Complete NMR assignment of succinimide, its detection and quantification in peptides and intact proteins

Luigi Grassi,[†,‡,+] Christof Regl,[†,§,+] Sabrina Wildner,[†,§] Gabriele Gadermaier,[†,§] Christian G. Huber,[†,§] Chiara Cabrele,[*,†,‡] and Mario Schubert[*,†,‡]

[†]Christian Doppler Laboratory for Innovative Tools for Biosimilar Characterization, University of Salzburg, Hellbrunner Strasse 34, 5020 Salzburg, Austria

[‡]Department of Molecular Biology, University of Salzburg, Billrothstrasse 11, 5020 Salzburg, Austria

[§]Department of Molecular Biology, University of Salzburg, Hellbrunner Strasse 34, 5020 Salzburg, Austria

KEYWORDS: Succinimide, posttranslational modifications, NMR spectroscopy, mass spectrometry, isoaspartate, aspartimide, proteomics.

ABSTRACT: Detecting and quantifying posttranslational modifications (PTMs) in full-length proteins is a challenge, especially in the case of spontaneously occurring, non-enzymatic PTMs. Such a PTM is the formation of succinimide (Snn) in a protein that occurs spontaneously in prone primary sequences and leads typically to an equilibrium between Snn and its hydrolysis products isoaspartate (isoAsp) and aspartate. In order to detect these modifications in proteins by NMR spectroscopy chemical shift assignments of reference compounds are required. We used peptide synthesis and 2D NMR spectroscopy to assign all ^{1}H and ^{13}C chemical shifts of Snn and isoAsp and found characteristic chemical shift correlations. To provide chemical shift reference data suitable for comparison with data of denatured proteins, we repeated the assignment in 7 M urea (pH 2.3) and in DMSO. Most characteristic of Snn are the two downfield shifted carbonyl chemical shifts, the chemical shift correlations of Cβ-Hβ of Snn and Cα-Hα of the succeeding residue which are clearly distinct from random coil chemical shift correlations. The characteristic 2D NMR fingerprints of Snn were used to detect and quantify this PTM in the model protein lysozyme, the biotherapeutic filgrastim and the Fc part of immunoglobulin G1. Mass spectrometry (MS) was applied as an additional independent method. The orthogonality of the NMR and MS techniques allows cross-validation, which is especially important to search for subtle PTMs in proteins. Studying PTMs by NMR spectroscopy is a promising method to analyze proteins and peptides either from natural sources, recombinant expression or chemical synthesis.

The detection and quantification of post-translational modifications (PTMs) is of utmost importance for proteomics and the characterization of natural and recombinant proteins, including protein- and peptide-based biopharmaceuticals. Mass spectrometry (MS) combined with separation methods is the most common approach to study PTMs. In a bottom-up approach, proteins are digested with specific proteases (*e.g.* trypsin), followed by separation using reversed-phase HPLC coupled to MS. Subsequently, relative quantification is performed based on the UV-chromatogram as well as on extracted ion chromatograms of modified and non-modified peptides.[1,2] Alternatively, top-down mass spectrometry without previous proteolysis has been applied for detection as well as quantification of PTMs in biopharmaceuticals,[3,4] allowing for analysis with minimal sample preparation.

However, MS has also certain disadvantages. Top-down approaches are limited in resolution due to the natural isotope distribution. Therefore, small mass differences, *e.g.* due to deamidation (+1 Da), are difficult to detect in large proteins. Bottom-up approaches face other difficulties like challenges with quantification and incomplete coverage. Some PTMs are not detectable due to instability in the gas phase, poor ionization efficiency, or very small mass differences.[5] We recently demonstrated that NMR spectroscopy is a valuable technique, complementary to MS, for the detection of PTMs in proteins, including the unambiguous identification of a certain PTM, its exact topology and stereochemistry, as demonstrated by the analysis of glycosylation.[6] Although this NMR approach of analyzing the proteins under denaturing conditions cannot localize a PTM in a protein sequence, it is in principle able to detect and quantify any PTM in a sample, including so far unknown modifications. A combination of NMR and MS techniques with their complementary strengths seems very promising. In order to unambiguously identify certain types of PTMs by NMR spectroscopy, NMR data of reference compounds are essential. Therefore, we aim at compiling reference chemical shifts applicable for detecting a large variety of PTMs in the future.

Here we focus on succinimide formation and its hydrolysis products, which may occur spontaneously in Asp- and Asn-containing proteins and peptides.[7-9] Depending on the amino acid following Asp (Asp-Xaa), Asp-containing peptides are susceptible to acid- or base-catalyzed cyclization,[10,11] in which the backbone nitrogen of the next residue forms a covalent bond to the Cγ of Asp under loss of a water molecule (Figure 1a). The cyclic product, called succinimide (Snn) or aspartimide, can be hydrolyzed by water to form a mixture of isoAsp- and Asp-containing sequences. In such an equilibrium the chiral center of Asp can even racemize, thus decreasing the product homogeneity further.[7,12] The motif Asp-Gly is espe-

cially prone to Snn formation,[7,13] followed by Asp-Ser, which is slightly less susceptible.[14] For Asp-Gly the kinetics of the equilibrium between Asp, Snn and isoAsp has been analyzed and half-lives of few days have been reported.[9] Studies of synthesized peptides show that Asp-Gly readily undergoes Snn formation followed by epimerization and a mixture of α- and β-peptides.[13] However, the three-dimensional structure of a protein has a strong effect on Snn formation, either preventing or promoting it.

Snn formation is also a product of a second mechanism - the spontaneous deamidation of Asn-containing peptides/proteins, in which cyclization occurs under loss of NH_3.[8,15] The deamidation rate is determined by the protein sequence, the solution conditions, especially the pH, and also the three-dimensional structure. Most sensitive peptide sequences prone to deamidation are Asn-Gly, Asn-His and Asn-Ser, for which deamidation half-times of ~1 day, 10 and 16 days were measured within short peptides at neutral pH and 37°C.[16]

