analytical chemistry

Article

Cite This: *Anal. Chem.* 2018, 90, 14355−14362

pubs.acs.org/ac

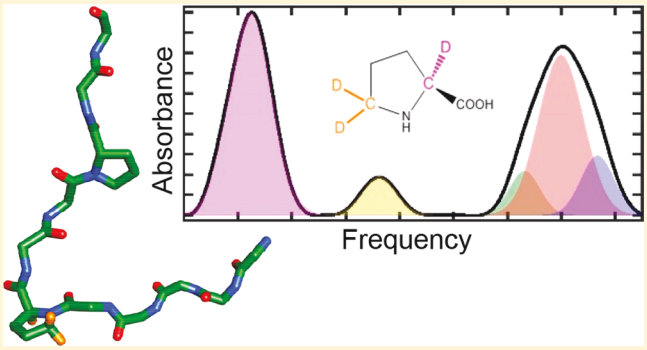
Residue-Specific Conformational Heterogeneity of Proline-Rich Sequences Uncovered via Infrared Spectroscopy

Gregory S. Bukowski and Megan C. Thielges*

Department of Chemistry, Indiana University, 800 East Kirkwood, Bloomington, Indiana 47405, United States

Supporting Information

ABSTRACT: Conformational heterogeneity is critical to understanding protein function but challenging to quantify. Experimental approaches that can provide sufficient temporal and spatial resolution to measure even rapidly interconverting states at specific locations in proteins are needed to fully elucidate the contribution of conformational heterogeneity and dynamics to function. Infrared spectroscopy in combination with the introduction of carbon deuterium bonds, which provide frequency-resolved probes of their environments, can uncover rapidly interconverting states with residue-specific detail. Using this approach, we quantify conformational heterogeneity of proline-rich peptides associated with different proline backbone conformations, as well as reveal their dependence on the sequence context.



The importance of conformational heterogeneity and its dynamics to a complete molecular description of protein function is now well-appreciated. For example, it can play critical roles in mechanisms of biological molecular recognition, catalysis, and allostery.[1−4] Furthermore, many proteins are intrinsically disordered, such that conformational dynamics are required to attain the functional form.[5,6] However, because proteins can adopt a wide range of conformations that interconvert on a broad range of time scales, quantifying conformational heterogeneity can be experimentally challenging. Investigating rapidly fluctuating states is particularly difficult because of the need for experimental approaches with sufficiently fast time scales. NMR spectroscopy, for instance, is among the most powerful approaches that has contributed immensely to our current understanding of protein conformational changes and their role in function;[2,4,7,8] however, it has an inherently slow (ms−μs) time scale. If conformational states interconvert more rapidly than this time scale, the resonances associated with the states coalescence, and thus the presence of the multiple states could remain undiscovered. The complex architecture of proteins creates a second challenge, as conformational heterogeneity could differ among locations in the structure. Thus, an experimental approach that can provide both high spatial and temporal resolution is needed to fully characterize the conformations of a protein and thereby help elucidate their potential contribution to biological function.

Infrared (IR) spectroscopy has an inherently fast time scale and the potential for high spatial resolution. The frequency of an IR absorption is sensitive to the chromophore environment, including local electrostatics, hydrogen bonding, and packing interactions.[9,10] The appearance of more than the expected number of absorption bands indicates that the chromophore experiences multiple environments, providing evidence for local conformational heterogeneity. Band coalescence due to rapid interconversion is not a significant concern for IR spectroscopy, as, for instance, conformers associated with differences in vibrational frequencies of several cm^{-1} would have to interconvert faster than the ps time scale to result in band coalescence. Unfortunately, massive spectral congestion hinders the application of IR spectroscopy to characterize specific sites in proteins. This issue has motivated extensive efforts to develop IR probe groups that can be incorporated into proteins and that absorb in a "transparent" frequency window of the protein's IR spectrum (∼1900−2300 cm^{-1}).[11] A number of such probe groups, such as cyano, azido, and metal−carbonyl-functionalized amino acids, have been explored. However, these are extrinsic probes that do not directly probe protein vibrations and that might themselves be perturbative. Whether the spectroscopy measures a native feature or one artificially introduced by the probe can be difficult to determine. Carbon deuterium (C−D) bond incorporation is nonperturbative, and although C−D bonds are less widely utilized because of their relatively weak absorptions, they are otherwise ideal IR probes of proteins.[12−14] Importantly, in contrast to extrinsic IR probes, C−D bonds provide intrinsic probes that report on vibrations of the protein itself and thus can provide a direct report of its conformations and dynamics.

Received: August 21, 2018
Accepted: November 21, 2018
Published: November 21, 2018

DOI: 10.1021/acs.analchem.8b03813
Anal. Chem. 2018, 90, 14355−14362

Previously, we characterized site-specifically incorporated C−D probes with IR spectroscopy to study the conformational heterogeneity and dynamics of a peptide containing the proline-rich (PR) recognition motif, P*xx*P (where *x* is typically a hydrophobic amino acid),[15] a motif recognized by Src homology 3 (SH3) domains.[16−18] This and other PR motifs are abundant in eukaryotic proteomes and are often found in intrinsically disordered regions, whose conformational dynamics are likely key to function.[5,6] The IR spectra of C−D probes introduced at the two conserved proline residues of the PR motif of the protein Pbs2 (pPbs2) showed multiple absorption bands for single vibrational modes, indicating that they experienced multiple distinct microenvironments; thus, the IR spectra provided evidence for peptide conformational heterogeneity. Interestingly, the same bonds in the peptide that showed multiple IR absorptions showed only single resonances when characterized via NMR spectroscopy, implying that the states resolved by IR spectroscopy are in rapid exchange on the NMR time scale. These results demonstrated how IR spectroscopy, because of its inherently faster time scale, makes possible direct detection of conformational heterogeneity that has been missed thus far because of the rapid dynamics. However, the relationship between the bands and molecular differences among the local environments of the C−D probes was not yet understood.

