A Dataset of Simple 1-D and 2-D NMR Spectra of Peptides, Including all Encoded Amino Acids, for Introductory Instruction in Protein Biomolecular NMR Spectroscopy

Noah J. Daniecki, Nicholas V. Costantini, Geoffrey M. Sametz*, and Neal J. Zondlo*

Department of Chemistry and Biochemistry

University of Delaware

Newark, DE 19716

United States

* To whom correspondence should be addressed. email: *sametz@udel.edu*, *zondlo@udel.edu*.

Abstract

NMR spectroscopy is the most important technique for understanding the structure of peptides and proteins in solution, providing information at the single-residue and single-atom level. However, written instruction in the interpretation of NMR spectra of peptides and proteins is generally focused on advanced techniques and highly complex spectra, with a lack of simple spectra and guides available for beginning students. In order to address this instructional limitation, we have generated a dataset of ^{1}H NMR spectra of a series of simple peptides that include all canonical amino acids. Peptides examined include Ac-**X**(S/pS)-NH_2, Ac-**X**(T/pT)-NH_2, and Ac-**X**PPGY-NH_2, where **X** = all encoded amino acids, pS = phosphorylated Ser, and pT = phosphorylated Thr. The characterization of each peptide includes a 1-D spectrum and a TOCSY spectrum, with both the raw and processed available. The spectra can be used for instructional applications including analysis of regions of the spectra (e.g. amide H^N, aromatic, H□, and aliphatic regions); identification of spin systems and residue assignment via TOCSY spectra; analysis of conformational features including amide H_N chemical shift dispersion and changes due to hydrogen bonding or post-translational modifications; the $^{3}J_{□N}$ coupling constant that reports on the ϕ torsion angle and on order versus disorder at a given residue; conformational preferences at H□ via chemical shift index analysis; understanding of diastereotopic hydrogens; dynamic processes, including hydrogen exchange; and identification of proline *cis-trans* isomerism. In addition, for a limited number of peptides, NOESY spectra are included to allow sequential resonance assignment and for assignment of *trans* versus *cis* proline conformations. Spectra from closely related peptides allow the analysis of the relative effects of single amino acid changes. The paper is written to be directly accessible to students as a tutorial guide. In addition, the data can be used by instructors for problem sets and exams.

Introduction

NMR spectroscopy is the primary high-resolution technique to understand the structure of peptides and proteins in solution.[1] NMR spectroscopy provides details on structure at the single-residue level and indeed at the single-atom level, as each resonance indicates the environment at a single hydrogen. In contrast, bulk techniques such as circular dichroism spectroscopy give global details about secondary structure, while providing no information at the individual residue level.[2] NMR spectroscopy also provides information about the dynamics of molecules in solution that is not typically possible via techniques such as X-ray crystallography, which normally exhibits a single conformation that may or may not be the dominant conformation in solution.[3] NMR spectroscopy is particularly valuable in understanding structure in intrinsically disordered regions of proteins, which are typically not amenable to X-ray crystallography.[4-14] Importantly, nearly all students in chemistry, the biological sciences, and biomedically oriented fields in engineering have a basic understanding of NMR spectroscopy due to its applications in introductory organic chemistry courses and associated laboratory courses, as well as from dedicated courses in spectroscopy.

However, despite the broad importance of NMR spectroscopy in the analysis of proteins, the technique in practice often has a significant barrier to entry. Most importantly, organic chemistry classes do not typically significantly cover NMR spectroscopy as applied to peptides and proteins. Indeed, peptide structure is typically a late or optional topic in organic chemistry classes. In addition, TOCSY spectra that are central to peptide and protein resonance assignments are not commonly taught in the introductory organic chemistry spectroscopy curriculum. As such, students are often dependent on specialized classes or research to have access to instruction in this technique, inherently and unnecessarily limiting its applications.

Indeed, much instruction in these techniques is passed on to students in their individual research groups. However, because essentially all students in the biological sciences have early instruction in the basics of NMR spectroscopy, this technique should inherently be highly accessible.

Wüthrich wrote a landmark book on using NMR spectroscopy to understand structure and dynamics in proteins.[1] In addition, many subsequent textbooks, review articles, and original research articles have been written on the application of NMR spectroscopy to understanding protein structure.[3,5,7,13,15] However, early articles include representations of the spectra that lack either modern graphical analysis or access to the underlying spectra. Recent articles, in contrast, focus on larger proteins, whose NMR spectra are functionally inaccessible to beginning students due to spectral complexity. While an exceptional database of NMR data is available at BioMagResBank,[16] these data are better appreciated by advanced students and researchers than they are accessible to students in their introductory work in NMR spectroscopy of peptides and proteins.

In our instruction both of undergraduate students in the teaching laboratory and of starting undergraduate students and graduate students in the research laboratory, we realized that additional teaching resources with more manageable NMR spectra and instructional guides could simplify instruction and accelerate student understanding of the basic elements of peptide and protein NMR spectroscopy. We recently developed a teaching laboratory module for introductory (second-semester) organic chemistry laboratory students to learn about techniques in solid-phase peptide synthesis and conformational analysis.[17] In the context of that work, the students synthesized a series of peptides Ac-**X**(S/pS)-NH_2, Ac-**X**(T/pT)-NH_2, and Ac-**X**PPGY-NH_2, where **X** = all 20 canonical amino acids, S = serine, pS = phosphorylated serine, T =

threonine, and pT = phosphorylated threonine (Figure 1). That work was focused on the research questions of understanding either (a) the effects of local amino acids on phosphorylation-induced structure in peptides or (b) the propensities of amino acids for nucleation of polyproline II helix (PPII) structure.[18-21] The discussion of those residue-specific effects on peptide structure will be reported separately. However, the sets of simple NMR spectra from these peptides represent a potentially useful general education tool in biochemistry, biophysics, chemical biology, and molecular biology. Herein, we report a dataset of simple NMR spectra for instruction in peptide and protein NMR spectroscopy, along with a general set of guidelines for their interpretation. These data include all canonical (encoded) amino acids, with each amino acid present in at least 5 different peptide spectral contexts. Because the peptide sequences are quite short, the NMR spectra are simpler than typical peptide and protein NMR spectra. These spectra can be used in introductory classroom and laboratory contexts, where the simple spectra can be readily interpreted by students. This manuscript is also intended to function as a basic tutorial guide for students in this instruction, which can be complemented by traditional in-class teaching, textbooks, and laboratory instruction manuals, as well as by various web-based resources.

Results

1-D ^{1}H NMR spectra: regions of the NMR spectrum, amide H_N chemical shifts, and $^{3}J_{aN}$ coupling constants. One challenge of peptide and protein NMR spectroscopy is the preponderance of peaks in the NMR spectra, in contrast to NMR spectra of simple organic molecules. The most basic task in any NMR analysis is peak identification and assignment. The NMR spectra of peptides and proteins have a number of key regions that can be readily identified by experienced practitioners, and that students can be taught to identify and interpret. The

starting point here is to make sense of the general spectrum, in order to reduce the complexity to individual spectral regions that are manageable with guidance.

The key spectral regions are the amide H_N resonances (~7–9 ppm), aromatic H resonances (~6.5–8.5 ppm), H□ (~4–5 ppm), and resonances from aliphatic hydrogen C–H bonds (~1–4 ppm). These spectral regions are broadly annotated in Figure 2, while also emphasizing the resonance from H_2O at 4.8 ppm and the overlap that occurs between these spectral regions.

Several key NMR parameters can be obtained from simple 1–D NMR spectra. First, amide chemical shift dispersion (overall range of the amide chemical shifts) is inherently related to structure in peptides and proteins (Figure 3). Thus, disordered proteins exhibit relatively little amide chemical shift dispersion (a small range of amide chemical shifts). In contrast, folded proteins and ordered regions of peptides exhibit much greater amide chemical shift dispersion (a wider range of amide chemical shifts). In related peptides, differences in H_N chemical shift dispersion can be readily identified. In particular, differences are observed herein in related peptides as a function of phosphorylation.

More generally, in most of the peptides, the distinct amide hydrogens are all well resolved. In these H_N, the coupling constant between H_N and H□ ($^3J_{□N}$) can be determined directly from the 1-D NMR spectrum (Figure 4). $^3J_{□N}$ correlates with the ϕ main chain torsion angle, via a parametrized Karplus relationship.[15] In disordered peptides, it is unlikely that a single conformation in ϕ is adopted, and $^3J_{□N}$ broadly represents the population-weighted average of all conformations present in solution. However, $^3J_{□N}$ correlates overall with ϕ, with values around 7 Hz indicating random coil conformations. Smaller values of $^3J_{□N}$ (especially < 6 Hz) indicate more compact conformations in ϕ (e.g., more □-helical), while larger values of $^3J_{□N}$ (especially > 8 Hz) indicate more extended conformations in ϕ (e.g., more □-like) Most generally, deviations

from random coil values of $^3J_{\square N}$ (6–8 Hz) indicate significant ordering in peptides. In addition, changes in $^3J_{\square N}$ as a result of variations in peptide sequence or post-translational modification more generally can be correlated with changes in biases in peptide main-chain conformation, including the propensities of amino acids for different regions of Ramachandran space.[19,22] Shifts to smaller values of $^3J_{\square N}$ indicate adoption of more compact conformations, while shifts to larger values of $^3J_{\square N}$ indicate changes to more extended conformations. For example, in Ac-XPPGY-NH_2 peptides, $^3J_{\square N}$ = 6.0 Hz for Ala, but 7.7 Hz for Val, consistent with Ala preferring more compact conformations, but Val preferring more extended conformations. As an example of how post-translational modifications can change protein structure,[20,21,23,24] $^3J_{\square N}$ changes from 8.3 Hz for Thr to 6.4 Hz for pThr in Ac-Y(T/pT)-NH_2 peptides.

Identification of amino acid spin systems and resonances in NMR spectra using TOCSY spectra. Even in simple peptide NMR spectra, the identity of the amide H_N and other chemical shifts cannot typically be definitively determined solely from the 1-D spectra. TOCSY spectra are the predominant method to assign resonances in peptide and protein NMR spectra.[25,26] TOCSY spectra are, to a first approximation, similar to COSY spectra that also exhibit longer-range through-bond coupling, albeit without providing coupling constants. COSY spectra are commonly taught in organic spectroscopy classes, and indicate which resonances have short-range (2-bond or 3-bond) through-bond (scalar) coupling (^2J or ^3J). In contrast, in TOCSY spectra, longer-range correlations are observed, in addition to the correlations seen in COSY spectra. In TOCSY spectra, no longer-range correlations are observed that extend through amide bonds, and thus the TOCSY spectra represent correlations within single residues of a peptide or protein. Peaks in TOCSY spectra also exhibit "self-correlation" peaks, that is, crosspeaks of a

resonance with itself, which is manifested in the "diagonal" that extends from the lower left to the upper right (i.e. the ⁄ pattern that appears to divide the spectrum in half) of the TOCSY spectrum.

TOCSY spectra exhibit correlations within the entire spin systems of single amino acids. As such, in the TOCSY spectra, correlations are typically observed between H_N and all of Hα, Hβ, and (if present) even Hγ, Hδ, and Hε (Figure 5). The "fingerprint region" of TOCSY spectra (upper left quadrant) thus connects all of these hydrogens from a single residue in a single vertical slice centered on the chemical shift of H_N. Because every amino acid has distinct sets of hydrogens, with characteristic typical ranges of chemical shifts,[1] the amino acid identity of any given amide hydrogen resonance can often be clearly determined directly from the fingerprint region of the TOCSY spectrum (Figure 6, Table 1). This assignment then inherently includes the more difficult assignments of (often overlapping) peaks in the aliphatic region of the NMR spectra. Each crosspeak in the TOCSY spectrum indicates the correlation between any two hydrogens in the same amino acid, and thus also gives the chemical shift of each of those hydrogens.

Considering the peptide Ac-MT-NH_2 (Figure 5), the amide resonance at 8.41 ppm can be identified as the H_N of Met via its correlation to peaks at 4.52 ppm (Hα), 2.64 ppm (Hβ), and 2.03 ppm (terminal CH_3 Hε), which represents a characteristic pattern for Met. In contrast, the amide resonance at 8.16 ppm can be identified as Thr due to its correlation with hydrogens at 4.36 ppm (Hα), 4.28 ppm (Hβ), and 1.21 ppm (CH_3 Hγ). Notably, the side-chain hydrogens are not typically identified via their coupling constant patterns, the approach that is typical in organic chemistry NMR spectroscopy, due both to complex splitting patterns (multiple adjacent hydrogens coupled) and to spectral overlap in the aliphatic region. In addition, Hα and other

peaks near the water resonance are often greatly reduced in magnitude or not observed due to water suppression techniques (i.e., the methods necessary to permit the observation of peptide peaks present at μM–mM concentrations in 55 M H_2O), but are readily identifiable in TOCSY spectra.

The simplest TOCSY correlations for residues with aliphatic side chains are those of Ala, Cys, Asp, Asn, and Ser (Figure 6, which includes an example of the correlations observed for each amino acid). In their TOCSY fingerprint regions, these residues exhibit correlations between H_N, Hα, and Hβ. In each case, the chemical shift of Hα correlates with the electron-withdrawing nature of the group attached to Hα (Ser Hα ~ 4 ppm; Cys Hα ~ 3 ppm; Asp/Asn Hα ~ 2.5–3 ppm; Ala Hα ~ 1.3 ppm; see also discussion of diastereotopic hydrogens, below). Thr, Glu, and Gln fingerprint regions are also fairly straightforward to identify, with correlations out to Hγ. Importantly, the fingerprint regions of Asp *versus* Asn, or of Glu *versus* Gln, are not readily distinguished via the TOCSY spectra, because the patterns of chemical shifts are similar, and no correlations are typically observed from H_N to the carboxamide hydrogens of Asn or Gln.

More complex fingerprint regions are observed for amino acids with longer side chains (Ile, Lys, Leu, Met, Arg, Val) (Figure 6, Figure 7). In particular, correlations with the hydrogens most distant from H_N, or with 3° methine (C–H) hydrogens, are sometimes weak or not observed to be above the noise in the fingerprint region. In these cases, the aliphatic region (upper right quadrant) of the TOCSY spectrum can be used to assign all resonances within the spin system of these amino acids. For example, in Ac-IpS-NH_2, correlations from H_N are only seen to some of the aliphatic side-chain hydrogens of Ile. However, all of these hydrogens exhibit strong correlations to each other, with several also correlating to Hδ, as observed within the upper right (aliphatic quadrant), confirming the resonances of the entire Ile spin system.