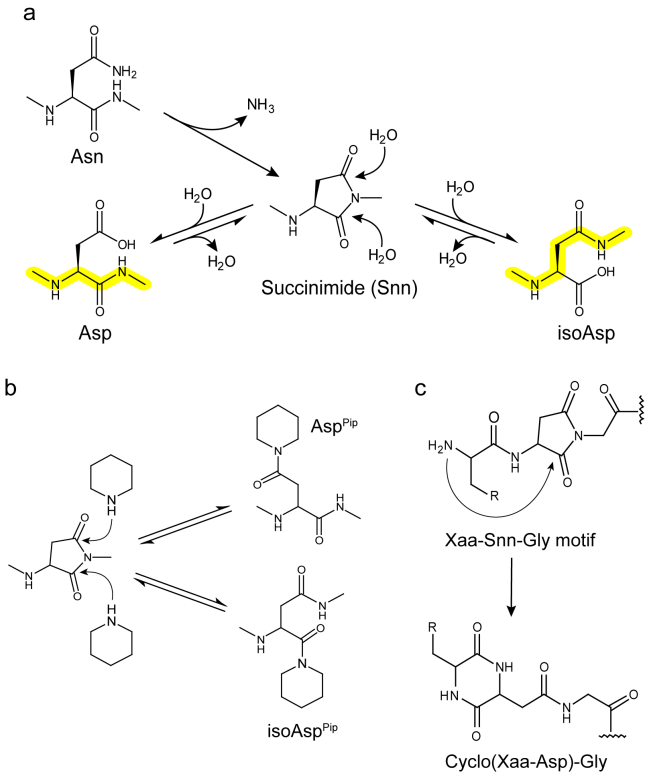

Figure 1. Chemical reactions involving succinimide formation, hydrolysis and byproducts. (a) Summarized spontaneous modifications that can occur in certain Asn- and Asp-containing protein sequences. The protein backbone containing Asp or isoAsp is indicated on yellow background. (b) Piperidides of Asp and isoAsp are observed as byproducts in solid-phase peptide synthesis with Fmoc-chemistry, using piperidine as a base (only the piperidide of Asp was observed in this work). (c) Formation of diketopiperazine (DKP) as a side reaction during chemical synthesis of Snn-containing peptides.

Whereas Snn can be formed as side-product during the chemical synthesis of Asp-Xaa-containing peptides,[10,14,17] especially when the Asp side chain is protected by an allyl group,[18] its presence in proteins is rare.[19] Nevertheless, 22 crystal structures containing a Snn and 44 structures containing an isoAsp moiety can be found in the Protein Data Bank (PDB) (Tables S-1 and S-2). A well-studied case is lysozyme that contains an Asp-Gly motif susceptible to Snn formation, and for which the crystal structures of both the Snn- and isoAsp-containing forms are available.[20]

Although Snn formation has been known since decades and represents a major concern in the preparation of synthetic or recombinant peptides and proteins, the NMR characterization of Snn and its hydrolysis products is still incomplete, probably due to the complexity of the equilibrium between Asp, Snn and isoAsp. The chemical shift database for proteins, the BioMagResBank,[21] contains only few entries with isoAsp and not a single one with Snn. Examples for NMR assignments of isoAsp are available from bovine calbindin D_{9K},[22] excisionase from bacteriophage HK022,[23] the molecular chaperone DnaK from *Thermus Thermophilus*,[24] malate synthase G from *Escherichia coli*,[25] RNase B,[26] and small peptides.[27] Unfortunately, the reported chemical shifts of isoAsp do not reveal unique values, which likely reflect the distinct structural environments in a folded protein. However, there might be another explanation: since the entries contain typically only one set of chemical shifts per residue and do not mention signal sets of minor populated species, there is the risk that chemical shifts of several species are mixed up.

Although peptides with Snn were already investigated by NMR spectroscopy,[28,29] a complete assignment of all ^{1}H, ^{13}C chemical shifts is not available so far. Because of the fact that Snn occurs typically in equilibrium with isoAsp and Asp, the mixing-up of chemical shifts must be avoided; moreover, because of the presence of contradicting chemical shifts, a resonance assignment based on redundant chemical shift correlations with 2D NMR techniques is required. Here, we used 2D NMR spectroscopy and completely assigned all ^{1}H, ^{13}C chemical shifts of peptides containing Snn and its hydrolysis products, and identified characteristic $^{1}H/^{13}C$ chemical shift correlations for Snn and isoAsp under various conditions including denaturing conditions. Interestingly, the chemical shifts of Snn and isoAsp are very different from the random coil chemical shifts of the 20 standard amino acids. Therefore, they can be used in combination with our recently proposed method to detect PTMs in intact proteins.[6] Indeed, by using lysozyme, filgrastim (the biotherapeutic Neupogen®) and the CH2-CH3 domains of IgG1 (Fc/2) as model systems, the characteristic chemical shifts of Snn enabled its detection and quantification in the denatured intact proteins.

EXPERIMENTAL SECTION

Peptides. Peptides were synthesized by Fmoc-chemistry using solid-phase peptide synthesis (SPPS) and analyzed by HPLC and MALDI-TOF-MS as described in the Supporting Information. For NMR analysis the crude products were used, in order to detect possible side products: 1.2-2 mg of peptide were lyophilized and subsequently dissolved in 500 µL solvent. The following solvents were used: D_2O (Armar Chemicals), uncorrected pH (pH*) was adjusted to 2.5 with DCl (Armar Chemicals); 93% H_2O/7% D_2O pH-adjusted to 2.5; 7 M urea-d_4 (Armar Chemicals) in D_2O, pH* was adjusted to 2.3 with DCl in D_2O; 7 M urea 93% H_2O/7% D_2O, pH-adjusted to 2.3; DMSO-d_6.

Succinimide formation in lysozyme. Lysozyme (Sigma-Aldrich) was dissolved in 150 mM sodium acetate buffer (pH 4) and incubated at 40°C for 7 weeks. For the NMR analysis 500 µL of the sample were lyophilized overnight, dissolved in 500 µL of a 7 M urea-d_4 in D_2O solution containing 25

mmol·L^{-1} DTT-d_{10} (Cambridge Isotope Laboratories), incubated at room temperature for 16 h (pH* ~7.0, still not fully denatured as seen in Figure S-1). Finally the pH* was adjusted with DCl in D_2O to 2.3, which resulted in a complete unfolding.

NMR spectroscopy. All spectra were recorded on a 600 MHz Bruker Avance III HD spectrometer equipped with a $^{2}H/^{13}C/^{15}N/^{31}P$ quadruple-resonance probe at 298 K, except one experiment with the Fc/2 sample that was measured on a Bruker Avance III HDX 700 MHz NMR spectrometer using quadruple resonance inverse cryoprobe (QCI-F: ^{1}H, ^{13}C, ^{15}N, ^{19}F). Sample volumes of 500 µL in standard 5 mm NMR tubes were used. Standard 2D [^{1}H, ^{13}C]-HSQC, [^{1}H, ^{13}C]-HMBC and [^{1}H, ^{1}H]-TOCSY experiments, were recorded together with a 2D [^{1}H, ^{13}C]-HMQC-COSY[6,30] and a [^{1}H, ^{13}C]-HMBC optimized for correlations with carbonyl resonances using shaped pulses (more details in Supplementary Methods). All spectra were processed with Topspin 3.2 (Bruker Biospin), referenced to 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) and further analyzed in Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco, USA). For the measurements under denaturing conditions external referencing was achieved using a sample with 0.5 mM DSS dissolved in 7 mol·L^{-1} urea-d_4 in D_2O at pH* 2.3. ^{1}H and ^{13}C chemical shift assignments of Snn and related residues are summarized and compared to the limited previous data in Table S-3, of side products in Table S-4 and complete assignments of all peptides are found in Tables S-5 to S-17.