To more completely elucidate how the spectral signatures inform on the molecular environment of the C−D probes, and thereby further develop the approach for characterizing conformational dynamics, we have now extended our investigation of PR peptides to include a more systematic study. In addition to revisiting C−D probes previously characterized at three proline residues in the PR motif of pPbs2, we investigated C−D probes introduced at the fourth proline residue in the sequence (Table 1). We also extended

Table 1. Sequences of Peptides and Residue Numbering[a]

PR motif			P	*x*	*x*	P	
	P−2	P−1	P0	P1	P2	P3	
pNS5A							
H_2N-RKRR	**P**	**P**	**P**	I	**P**	**P**	A-Ac
pPbs2							
Ac-VNK	**P**	L	**P**	**P**	L	**P**	VA-NH_2
P(−2)A/P1A pPbs2							
Ac-VNK	A	L	**P**	A	L	**P**	VA-NH_2

[a]Based on consensus numbering for this PR motif.[21]

our efforts to elucidate the relationship between peptide sequence and conformational heterogeneity through characterization of a variant of pPbs2 in which the two nonmotif proline residues were mutated to alanine (P(−2)A/P1A pPbs2)[19] to examine the effects of increased conformational freedom. Finally, we investigated a third peptide that contains the PR recognition motif of the nonstructural protein 5A from hepatitis C virus (pNS5A).[20] Of the peptides, pNS5A has the greatest proline content, with five proline residues in close proximity in the sequence. Overall, we characterized C−D probes incorporated at a total of 11 proline residues. We observe multiple absorption bands for single vibrational modes, revealing widespread conformational heterogeneity. Moreover, we find that the C−D probes can be used to characterize the ensemble of backbone conformations in a residue-specific manner, enabling us to uncover sequence-dependent variation in the heterogeneity of local secondary structure.

EXPERIMENTAL SECTION

Sample Preparation. Peptides with sequences Ac-VNKPLPPLPVA-NH_2 (pPbs2), Ac-VNKALPALPVA-NH_2 (P(−2)A/P1A pPbs2), and Ac-APPIPPPRRKR-NH_2 (pNS5A) were synthesized by Fmoc (fluorenylmethoxycarbonyl) solid-phase peptide synthesis. They were acetylated and amidated at the N- and C-termini, respectively. Complete details of synthesis are provided in the Supporting Information. We prepared in total 14 peptides, one unlabeled variant of each peptide and variants with (C_α-*d*, C_δ-d_2)-proline incorporated at each of the proline residues in each peptide: P(−2), P0, P1, P3 in pPbs2; P0, P3 in P(−2)A/P1A pPbs2; and P(−2), P(−1), P0, P2, P3 in pNS5A.

Fourier Transform-Infrared (FT-IR) Spectroscopy. Peptide concentrations were determined by UV/vis absorbance spectroscopy (pPbs2 ε_{205} = 43318.56 M^{-1} cm^{-1}; P(−2)A/P1A pPbs2 ε_{205} = 40737.27 M^{-1} cm^{-1}; pNS5A ε_{220} = 17450 M^{-1} cm^{-1})[15,20] and adjusted to 4 mM in 50 mM phosphate, pH 7.0, 100 mM NaCl. FT-IR spectra were recorded using a dry $N_2(g)$-purged Agilent Cary 670 FT-IR spectrometer with a $N_2(l)$-cooled MCT detector. A 4-term Blackman Harris apodization function and zero-filling factor of 8 were applied to process all interferograms, which were averages of 8000 scans. IR absorption spectra of the C−D probes in each peptide were generated using transmission spectra of unlabeled and labeled peptide acquired under identical conditions. To correct for slowly varying residual background absorbance, a polynomial fit was then subtracted from the spectra. Each background-subtracted absorption spectrum was fit to a Gaussian function or sum of Gaussian functions to determine the number, relative intensities, center frequencies, and full width at half-maximum line widths of the absorption bands (see Supporting Information). The reported averages and standard deviations were obtained from analysis of spectra of independently prepared samples.

Circular Dichroism. Peptides were dissolved in sterile water and adjusted to a final concentration of 12 μM in 10 mM phosphate, 140 mM NaF, pH 7.0. Circular dichroism spectra were collected with a Jasco J-715 Circular Dichroism Spectropolarimeter from 185 to 250 nm in 0.5 nm steps at 25°C. All spectra were collected in a 0.1 cm quartz cuvette and are averages of 100 scans.

Molecular Dynamics Simulations. Molecular dynamics (MD) simulations were performed with the AMBER16[22] software package and implemented the ff14SB force-field[23,24] and TIP3P potential for water molecules.[25] The initial coordinates and parameters for the peptides were generated using xLeap from AmberTools16. The peptides were solvated in a cubic (10 × 10 × 10 $Å^3$) box. Periodic boundary conditions were used with a distance cutoff of 10 Å for nonbonded interactions. All bonds involving hydrogen atoms were constrained by the SHAKE algorithm.[26] The peptides were energy-minimized with 1000 steps of the steepest descent and 1000 subsequent steps of the conjugate gradient algorithms. The peptides were heated to 300 K over 30 ps with 0.5 fs steps, then equilibrated for 595 ps at 300 K, and controlled by a Langevin thermostat. Seven trajectories of 12 ns duration with 2 fs steps were run and analyzed with CPPTRAJ[27] from AmberTools16. Additional details are provided in the Supporting Information.