Spin systems in aromatic amino acids. The side chains of all of the aromatic amino acids are distinct (Figure 8 for a comparison of the aromatic regions of 1-D spectra; Figure 9 for TOCSY spectra). The usual correlation observed in the fingerprint region is between H_N, Hα, and Hβ. This pattern is similar to those of Cys, Asp, Asn, and Ser, with Hβ on aromatic amino acids typically ~ 3 ppm (similar to that of Cys). While the aromatic hydrogens do not typically exhibit a cross-correlation to the amino acid H_N in the TOCSY spectra, the aromatic hydrogens sometimes exhibit TOCSY cross-peaks to the Hβ of the same residue (Figure 9b), which can then be correlated to H_N. Independently, the aromatic side-chain spin systems of Phe, Tyr, Trp, and His are readily differentiated, and identifiable in both the 1-D spectra (Figure 8) and in the cross-correlations in the TOCSY spectra (Figure 9). Tyr exhibits a simple pair of apparent doublets, representing the Hδ (H_{ortho}) and Hϵ (H_{meta}) that are coupled to each other, each with a multiplicity of 2. Phe exhibits a more complicated spectrum due to Hζ (H_{para}) being coupled to two Hϵ (H_{meta}), observed as a triplet, and due to the Hϵ (H_{meta}) being coupled to both Hδ (H_{ortho}) and Hζ (H_{para}). Trp is distinct due to its indole H at 10.0-10.5 ppm, in addition to 5 other aromatic hydrogens with different coupling patterns. All of these aromatic resonances can be correlated to each other in the aromatic region of the TOCSY spectra.

Diastereotopic hydrogens. In achiral compounds, the individual hydrogens of each methylene (–CH_2–) group are chemically equivalent, and appear as a single hydrogen in the NMR spectrum. In contrast, in compounds with at least one stereocenter, the individual hydrogens of each methylene are diastereotopic, and thus *may* have distinct chemical shifts. As such, within peptides and proteins, the hydrogens on all methylene groups are diastereotopic (Figure 10). The chemical shift dispersion between any diastereotopic hydrogens can directly indicate order at those atoms, as it indicates different electronic and chemical environments of

those hydrogens, which can occur as a result of defined conformational preferences, instead of all conformations being of equivalent energy energy, which would generate similar time-weighted chemical environments.

Diastereotopic hydrogens are an inherently difficult concept for students to understand. This challenge is in part due to a limited number of examples in their coursework. Independent of this difficulty, the diastereotopic hydrogens in peptides and proteins *often* have distinct chemical shifts (but also frequently do not). Examples of diastereotopic hydrogens that potentially could exhibit distinct chemical shifts include Hα of Gly; Hβ of all amino acids except Ala, Ile, Thr, and Val; Hγ of Glu, Lys, Met, Pro, Gln, and Arg; Hδ of Lys, Pro, and Arg; and Hϵ of Lys. In general, the closer a pair of diastereotopic hydrogens is to the peptide backbone, the more likely that different chemical shifts will be observed, although *any* ordering in the predominant 3-D structure present in solution can lead to greater differences in the chemical shifts of diastereotopic hydrogens. In addition, the methyl ($-CH_3$) groups of Val and Leu are diastereotopic, and can exhibit distinct chemical shifts.

In the fingerprint region spectral slices of Figure 6, differences in chemical shifts of the diastereotopic Hα can readily be seen for Asp, Glu, Phe, and Ser. These can also be observed in well-resolved 1–D spectra (Figure 10). For example, in Ac-DS-NH_2, the diastereotopic Hβ both of Ser and of Asp exhibit distinct chemical shifts. In Ac-WS-NH_2, the diastereotopic Hβ of Ser are distinct, while the diastereotopic Hβ of Trp become equivalent due to the having the same chemical shift. Similarly, the diastereotopic β-CH_3 groups of Val are distinct in Ac-VpS-NH_2, but are equivalent in Ac-VT-NH_2.

Carboxamide hydrogens. The hydrogens of carboxamide groups ($-C(O)-NH_2$), present in Asn, Gln, and at the C-terminal carboxamide in peptides, exhibit distinctive patterns in the NMR

spectrum (Figure 11). Because amide bond rotation is slow on the NMR timescale, these hydrogens have distinct resonances. In addition, due to the carbonyl, these resonances do not exhibit coupling that is observable within the fingerprint region with other resonances in the amino acid. Thus, carboxamide resonances appear as a pair of broad singlets at ~ 7 ppm in the 1–D NMR spectrum, with correlation between these resonances observed in the TOCSY spectrum. Correlations are sometimes observed between one of the carboxamide hydrogens and the nearest -CH_2- hydrogens (Figure 11b).

Proline resonances. Unlike other canonical amino acids, proline lacks an amide hydrogen. As such, there are no crosspeaks associated with proline in the fingerprint region of the TOCSY spectrum. Proline resonances are identified in the aliphatic region of the TOCSY spectrum, via a pattern of correlations between Hα (4.0–4.5 ppm), Hβ (2.0–2.5 ppm), Hγ (2.0–2.5 ppm), and Hδ (3.0–3.5 ppm) (Figure 12). This set of cross-correlations is typically distinct from those of any other amino acid. Glu, Gln, and Met can exhibit similar patterns of correlations between Hα and more upfield aliphatic resonances, but each of those will also exhibit a crosspeak with an amide H_N. Notably, Pro Hβ, Hγ, and Hδ are all diastereotopic, and thus each of these positions can exhibit distinct chemical shifts for the diastereotopic hydrogens.

Proline cis-trans isomerism. Amide bonds between amino acids are in slow exchange between the *trans* and *cis* amide conformations, due to the partial double-bond character of amides (barrier to bond rotation ~ 20 kcal mol^{-1}, corresponding to half-lives of seconds).[27-30] For all amino acids other than proline, only the *trans* amide bond is typically observed, due to the large steric clash present between the Cα of adjacent residues when a *cis* amide bond is present. In contrast, the amide bonds at proline can be present as either a *trans* or a *cis* amide bond due to similar potential steric clashes of the pre-proline residue with either proline Cδ (Cα•••Cδ clash,

cis-proline) or proline Cδ (Cδ•••Cδ clash, *trans*-proline, a steric clash not present in non-proline amino acids) (Figure 12b). In folded proteins, only a single amide conformation (*trans* OR *cis*) is typically observed at each amide bond, with *cis*-proline observed at approximately 5% of all proline residues, and the tertiary structure stabilizing the observed proline amide conformation. Notably, identifying whether a proline amide bond is *trans* or *cis* is not trivial by ^{1}H NMR spectroscopy (see below), particularly in complex proteins.[13,31-34]

In contrast, within peptides and in intrinsically disordered proteins, *both trans*-proline and *cis*-proline are commonly observed, leading to more complicated NMR spectra with separate sets of resonances for each conformation, due to the interconversion being in slow exchange ($t_{1/2}$ ~ seconds to minutes) on the NMR timescale. Typically, the major species has *trans*-proline, while the minor species has *cis*-proline (Figure 13). The identity of local residues or local structure can dramatically impact the amount of *cis*-proline observed.[27,35-38] Notably, if multiple proline residues are present within a peptide or protein, the total number of possible species observable is 2^n, where n = number of proline residues.[19,36]

In peptides in this dataset with a proline residue, proline *cis-trans* isomerism is observable in both the 1-D and TOCSY NMR spectra. Peptides with aromatic-proline, proline-proline, or proline-aromatic sequences are most likely to exhibit increased populations of *cis*-proline.[27,28,35,37,39-43] The equilibrium contrast ($K_{trans/cis}$) between *trans*-proline and *cis*-proline can be directly determined via integration of relevant peaks. Typically, due to the large differences in local structure between *trans*-proline and *cis*-proline, all nearby hydrogens exhibit separate resonances for the species with *trans*-proline *versus* the species with *cis*-proline. Notably, the chemical shifts and $^{3}J_{\alpha N}$ of all resonances can be quite distinct in the *trans*-proline versus *cis*-

proline conformations, consistent with the dramatically different nature of these conformations and the ability of proline *cis-trans* isomerism to act as a functional switch in proteins.[29,30,44]

The identities of *cis*-proline versus *trans*-proline can be definitively determined via NOEs, through-space correlation of nuclei which occurs via magnetization transfer. In practice, this determination of NOEs is accomplished through 2-D NOESY or ROESY spectra. For *trans*-proline, Hα of the residue prior to proline is close to the Pro Hδ, and thus a strong NOE is observed between these resonances. In contrast, in *cis*-proline, Hα of the residue prior to proline is close to the Pro Hα, and thus a strong NOE is instead observed between these resonances.

Identification of dynamics: hydrogen exchange. In addition, the effects of hydrogen exchange can be identified in some of the pH-dependent NMR spectra (Figure 14). The rate of hydrogen exchange on amide hydrogens is greater with increasing pH, due to rate = k_{ex}[amide][$^-$OH], with solvent-exposed amides exhibiting broadening and exchange at pH ≥ 7.[45] Amide exchange is an important technique for understanding structure and solvent accessibility of amides.[46-49] Protected amide hydrogens (e.g. those within regular secondary structures or in the buried core of proteins) are typically observed at pH ≥ 7, but those that are solvent-exposed and not in hydrogen-bonded structures exchange at or faster than the NMR timescale, resulting in broadening or disappearance of these amide resonances. In these peptides, only minimal structure is expected, and thus the NMR spectra at lower pH exhibit all amide hydrogens, while at higher pH the amide hydrogens are partially or fully exchanged, and in the latter case they are thus not observed.

In addition, in peptides with Ser, Thr, or Tyr, the side chain hydroxyl H is not observed at any pH, as a result of rapid hydrogen exchange due to hydrogen bonding with water. Similarly, in peptides with the basic residues Lys or Arg, the hydrogens on side-chain nitrogens

(ammonium or guanidinium) are typically either absent or broadened in the NMR spectra due to rapid exchange with water. In contrast, aliphatic and aromatic hydrogens are observed under all spectral conditions, due to the low acidity of the C–H bonds. These spectra emphasize the inherently dynamic nature of functional groups in water, with rapid, intermediate, slow, or no hydrogen exchange at a given functional group, depending on its solvent exposure, hydrogen bonding, and acidity.

In order to observe TOCSY correlations with amide hydrogens, NMR spectra are recorded in solutions with 90% H_2O and 10% D_2O, with water-suppression techniques used in order to minimize the signal associated with the water resonance at 4.8 ppm. In 100% D_2O, all solvent-exposed amide hydrogens exchange with deuterium, resulting in the disappearance of peaks associated with solvent-exposed amide hydrogens (Figure 14). Amide H–D exchange is also an important technique for understanding structure in folded proteins, in which amide hydrogens (as well as other exchangeable hydrogens including on alcohols, ammoniums, or guanidiniums) that are solvent-exposed exhibit rapid H–D exchange, while those that are protected from solvent (e.g. part of the stably folded secondary structures or otherwise buried) exhibit slower exchange. Amide H–D exchange can be observed quantitatively as a function of time, both by NMR spectroscopy (on a residue-specific basis) and by mass spectrometry (based on bulk changes in numbers of ^{1}H-hydrogens replaced with ^{2}H-deuterium).

Sequential resonance assignment using NOESY/ROESY spectra. In simple peptides, where there is at most 1 copy of any amino acid type (resonances with similar spin systems and chemical shifts) in the sequence, definitive resonance assignments can typically be made directly from the TOCSY spectra. In contrast, in complex peptides and proteins, the presence of more than one copy of a given amino acid type in the sequence generates ambiguity about, for

example, which alanine resonances are associated with any one alanine residue in a peptide or protein with multiple alanines. This assignment problem is accomplished using NOESY spectra[50,51] via sequential resonance assignments (a resonance walk), which correlate the resonances of one amino acid with those of the adjacent residue in the sequence. ROESY spectra are used alternatively to NOESY spectra to obtain equivalent through-space correlations of nuclei.[52-54] In order to provide a context for sequential resonance assignments, NOESY spectra were recorded for several peptides with clear resolution of all relevant resonances (Figure 15). In sequential assignment, Hα of the first residue is correlated with H_N of the subsequent residue. That H_N then correlates with Hβ of the same residue (assignments that are confirmable via the TOCSY spectra), which then correlates with H_N of the subsequent residue. Sequential resonance assignments are broken at proline residues due to the absence of an amide hydrogen H_N. However, the proline Hα can be correlated to H_N of the subsequent residue. Sequential resonance assignments in these peptides can also assign the resonances of the N-terminal (acetyl, equivalent to an Hα position on the residue prior to the first residue) and C-terminal (carboxamide, equivalent to H_N on the residue after the final residue) capping groups.

Thus, for example, in the data herein, the sequential resonance assignments allow differentiation between isomeric Ac-ST-NH_2 and Ac-TS-NH_2 peptides, confirming or determining peptide sequence (Figure 16). The key correlations for sequential assignment are Hα of one residue (or the acetyl group) to H_N of the next residue. Thus, in Ac-ST-NH_2 the acetyl resonance exhibits an NOE with Ser H_N, while in Ac-TS-NH_2 the acetyl resonances has an NOE with the Thr H_N. The NOE between Ser Hα and Thr H_N (Figure 16a) confirms the Ac-ST-NH_2 sequence, while the NOE between Thr Hα and Ser H_N (Figure 16b) confirms the Ac-TS-NH_2 sequence. In both cases, the C-terminal residue shows a strong NOE to one carboxamide H_{N1} and

a weaker NOE to the other carboxamide H_{N2}, confirming both the identity of the C-terminal residue and which carboxamide hydrogen is cisoid (stronger NOE, equivalent to the position of an amide hydrogen in a subsequent amino acid) and which is transoid (weaker NOE, equivalent to the position of Cα in a subsequent amino acid).

Inclusion of raw and processed data. For all spectra within this manuscript, we provide the data in multiple formats. First, and most simply, we provide the processed spectra in a typical manner, so that students can analyze all of the data described above, for any given peptide, on their own. These spectra could be employed as problem set or exam questions, in which students are asked to make spectral assignments of key resonances, or are asked to identify the amino acids present within a given spectrum. Similarly, if students are given the values of "random coil" chemical shifts, for example as in the widely used tables from Wüthrich,[1] they can identify the presence and nature of any ordering present in the peptides via chemical shifts that deviate substantially from random coil values. $^{3}J_{\alpha N}$ values can also be employed similarly in instruction to identify order at individual amino acids.[5,15,55]

In our experience, one source of confusion for students is that observed chemical shifts normally do not exactly match the random-coil chemical shifts in the Wüthrich table.[1] In Table 1, we provide the chemical shift ranges observed for the peptides herein, emphasizing that typical chemical shifts in peptides and proteins have a much wider range than indicated single-value random-coil chemical shifts. Moreover, we emphasize that in more complex peptides and proteins, a substantially wider range of chemical shifts will be observed than are indicated here, due to the impacts of structure and interactions with those functional groups in peptides and proteins.

Alternatively, all spectra are also provided as raw FID data. These FID data could also be used as data sets to give students practice in topics such as baseline correction, apodization functions, zero filling, referencing, and other aspects of processing of 1-D or 2-D NMR data.

Conclusion

Herein, we report sets of 1-D and TOCSY ^{1}H NMR spectra of simple peptides (2-5 amino acids). NMR data are included for every amino acid within at least 5 different peptide contexts. The simple NMR spectra are intended to allow students to obtain an introductory understanding and familiarity with using NMR spectroscopy for peptides and proteins, using simple spectra that can be readily fully assigned and that lack the complications in NMR spectra of peptides and proteins. This manuscript was written to allow its application in teaching NMR spectroscopy in the classroom or for students in the research lab, at both the undergraduate and graduate levels. These data were originally collected from NMR spectra of peptides synthesized by students at the University of Delaware in the CHEM334 laboratory class ("Organic Chemistry Laboratory 2"), with NMR data interpretation conducted by students on peptides that they had synthesized. Additional discussion and the interpretation of the data across each peptide series (e.g., what trends were observed and what do those trends mean for protein structure) will be reported elsewhere.