Mass spectrometry. Analysis was carried out on an HPLC system (UltiMate 3000, Thermo Fisher Scientific, Germering, Germany). Heptapeptides were separated on a Hypersil Gold aQ column (100 × 1.0 mm i.d., 1.9 µm particle size, 175 Å pore size, Thermo Fisher Scientific, Sunnyvale, CA, USA), for lysozyme samples a Waters XBridge Protein BEH C4 column (150 × 2.1 mm i.d., 3.5 µm particle size, 300 Å pore size, Waters, Milford, MA, USA) was used. Mass spectrometry was conducted on a Thermo Scientific™ Q Exactive™ Hybrid-Quadrupole-Orbitrap mass spectrometer equipped with an Ion Max source with a heated electrospray ionization (HESI) probe (Thermo Fisher Scientific, Bremen, Germany). For more experimental details see Supporting Information.

RESULTS AND DISCUSSION

Chemical synthesis of the model peptides and side reactions. In an effort to detect Snn in immunoglobulins by using 2D NMR spectroscopy, we decided to study a short decapeptide sequence from the IgG1 Fc domain containing an Asn-Gly motif (peptide **1**). For this purpose we synthesized analogs with Snn (peptide **2**) and isoAsp (peptide **3**) using solid-phase peptide synthesis (SPPS) and Fmoc-chemistry (Figure 2). As control we synthesized the wild type sequence **1** and the deamidated form with Asp (peptide **4**) and D-Asp (peptide **5**) for the case of epimerization. Peptides **1**, **3-5** were prepared by using the building blocks Fmoc-Asn(trityl)-OH, Fmoc-Asp-OtBu, Fmoc-Asp(OtBu)-OH, and Fmoc-D-Asp(OtBu)-OH, respectively. Peptide **2** was prepared by using Fmoc-Asp(OAll)-OH that is known to strongly promote Snn formation, in particular of the Asp-Gly motif under basic conditions.[18] The syntheses of peptides **1**, **3-5** yielded quite homogeneous products containing the Asn (95%), isoAsp (83%, and 14% for Snn), Asp (91%), and D-Asp (93%) residues, respectively, as shown by the analytical HPLC traces of the crude products (Figure S-2 and Table S-18). In contrast, the synthesis of peptide **2** always resulted in a mixture of the desired Snn-containing peptide (**2**), and of an Asp(piperidide) byproduct ($\mathbf{2}^{Pip}$) in a ratio of 6:1 (Figure S-2). Asp(piperidide) is a common impurity of peptides prepared by SPPS with Fmoc-chemistry, which contain Asp-Xaa motifs prone to Snn formation. The use of piperidine during the Fmoc-cleavage promotes both Snn formation and aminolysis (Figure 1b). This is supported by a prolonged treatment of the mixture containing the Snn-containing peptide **2** with piperidine, which led to an almost complete conversion of Snn to Asp(piperidide) (Figure S-3 and Table S-3).

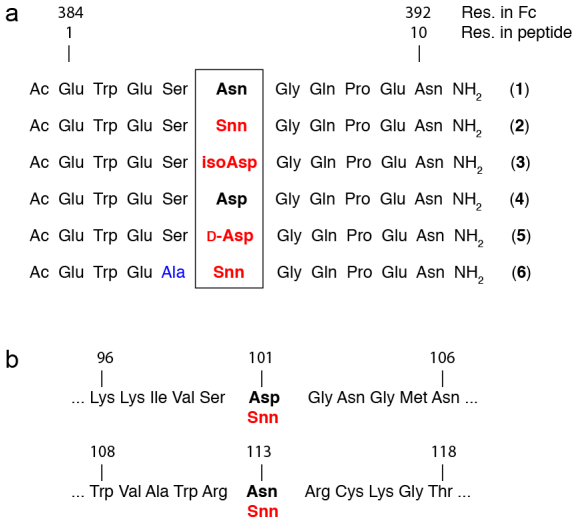

Figure 2. Peptide and protein sequences used in this study. (a) Synthesized peptides based on the IgG1 Fc fragment 384-392. (b) Two regions of lysozyme, where Snn formation was detected in this study, in contrast to previous studies where Snn was only detected at position 101.[34]

Quite unexpectedly, an additional side product was identified by NMR spectroscopy (see next section) as an N-terminally truncated fragment ($\mathbf{2}^{DKP}$) containing diketopiperazine (DKP) formed by Ser-4 and Asp-5 (~36% compared to ~56% of **2** and ~8% of $\mathbf{2}^{Pip}$). DKP is formed after Fmoc-deprotection of the α-amino group of Ser(tBu): the intramolecular attack of the free amine to the α-carbonyl of Snn is highly favored by the basic conditions and the formation of the six-membered ring (Figure 1c). To verify if the residue preceding Asp might affect the DKP formation, we synthesized a Snn-containing peptide analog, in which Ser-4 was replaced by the sterically less demanding Ala (peptide **6**). In this case, NMR analysis of the product of the synthesis revealed the presence of ~43% Snn-containing peptide (**6**), together with the impurities containing the DKP formed by Ala-4 and Asp-5 ($\mathbf{6}^{DKP}$) (~49%) and the Asp(piperidide) ($\mathbf{6}^{Pip}$) (~8%). Thus, the formation of (Ala-Asp)-DKP seems to be more favored than that of (Ser(tBu)-Asp)-DKP during SPPS with Fmoc-chemistry, suggesting that less hindered residues might facilitate the intramolecular mechanism of Snn aminolysis. The rearrangement of Snn-containing sequences into DKP-containing ones has been described previously for N-terminal motifs of Xaa-Asp/Asn-Yaa[31-33] following the previously proposed mechanism.[31] However, none of the resulting DKP-containing side products has ever been characterized by NMR spectroscopy so far. Further confirmation of the presence of DKP by chemical treatment, HPLC, MS and MS^2 spectrometry is provided in Figure S-4.