DOI: 10.1021/acs.analchem.8b03813
Anal. Chem. 2018, 90, 14355−14362

Density Functional Theory Calculations. Density functional theory (DFT) calculations for Ac-L(d_3)PA-NMe in the gas phase were performed using Spartanv16 software.[28] The B3LYP hybrid functional was implemented with the 6-31G* basis set. An energy profile was obtained by stepping the ψ angle from −80° to 180° in 2° increments, while the φ angle was constrained to −66°. Two energy minima were observed, and the IR frequencies of the C−D probes for the proline structures at the each minimum were calculated.

RESULTS AND DISCUSSION

Proline residues are known to promote the formation of polyproline II (PPII) secondary structure,[29,30] an extended left-handed helix with threefold symmetry characterized by backbone conformations with φ and ψ angles around −75° and 145°, respectively. The varying proline content of pPbs2, P(−2)A/P1A pPbs2, and pNS5A is expected to result in disparate secondary structures. To examine the overall secondary structure of the PR peptides, we synthesized each peptide via solid-phase peptide synthesis and characterized them via visible circular dichroism (CD) spectroscopy (Figure 1). The CD spectrum for polyproline, which is expected to

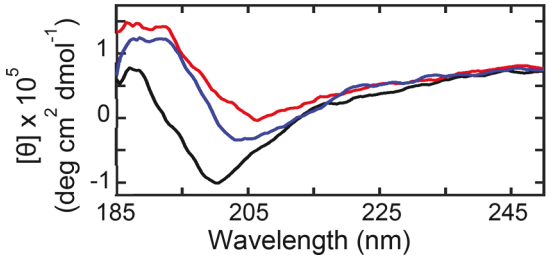


Figure 1. Circular dichroism spectra of pNS5A (blue), pPbs2 (red), and P(−2)A/P1A pPbs (black).

adopt PPII structure, is characterized by two positive maxima at 185 and 220 nm and a large negative band at 206 nm.[31] As anticipated, the CD spectra of the peptides showed variations in the shape and location of bands. Among the peptides, the CD spectrum for only pNS5A showed all three features of PPII structure. In contrast, the CD spectrum for P(−2)A/P1A pPbs2 showed a small maximum at 185 nm but a negative band at lower wavelength, closer to that characteristic of unfolded structure (197 nm).[31] While the CD spectrum for pPbs2 is similar to that determined for pNS5A, with positive amplitude near 185 nm and negative amplitude near 206 nm, it does not show the positive band at 220 nm characteristic of PPII structure. Thus, the CD spectrum for pNS5A, which showed the most features indicative of PPII structure, was consistent with the greatest PPII content. In addition, the substitution of the two proline residues in P(−2)A/P1A pPbs2 led to a CD spectrum more like that found for unfolded structure, indicating the disruption of PPII structure found in the native sequence.

We then synthesized each peptide with each proline individually deuterated (C_α-d, C_δ-d_2; Table 1 and Figure 2), and the C−D probes in each peptide were then characterized via IR spectroscopy (Figure 3 and Figure S-1). The spectra arise from three vibrational transitions. In each case, the absorption at lowest frequency is assigned to the $C_\delta D_2$ symmetric stretch and the absorption at highest frequency is assigned to the $C_\delta D_2$ asymmetric stretch, while the weak

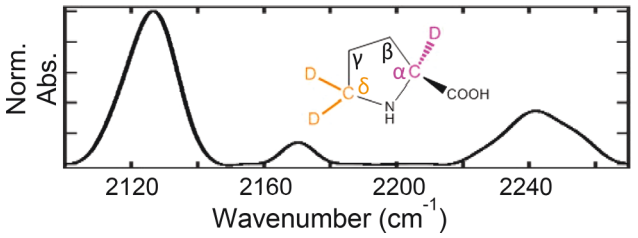


Figure 2. Structure of (C_α-d, C_δ-d_2)-proline showing location of $C_\delta D_2$ (orange) and $C_\alpha D$ (magenta) bonds and an example FT-IR spectrum for labeled residue P2 in pNS5A.

absorption at intermediate frequency is assigned to the $C_\alpha D$ stretch.[13] As in our previous study of pPbs2,[15] we noticed significant asymmetry in the $C_\delta D_2$ absorption band shapes that reflect the presence of multiple underlying bands.

Supporting this possibility, the second derivative spectra show multiple minima for many of the $C_\delta D_2$ absorptions (Figure S-2). Moreover, attempts to model the absorptions with a single Gaussian band resulted in systematic residuals among sets of replicate spectra and among differently labeled sites (Figures S-3 and S-4). We then fit the $C_\delta D_2$ absorptions to a superposition of Gaussian bands with the significance of including an additional component evaluated by F-tests (Supporting Information). This analysis yielded a consistent set of bands that well-modeled the spectra (Table 2 and Figure S-1). The absorption of the asymmetric stretch for all sites was well-modeled by a superposition of up to three bands separated from each other by 12−15 cm^{-1} with the band at intermediate frequency making the greatest contribution to the total absorbance (Figure 3A). The absorption of the symmetric stretch for all sites was well-modeled by the superposition of up to two bands: one at 2124−2126 cm^{-1} and a second at 2112−2114 cm^{-1} that contributed at most 15% of total absorbance (Figure 3B).