Methods

Peptide synthesis. The synthesis, purification, and characterization of all peptides were described previously.[17]

1-D ^{1}H NMR spectroscopy. 1-D spectra were collected on either a Brüker 600 MHz NMR spectrometer equipped with a 5-mm Brüker SMART probe or a Brüker 600 MHz NMR spectrometer equipped with a cryogenic QNP probe. Experiments were conducted with water suppression using an excitation sculpting pulse sequence with gradients (zgesgp) and a relaxation delay of 1.7–2.0 s.[56] 32,768 or 65,536 data points were collected. Peptides were allowed to dissolve in buffer containing 5 mM phosphate, 25 mM NaCl, 90% H_2O/10% D_2O, and 10 μM TSP. Spectra were recorded on solutions at pH 4.0 (+/- 0.1) unless otherwise specified. 1-D spectra were recorded at 298 K unless the amide hydrogens did not resolve, in which case a temperature of 278 K was used. Spectra were zero-filled to fit the FID size. Exponential and gaussian apodization functions were fitted to the data, the Hz values for each of these functions were varied per spectra and ranged from 0-0.3 Hz for the exponential function and from 0-2.5 Hz for the gaussian function.

TOCSY NMR spectroscopy. Most TOCSY spectra were collected using a Brüker 600 MHz spectrometer equipped with a cryogenic QNP probe. Experiments were conducted using homonuclear Hartman-Hahn transfer using the MLEV17 sequence for mixing, using two power levels for excitation and spinlock, phase sensitive method. Water suppression was accomplished using a 3-9-19 pulse sequence with gradients, with 25% nonuniform sampling, and a relaxation delay of 1.7–2.0 s.[26,57,58] 256 data points were collected in the F1 dimension and 2048 data points were recorded in the F2 dimension. A smaller subset of TOCSY spectra were collected using a Brüker 600 MHz spectrometer equipped with a 5-mm Brüker SMART probe. Experiments were conducted using homonuclear Hartman-Hahn transfer using the DIPSI2 sequence for mixing, phase sensitive method. Water suppression was achieved using excitation sculpting with gradients, with 50% nonuniform sampling, and a relaxation delay of 1.7–2.0 s.[56,59] 512 data

points were recorded in the F1 dimension and 2048 data points were collected in the F2 dimension. Spectra were recorded at 298 K.

NOESY NMR spectroscopy. NOESY spectra were collected on a Brüker 600 MHz spectrometer equipped with a cryogenic QNP probe. Experiments were conducted using the homonuclear correlation via dipolar coupling, phase-sensitive method. Water suppression was achieved using a 3-9-19 pulse sequence with gradients, with 50% nonuniform sampling, and a relaxation delay of 2.0 s.[57,58] 400 data points were recorded in the F1 dimension and 2048 datapoints were collected in the F2 dimension. Spectra were recorded at 298 K. NOESY spectra were collected for the peptides Ac-ST-NH_2, Ac-DT-NH_2, Ac-PS-NH_2, Ac-QS-NH_2, Ac-TS-NH_2, Ac-WS-NH_2, Ac-SpT-NH_2, Ac-ApS-NH_2, Ac-WPPGY-NH_2, and Ac-NPPGY-NH_2.

Acknowledgements

We thank the students who synthesized the peptides herein, and who conducted initial data interpretation and whose feedback was central to the preparation and organization of this manuscript. Those students will be acknowledged individually in subsequent manuscripts on the scientific interpretation of the data. We thank NSF (CHE-2004110 and BIO-1616490) for funding. Instrumentation support was provided by NIH (GM110758, S10OD025185, P20GM104316, and S10RR026962) and NSF (CHE-1229234).

Supporting Information Available

Additional representative NMR spectral comparisons; and 1-D and TOCSY NMR spectra for all peptides, available both as processed spectra and as unprocessed data. This material is available free of charge via the journal web site.

Figures

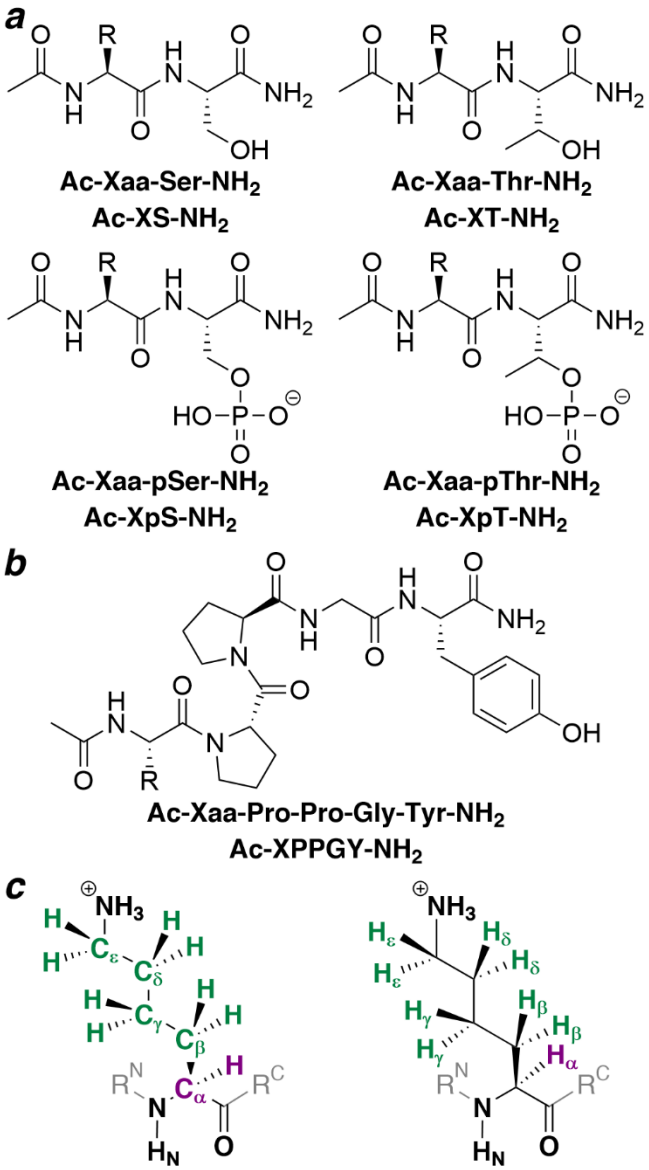

Figure 1. **Peptide nomenclature and identities**. (a) N-acetylated, C-terminal carboxamide dipeptides with serine (Ser/S), phosphoserine (pSer/pS), threonine (Thr/T), and phosphothreonine (pThr/pT). R indicates the Xaa/X amino acid side chain. (b) Pentapeptide backbone of the peptide Ac-Xaa-Pro-Pro-Gly-Tyr-NH_2 (Ac-XPPGY-NH_2). (c) Greek letter nomenclature for amino acid side-chain carbons (left) and hydrogens (right) with lysine (Lys/K) shown as an example. H_N = amide hydrogen, R^N = N-terminal amino acid sequence, R^C = C-terminal amino acid sequence.

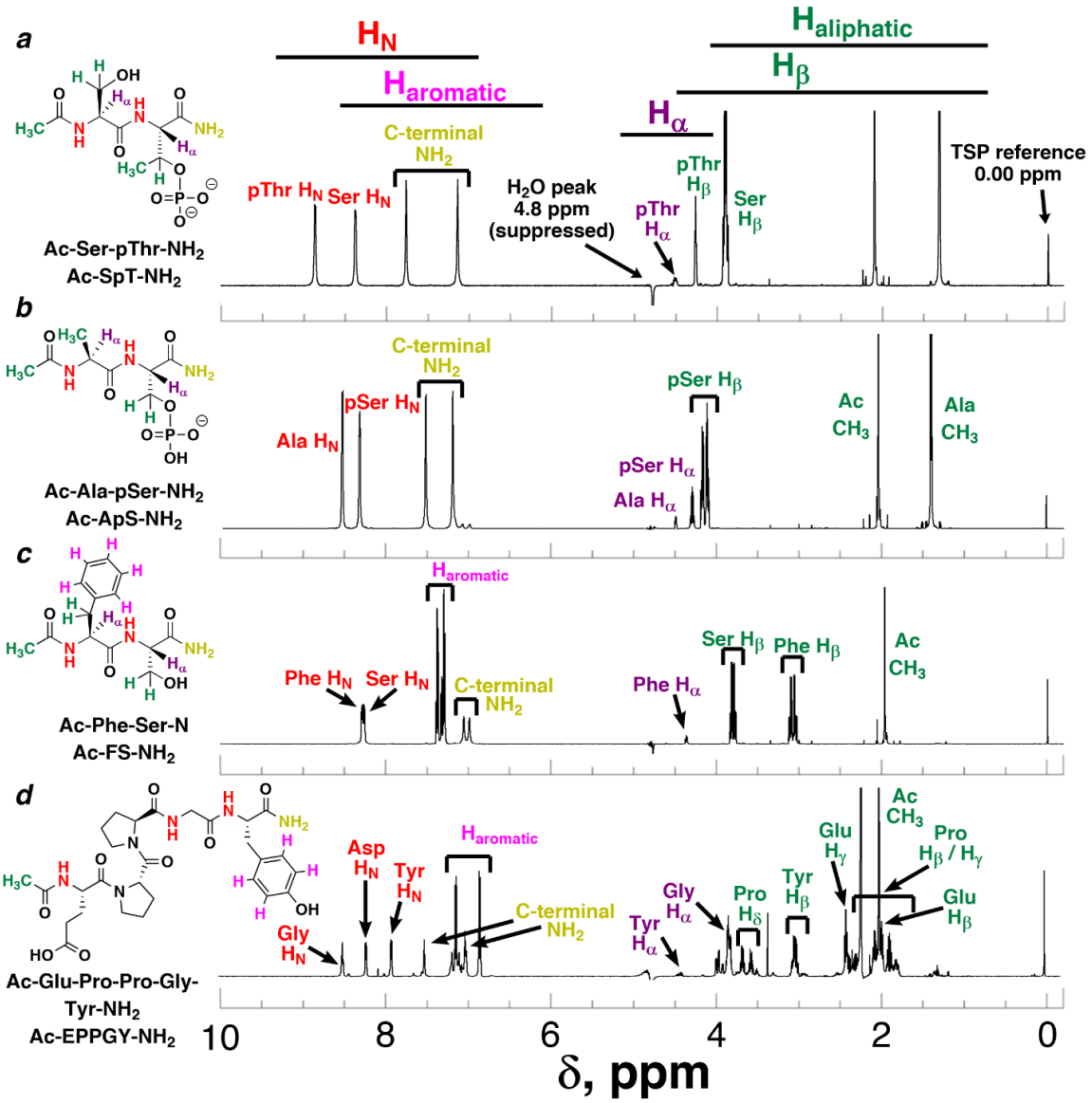

Figure 2. 1–D ^{1}H NMR spectra of selected peptides, with key spectra regions indicated and all identifiable resonances annotated. (a) NMR spectrum of the dipeptide Ac-serine-phosphothreonine-NH_2 (Ac-Ser-pThr-NH_2/Ac-SpT-NH_2) at pH 6.8. The following regions are shown above the 1–D spectrum: amide hydrogens (red), aromatic hydrogens (magenta), alpha hydrogens (purple), beta hydrogens, and aliphatic hydrogens (green), represented as H_N, $H_{aromatic}$, $H_{\square}$, $H_{\square}$, and $H_{aliphatic}$ respectively. Peaks from water and 3-(trimethylsilyl)propionic-2,2,3,3-2-d_4 acid sodium salt (TSP), which is used for referencing to 0.00 ppm, are also annotated. (b) Spectrum of Ac-alanine-phosphoserine-NH_2 (Ac-Ala-pSer-NH_2/Ac-ApS-NH_2) at pH 4.0. (c) Spectrum of Ac-phenylalanine-serine-NH_2 (Ac-Phe-Ser-NH_2/Ac-FS-NH_2) at pH 4.0. (d) Spectrum of Ac-glutamate-proline-proline-glycine-tyrosine-NH_2 (Ac-Glu-Pro-Pro-Gly-Tyr-NH_2/Ac-EPPGY-NH_2) at pH 4.0. All NMR spectra were obtained using a solution of the peptide in 90% H_2O/10% D_2O containing 5 mM sodium phosphate pH 4.0 or 6.8 and 25 mM NaCl at 298 K, with 100 μM TSP as an internal reference.

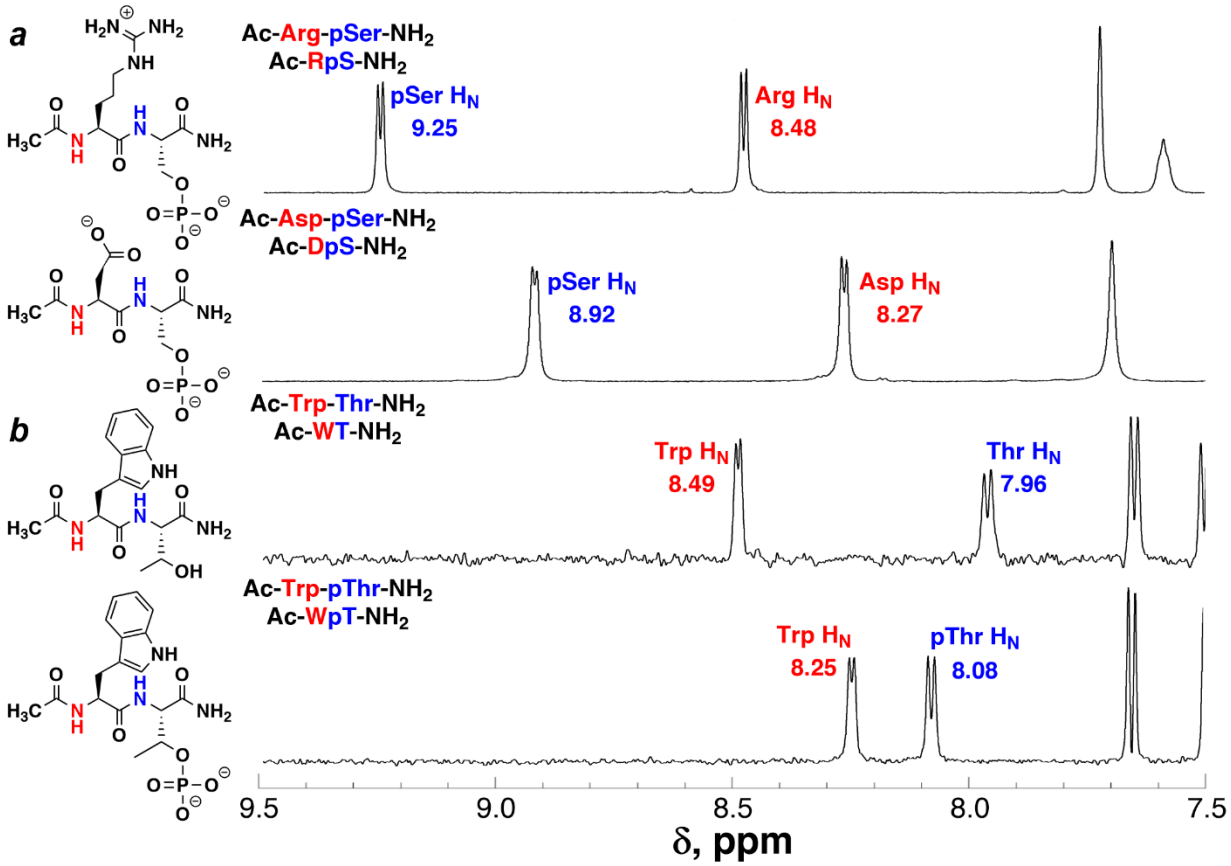

Figure 3. **Amide hydrogen chemical shift dispersion**. (a) Amide regions from the ^{1}H NMR spectra of Ac-Arg-pSer-NH_2 (Ac-RpS-NH_2) and Ac-Asp-pSer-NH_2 (Ac-DpS-NH_2) at pH 6.8, with amide resonances identified by amino acid. These spectra demonstrate large chemical shift differences between the amide hydrogens of the phosphoserine and the N-terminal amino acid residue, which is associated with order at phosphoserine. (b) Amide regions from the ^{1}H NMR spectra of Ac-Trp-Thr-NH_2 (Ac-WT-NH_2) and Ac-Trp-pThr-NH_2 (Ac-WpT-NH_2) at pH 4.0 and 6.8, respectively. Amide resonances are labeled with their corresponding amino acid. These spectra demonstrate smaller chemical shift dispersion between the amide hydrogens, but also indicate how amide hydrogen chemical shifts can change due to changes in local structure (e.g. as a result of changes to the adjacent amino acid).

