The denatured state of HIV-1 protease under native conditions

Rösner, Heike I.; Caldarini, Martina; Potel, Gregory; Malmodin, Daniel; Vanoni, Maria A.; Aliverti, Alessandro; Broglia, Ricardo A.; Kragelund, Birthe B.; Tiana, Guido

Published in:
Proteins: Structure, Function, and Bioinformatics

DOI:
10.1002/prot.26189

Publication date:
2022

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
The denatured state of HIV-1 protease under native conditions

Heike I. Rösner1,2 | Martina Caldarini3 | Gregory Potel4 | Daniel Malmodin2 | Maria A. Vanoni5 | Alessandro Aliverti5 | Ricardo A. Broglia3,6 | Birthe B. Kragelund2 | Guido Tiana3,7

1BRIC, University of Copenhagen, Copenhagen N, Denmark
2Structural Biology and NMR Laboratory (SBNiNlab), 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

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 conditions is important to determine their folding, stability, aggregation, and misfolding properties that can have direct implication for disease states. Except for a few specific proteins, $D_0$ is so poorly populated that it escapes experimental observation. To overcome this problem, induced denatured states can be stabilized by chemical agents like urea, guanidinium hydrochloride (GdmCl) or acids, populating the states $D_{\text{urea}}, D_{\text{GdmCl}},$ and $D_{\text{acid}}$, 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.

Received: 10 June 2021 Revised: 21 July 2021 Accepted: 22 July 2021 DOI: 10.1002/prot.26189
are not necessarily similar to $D_0$ and which show variation among themselves. However, from a thermodynamic point of view, calorimetry experiments 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.

One could then ask whether the conformational properties of the different denatured states $D_{\text{urea}}$, $D_{\text{GdmCl}}$, $D_{\text{acid}}$, and $D_0$ are similar as well. Although these states were originally believed to be randomly disordered, recent studies have revealed them to contain transient secondary and even tertiary structures. 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-PR1–95), a protein necessary for HIV-1 to replicate in infected cells. 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. Moreover, the native conformation of mHIV-1-PR1–95 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, quantifying the nonlocality of native contacts, is 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 homodimer (Figure 1A). Analysis of its folding kinetics identified a monomeric intermediate that associates to form the native dimer structure. Deletion of the last four C-terminal residues stabilizes a monomeric, fully folded form. In fact, the native structure of this mHIV-1-PR1–95, which predominantly contains β-sheet structure and a C-terminal α-helix, 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-PR1–95 is of the order of a minute. Also, mechanical unfolding experiments suggest the presence of folding and unfolding intermediates. Interestingly, mHIV-1-PR was shown to display cold denaturation well above zero degrees Celsius, a feature that allowed us to compare the denatured states $D_{\text{urea}}$, $D_{\text{GdmCl}}$, and $D_{\text{acid}}$ to a further state $D_{\text{cold}}$.

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. 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, a wild-type dimer in acetic acid and HIV-1-protease embedded in its viral precursor protein in urea 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-PR1–95. (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-PR1–95 (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-PR$_{1-95}$ 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-PR$_{1-95}$ 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-PR$_{1-95}$, 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 $^{15}$N and $^{13}$C 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. 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 $^{15}$N-$^{13}$C-labeled ~200 μM protein solution in 20 mM sodium phosphate, pH 6.0, and 10% D$_2$O (v/v), 125 μM 2,2-dimethyl-2-silanepentane-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 $^{15}$N-labeled ~200 μM protein solution. The backbone nuclei were assigned using HSQC, HNCA, HNCO, HNCACB, HNCCD, and $^{15}$N-edited NOESY-HSQC spectra recorded at 25°C for the samples containing 4 and 8 M urea, and using only HSQC, HNCA, HNCO, HNCCD, HNCA, and $^{15}$N-edited NOESY-HSQC spectra for 1 M GdmCl and 25% (v/v) acetic acid. For the remaining samples, only the HSQC, HNCA, HNCCD, 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 $^{15}$N-labeled protein in 20 mM sodium phosphate, pH 6.0, and 10% D$_2$O (v/v), 125 μM DSS, also containing 4 or 8 M urea, 1 M GdmCl or 25% (v/v) acetic acid, at 25°C. 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. 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 $^1$H spectra recorded immediately before the remaining experiments. Indirect referencing
was used in the $^{15}$N and $^{13}$C dimensions by use of conversion factors.$^{55}$ The spectra were processed using nmrPipe$^{46}$ 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 $(\delta_C$, $\delta_N$, and $\delta_{\alpha})$ analyzed by reduced spectral density mapping using relax.$^{49,50}$

2.3.4 | DOSY processing

Each set of 60 1D-$^1$H spectra was separately processed and analyzed using The DOSY Toolbox$^{51}$ 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 $(\Delta(\delta_C) + \Delta(\delta_N) - 0.5*\Delta(\delta_{\alpha}))$.$^7$

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 chi$^2$ lower than 5.