The IR spectra of the C−D probes reveal substantial differences in the environments experienced by the proline residues. Most significantly, we observed variation in the relative contributions to the total absorbance of the component bands of the $C_\delta D_2$ modes (Table 2; Figure 3). In addition, the frequencies of the asymmetric stretch bands were remarkably sensitive to their local microenvironment, varying among the residues by as much as 13 cm^{-1}. In comparison to the asymmetric stretch, the frequencies of the symmetric stretch bands were only moderately sensitive to the probe's location in the peptide, differing by only ∼2 cm^{-1} among the sites. Similarly, with the exception of P(−2)A/P1A pPbs2, the band for the $C_\alpha D$ stretch varied by only up to 2 cm^{-1}, although we note that previous computation studies have indicated that this mode could be utilized for probing the structure of amino acids.[32]

To gain insight into the origins of the different microenvironments probed by the C−D bonds, we performed MD simulations of each peptide. Proline is well-known to adopt two stable conformations, the α conformation, which corresponds to the structure found in right-handed α helices, and the β conformation, which is found in PPII secondary structure.[33,34] As expected from the high proline content, each peptide primarily adopted a PPII secondary structure showing backbone conformations with φ and ψ angles around −75° and 145°, respectively (Figure 4A,B and 5).[33,34] For the majority of the residues, we also observed a small population of

DOI: 10.1021/acs.analchem.8b03813
Anal. Chem. 2018, 90, 14355−14362

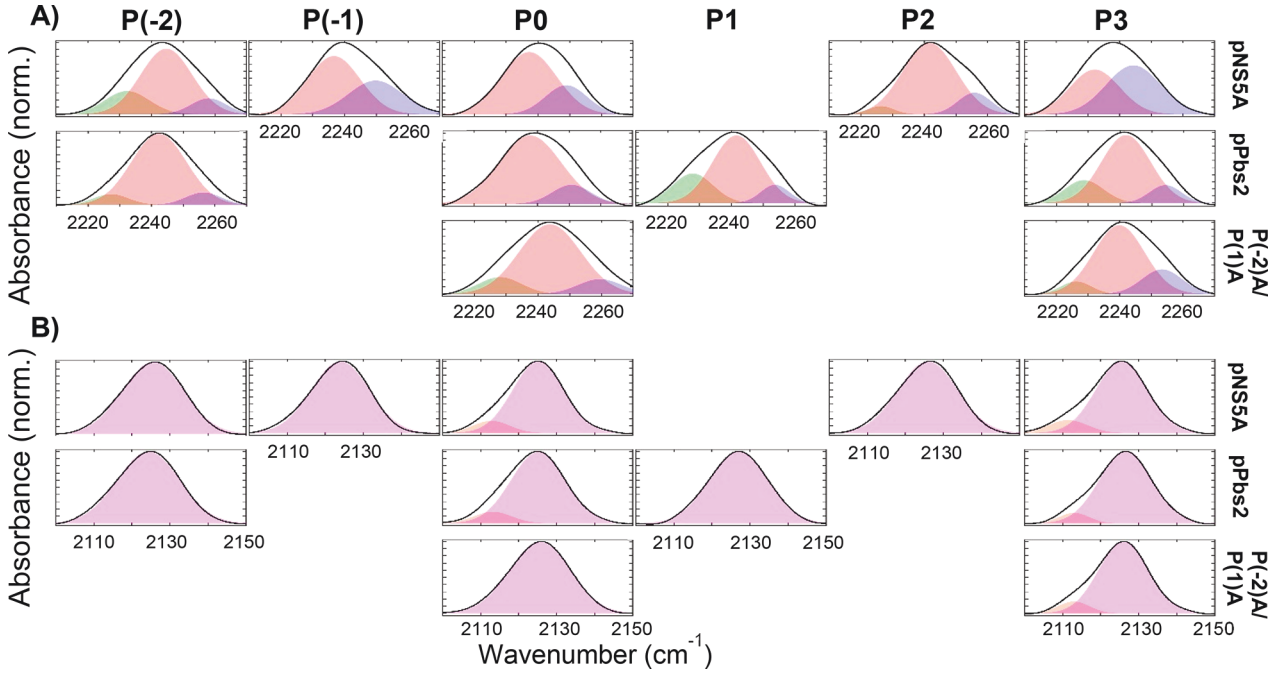


Figure 3. $C_\delta D_2$ absorption (black line) of the asymmetric stretch (A) and symmetric stretch (B) from FT-IR spectra of C−D probes incorporated at each proline of pNS5A (top row), pPbs2 (middle row), and P(−2)A/P1A pPbs2 (bottom row). Component bands from spectral modeling are shown in green, red, and blue for the asymmetric stretch and magenta and orange for the symmetric stretch. Shown are the average spectra.

α conformation (with φ and ψ angles around −70° and −25°, respectively). Its adoption by a proline residue in the MD simulations resulted in a kink in the global peptide structure (Figure 5).

Notably, all the IR spectra of those residues found in the MD simulations to adopt the α conformation also showed the low frequency $C_\delta D_2$ asymmetric stretch band, indicating that the band might reflect the associated backbone structure. Interestingly, we also observed that the histograms of the ψ angles populated within the PPII ensemble were asymmetric, suggesting the presence of two distinct subpopulations that might be associated with the two bands at highest frequency found for the $C_\delta D_2$ asymmetric absorption (Figure 4C). Supporting this interpretation, the relative area of the band at low frequency compared to the total area of the two bands at higher frequency correlates remarkably well with the relative occurrence of the α and PPII structure for the proline residues in the MD simulations (Figure 6). We note, however, that the intention of the simulations was to assist spectral assignment, rather than quantitatively capture equilibrium populations of the conformers. Although population of the cis in addition to the trans backbone conformation is another possible origin for the multiple bands, we found no evidence for the cis state in the MD simulations. In addition, we found no evidence of cis population in our previous study of pPbs2,[15] where the slow time scale typical of cis/trans isomerization would have resulted in resolved states in the NMR spectra.[35,36] Thus, we assign the two bands of the $C_\delta D_2$ asymmetric stretch at higher frequency to two populations with PPII conformation and the band at low frequency to a population with the α conformation.