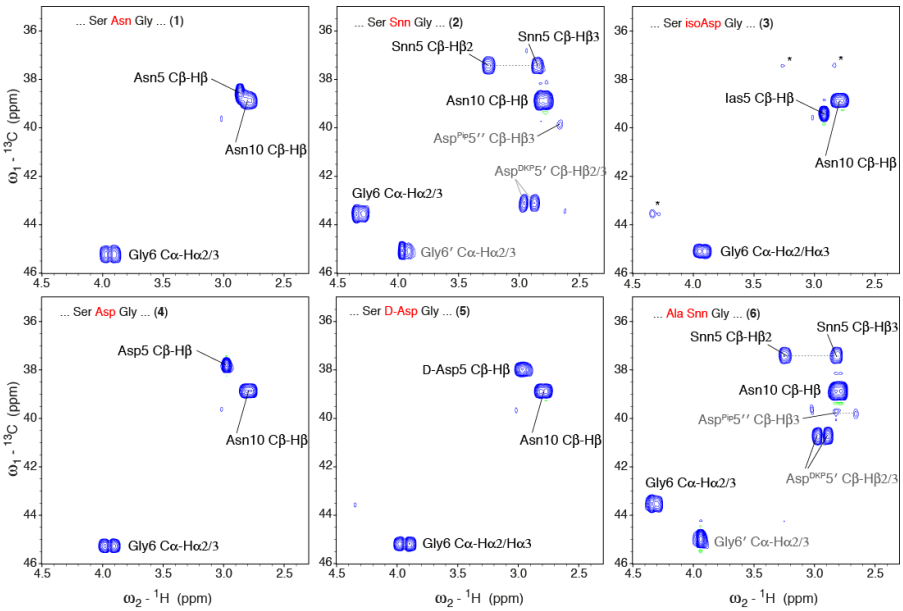

Figure 3. Overview of all synthesized peptides focusing on the main differences between them in natural abundance ^{1}H-^{13}C-HSQC spectra recorded at 600 MHz, 298 K with typically 100 transients and concentrations of 2 mM in D_2O. Shown is the fingerprint region of Cβ-Hβ correlations of Asn, Asp, isoAsp (indicated with the three-letter code Ias) and Snn together with Cα-Hα correlations of Gly. Spectra of peptides containing Snn show a second set of signals that originate from a truncated peptide lacking Glu1-Glu3 in which residue 4 and Asp5 form a cyclic diketopiperazine (DKP), a side product of the synthesis. The second set of signals is labeled in grey and the residue numbers are marked with a prime. Asterisks indicate signals of small amounts of Snn-Gly.

Chemical shift assignments in aqueous solutions, in 7 mol·L^{-1} urea and in DMSO. Initial NMR assignments of the peptides were performed in aqueous solutions (93% H_2O/7% D_2O or 100% D_2O) without any buffer to obtain highest sensitivity (at pH 2.3). The natural abundance of ^{13}C was used to measure 2D ^{1}H-^{13}C HSQC (Figure 3), ^{1}H-^{13}C HMBC, and ^{1}H-^{13}C HMQC-COSY spectra as described previously.[6] 2D ^{1}H-^{15}N HSQC spectra using the natural abundance of ^{15}N were recorded, which showed a nicely separated fingerprint of each peptide, where each residue is represented by its backbone N-H correlation (Figure S-5). To establish the topology of each spin system we used, in addition to the ^{13}C-correlations, two ^{1}H-^{1}H TOCSY spectra with mixing times of 120 ms and 12 ms. For the sequential assignment 2D ^{1}H-^{1}H ROESY spectra were used, as illustrated by a sequential walk of peptide **4** in Figure S-6.

To confirm the presence of Snn and to verify our sequential assignments by an independent through-bond approach, we used a long-range ^{1}H-^{13}C correlation experiment, optimized for carbonyl resonances. Details are described in the Supporting Information. We observed correlations to C and side chain carbonyls for Hα, Hβ, and H^N protons (Figure 4). Spectra were recorded either in D_2O, to detect Hα correlations close to the water signal with high sensitivity, or in 93% H_2O/7%D_2O, to detect H^N correlations. The presence of Snn in the preparation of peptide **2** was confirmed by through-bond, long-range correlations from Gly Hα resonances to both C′ and Cγ of Snn. The absence of a Gly ^{1}H-^{15}N correlation in the ^{15}N HSQC spectrum reflects the lack of an amide proton due to the involvement of the Gly nitrogen in the cyclisation.

The NMR characterization of the second major component in sample **2**, the diketopiperazine **2^{DKP}**, revealed very unusual weak ^{1}H-^{1}H TOCSY correlations between the two Hα nuclei within the six-membered ring (Figure S-7). These cross-peaks originate from small long range $^{5}J_{H\alpha H\alpha}$ scalar couplings that were observed in similar cyclic dipeptides.[35] The byproduct **2^{Pip}** could be distinguished from the other products by its ^{1}H-^{13}C correlations (comparing Figure S-8 with Figure 3). The chemical shift assignments of **2^{DKP}** and **2^{Pip}** are given in Table S-4. The fact that the ^{1}H-^{13}C HSQC spectra of the peptides (Figure 3) did not change after hours/days indicated that all peptides were sufficiently stable at 25°C under these conditions. Only the isoAsp-containing peptide **3** showed weak signals of Snn and Asp (Figures 3 and S-2c).

As the reference chemical shifts of the natural amino-acid residues in peptides and proteins are given for denaturing conditions, all essential 2D spectra were re-measured with the peptides dissolved in 7 mol·L^{-1} urea at pH 2.3, in order to provide reference chemical shifts for Snn and isoAsp. The chemical shifts were very similar to the values measured in plain D_2O and H_2O/D_2O (Tables S3 and S4). We also measured the chemical shifts in DMSO to obtain reference values for this alternative method to dissolve and denature proteins (Tables S3 and S4).

Comparison of Snn and isoAsp chemical shifts with the random coil shifts of a representative denatured protein. In order to judge if the obtained chemical shifts and, in particular, the specific ^{1}H-^{13}C correlations were unique for Snn and isoAsp, we compared these correlations with the random coil correlations of a denatured protein. We chose the well-characterized chicken egg lysozyme as model protein, because chemical shift assignments were available for folded and denatured forms of lysozyme.[36,37] We unfolded the protein first by reducing the disulfide bonds at neutral pH* followed by denaturation in 7 mol·L^{-1} urea-d_4 in D_2O at pH* 2.3. A very simplified ^{1}H-^{13}C HSQC spectrum was obtained (Figure S-9). Basically, only signals at random coil chemical shifts are visible, which can be readily assigned with tabulated values of random coil chemical shifts[38-40] together with few ^{1}H-^{1}H TOCSY and HMQC-COSY spectra. Small deviations of backbone Cα-Hα correlations were observed, which is typical for neighbors of Phe, Pro, Trp, Tyr, and in some cases of Ile, Val, Asp, and Asn.[40]

