*. 1998;283:475-488.
60. Levy Y, Cafflisch A, Onuchic JN, Wolynes PG. The folding and dimerization of HIV-1 protease: evidence for a stable monomer from simulations. *J Mol Biol*. 2004;340:67-79.

61. Louis JM, Ishima R, Aniana A, Sayer JM. Revealing the dimer dissociation and existence of a folded monomer of the mature HIV-2 protease. *Protein Sci.* 2009;18:2442-2453.
62. Schwalbe H, Fiebig KM, Buck M, et al. Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. 1997;36:8977-8991.
63. Farrow N, Zhang O, Szabo A, Torchia D, Kay LE. Spectral density function mapping using ¹⁵N relaxation data exclusively. *J Biomol NMR.* 1995;6:153-162.
64. Nygaard M, Kragelund BB, Papaleo E, Lindorff-Larsen K. An efficient method for estimating the hydrodynamic radius of disordered protein conformations. *Biophys J.* 2017;113:550-557.
65. Karplus M, Weaver DL. Protein-folding dynamics. *Nature.* 1976;260:404-406.
66. Broglia R, Tiana G. Hierarchy of events in the folding of model proteins. *J Chem Phys.* 2001;114:7267-7273.
67. Broglia RA, Tiana G, Pasquali S, Roman HE, Vigezzi E. Folding and aggregation of designed proteins. *Proc Natl Acad Sci U S A.* 1998;95:12930-12933.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Rösner HI, Caldarini M, Potel G, et al. The denatured state of HIV-1 protease under native conditions. *Proteins.* 2022;90(1):96-109. <https://doi.org/10.1002/prot.26189>