To further assess the potential connection between the multiple bands found for the $C_\delta D_2$ asymmetric stretch and residue conformation, we performed DFT calculations of a trimer peptide model for the sequence context of P0 in P(−2)A/P1A pPbs2 to determine the energy profile as a function of the ψ angle, with φ angle fixed at −66°, the average from the MD simulations. The energy profile showed two distinct minima with ψ angles of −93° and −25° (Figure 4D), characteristic of the PPII and the α conformation, respectively.[34] The frequency of the $C_\delta D_2$ asymmetric stretch was calculated for each of these two structures and found to be lower by ∼11 cm^{-1} for the α conformation (Table S-4), in excellent agreement with the difference of 12−15 cm^{-1} determined experimentally (Table 2), further supporting the assignment of the band at low frequency to the α conformation.

Finally, we experimentally characterized C−D probes at residue P0 of pPbs2 and P(−2)A/P1A pPbs2 at increasing temperatures (Figure 7). For both peptides, but more dramatically for the native sequence, as the temperature was increased, the absorption for the $C_\delta D_2$ asymmetric stretch shifted to lower frequency and the amplitude at the high-frequency side decreased. These spectral changes are consistent with a loss of PPII structure, particularly the subpopulation associated with the lower intensity, highest frequency band. Although the peptides appear to retain some secondary structure at the highest temperature investigated (70 °C), the spectra for the two peptides become more similar with increasing temperature (Figure S6), which further supports that the differences at lower temperatures are due to distinct secondary structures. Thus, all together the data indicate that the $C_\delta D_2$ asymmetric stretch can be used as a local reporter of residue conformation.

Analysis of the asymmetric stretch absorptions provides unprecedented insight into residue-specific conformational heterogeneity of the peptides (Figure 6). The IR data indicate that for pNS5A the core motif residues P0 and P3 adopt exclusively the PPII structure, and the residues P(−1) and P2 flanking the core motif show little or no population of α conformation. In comparison, for pPbs2 the core motif residue P0 similarly adopts only PPII structure, but P3 and the

DOI: 10.1021/acs.analchem.8b03813
Anal. Chem. 2018, 90, 14355−14362

Table 2. Parameters from Modeling of C−D Bond Absorptions

	$C_\delta D_2$ symmetric stretch		$C_\alpha D$ stretch	$C_\delta D_2$ asymmetric stretch	
	ν (cm^{-1})	area (%)	ν (cm^{-1})	ν (cm^{-1})	area (%)
pPbs2					
P(−2)	2124.2 ± 0.2		2170.1 ± 1.4	2226.8 ± 1.2	6 ± 4
				2242.4 ± 0.4	85 ± 6
				2256.7 ± 1.4	9 ± 6
P0	2125.2 ± 0.01	87 ± 7	2168.3 ± 0.1	2237.6 ± 0.5	84 ± 8
	2113.6 ± 0.2	13 ± 7		2250.6 ± 1.1	16 ± 8
P1	2126.9 ± 0.2		2168.7 ± 0.5	2228.0 ± 0.9	24 ± 3
				2241.8 ± 0.6	66 ± 3
				2253.9 ± 0.2	10 ± 0.3
P3	2126.5 ± 0.3	92 ± 2	2168.7 ± 0.2	2228.8 ± 0.9	18 ± 12
	2113.2 ± 0.8	8 ± 2		2241.9 ± 0.5	70 ± 9
				2254.2 ± 0.6	12 ± 4
P(−2)A/P1A pPbs2					
P0	2125.3 ± 0.2		2165.3 ± 0.2	2229.1 ± 1.4	27 ± 8
				2243.3 ± 0.8	63 ± 9
				2255.8 ± 0.7	10 ± 4
P3	2126.0 ± 0.3	90 ± 2	2168.5 ± 0.2	2226.3 ± 0.8	8 ± 1
	2112.9 ± 0.6	10 ± 2		2239.9 ± 0.6	72 ± 3
				2253.3 ± 0.8	20 ± 3
pNS5A					
P(−2)	2125.6 ± 0.1		2168.0 ± 0.5	2232.8 ± 1.0	24 ± 8
				2244.8 ± 1.2	64 ± 8
				2257.7 ± 0.7	12 ± 3
P(−1)	2124.0 ± 0.6		2170.7 ± 0.1	2238.2 ± 1.2	77 ± 8
				2252.5 ± 1.6	23 ± 8
P0	2125.2 ± 0.02	88 ± 3	2168.8 ± 0.2	2237.4 ± 0.4	75 ± 5
	2113.0 ± 0.1	12 ± 3		2249.0 ± 0.6	25 ± 5
P2	2125.9 ± 0.4		2170.3 ± 0.1	2226.4 ± 0.6	4 ± 1
				2241.7 ± 0.6	81 ± 2
				2255.9 ± 0.5	14 ± 2
P3	2125.5 ± 0.3	90 ± 7	2170.0 ± 0.2	2232.5 ± 1.7	48 ± 12
	2111.6 ± 0.6	10 ± 7		2244.6 ± 1.7	52 ± 12

adjacent residue P1 show substantial population of the α conformation. In addition, the spectral data indicate that substitution of the two nonmotif proline residues for alanine in P(−2)A/P1A pPbs2 leads to perturbation of the local structure most significantly at P0, resulting in an increase in population of the α conformation. Thus, the IR spectra indicate the greatest overall content of PPII structure for pNS5A and the greatest content of α structure for P(−2)A/P1A pPbs2. This conclusion is in agreement with the known tendency of proline to promote the PPII structure[29,30] and with the CD spectra of the peptides but, importantly, provides residue-specific detail. For example, inspection of the sequence context reveals that only those residues that are not followed in sequence by another proline adopt the α conformation. Such influence of proline on the conformation of the preceding residue is well-known.[33,34] Steric hindrance between the side chain of proline near $C_\delta H_2$ and the $C_\beta H_2$ of the side chain of the next amino makes the α conformation unfavorable when a proline follows in the sequence, leading to its known behavior as a "helix breaker".