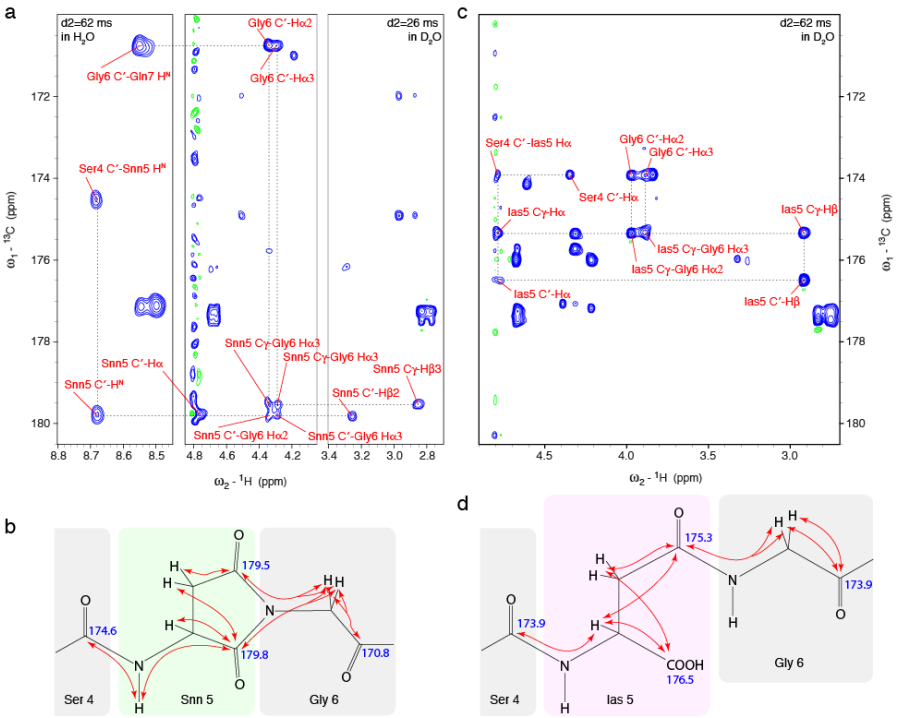

Figure 4. Assignment of carbonyl resonances of succinimide and isoaspartate in peptide environment. (a) Two dimensional H-CO spectra of the peptide **2** containing Snn measured in pure D_2O or H_2O/ D_2O. The spectra were measured at natural abundance in either 93% H_2O/ 7% D_2O (left) or 100% D_2O (right) at 298 K with peptide concentrations of 2 mmol·L^{-1} and 600 or 256 transients, respectively. Long range H-CO correlations are labeled in red. (b) Chemical structure of the Snn and its neighboring residues with the long range H-CO correlations shown by red arrows. The assigned chemical shifts of the carbonyls are given in blue in ppm. (c) H^N-CO spectrum of the peptide **3** containing isoAsp (indicated here with the three-letter code Ias). The spectrum was measured in 100% D_2O at 298 K with a peptide concentration of 2 mM and 500 transients. (d) Chemical structure of the isoAsp and its neighboring residues with the long range H-CO correlations are indicated by red arrows.

Superimpositions of the ^{1}H-^{13}C HSQC spectra of the peptide mixture **2**, peptide **3** and of denatured lysozyme are shown in Figure 5a. The Cβ-Hβ2, Cβ-Hβ3 correlations of Snn and the Cα–Hα correlation of Gly involved in the five-membered ring of Snn were clearly distinct from the correlations displayed in the denatured lysozyme sample, which represent the random coil chemical shifts of the 20 natural amino acids. Therefore, these correlations are particularly suited to detect the presence of Snn. The Cα-Hα signal of Gly is more intense than the two Cβ-Hβ signals of Snn, since it originates from two protons. Although the signals of isoAsp are slightly different from those of Asp or Asn, it was not possible to clearly distinguish them from the random coil chemical shifts of the canonical amino acids in lysozyme. In summary, this characteristic fingerprint region of an ^{1}H-^{13}C HSQC spectrum displays Snn Cβ-Hβ correlations and the Cα-Hα correlation of Gly as part of Snn-Gly that are well suited to detect Snn with a background of random coil protein chemical shifts.

Beside these unique chemical shifts in an ^{1}H-^{13}C HSQC spectrum we identified other characteristic resonances of Snn and isoAsp that are summarized in Figure 6 and Table S-3. For example the two carbonyl ^{13}C chemical shifts of Snn at ~180 ppm are most characteristic and clearly different from Asp, isoAsp and Asn, but at ^{13}C natural abundance they are only accessible by insensitive heteronuclear multiple bond correlation (HMBC) experiments using weak long-range $^nJ_{C\text{-}H}$ scalar couplings. The chemical shifts for isoAsp are less characteristic, but specific long range H^N-CO correlations are unique, *e.g.* the amide of isoAsp with its own carboxyl resonance and, more importantly, the Cγ resonance of isoAsp with the H^N of the following residue.[23,25] The chemical shift correlations of D-Asp are very similar to L-Asp within our model peptide, so that racemization will not be detected easily with NMR spectroscopy. However, the application of an additional chiral shift reagent or chiral solvating agent[41] might be helpful to distinguish the two species in the future.

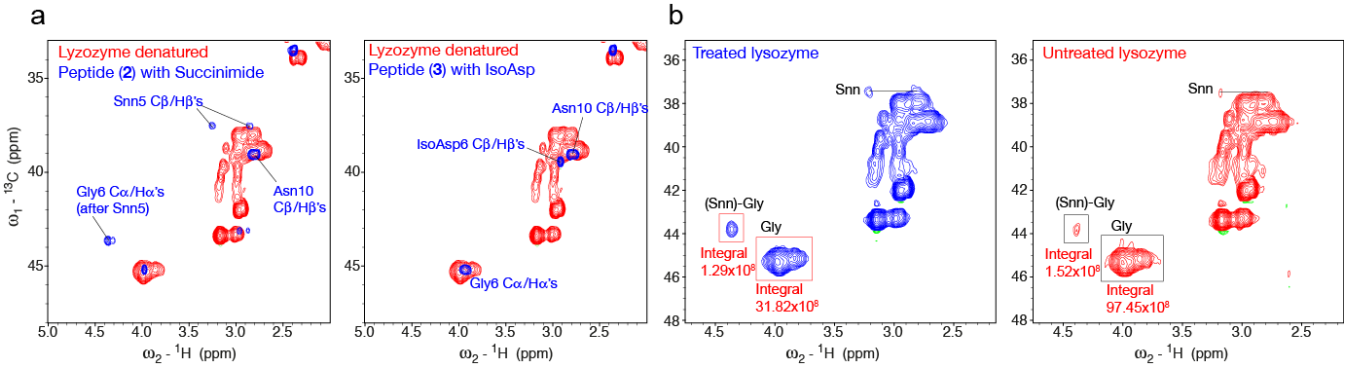

Figure 5. Detection of Snn by unique chemical shift correlations. (a) Characteristic regions in a ^{13}C-HSQC at natural abundance of the peptides containing Snn (**2**) and isoAsp (**3**), overlaid with random coil correlations of denatured lysozyme. The spectrum of untreated 2.2 mmol·L^{-1} lysozyme was recorded at 298 K in D_2O containing 7 mol·L^{-1} urea-d_4 (pH* 2.3)using 100 transients. (b) Detection and quantification of Snn in treated and untreated lysozyme dissolved in completely deuterated 7 mol·L^{-1} urea-d_4 (pH* 2.3). 156 and 100 transients were applied for treated and untreated lysozyme, respectively, using otherwise comparable parameters.