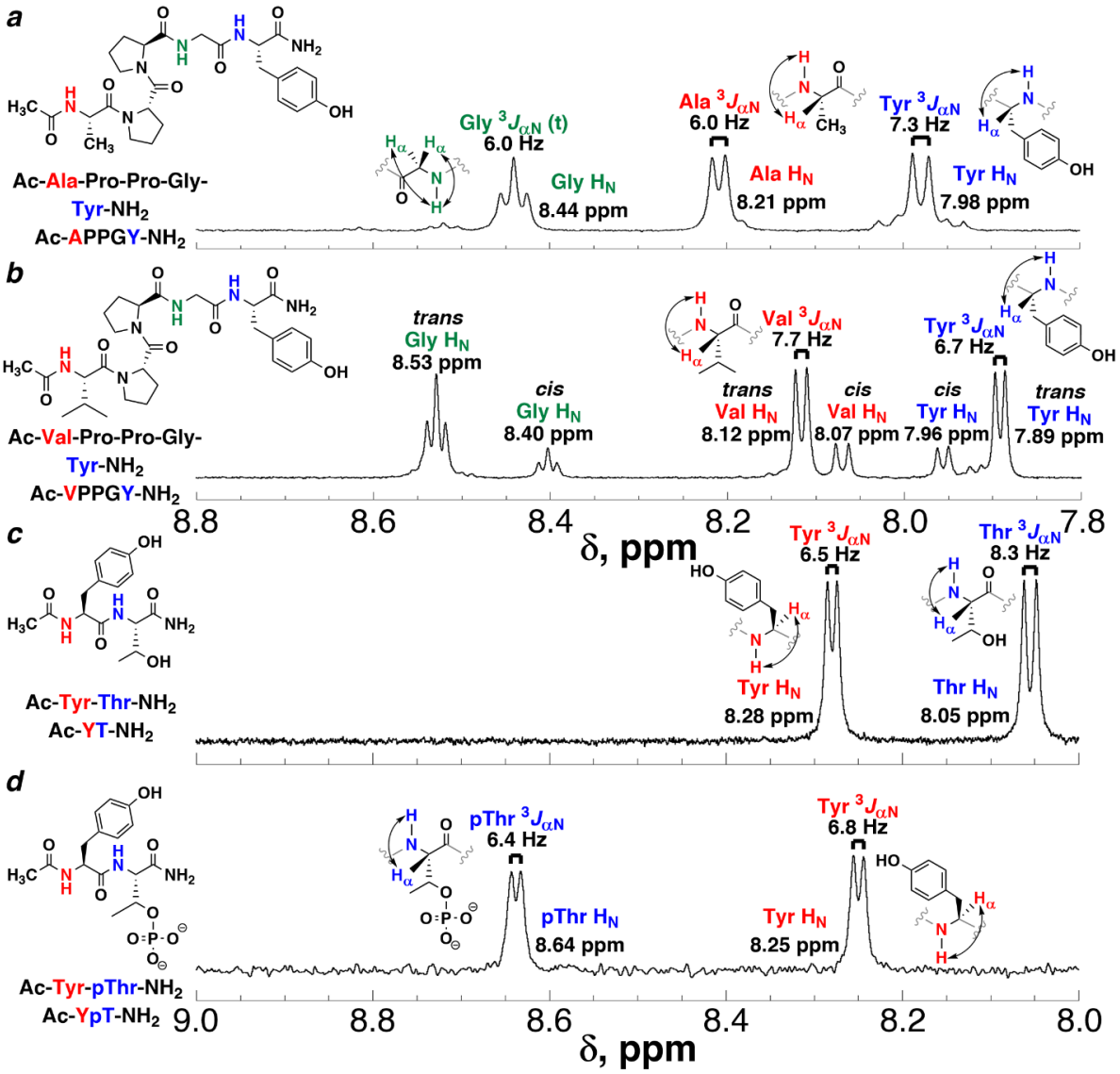

Figure 4. H_N–H_α $^3J_{\alpha N}$ coupling constants. (a) Amide-aromatic regions of the ^{1}H NMR spectrum of Ac-Ala-Pro-Pro-Gly-Tyr-NH_2 (Ac-APPGY-NH_2) with the chemical shifts of all amide hydrogens annotated. H_α–H_N coupling constants ($^3J_{\alpha N}$) are listed, with brackets depicting how this coupling constant was measured. Diagrams shown demonstrate the source of $^3J_{\alpha N}$ coupling and explain the multiplicities observed for each signal. $^3J_{\alpha N}$ values serve as an indication of average ϕ angle values at a particular residue, via a parametrized Karplus relationship (Figure S2).[15] (b) Spectrum of Ac-Val-Pro-Pro-Gly-Tyr-NH_2 (Ac-VPPGY-NH_2). This peptide exhibits distinct signals from both the *trans*-proline and *cis*-proline amide bond conformational isomers, resulting in two peaks corresponding to each amide hydrogen found in the structure. (c) Spectrum of Ac-Tyr-Thr-NH_2 (Ac-YT-NH_2). (d) Spectrum of Ac-Tyr-pThr-NH_2 (Ac-YpT-NH_2). In conjunction with (c), this spectrum serves to highlight the effect of phosphorylation on $^3J_{\alpha N}$ values and therefore the role of phosphorylation in changing protein structure.

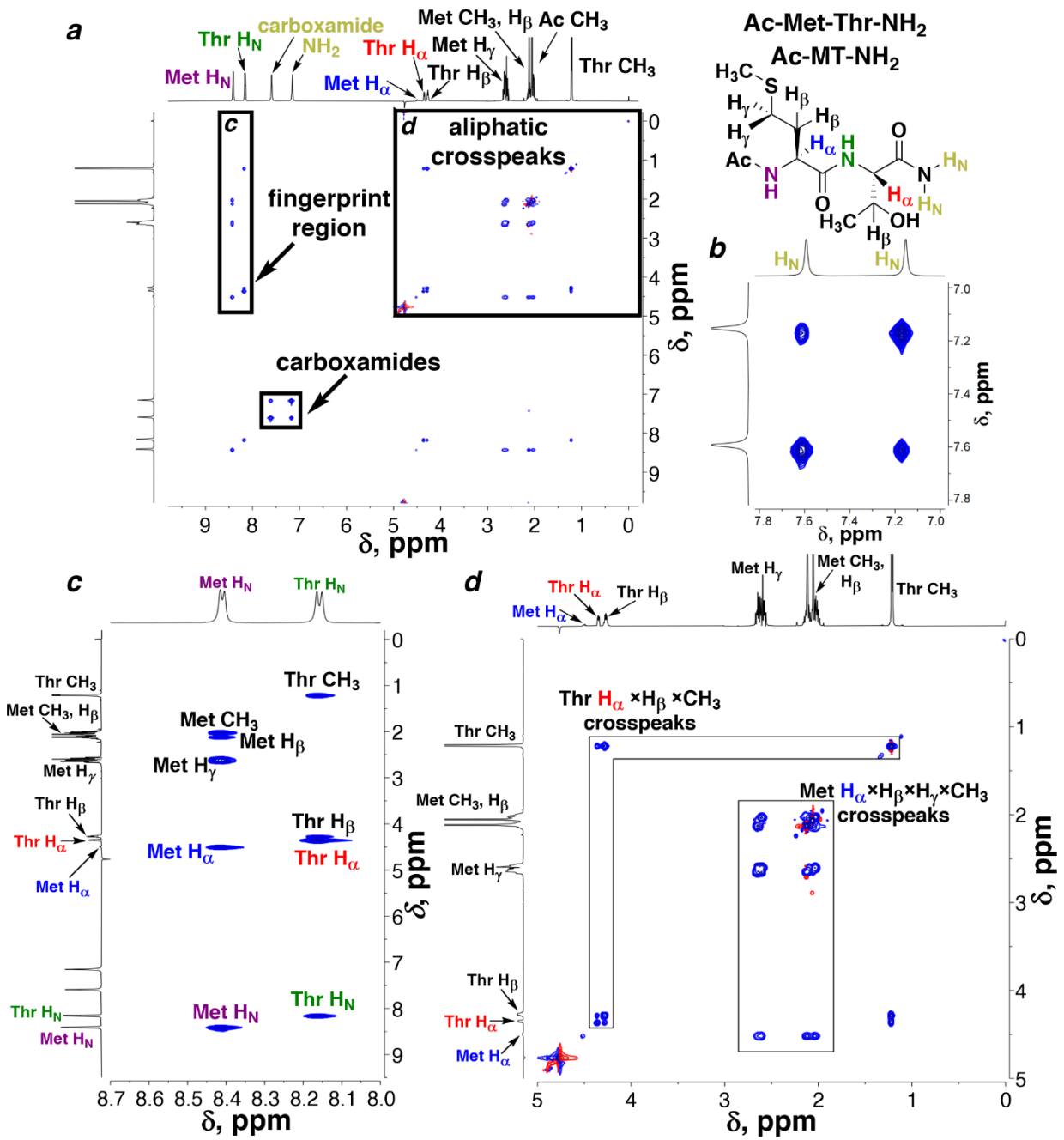

Figure 5. **Interpretation of total correlation spectroscopy (TOCSY) spectra**. (a) Full TOCSY spectrum of Ac-Met-Thr-NH_2 (Ac-MT-NH_2), with the 1-D ^{1}H NMR spectrum annotated on the x-axis. Boxes indicate regions of the spectrum further analyzed in subsequent parts of this Figure. (b) Carboxamide hydrogens in the TOCSY spectrum, showing the characteristic pattern of these hydrogens. (c) Fingerprint region of the TOCSY spectrum, showing each amide hydrogen and the crosspeaks corresponding to the side chains of each amino acid. (d) Aliphatic region of the TOCSY spectrum, with crosspeaks from each side chain indicated in separate boxes.

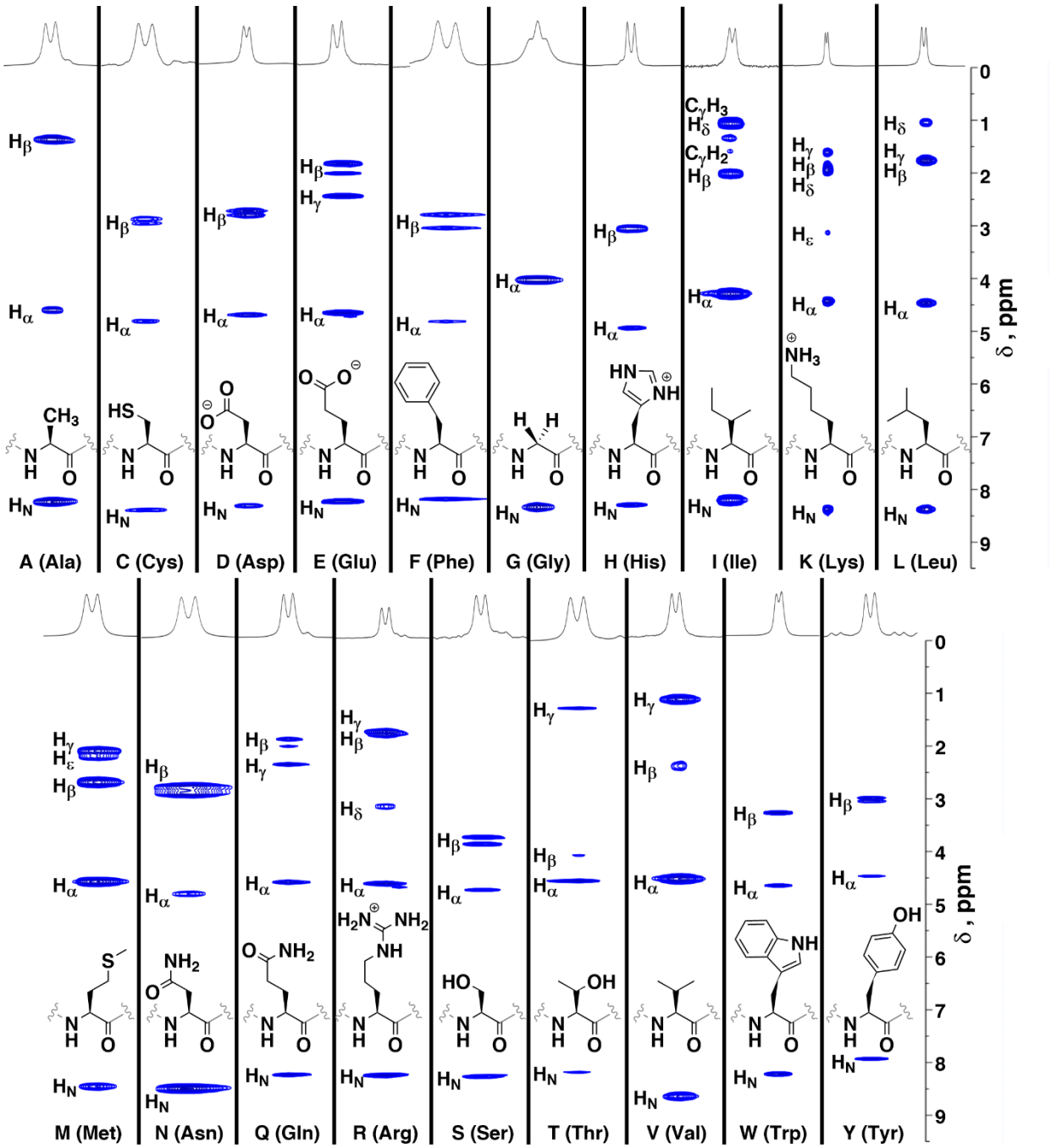

Figure 6. **Side-chain correlations to the amide hydrogen observed in the fingerprint region of TOCSY spectra**. Vertical slices at the H_N resonance from the TOCSY spectra of all 19 amino acids bearing an amide hydrogen. These slices show the characteristic crosspeak patterns observed in the fingerprint region of TOCSY spectra for each non-proline amino acid.