Whereas we find that the absorption of the $C_\delta D_2$ asymmetric stretch reports on the residue backbone conformation, the absorption of the symmetric stretch appears sensitive to a different aspect of the C−D probe environment. In several cases, the symmetric stretch mode shows a second absorption band at lower frequency (Figure 3B), but no correlation is found with the asymmetric stretch with regard to either the number of bands or their relative intensities (Table 2; Figure S-5). Thus, the second band found for the symmetric stretch absorption does not similarly report on a population differentiated by the backbone conformation. Rather, the band was present even when a residue adopted exclusively a PPII structure (e.g., P0 of pPbs2). In addition, the band appeared for P0 of pPbs2, but disappeared with introduction of the P(−2)A/P1A mutations, which is found by analysis of the asymmetric stretch to disrupt PPII structure at this site (Figure 6 and Figure S-5). Furthermore, it is notable that the second band was present for only the conserved proline residues of the recognition motif and, moreover, was observed for all of these residues except P0 of P(−2)A/P1A pPbs2 (Figure 3B). For pPbs2 and pNS5A the band contributed a similar 10−15% of the integrated absorbance for both P0 and P3 (Figure S-5), suggesting that the bands at both sites might reflect a population of a single structure adopted by the peptide. This possibility would be facilitated by the threefold symmetry of the PPII structure that places the two motif proline residues adjacent in space, where the C−D probes likely would be sensitive to a common feature of their environment. In any

DOI: 10.1021/acs.analchem.8b03813
Anal. Chem. 2018, 90, 14355−14362

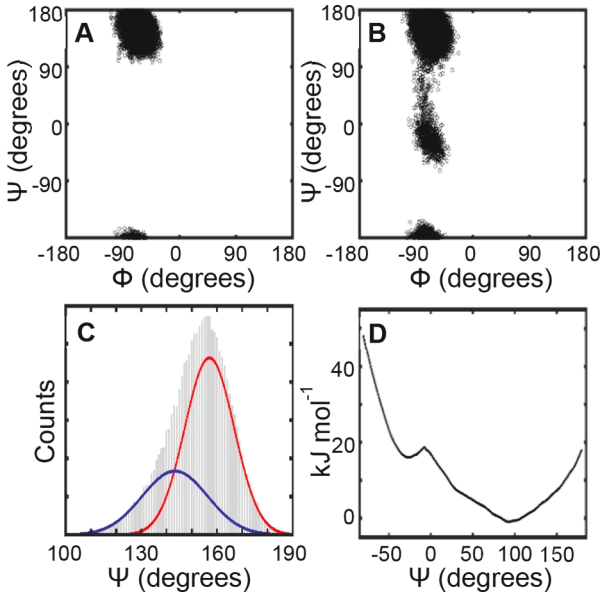


Figure 4. Analysis of MD simulations of PR peptides. Ramachandran plots showing ψ and φ backbone angles adopted for (A) P0 of pPbs2 and (B) P0 of P(−2)A/P1A pPbs2. (C) Histogram of ψ angles associated with PPII structure for d_3P0 of pPbs2, and the fit to a sum of two Gaussian functions (blue and red lines) that reflect two subpopulations of PPII structures. (D) Energy profile as a function of the ψ angle with φ set to −66° from DFT calculations of Ac-L(d_3)PA-NMe.

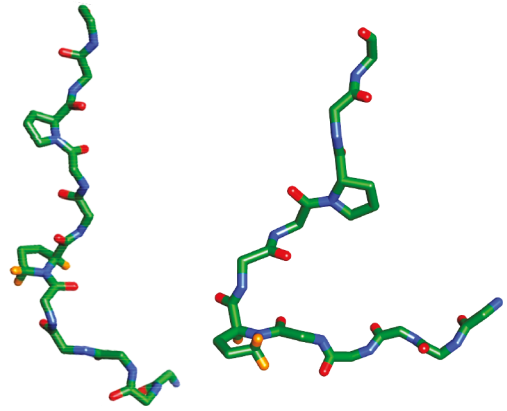


Figure 5. Average structures from MD simulation of P(−2)A/P1A pPbs2 for frames in which residue P0 adopts ψ and φ backbone angles of the PPII conformation (left) and for frames in which P0 adopts the α conformation (right), leading to a kink in the peptide. For clarity, the side chains are omitted for all residues except proline, and the location of the $C_\delta D_2$ bonds is shown in orange.

case, the appearance of multiple bands for the symmetric stretch further reveals the complex conformational heterogeneity of PR sequences, and moreover, the observation of the similar conformational distributions distinctly sensed by motif residues P0 and P3 suggests the ensemble population might be important for their recognition by cognate SH3 domains.

CONCLUSION

We have demonstrated the utility of C−D bond vibrations as probes of the conformational heterogeneity along a peptide with residue-specific spatial detail. The multiple bands observed for the $C_\delta D_2$ asymmetric stretch mode inform on