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

2.4 | Molecular dynamics simulations

The mHIV-1-PR$_{1-95}$ system was described with the Amber 99SBdisp force field$^{26}$ in Tip4/pd water and simulated with Gromacs 2020.4.$^{53}$ The protein was prepared in a dodecahedral box of 571 nm$^3$ 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 $R_2$ 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-PR$_{1-95}$ and cold-denatured mHIV-1-PR$_{1-95}$ (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 $\beta$-sheet protein. To monitor the unfolding temperature of mHIV-1-PR$_{1-95}$, 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 $T_m$ 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-PR$_{1-95}$ 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-PR$_{1-95}$ showed strong visible aggregation making reliable measurements below 4 M urea.
impossible. In all NMR experiments, the protein was >95% unfolded as judged from the CD signal. Monitoring the hydrodynamic radius \( R_h \) by pulsed-filed-gradient (PFG) NMR experiments at 4 M urea showed that the hydrodynamic radius, \( R_h = 27.2 \pm 0.5 \) Å, 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 ± 0.6 Å (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 ± 0.5 Å at 0.75 M GdmCl to 26.2 ± 0.5 Å at 2 M GdmCl (Table 1).
TABLE 1  

<table>
<thead>
<tr>
<th>Urea</th>
<th>GdmCl</th>
<th>Acetic acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>c (urea)</td>
<td>R₀ (Å)</td>
<td>c (GdmCl)</td>
</tr>
<tr>
<td>4 M</td>
<td>27.2</td>
<td>0.75 M</td>
</tr>
<tr>
<td>6 M</td>
<td>27.7</td>
<td>1 M</td>
</tr>
<tr>
<td>8 M</td>
<td>28.0</td>
<td>2 M</td>
</tr>
</tbody>
</table>

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

*The 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 protein 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. 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 R₀ from 27.2 Å ± 0.5 at 9% acetic acid to 29.8 ± 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. In addition, the wild-type protein folds through a monomeric phase before dimerization. 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 monomer 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 D₀.

### 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 non-denaturing conditions, Cα, C', N, and HN chemical shifts from individual titration series were extrapolated to the low intensity peaks observed at zero denaturation, 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 Cα, C', N, and HN 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 R2 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 R₁ values remained more or less constant throughout the sequence, with average values of 1.44 ± 0.04 (5°C), 1.48 ± 0.01 (4 M urea), 1.50 ± 0.01 (8 M urea), 1.56 ± 0.03 (1 M GdmCl), and 1.36 ± 0.02 ms⁻¹ (25% acetic acid), respectively (Figure 5). The N- and the C-termini showed slower 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 R₁ 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 R₂ value is usually the most informative parameter for denatured proteins as it can reveal regions that undergo chemical
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. 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 R\textsubscript{8}-L\textsubscript{24}, V\textsubscript{32}-G\textsubscript{48}, V\textsubscript{56}-G\textsubscript{68}, and G\textsubscript{78}-A\textsubscript{95} (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\textsubscript{1-95} showed a more elongated conformation, as testified by the corresponding $R_h$ value (Table 1).

The probability function of finding motions at a given angular frequency $\omega$ 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. 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 ($0$, $\omega_\text{H}$, and $\omega_\text{N}$, 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.
3.4 Experimental validation of molecular dynamics simulation of $D_0$

A replica-exchange simulation of 68 μs 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 Sparta+ 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 $R_g$ can be
estimated 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 ($p = 10^{-5}$), 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 μs-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 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).

**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 $(2.7 - 2*RMSF/nm)/ms$, 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 $(8.7 - 3*SASA/nm^2)/ms$.
3.5 Analysis of the ensemble of conformations of D₀

The ensemble of conformations generated by the replica-exchange algorithm at 300 K was further analyzed to characterize D₀. The average $R_\text{g}$ 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 $\approx -1.8$ (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 $q_\text{N}$ of native contacts is low ($\approx 0.018 \pm 0.011$). To be noted the fraction $q_\text{N}$ 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 $q_\text{N}$ 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 D₀ 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 $\alpha$-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 D₀, (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 $\beta$-hairpin involving residues 39–45 and the native, C-terminal $\alpha$-helix. Cluster C displays a $\beta$-hairpin involving residues 56–63, a $\beta$-turn 80–83, and tertiary contacts between this and the N-terminal region 4–6.

In the other clusters (cf. Figure S15), particular recurrent non-native contacts in the region 40–50 and the native $\alpha$-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-PR$_{1-95}$.

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 experiments$^9$ 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 field$^{54}$ 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...
In the case of a viral protein as HIV-1-PR, such structures can also be relevant as targets of antiviral molecules.\textsuperscript{19-21} We found in D\textsubscript{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 \(\alpha\)-helix and a non-native \(\beta\)-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 \(\alpha\)-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\textsubscript{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\textsuperscript{\textprime} and C\textsuperscript{\textprime\prime} at few places along the chain (Figure 4), suggesting the presence of transient helicity in regions that in the native structure form \(\beta\)-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. Rösnér 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_PS1_SEED_2019_GTIAN), Open Access Funding provided by Universita 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.

REFERENCES

3. Shortle D. The denatured state (the other half of the folding equation) and its role in protein stability. FASEB J. 1996;10:27-34.
16. 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.

DATA AVAILABILITY STATEMENT

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


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