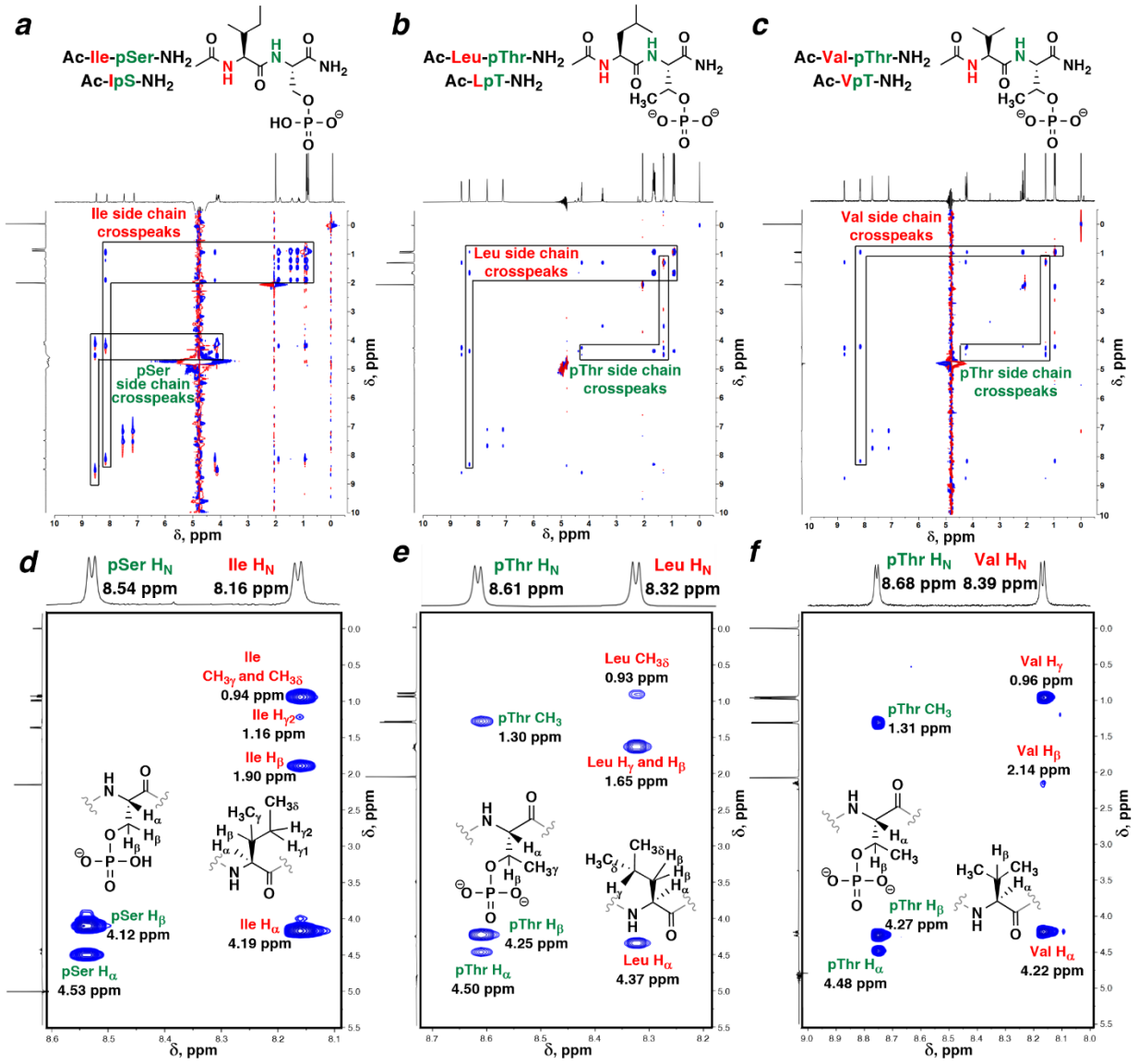

Figure 7. **Identification and differentiation of amino acids with aliphatic side chains**. (a-c) Full TOCSY spectra with side-chain crosspeaks identified by amino acid. Spectra were recorded at pH 6.8 for Ac-IpS-NH_2 and Ac-VpT-NH_2, or at pH 4.0 for Ac-LpT-NH_2. (d-f) Expansion of the fingerprint regions to highlight individual correlations. Ac-IpS-NH_2 shows a peak from $H_{\square}$ at 1.47 ppm in the aliphatic region of the TOCSY spectrum, but this hydrogen is not observed as a crosspeak with the isoleucine amide hydrogen. Leucine crosspeaks show overlap of peaks from $H_{\square}$ and $H_{\square}$, resulting in a complex pattern of peaks at this chemical shift in the ^{1}H NMR spectrum.

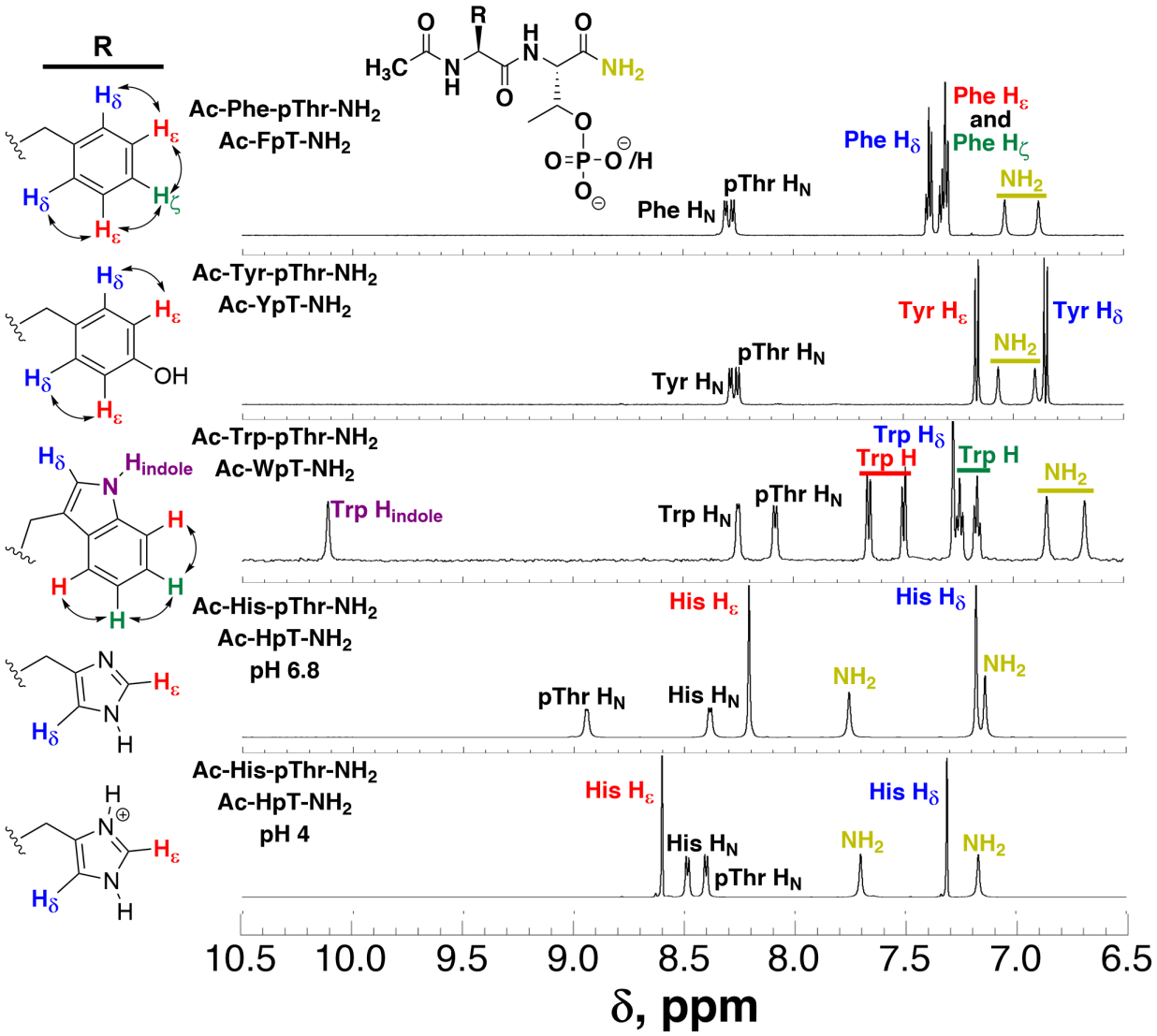

Figure 8. **Aromatic regions of the ^{1}H NMR spectra for all four aromatic amino acids, showing the typical chemical shifts and distinct coupling patterns observed for each**. At pH 6.8, His is present as a mixture of cationic and neutral forms, while at pH 4 His is almost exclusively in the cationic form; the differences in chemical shifts of the hydrogens are consistent with changes in populations of the His ionization states.

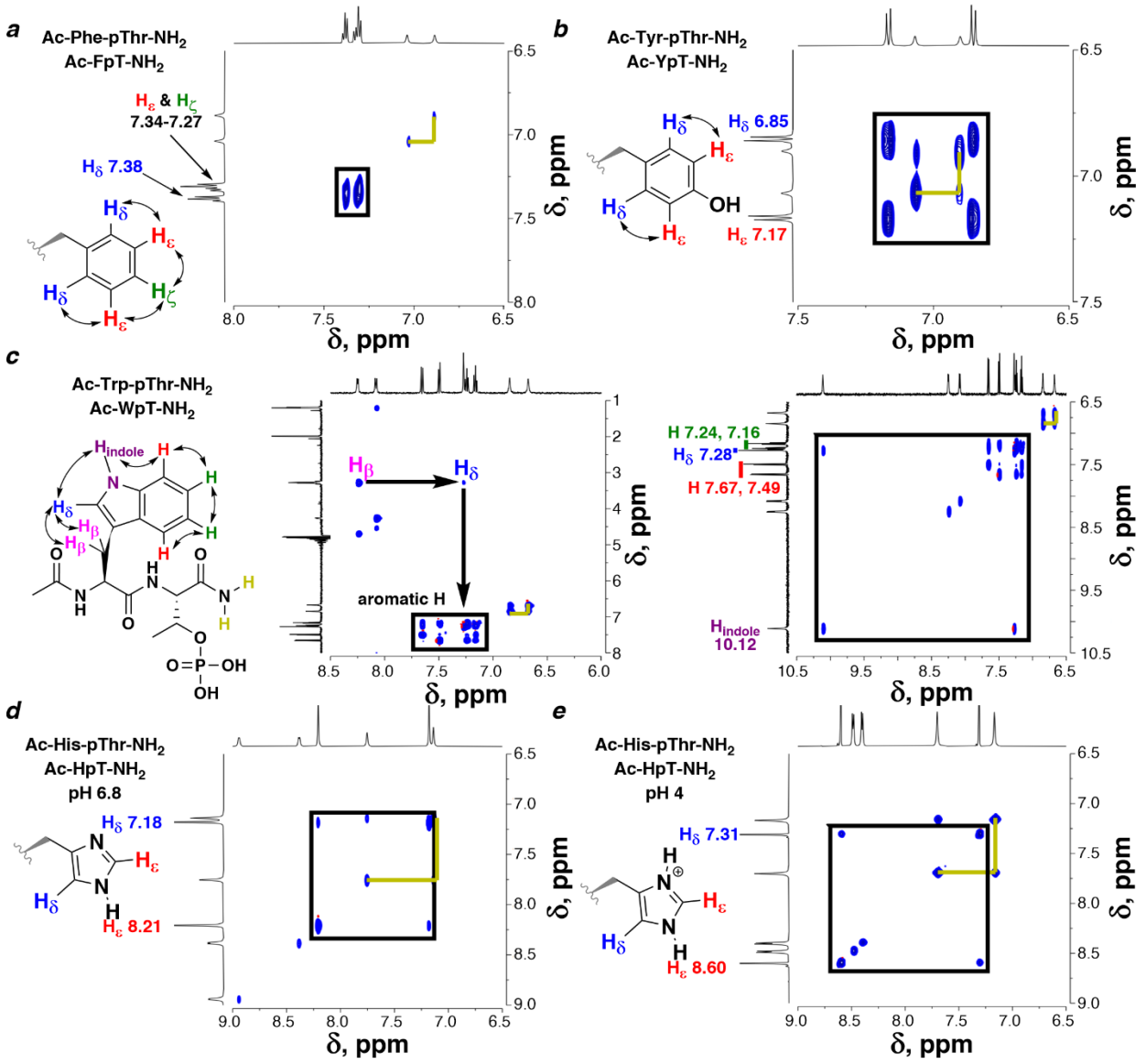

Figure 9. **TOCSY spin systems of aromatic amino acid side chains**. (a,b) Aromatic regions from the TOCSY spectra of Ac-FpT-NH_2 and Ac-YpT-NH_2 (pH 4.0). The aromatic region of the TOCSY spectrum demonstrates the signature crosspeak patterns for these amino acids. (c) Spectra corresponding to Ac-WpT-NH_2 (pH 4.0) showing the crosspeaks between $H_{\square}$ and $H_{\square}$ (left), allowing for definitive determination of aromatic amino acids via correlation to peaks from the TOCSY fingerprint region. (d,e) Aromatic regions from the TOCSY spectra of Ac-HpT-NH_2 at pH 6.8 and 4.0, respectively, which show the effects of changes in ionization state at His residues on the chemical shifts. Carboxamide peaks often overlap with aromatic peaks; as such, the carboxamide resonances have been indicated by gold labels in the spectra, showing one side of the box-like signals observed from these hydrogens.