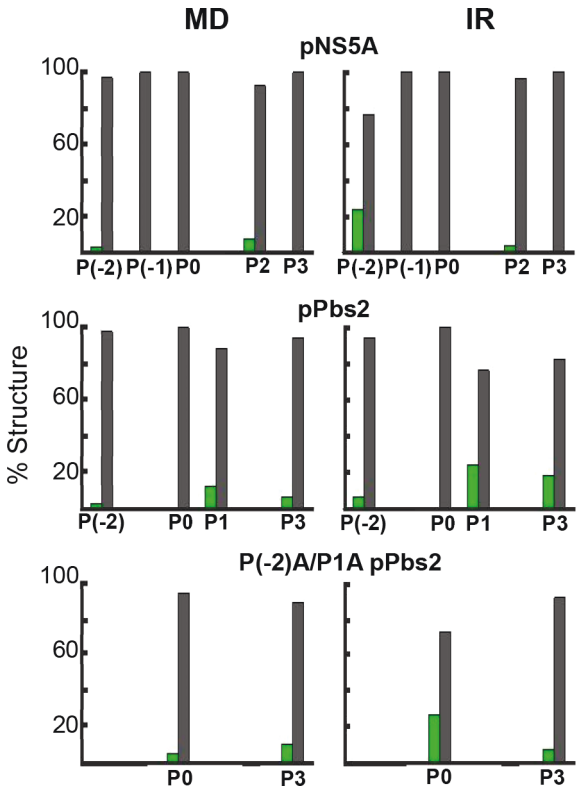


Figure 6. Left panel: bar graph showing relative population of PPII (gray) and other (green) secondary structures in MD simulations for each proline residue of pNS5A (top), pPbs2 (middle), and P(−2)A/P1A pPbs2 (bottom). Right panel: bar graph showing the relative integrated absorbance of the two IR bands at high frequency (gray) and that at low frequency (green) for the $C_\delta D_2$ asymmetric stretch for each proline residue of pNS5A (top), pPbs2 (middle), and P(−2)A/P1A pPbs2.

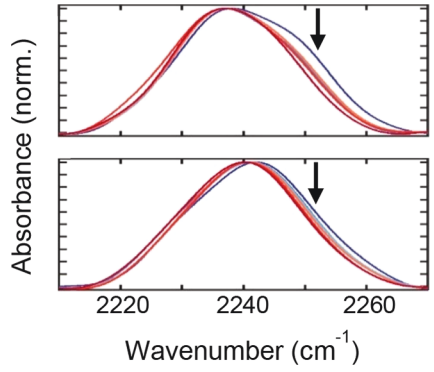


Figure 7. FT-IR spectra of the $C_\delta D_2$ asymmetric stretch of d_3P0 in pPbs2 (top) and P(−2)A/P1A pPbs2 (bottom) at varying temperatures. The arrows demark the direction of spectral change with increasing temperature, from 25 °C (blue) to 70 °C (dark red). Shown are the average spectra.

conformational states of the backbone that are expected to be highly dynamic. Indeed, previous characterization of the same bonds of P0 and P3 in pPbs2 via NMR spectroscopy found the states to be unresolvable because of rapid interconversion on the NMR time scale,[15] demonstrating the importance of the fast inherent time scale of IR spectroscopy to ensure detection of all potentially important conformational heterogeneity. Not only do we uncover such states in the peptides, we also find

DOI: 10.1021/acs.analchem.8b03813
Anal. Chem. 2018, 90, 14355−14362

that they are residue-specific and depend on the sequence context. We anticipate that such conformational heterogeneity is likely to contribute to the biophysics underlying the function of PR sequences and other recognition motifs often found in intrinsically disordered sequences. For instance, the conformational dynamics of PR sequences impact entropic changes that can determine the thermodynamics of protein recognition.[37,38] Proteins with this and other PR motifs and their cognate ligands such as SH3, WW, and EVH1 domains are highly abundant in eukaryotic systems, and their interactions play key roles in signaling and other cellular processes.[17,39,40] Elucidating the role of conformational heterogeneity in their molecular recognition should help to unravel how such sequences are evolved to mediate specificity in the complex interaction networks that underlie cellular function. In addition to small linear recognition motifs, PR regions are often found as highly repetitive sequences in proteins, such as collagen, myosin light-chain kinase, and the C-terminal tail of RNA polymerase II.[39] The extended PPII structure adopted by PR sequences not only facilitates binding interactions but also can serve as a spacer to define the spatial separation of protein domains. Moreover, PPII structure is a major constituent of denatured sequences and the PPII conformation is commonly adopted by individual residues in structured proteins, with estimates suggesting as much as 10% of residues.[41,42] While the present study demonstrates the utility of C−D bonds as probes of the conformations of proline, C−D bonds placed near the backbone of other amino acid residues could expand the tool set. For example, previous characterization of an SH3 domain via C−D bonds incorporated at glycine residues showed evidence for conformational heterogeneity attributed to backbone structure.[43] Continued development of IR spectroscopy of C−D probes will afford a general approach for experimentally characterizing rapidly fluctuating conformations that makes possible the investigation of their role in biological function.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.8b03813.

Experimental details of peptide synthesis and sample preparation; additional spectra, additional modeling parameters, and evaluation of fits of spectra (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: thielges@indiana.edu.

ORCID

Megan C. Thielges: 0000-0002-4520-6673

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

G.S.B. and M.C.T. thank Indiana University and the National Science Foundation (Grant No. MCB-1552996) for funding.

■ REFERENCES

(1) Tsai, C. J.; del Sol, A.; Nussinov, R. *J. Mol. Biol.* **2008**, *378* (1), 1−11.

(2) Petit, C. M.; Zhang, J.; Sapienza, P. J.; Fuentes, E. J.; Lee, A. L. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (43), 18249−18254.

(3) Hammes, G. G.; Benkovic, S. J.; Hammes-Schiffer, S. *Biochemistry* **2011**, *50*, 10422−10430.

(4) Wand, A. J.; Sharp, K. A. *Annu. Rev. Biophys.* **2018**, *47*, 41−61.

(5) Habchi, J.; Tompa, P.; Longhi, S.; Uversky, V. N. *Chem. Rev.* **2014**, *114* (13), 6561−6588.

(6) Berlow, R. B.; Dyson, H. J.; Wright, P. E. *FEBS Lett.* **2015**, *589* (19 Pt A), 2433−2440.