Whereas the distinct ^{13}C carbonyl chemical shifts of Snn and isoAsp are difficult to access in proteins with natural isotope abundance, in the case of $^{13}C/^{15}N$ labeled proteins, whose carbonyl chemical shifts are easily accessible using the triple-resonance experiments HNCO and HN(CA)CO together with HCCO, those characteristic shifts will be very valuable.

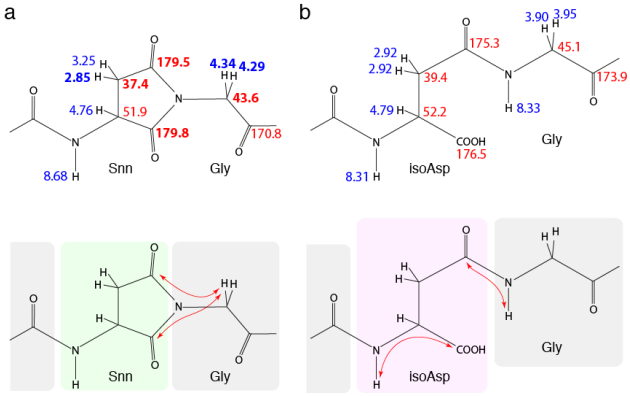

Figure 6. Random coil chemical shift assignment and typical correlations of Snn (a) and isoAsp (b). ^{1}H chemical shifts are shown in blue, ^{13}C chemical shifts in red. Chemical shifts leading to unique ^{1}H-^{13}C correlations are indicated in bold. Unique long range ^{1}H-^{13}C correlations indicative for either Snn or isoAsp are shown at the bottom as red arrows.

Controlled formation of Snn in lysozyme – a model system to detect Snn by NMR spectroscopy and mass spectrometry. After the identity of commercially obtained lysozyme was verified by comparing the ^{1}H-^{13}C correlations (Figure S-9a) with previously reported chemical shift assignments,[42] Snn formation was promoted by incubation at pH 4 in a 150 mM sodium acetate buffer at 40°C for 7 weeks, according to Tomizawa.[34] A ^{1}H-^{13}C-HSQC spectrum of the treated and lyophilized protein dissolved in deuterated denaturing buffer, followed by disulfide-bond reduction using dithiothreitol (DTT)-d_{10} and pH* adjustment to 2.3, is shown in Figure 5b. Both lysozyme samples were completely denatured, and random coil chemical shifts are visible. Most interestingly, the treated sample displayed additional signals corresponding to the characteristic Cβ–Hβ correlations of Snn and the Cα/Hα correlations of Gly following the Snn, confirming the presence of Snn in the intact protein. The presence of isoAsp could not be unambiguously confirmed due to spectral overlap. The integrals of Gly Cα–Hα correlations were used to quantify the content of Snn (Figure 5b): the Gly(Snn) signal integral corresponds to 3.9 % compared to the combined integrals of all Gly Cα–Hα signals (lysozyme contains 12 Gly residues). On average these integrals indicate a Snn population of 46% per molecule, which agrees roughly with the value previously reported for lysozyme after treatment at pH 4 in 150 mM acetate buffer.[34] Some Snn was also detected in the untreated sample, resulting in 18% per molecule.

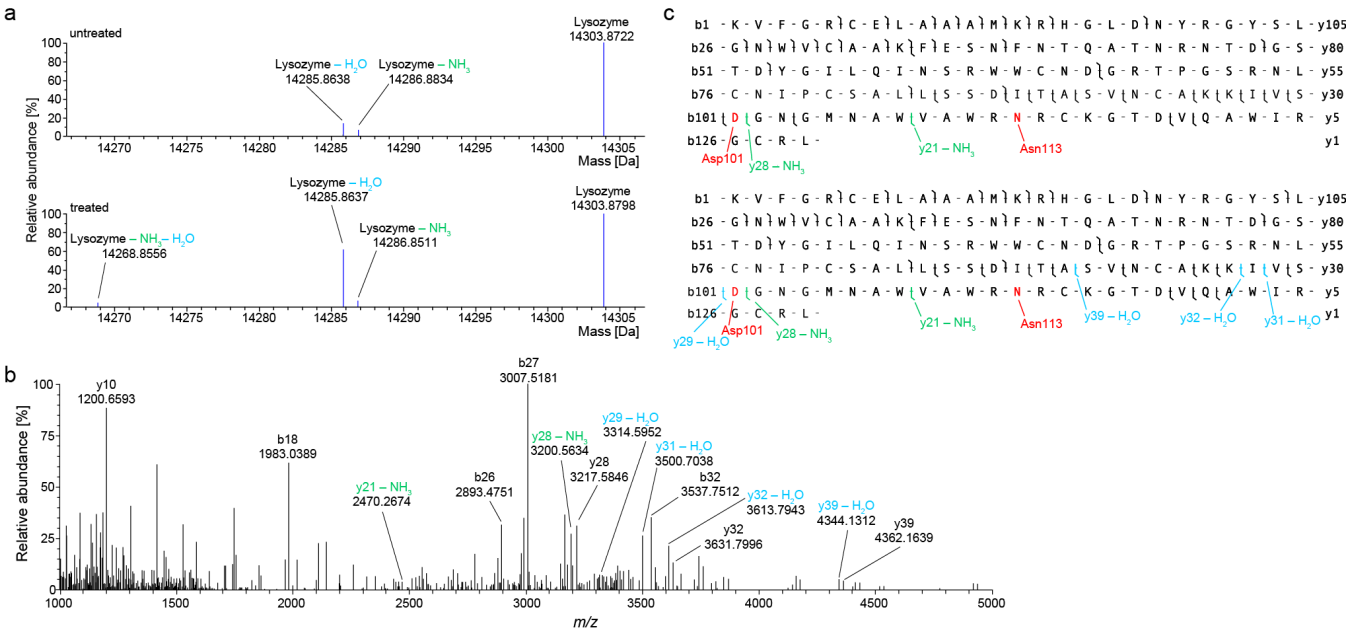

Figure 7. Detection of succinimide in lysozyme by top-down mass spectrometry. (a) Deconvoluted mass spectra of reduced lysozyme as control (top) and of lysozyme treated at pH 4 (bottom): Intact mass spectrometry revealed lysozyme species with NH_3 or H_2O loss already in the control sample. H_2O loss indicating succinimide formation was detected at higher extent in the sample treated at pH 4. (b) Top-down fragment ion spectrum of the pH 4-treated lysozyme. Fragments diagnostic for Snn-formation due to deamidation of Asn113 are labeled green, fragments diagnostic for Snn formation at Asp101 in light blue. (c) Sequence coverage map of untreated (top) and pH 4 treated lysozyme (bottom) generated with ProSight Lite. The diagnostic fragments y21 and y28 indicated in green show Snn formation with ammonia loss (-17 Da) at Asn113 (red), whereas fragments y29, y31, y32 and y39 shown in cyan indicate Snn formation with water loss (-18 Da) at Asp101 (red).