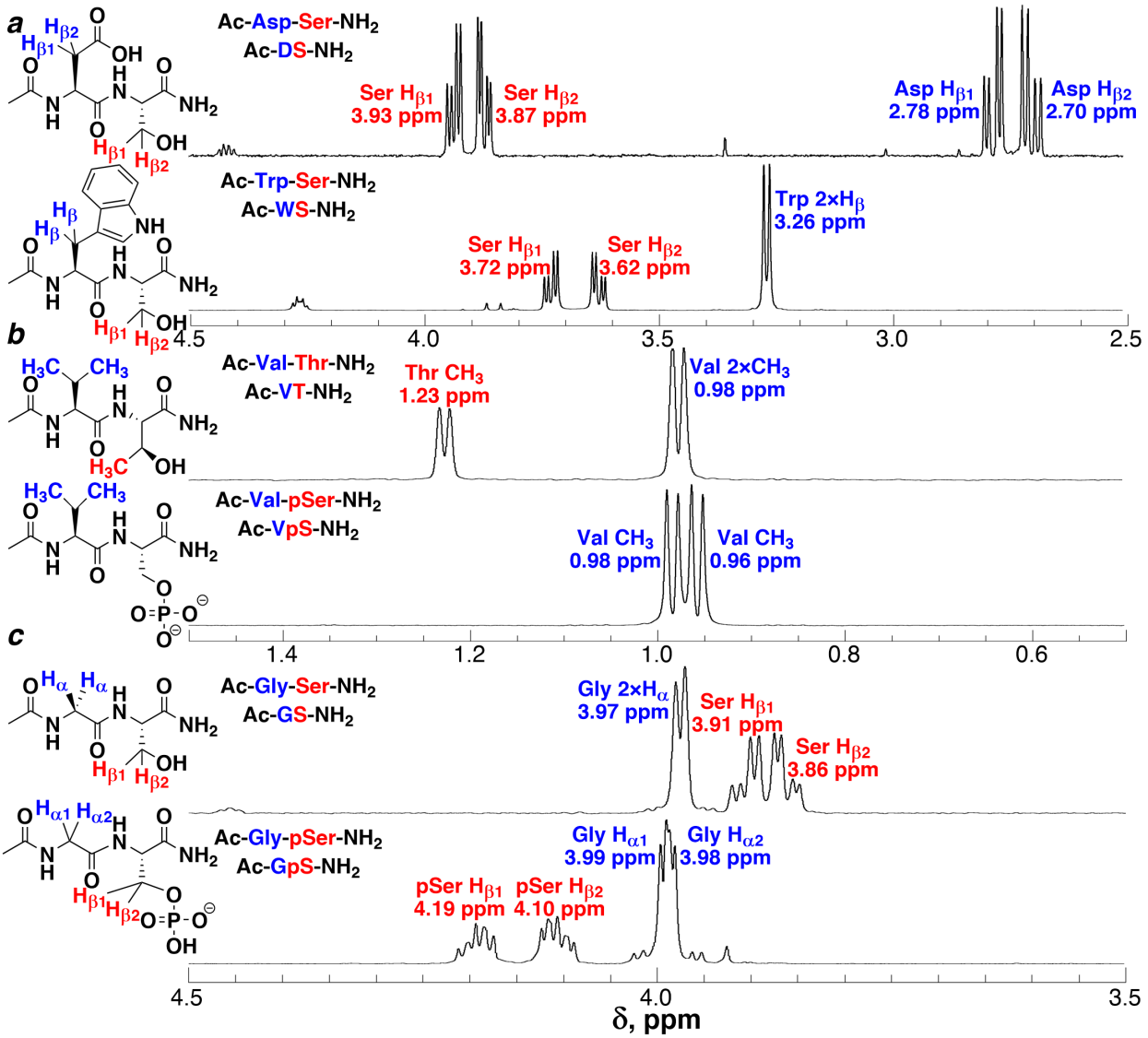

Figure 10. **Diastereotopic hydrogens and diastereotopic methyl groups**. (a) $H_{\square}$ of selected amino acids, demonstrating two distinct signals at Ser and Arg, with the diastereotopic $H_{\square}$ readily distinguishable for each. However, the $H_{\square}$ of tryptophan (Trp) are indistinguishable from one another. These spectra highlight the two limiting cases observed for any set of diastereotopic hydrogens. (b) The ^{1}H NMR spectrum of Ac-Val-Thr-NH_2 shows the methyl groups of valine as one doublet. However, in the spectrum of Ac-Val-pSer-NH_2, these diastereotopic methyl groups show distinct signals as two separate doublets. (c) The $H_{\square}$ of glycine (Gly) are also diastereotopic and can present as a single doublet (in Ac-Gly-Ser-NH_2), or as a pair of doublets, overlapping in Ac-Gly-pSer-NH_2; in both of these spectra, second-order effects are also observed. The $H_{\square}$ of Ser are distinct in both glycine-containing spectra. The NMR spectra of Ac-Asp-Ser-NH_2, Ac-Trp-Ser-NH_2, Ac-Val-Thr-NH_2, Ac-Gly-Ser-NH_2, and Ac-Gly-pSer-NH_2 were recorded at pH 4.0, while the spectrum of Ac-Val-pSer-NH_2 was recorded at pH 6.8. Assignments of H_1 and H_2 correspond to the signal with the most downfield and most upfield chemical shift, respectively, and are not an indication of stereochemistry.

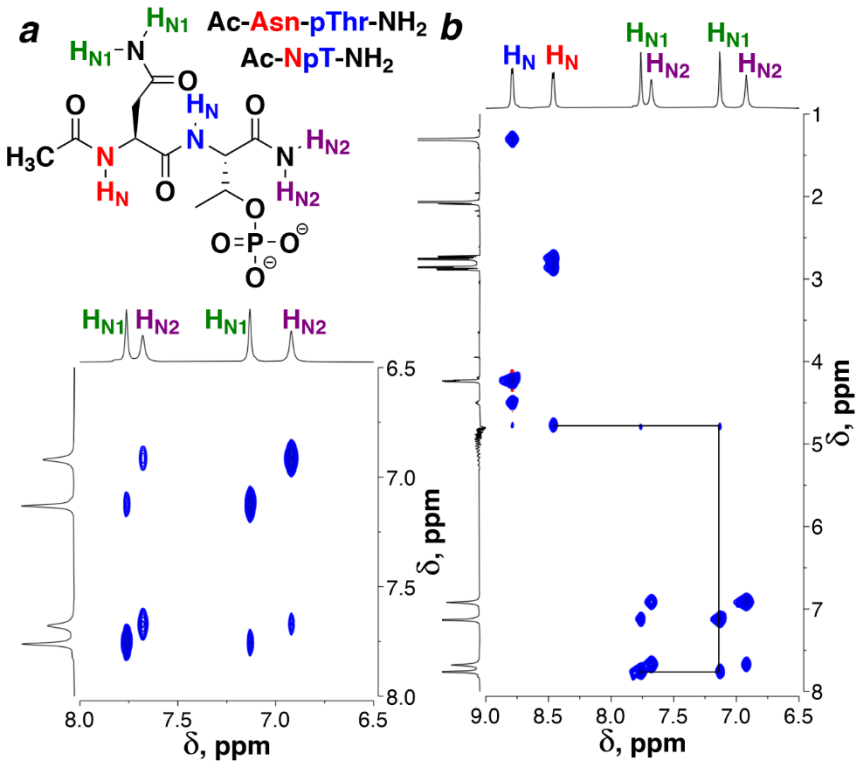

Figure 11. **Carboxamide hydrogens in the TOCSY spectrum**. (a) Carboxamide region of the TOCSY spectrum of Ac-Asn-pThr-NH_2, where correlations between both pairs of carboxamide hydrogens can be seen clearly. C-terminal carboxamide peaks are shown in purple. (b) Fingerprint region of the TOCSY spectrum, with crosspeaks between $H_{\square}$ of Asn and its carboxamide hydrogens (green).

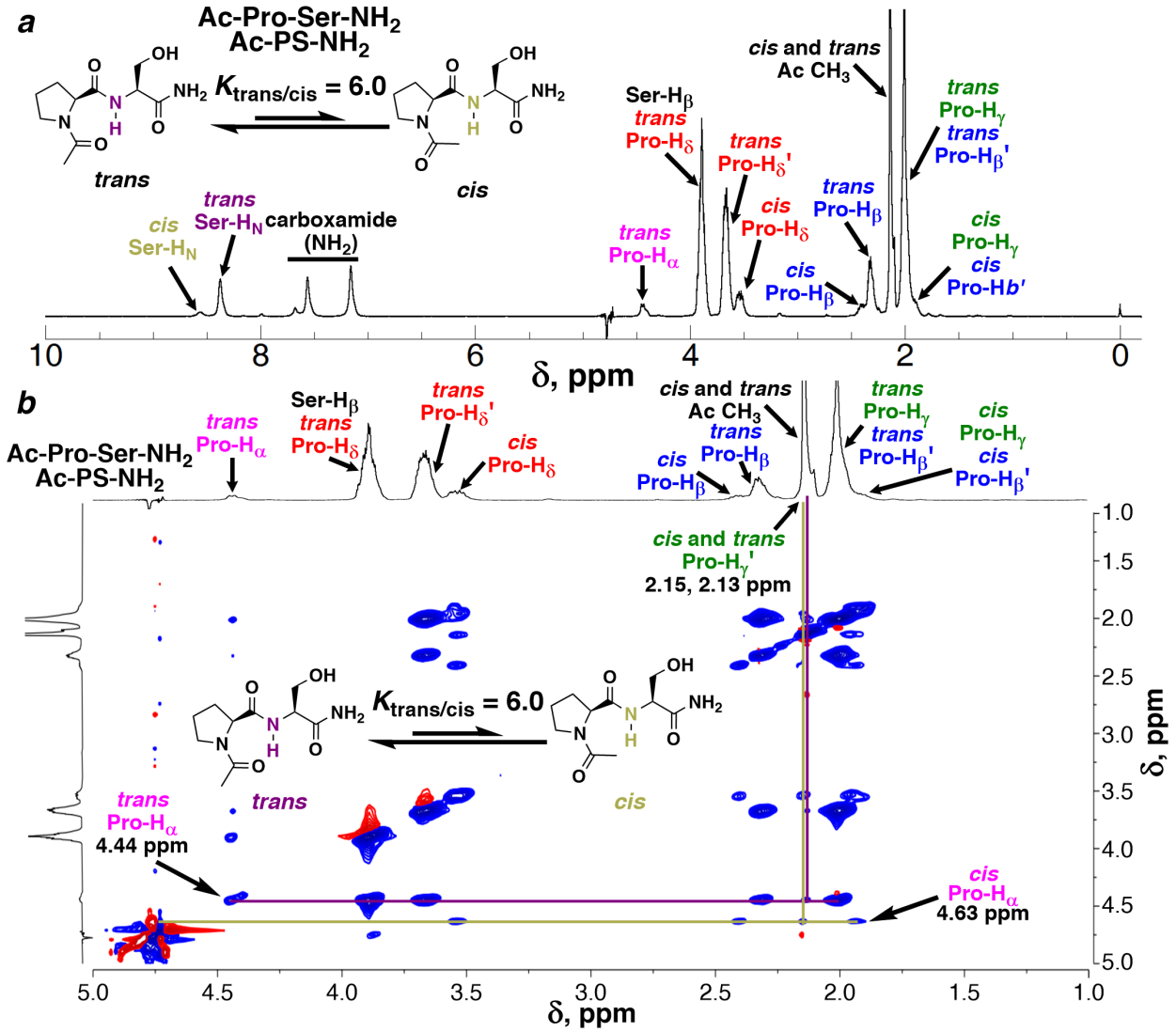

Figure 12. NMR spectra of proline. (a) ^{1}H NMR spectrum of Ac-Pro-Ser-NH$_2$, showing signals from conformations with both *trans*-proline and *cis*-proline. The equilibrium constant $K_{trans/cis}$ was calculated from the ratio of integration of resolved peaks from *trans*-proline to peaks from *cis*-proline. (b) Aliphatic region of the ^{1}H NMR spectrum of Ac-Pro-Ser-NH$_2$ with *cis* (gold) and *trans* (purple) H$_{\square}$ resonances identified. Crosspeaks allow for the determination of chemical shifts of peaks obscured in the 1-D spectra, as is seen with *cis* H$_{\square}$, *cis* H$_{\square}$' and *trans* H$_{\square}$'.

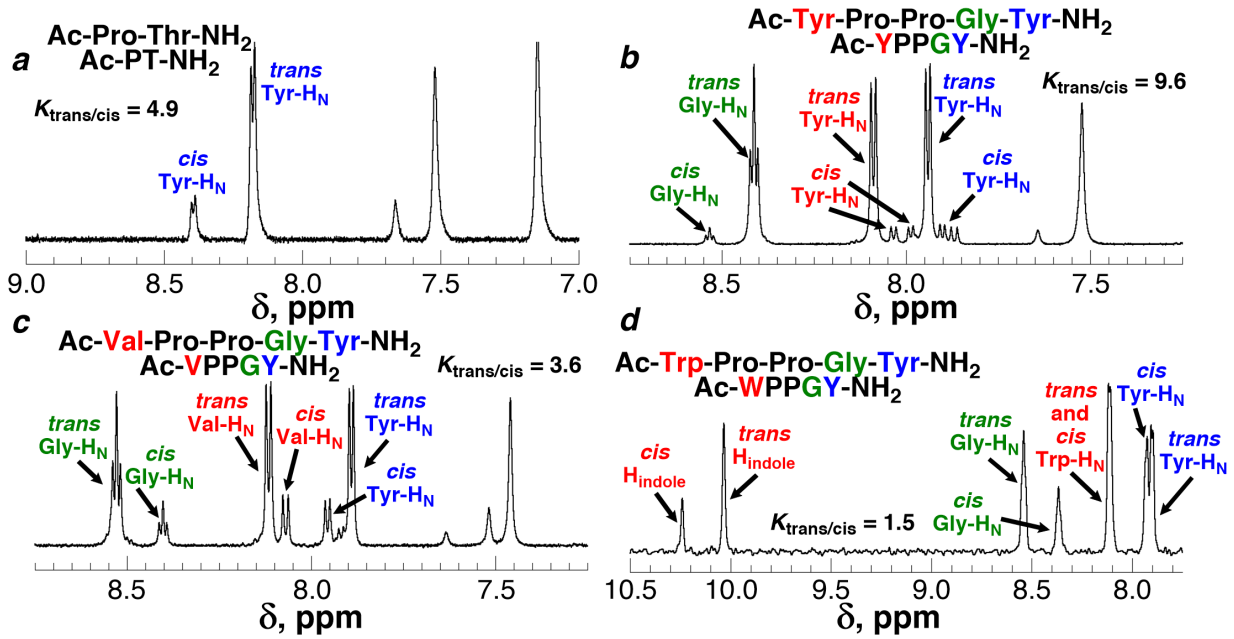

Figure 13. **Proline *cis-trans* isomerism**. Amide regions of the ^{1}H NMR spectra of the peptides Ac-Pro-Thr-NH_2, Ac-Tyr-Pro-Pro-Gly-Tyr-NH_2, Ac-Val-Pro-Pro-Gly-Tyr-NH_2 and Ac-Trp-Pro-Pro-Gly-Tyr-NH_2, showing peaks from both conformational isomers of the proline amide bond. These spectra show the ranges of values of $K_{trans/cis}$ that might be observed in NMR spectra of peptides containing proline.