(7) Palmer, A. G., III *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 129−155.

(8) Stollar, E. J.; Lin, H.; Davidson, A. R.; Forman-Kay, J. D. *PLoS One* **2012**, *7*, No. e51282.

(9) Blasiak, B.; Londergan, C. H.; Webb, L. J.; Cho, M. *Acc. Chem. Res.* **2017**, *50* (4), 968−976.

(10) Kim, H.; Cho, M. *Chem. Rev.* **2013**, *113* (8), 5817−5847.

(11) Adhikary, R.; Zimmermann, J.; Romesberg, F. E. *Chem. Rev.* **2017**, *117* (3), 1927−1969.

(12) Chin, J. K.; Jimenez, R.; Romesberg, F. E. *J. Am. Chem. Soc.* **2001**, *123* (10), 2426−2427.

(13) Adhikary, R.; Zimmermann, J.; Liu, J.; Forrest, R. P.; Janicki, T. D.; Dawson, P. E.; Corcelli, S. A.; Romesberg, F. E. *J. Am. Chem. Soc.* **2014**, *136* (39), 13474−13477.

(14) Le Sueur, A. L.; Schaugaard, R. N.; Baik, M.-H.; Thielges, M. C. *J. Am. Chem. Soc.* **2016**, *138* (22), 7187−7193.

(15) Horness, R. E.; Basom, E. J.; Mayer, J. P.; Thielges, M. C. *J. Am. Chem. Soc.* **2016**, *138* (4), 1130−1133.

(16) Zarrinpar, A.; Bhattacharyya, R. P.; Lim, W. A. *Sci. Signaling* **2003**, *2003* (179), re8.

(17) Ball, L. J.; Kuhne, R.; Schneider-Mergener, J.; Oschkinat, H. *Angew. Chem., Int. Ed.* **2005**, *44* (19), 2852−2869.

(18) Li, S. S. *Biochem. J.* **2005**, *390* (Pt 3), 641−653.

(19) Zarrinpar, A.; Park, S. H.; Lim, W. A. *Nature* **2003**, *426* (6967), 676−680.

(20) Martin-Garcia, J. M.; Luque, I.; Ruiz-Sanz, J.; Camara-Artigas, A. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2012**, *68*, 1030−1040.

(21) Yu, H.; Chen, J. K.; Feng, S.; Dalgarno, D. C.; Brauer, A. W.; Schreiber, S. L. *Cell* **1994**, *76*, 933−945.

(22) Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N. et al. *AMBER 2016*; University of California: San Francisco, 2016.

(23) Price, D. J.; Brooks, C. L. *J. Chem. Phys.* **2004**, *121* (20), 10096−10103.

(24) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696−3713.

(25) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79* (2), 926−935.

(26) Lambrakos, S. G.; Boris, J. P.; Oran, E. S.; Chandrasekhar, I.; Nagumo, M. *J. Comput. Phys.* **1989**, *85* (2), 473−486.

(27) Roe, D. R. C.; Cheatham, T. E. *J. Chem. Theory Comput.* **2013**, *9*, 3084−3095.

(28) Shao, Y.; Molnar, L. F.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P.; et al. *Phys. Chem. Chem. Phys.* **2006**, *8* (27), 3172−3191.

(29) Creamer, T. P. *Proteins: Struct., Funct., Genet.* **1998**, *33* (2), 218−226.

(30) Kelly, M. A.; Chellgren, B. W.; Rucker, A. L.; Troutman, J. M.; Fried, M. G.; Miller, A. F.; Creamer, T. P. *Biochemistry* **2001**, *40* (48), 14376−14383.

(31) Lopes, J. L. S.; Miles, A. J.; Whitmore, L.; Wallace, B. A. *Protein Sci.* **2014**, *23* (12), 1765−1772.

(32) Mirkin, N. G.; Krimm, S. *J. Phys. Chem. A* **2007**, *111* (24), 5300−5303.

(33) Schimmel, P. R.; Flory, P. J. *J. Mol. Biol.* **1968**, *34* (1), 105−120.

DOI: 10.1021/acs.analchem.8b03813
Anal. Chem. 2018, 90, 14355−14362

(34) Macarthur, M. W.; Thornton, J. M. *J. Mol. Biol.* **1991**, *218* (2), 397−412.
(35) Sarkar, S. K.; Young, P. E.; Sullivan, C. E.; Torchia, D. A. *Proc. Natl. Acad. Sci. U. S. A.* **1984**, *81* (15), 4800−4803.
(36) Schubert, M.; Labudde, D.; Oschkinat, H.; Schmieder, P. *J. Biomol. NMR* **2002**, *24* (2), 149−154.
(37) Ferreon, J. C.; Hilser, V. J. *Protein Sci.* **2003**, *12* (3), 447−457.
(38) Zeng, D. Y.; Shen, Q. L.; Cho, J. H. *Biochem. Biophys. Res. Commun.* **2017**, *484* (1), 21−26.
(39) Williamson, M. P. *Biochem. J.* **1994**, *297*, 249−260.
(40) Kay, B. K.; Williamson, M. P.; Sudol, P. *FASEB J.* **2000**, *14* (2), 231−241.
(41) Tiffany, M. L.; Krimm, S. *Biopolymers* **1968**, *6* (9), 1379−1382.
(42) Sreerama, N.; Woody, R. W. *Biochemistry* **1994**, *33* (33), 10022−10025.
(43) Cremeens, M. E.; Zimmermann, J.; Yu, W.; Dawson, P. E.; Romesberg, F. E. *J. Am. Chem. Soc.* **2009**, *131* (16), 5726−5727.

DOI: 10.1021/acs.analchem.8b03813
Anal. Chem. 2018, 90, 14355−14362