Table 1. Mass spectrometry of untreated and pH 4-treated lysozyme displaying the deconvoluted masses of intact and Snn-containing protein, an estimation of abundance and mass deviation.

Lysozyme	Measured mass [Da]	Mass deviation [Da]	Intensity [counts]	Abundance [%]	Identification	Mass deviation [ppm]
Untreated	14303.88	0.00	17320529.44	84.22	Lysozyme	0.18
Untreated	14286.87	-17.01	2221494.73	10.80	Lysozyme $-NH_3$	1.60
Untreated	14285.87	-18.01	1023361.81	4.98	Lysozyme $-H_2O$	0.17
Treated	14303.87	0.00	12633377.34	58.52	Lysozyme	-0.55
Treated	14285.86	-18.01	7742555.40	35.86	Lysozyme $-H_2O$	-0.44
Treated	14286.85	-17.02	777226.17	3.60	Lysozyme $-NH_3$	0.08
Treated	14268.86	-35.02	436620.28	2.02	Lysozyme $-H_2O$ $-NH_3$	1.09

To confirm the NMR results on the Snn detection in lysozyme, we applied top-down MS on both the treated and untreated protein. In contrast to Tomizawa,[34] who analyzed a trypsin digest of treated lysozyme, we analyzed the sample directly, thus circumventing an extensive sample preparation that might have introduced artificial deamidation.[43] Mass spectra of the intact protein showed two additional products with a mass difference of -18 Da or -17 Da, respectively, which indicated the formation of Snn involving not only an Asp, but also an Asn residue (Figure 7a, Table 1). Indeed, top-down sequencing could detect diagnostic fragments (y29 and y21, Figure 7b,c), which localize Snn formation at Asp101 (-18 Da at y29) and Asn113 (-17 Da at y21). Intriguingly, the observed abundances are in good agreement with the NMR data: in total 15.8% Snn formation is observed in the initial lysozyme sample and 41.5% in the sample treated at pH 4. In NMR 18 % and 46% Snn was detected, respectively, showing good correlation between the two analytical methods.

Detection of Snn in the biopharmaceutical drug filgrastim and in the CH2-CH3 tandem domains of human IgG1 (Fc/2). To illustrate the applicability of our methodology to other proteins, we examined a biotherapeutic, filgrastim, a 175-residue protein that contains an Asp-Gly motif. In contrast to untreated lysozyme, a 2014 expired sample of untreated filgrastim did not show any Snn content either by NMR spectroscopy or mass spectrometry (Figures S-10 and S-11). However, we observed Met oxidation, a known contaminant of this protein.[3] To stimulate Snn formation, we incubated the protein at pH 4.0 and 40°C for 2-40 h and could indeed induce Snn formation but only in the presence of TCEP that reduced the disulfide bonds and thus destabilized the structure. This stability and lack of Snn is in agreement with a previous report.[44] After incubation in the presence of TCEP for 65 h NMR analysis revealed the typical glycine signal indicating Snn with an abundance of ~23% per molecule (Figure S-11). Mass spectrometry of treated filgrastim revealed a content of a -18 Da species with an abundance of 7% (Figure S-10 and Table S-19). Here it should be noted that, besides the Snn formation (-18 Da), the occurrence of Met oxidation (+16 or +32 Da) and other spontaneous modifications (for example, -16 Da and -17 Da species, likely arising from Gln deamidation) hampers a proper quantification of total Snn content.

In addition to filgrastim, we examined a recombinant, monomeric version of the fragment crystallizable region (Fc region) of human immunoglobulin G1 (IgG1) consisting of the two domains CH2 and CH3 of the heavy chain, called in the following Fc/2. Two Asp-Gly motifs are present in the sequence (Figure S-12). The protein was expressed in *E. coli* yielding ~2 mg pure protein per liter expression culture. Purity of the protein preparation was investigated by gel electrophoresis detecting a single protein band around 25 kDa (Figure S-12a). The exact concentration of 0.2 mg/ml was determined by amino acid analysis and amino acid distribution was in good agreement with the theoretical values. Incubation of the protein at pH 4 and 40 °C for one week resulted in a substantial amount of Snn as revealed by NMR spectroscopy and mass spectrometry (Figures S-13 and S-14, Table S-20). The NMR data revealed 3.4% Snn-Gly, whereas the mass spectrometry showed a loss of H_2O with an abundance of 31% and the loss of two H_2O molecules at 6%. These results suggest that the monomeric Fc fragment is prone to water loss not only via Snn formation from an Asp-Gly motif, but also from other motifs, and probably, via other mechanisms, especially at moderately elevated temperatures and slightly acidic conditions. However, the lack of the N-glycan within the CH2 domain and its natural environment are potentially contributing to this chemical instability.

CONCLUSIONS

With this work we show that our NMR method to detect PTMs in denatured intact proteins is applicable not only to large modifications like glycosylation,[6] but also to very subtle changes in the protein structure, which may be challenging for MS. Indeed, whereas the quantification of Snn or the detection of Asp isomerization to isoAsp, is not trivial by MS,[45] we show here that our NMR method can quantify Snn formation and, in principle, detect the presence of isoAsp based on its Cβ–Hβ signal, unless signal overlap does not impair its observation. Further advantages of the here proposed NMR method is its applicability to natural proteins with no isotope labeling and no limit to the protein size due to the denaturing conditions. However, the amount of protein that can be dissolved in the NMR tube is also governed by the protein size. As an alternative to denature a protein with high concentrations of urea, DMSO can be used, which is

even applicable to aggregates and fibrils.[46,47] The here reported random coil values of Snn and isoAsp in DMSO will serve as reference, making the study of such PTMs in so far inaccessible protein states possible.

The reference chemical shifts for Snn, isoAsp and diketopiperazine both in DMSO and 7 mol L^{-1} urea denaturing conditions will be also useful for the careful characterization of synthetic peptides containing Asp-Xaa motifs, as the formation of Snn is very often an undesired side reaction during their chemical preparation,[10,14,17] especially when the Asp side chain is protected by the allyl group.[18] The NMR characterization of diketopiperazine, which might occur as a side product of peptide synthesis, might be very important for applications in proteomics, considering its occurrence in tryptic and chymotryptic digests,[33] which might lead to incomplete sequence coverage in bottom-up MS approaches and failure of tagging some newly generated N-termini after digestion.