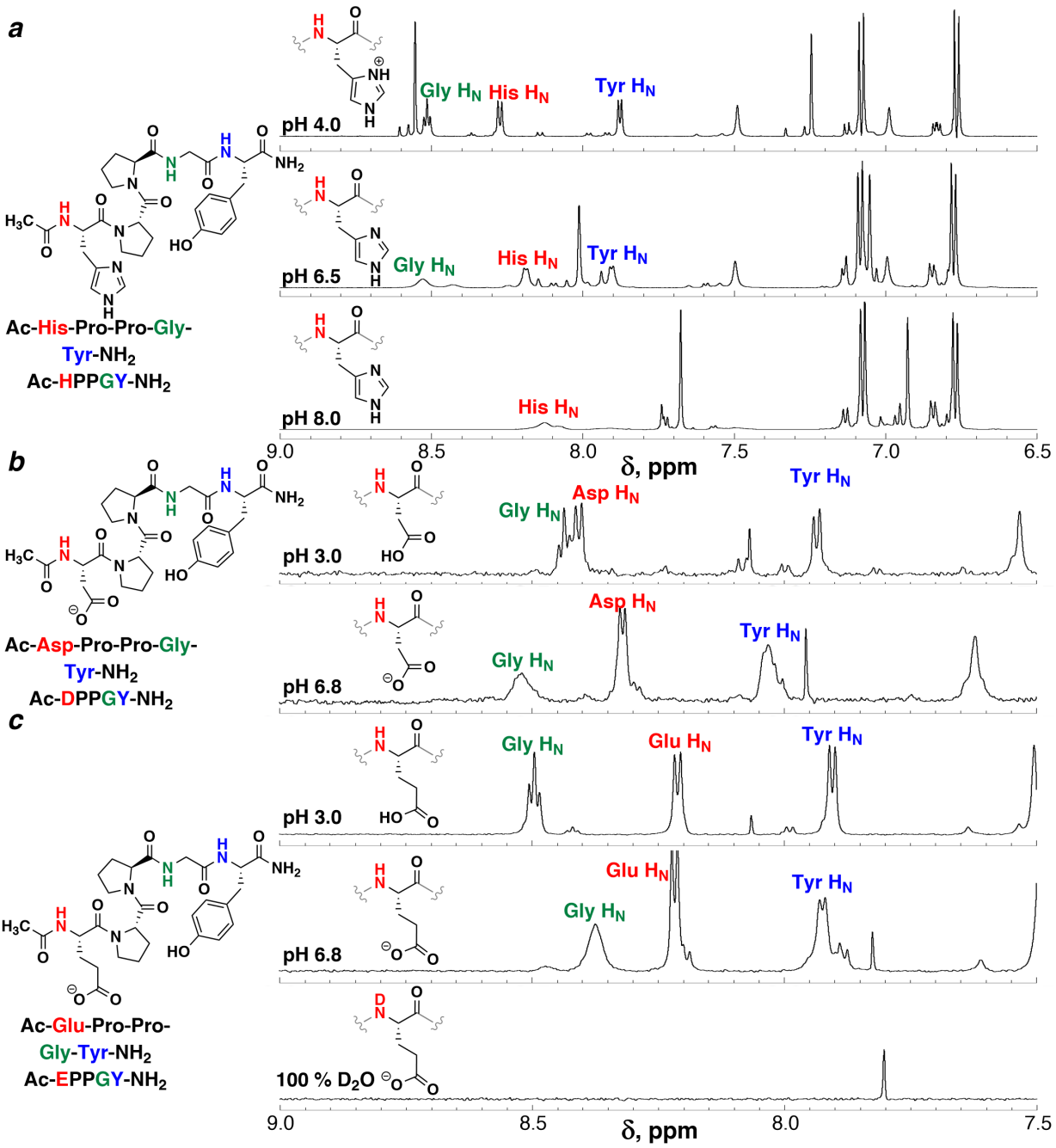

Figure 14. **Changes in ionization state and hydrogen exchange dynamics**. (a) Amide-aromatic region of the ^{1}H NMR spectrum of Ac-His-Pro-Pro-Gly-Tyr-NH_2 at pH 4.0, 6.5, and 8.0. Amide hydrogen exchange is faster with increasing [HO^-] (increasing pH), and results in broadening or disappearance of amide peaks due to intermediate and/or fast exchange of amide hydrogens on the NMR timescale. (b,c) Ionizable side chains, such as carboxylic acids, also exhibit chemical shift changes in all resonances due to changes in ionization state of the side chain. In 100% D_2O, no amide hydrogen resonances are observed, due to hydrogen-deuterium exchange (conversion of amide N–H to N–D).

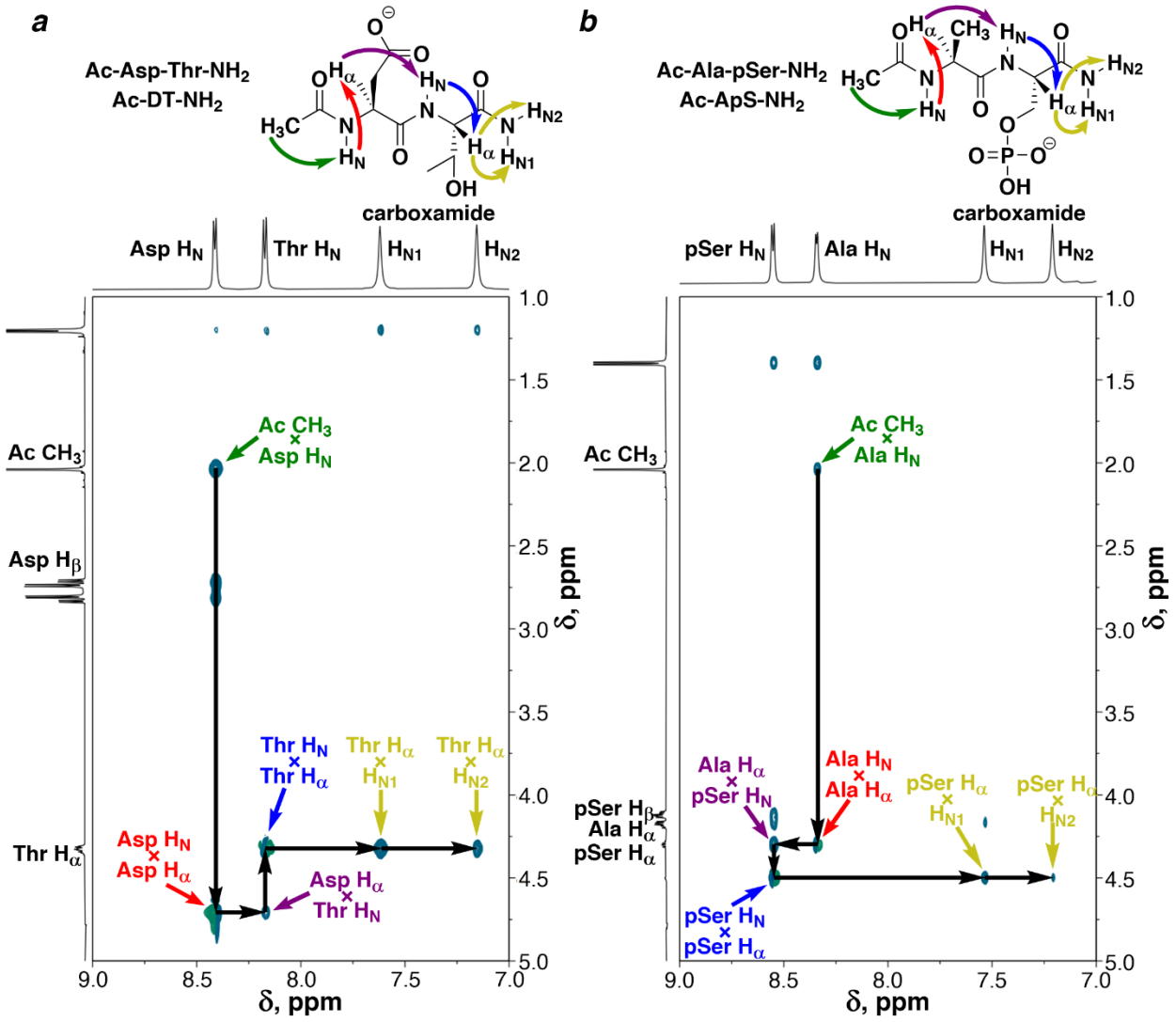

Figure 15. Sequential resonance assignment (resonance walks) using Nuclear Overhauser Effect Spectroscopy (NOESY) spectra. (a) NOESY spectrum of Ac-Asp-Thr-NH_2, where the crosspeaks correspond to through-space couplings that allow for correlations to be made between consecutive residues, which are close in space due to the covalent bonds between residues. These crosspeaks show correlations between the acetyl methyl group and the Asp H_N (green), between the Asp H_N and the Asp $H_{\square}$ (red, also seen in TOCSY spectra), between the Asp $H_{\square}$ and the Thr H_N (purple), between the Thr H_N and the Thr $H_{\square}$ (blue), and between the Thr $H_{\square}$ and the C-terminal carboxamide hydrogens (gold). The intensities of these crosspeaks are also indicative of the proximity of one hydrogen to another and allow for definitive assignments of each carboxamide signal (H_{N1} and H_{N2}, gold). (b) NOESY spectrum of Ac-Ala-pSer-NH_2, showing crosspeaks from the Ac- group to Ala H_N (green), Ala H_N to Ala $H_{\square}$ (red), Ala $H_{\square}$ to pSer H_N (purple), Ser H_N to pSer $H_{\square}$(blue), and from pSer $H_{\square}$ to the C-terminal carboxamide hydrogens H_{N1} and H_{N2} (gold), respectively.

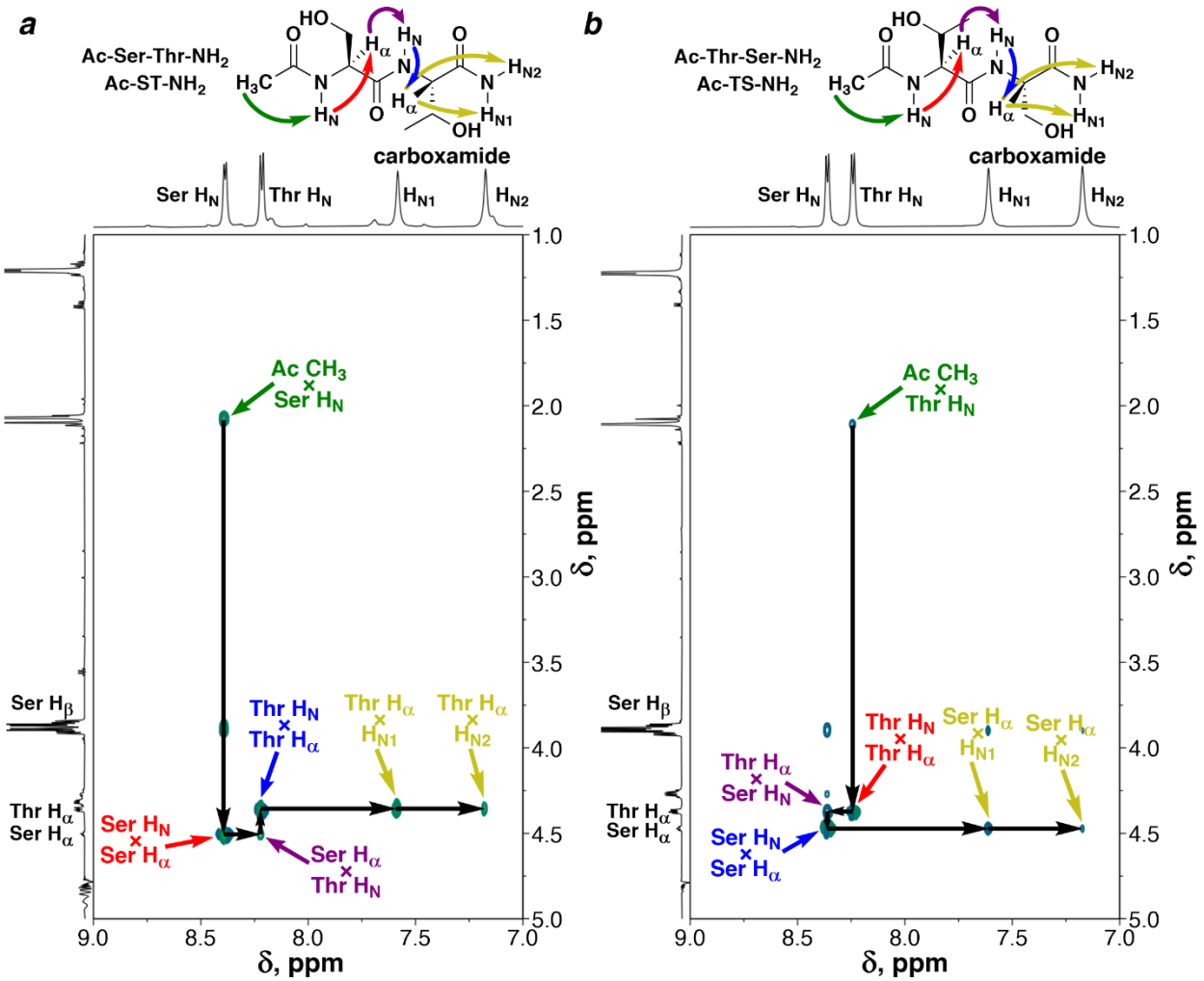

Figure 16. Differentiation of isomeric sequences via NOESY spectra and sequential resonance assignments. Through the systematic application of NOESY spectra, the order of amino acids in each peptide is identified, allowing for the differentiation of Ac-Ser-Thr-NH_2 from Ac-Thr-Ser-NH_2 via the crosspeaks observed in each NOESY spectrum.

Table

Table 1. Observed chemical shift ranges (ppm) for all amino acids in this library of peptide spectra.

residue	δH_N	δH_α	δH_β	δH_γ	δH_δ	δH_ε	other
A, Ala	8.5-8.1	4.6-4.2	1.5-1.2	–	–	–	–
C, Cys	8.5-8.4	4.6-4.3	3.0-2.7	–	–	–	–
D, Asp	8.5-8.2	4.9-4.6	2.9-2.3	–	–	–	–
E, Glu	8.4-8.2	4.7-4.2	2.2-1.7	2.5-2.0	–	–	–
F, Phe	8.4-8.1	4.8-4.6	3.3-2.7	–	–	–	–
G, Gly	8.6-8.2	4.1-3.8	–	–	–	–	–
H, His	8.5-8.0	4.9-4.6	3.3-2.7	–	7.4-6.9	8.6-7.6	–
I, Ile	8.5-8.0	4.4-4.1	2.0-1.7	1.6-1.1	1.0-0.8	–	CH_3 1.0-0.8
K, Lys	8.6-8.1	4.6-4.3	1.9-1.4	1.7-1.4	1.9-1.5	3.9-3.0	NH_3 7.6-7.4
L, Leu	8.4-8.2	4.4-4.3	1.7-1.4	1.7-1.4	1.0-0.8	–	–
M, Met	8.5-8.2	4.8-4.4	2.2-1.9	2.7-2.5	–	2.6-2.0	–
N, Asn	8.5-8.3	5.0-4.7	3.0-2.7	–	–	–	NH_2 7.6-6.8
P, Pro	–	5.0-4.3	2.5-1.8	2.2-1.5	4.0-3.5	–	–
Q, Gln	8.4-8.2	4.6-4.3	2.2-1.8	2.5-1.9	–	–	NH_2 7.6-6.8
R, Arg	8.4-8.1	4.6-4.2	2.0-1.6	1.8-1.5	3.3-3.0	–	
S, Ser	8.8-7.9	4.8-4.2	4.4-4.0	–	–	–	–
T, Thr	8.4-7.9	4.7-4.1	4.4-4.0	1.3-1.0	–	–	–
V, Val	8.3-7.8	4.5-4.0	3.1-2.1	1.0-0.9	–	–	–
W, Trp	8.3-7.9	4.8-4.6	3.4-3.1	–	7.3-7.2	–	$H_{aromatic}$ 7.7-7.1 H_{indole} 10.3-10.0
Y, Tyr	8.3-7.9	4.7-4.4	3.2-2.6	–	7.2-7.1	6.9-6.8	–
pS, pSer	9.3-8.0	4.9-4.3	4.4-4.0	–	–	–	–
pT, pThr	9.5-8.0	4.8-4.3	4.4-4.0	1.4-1.2	–	–	–
Ac-							2.1-1.9
-NH_2							7.9-6.9

References

(1) Wüthrich, K.: *NMR of Proteins and Nucleic Acids*; Wiley-Interscience, 1986.

(2) Johnson, W. C., Jr. Analysis of Circular Dichroism Data. *Methods Enzymol.* **1992**, *210*, 426-447.

(3) Cavanagh, J.; Fairbrother, W. J.; Palmer III, A. G.; Skelton, N. J.: *Protein NMR Spectroscopy*; Academic Press: San Diego, 1996.

(4) Vuister, G. W.; Delaglio, F.; Bax, A. An Empirical Correlation Between $^{1}J_{CaHa}$ and Protein Backbone Conformation. *J. Am. Chem. Soc.* **1992**, *114*, 9674-9675.

(5) Wishart, D. S.; Sykes, B. D.; Richards, F. M. The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **1992**, *31*, 1647-1651.

(6) Wishart, D. S.; Sykes, B. D. The C-13 Chemical Shift Index - A Simple Method for the Identification of Protein Secondary Structure Using C-13 Chemical-Shift Data. *J. Biomol. NMR* **1994**, *4*, 171-180.

(7) Eliezer, D. Biophysical characterization of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2009**, *19*, 23-30.

(8) Jensen, M. R.; Salmon, L.; Nodet, G.; Blackledge, M. Defining Conformational Ensembles of Intrinsically Disordered and Partially Folded Proteins Directly from Chemical Shifts. *J. Am. Chem. Soc.* **2010**, *132*, 1270-+.

(9) Jensen, M. R.; Ruigrok, R. W. H.; Blackledge, M. Describing intrinsically disordered proteins at atomic resolution by NMR. *Curr. Opin. Struct. Biol.* **2013**, *23*, 426-435.

(10) Theillet, F. X.; Smet-Nocca, C.; Liokatis, S.; Thongwichian, R.; Kosten, J.; Yoon, M. K.; Kriwacki, R. W.; Landrieu, I.; Lippens, G.; Selenko, P. Cell signaling, post-translational protein modifications and NMR spectroscopy. *J. Biomol. NMR* **2012**, *54*, 217-236.

(11) Theillet, F. X.; Binolfi, A.; Frembgen-Kesner, T.; Hingorani, K.; Sarkar, M.; Kyne, C.; Li, C. G.; Crowley, P. B.; Gierasch, L.; Pielak, G. J.; Elcock, A. H.; Gershenson, A.; Selenko, P. Physicochemical Properties of Cells and Their Effects on Intrinsically Disordered Proteins (IDPs). *Chem. Rev.* **2014**, *114*, 6661-6714.

(12) Gibbs, E. B.; Cook, E. C.; Showalter, S. A. Application of NMR to studies of intrinsically disordered proteins. *Arch. Biochem. Biophys.* **2017**, *628*, 57-70.

(13) Mateos, B.; Conrad-Billroth, C.; Schiavina, M.; Beier, A.; Kontaxis, G.; Konrat, R.; Felli, I. C.; Pierattelli, R. The Ambivalent Role of Proline Residues in an Intrinsically Disordered Protein: From Disorder Promoters to Compaction Facilitators. *J. Mol. Biol.* **2020**, *432*, 3093-3111.

(14) Bibow, S.; Ozenne, V.; Biernat, J.; Blackledge, M.; Mandelkow, E.; Zweckstetter, M. Structural Impact of Proline-Directed Pseudophosphorylation at AT8, AT100, and PHF1 Epitopes on 441-Residue Tau. *J. Am. Chem. Soc.* **2011**, *133*, 15842-15845.

(15) Vuister, G. W.; Bax, A. Quantitative J Correlation: A New Approach for Measuring Homonuclear Three-Bond J_{HNHa} Coupling Constants in ^{15}N-enriched Proteins. *J. Am. Chem. Soc.* **1993**, *115*, 7772-7777.

(16) Hoch, J. C.; Baskaran, K.; Burr, H.; Chin, J.; Eghbalnia, H. R.; Fujiwara, T.; Gryk, M. R.; Iwata, T.; Kojima, C.; Kurisu, G.; Maziuk, D.; Miyanoiri, Y.; Wedell, J. R.; Wilburn, C.; Yao, H.; Yokochi, M. Biological Magnetic Resonance Data Bank. *Nucl. Acids Res.* **2023**, *51*, D368-D376.

(17) Daniecki, N. J.; Costantini, N. V.; Zondlo, N. J.; Sametz, G. M. A Discovery-Oriented Lab in Solid-Phase Peptide Synthesis and Conformational Analysis. *to be submitted* **2024**.

(18) Bhatt, M. R.; Zondlo, N. J. Synthesis and Conformational Preferences of Peptides and Proteins with Cysteine Sulfonic Acid. *Org. Biomol. Chem.* **2023**, *21*, 2779-2800.

(19) Brown, A. M.; Zondlo, N. J. A Propensity Scale for Type II Polyproline Helices (PPII): Aromatic Amino Acids in Proline-Rich Sequences Strongly Disfavor PPII Due to Proline-Aromatic Interactions. *Biochemistry* **2012**, *51*, 5041-5051.

(20) Brister, M. A.; Pandey, A. K.; Bielska, A. A.; Zondlo, N. J. OGlcNAcylation and Phosphorylation Have Opposing Structural Effects in tau: Phosphothreonine Induces Particular Conformational Order. *J. Am. Chem. Soc.* **2014**, *136*, 3803-3816.

(21) Pandey, A. K.; Ganguly, H. K.; Sinha, S. K.; Daniels, K. E.; Yap, G. P. A.; Patel, S.; Zondlo, N. J. An Inherent Structural Difference Between Serine and Threonine Phosphorylation: Phosphothreonine Prefers an Ordered, Compact, Cyclic Conformation. *ACS Chem. Biol.* **2023**, *18*, 1938-1958.

(22) Shi, Z.; Chen, K.; Liu, Z.; Ng, A.; Bracken, W. C.; Kallenbach, N. R. Polyproline II propensities from GGXGG peptides reveal an anticorrelation with b-sheet scales. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 17964-17968.

(23) Bielska, A. A.; Zondlo, N. J. Hyperphosphorylation of tau induces local polyproline II helix. *Biochemistry* **2006**, *45*, 5527-5537.

(24) Elbaum, M. B.; Zondlo, N. J. OGlcNAcylation and Phosphorylation Have Similar Structural Effects in α-Helices: Post-Translational Modifications as Inducible Start and Stop Signals in α-Helices, with Greater Structural Effects on Threonine Modification. *Biochemistry* **2014**, *53*, 2242-2260.

(25) Braunschweiler, L.; Ernst, R. R. Coherence Transfer by Isotropic Mixing - Application to Proton Correlation Spectroscopy. *J. Magn. Reson.* **1983**, *53*, 521-528.

(26) Bax, A.; Davis, D. G. Mlev-17-Based Two-Dimensional Homonuclear Magnetization Transfer Spectroscopy. *J. Magn. Reson.* **1985**, *65*, 355-360.

(27) Grathwohl, C.; Wüthrich, K. X-Pro Peptide Bond as an NMR probe for conformational studies of flexible linear peptides. *Biopolymers* **1976**, *15*, 2025-2041.

(28) Grathwohl, C.; Wüthrich, K. NMR studies of the rates of proline cis-trans isomerization in oligopeptides. *Biopolymers* **1981**, *20*, 2623-2633.

(29) Fischer, G. Chemical aspects of peptide bond isomerisation. *Chem. Soc. Rev.* **2000**, *29*, 119-127.

(30) Schmidpeter, P. A. M.; Koch, J. R.; Schmid, F. X. Control of protein function by prolyl isomerization. *Biochim. Biophys. Acta Gen. Subj.* **2015**, *1850*, 1973-1982.

(31) Stewart, D. E.; Sarkar, A.; Wampler, J. E. Occurrence and role of cis peptide-bonds in protein structures. *J. Mol. Biol.* **1990**, *214*, 253-260.

(32) Pal, D.; Chakrabarti, P. Cis Peptide Bonds in Proteins: Residues Involved, their Conformations, Interactions and Locations. *J. Mol. Biol.* **1999**, *294*, 271-288.

(33) Shen, Y.; Bax, A. Prediction of Xaa-Pro peptide bond conformation from sequence and chemical shifts. *J. Biomol. NMR* **2010**, *46*, 199-204.

(34) Sebak, F.; Ecsedi, P.; Bermel, W.; Luy, B.; Nyitray, L.; Bodor, A. Selective H-1(alpha) NMR Methods Reveal Functionally Relevant Proline cis/trans Isomers in Intrinsically Disordered Proteins: Characterization of Minor Forms, Effects of Phosphorylation, and Occurrence in Proteome. *Angew. Chem., Int. Ed.* **2022**, *61*, e202108361.

(35) Dyson, H. J.; Rance, M.; Houghten, R. A.; Lerner, R. A.; Wright, P. E. Folding of Immunogenic Peptide-Fragments of Proteins in Water Solution .1. Sequence Requirements for the Formation of a Reverse Turn. *J. Mol. Biol.* **1988**, *201*, 161-200.

(36) Meng, H. Y.; Thomas, K. M.; Lee, A. E.; Zondlo, N. J. Effects of *i* and *i*+3 Residue Identity on Cis-Trans Isomerism of the Aromatic$_{i+1}$-Prolyl$_{i+2}$ Amide Bond: Implications for Type VI b-turn Formation. *Biopolymers (Peptide Sci.)* **2006**, *84*, 192-204.

(37) Thomas, K. M.; Naduthambi, D.; Zondlo, N. J. Electronic control of amide cis-trans isomerism via the aromatic-prolyl interaction. *J. Am. Chem. Soc.* **2006**, *128*, 2216-2217.

(38) Ganguly, H. K.; Majumder, B.; Chattopadhyay, S.; Chakrabarti, P.; Basu, G. Direct Evidence for CH...pi Interaction Mediated Stabilization of Pro-cisPro Bond in Peptides with Pro-Pro-Aromatic Motifs. *J. Am. Chem. Soc.* **2012**, *134*, 4661-4669.

(39) Yao, J.; Feher, V. A.; Espejo, B. F.; Reymond, M. T.; Wright, P. E.; Dyson, H. J. Stabilization of a Type-VI Turn in a Family of Linear Peptides in Water Solution. *J. Mol. Biol.* **1994**, *243*, 736-753.

(40) Reimer, U.; Scherer, G.; Drewello, M.; Kruber, S.; Schutkowski, M.; Fischer, G. Side-chain effects on peptidyl-prolyl cis/trans isomerization. *J. Mol. Biol.* **1998**, *279*, 449-460.

(41) Wu, W.-J.; Raleigh, D. P. Local Control of Peptide Conformation: Stabilization of cis Proline Peptide Bonds by Aromatic Proline Interactions. *Biopolymers* **1998**, *45*, 381-394.

(42) Zondlo, N. J. Aromatic-Proline Interactions: Electronically Tunable CH/pi Interactions. *Acc. Chem. Res.* **2013**, *46*, 1039-1049.

(43) Ganguly, H. K.; Elbaum, M. B.; Zondlo, N. J. Proline-Aromatic Sequences Stabilize Turns via C–H/π interactions in both cis-Proline and trans-Proline. *ChemRxiv* **2023**, DOI: 10.26434/chemrxiv-22023-26437nb26444.

(44) Andreotti, A. H. Native State Proline Isomerization: An Intrinsic Molecular Switch. *Biochemistry* **2003**, *42*, 9515-9524.

(45) Englander, S. W.; Mayne, L. Protein Folding Studied Using Hydrogen-Exchange Labeling and 2-Dimensional Nmr. *Ann. Rev. Biophys. Biomol. Struct.* **1992**, *21*, 243-265.

(46) Bai, Y. W.; Milne, J. S.; Mayne, L.; Englander, S. W. Primary Structure Effects on Peptide Group Hydrogen-Exchange. *Proteins Struct. Funct. Gen.* **1993**, *17*, 75-86.

(47) Bai, Y. W.; Sosnick, T. R.; Mayne, L.; Englander, S. W. Protein-Folding Intermediates - Native-State Hydrogen-Exchange. *Science* **1995**, *269*, 192-197.

(48) Englander, S. W. Protein folding intermediates and pathways studied by hydrogen exchange. *Ann. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 213-238.

(49) Krishna, M. M. G.; Hoang, L.; Lin, Y.; Englander, S. W. Hydrogen exchange methods to study protein folding. *Methods* **2004**, *34*, 51-64.

(50) Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. Investigation of Exchange Processes by 2-Dimensional Nmr-Spectroscopy. *J. Chem. Phys.* **1979**, *71*, 4546-4553.

(51) States, D. J.; Haberkorn, R. A.; Ruben, D. J. A Two-Dimensional Nuclear Overhauser Experiment with Pure Absorption Phase in 4 Quadrants. *J. Magn. Reson.* **1982**, *48*, 286-292.

(52) Bothnerby, A. A.; Stephens, R. L.; Lee, J. M.; Warren, C. D.; Jeanloz, R. W. Structure Determination of a Tetrasaccharide - Transient Nuclear Overhauser Effects in the Rotating Frame. *J. Am. Chem. Soc.* **1984**, *106*, 811-813.

(53) Bax, A.; Davis, D. G. Practical Aspects of Two-Dimensional Transverse Noe Spectroscopy. *J. Magn. Reson.* **1985**, *63*, 207-213.

(54) Hwang, T. L.; Shaka, A. J. Cross Relaxation without Tocsy - Transverse Rotating-Frame Overhauser Effect Spectroscopy. *J. Am. Chem. Soc.* **1992**, *114*, 3157-3159.

(55) Wishart, D. S.; Sykes, B. D.; Richards, F. M. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol.* **1991**, *222*, 311-333.

(56) Hwang, T. L.; Shaka, A. J. Water Suppression That Works - Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *J. Magn. Reson., Series A* **1995**, *112*, 275-279.

(57) Piotto, M.; Saudek, V.; Sklenar, V. Gradient-Tailored Excitation for Single-Quantum Nmr-Spectroscopy of Aqueous-Solutions. *J. Biomol. NMR* **1992**, *2*, 661-665.

(58) Sklenar, V.; Piotto, M.; Leppik, R.; Saudek, V. Gradient-Tailored Water Suppression for H-1-N-15 Hsqc Experiments Optimized to Retain Full Sensitivity. *J. Magn. Reson., Series A* **1993**, *102*, 241-245.

(59) Shaka, A. J.; Lee, C. J.; Pines, A. Iterative Schemes for Bilinear Operators - Application to Spin Decoupling. *J. Magn. Reson.* **1988**, *77*, 274-293.

Graphical Table of Contents