In conclusion we demonstrated that NMR spectroscopy can unambiguously prove the presence of succinimide in peptides and in intact denatured proteins. Quantification of Snn content agrees well with MS data. We estimate the error of our measurements in the order of 20% due to inaccuracies in measuring the integrals, an incomplete recovery of the equilibrium magnetization during the relaxation delays of 1.5 or 2 s. However, more accurate quantification can be achieved by using longer relaxation delays and more sensitive NMR equipment like cryo-probe technology. The latter will also lower the detection limit by a factor of 2-4, so that lower abundant modifications can be detected or less sample quantities are required. With our recently developed methodology of investigating denatured proteins in terms of glycosylation and here extended to Snn formation we provide an independent method that is orthogonal to MS and, therefore, highly appropriate for cross-validation of the two techniques. Although our NMR methodology is not suited for high-throughput analysis and would not be able to compete with MS workflows, it can unambiguously identify the chemical origin of H_2O loss from the Asp-Gly motif. In the near future NMR spectroscopy will especially be a very valuable tool to elucidate unknown modifications or degradation products in proteins, in particular biotherapeutics.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.
Supplementary Methods, Supplementary Figures S-1 to S-15 and Tables S-1 to S-21 (PDF).

AUTHOR INFORMATION

Corresponding Authors

* Email: mario.schubert@sbg.ac.at or cabrele@sbg.ac.at

Author Contributions

[+] These authors contributed equally to the work. MS, CC, LG, SW, GG and CR designed the experiments. LG synthesized the peptide samples. MS performed the NMR experiments, assigned the NMR resonances and performed database analysis. CR performed and interpreted the mass spectrometry experiments. SW and GG expressed, purified and characterized the Fc/2 fragment. CH supervised the mass spectrometry analysis. MS, CC, LG, CR wrote the manuscript.

Notes

The authors declare the following competing financial interest(s): The salary of Luigi Grassi and Sabrina Wildner was fully funded and Christian G. Hubers salary is partly funded by the Christian Doppler Laboratory for Biosimilar Characterization, which is partly supported by Sandoz GmbH and Thermo Fisher Scientific. The authors declare no other competing financial interest.

ACKNOWLEDGMENT

We acknowledge Sandoz GmbH for kindly providing expired samples of Neupogen®, Dr. Urs Lohrig, Dr. Johann Holzmann and Dr. Ines Forstenlehner from Sandoz GmbH and Dr. Kai Scheffler from Thermo Fisher Scientific for comments on the manuscript as well as scientific discussions. We thank Dr. Hanspeter Kählig and the University of Vienna for access to a Bruker 700 MHz spectrometer with a cryogenetic probe, Dr. Therese Wohlschlager (University of Salzburg) for helpful discussion, Dr. Jaime Prilusky (Weizmann Institute, Israel) for providing the OCA browser-database for protein structure/function data and Dr. Peter Schmieder (FMP Berlin) for helpful comments on the manuscript. M.S. is grateful to Noah E. Robinson for kindly sending his book Molecular Clocks. The financial support by the Austrian Federal Ministry of Science, Research, and Economy and by a Start-up Grant of the State of Salzburg is gratefully acknowledged.

REFERENCES

(1) Houde, D.; Kauppinen, P.; Mhatre, R.; Lyubarskaya, Y. *J Chromatogr A* **2006**, *1123*, 189-198.
(2) Hess, D.; Covey, T. C.; Winz, R.; Brownsey, R. W.; Aebersold, R. *Protein Sci* **1993**, *2*, 1342-1351.
(3) Holzmann, J.; Hausberger, A.; Rupprechter, A.; Toll, H. *Anal Bioanal Chem* **2013**, *405*, 6667-6674.
(4) Forstenlehner, I. C.; Holzmann, J.; Toll, H.; Huber, C. G. *Anal Chem* **2015**, *87*, 9336-9343.
(5) Parker, C. E.; Mocanu, V.; Mocanu, M.; Dicheva, N.; Warren, M. R. In *Neuroproteomics*, Alzate, O., Ed.; CRC Press/Taylor & Francis: Boca Raton (FL), 2010, pp 93-113.
(6) Schubert, M.; Walczak, M. J.; Aebi, M.; Wider, G. *Angew Chem Int Ed Engl* **2015**, *54*, 7096-7100.
(7) Geiger, T.; Clarke, S. *J Biol Chem* **1987**, *262*, 785-794.
(8) Robinson, A. B.; McKerrow, J. H.; Cary, P. *Proc Natl Acad Sci U S A* **1970**, *66*, 753-757.
(9) Capasso, S.; Kirby, A. J.; Salvadori, S.; Sica, F.; Zagari, A. *J Chem Soc Perk T 2* **1995**, 437-442.
(10) Nicolas, E.; Pedroso, E.; Giralt, E. *Tetrahedron Lett* **1989**, *30*, 497-500.
(11) Wade, J. D.; Mathieu, M. N.; Macris, M.; Tregear, G. W. *Lett Pept Sci* **2000**, *7*, 107-112.
(12) Li, B.; Borchardt, R. T.; Topp, E. M.; VanderVelde, D.; Schowen, R. L. *J Am Chem Soc* **2003**, *125*, 11486-11487.
(13) Korling, M.; Geyer, A. *Eur J Org Chem* **2015**, 2382-2387.
(14) Behrendt, R.; White, P.; Offer, J. *J Pept Sci* **2016**, *22*, 4-27.
(15) Robinson, A. B. *Proc Natl Acad Sci U S A* **1974**, *71*, 885-888.
(16) Robinson, N. E.; Robinson, A. B. *Proc Natl Acad Sci U S A* **2001**, *98*, 944-949.
(17) Lauer, J. L.; Fields, C. G.; Fields, G. B. *Lett Pept Sci* **1994**, *1*, 197-205.
(18) Kunz, H.; Waldmann, H.; Unverzagt, C. *Int J Pept Prot Res* **1985**, *26*, 493-497.
(19) Clarke, S. *Int J Pept Prot Res* **1987**, *30*, 808-821.
(20) Noguchi, S.; Miyawaki, K.; Satow, Y. *J Mol Biol* **1998**, *278*, 231-238.

(21) Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C. F.; Tolmie, D. E.; Kent Wenger, R.; Yao, H.; Markley, J. L. *Nucleic Acids Res* **2008**, *36*, D402-408.

(22) Chazin, W. J.; Kordel, J.; Thulin, E.; Hofmann, T.; Drakenberg, T.; Forsen, S. *Biochemistry* **1989**, *28*, 8646-8653.

(23) Rogov, V. V.; Lucke, C.; Muresanu, L.; Wienk, H.; Kleinhaus, I.; Werner, K.; Lohr, F.; Pristovsek, P.; Ruterjans, H. *Eur J Biochem* **2003**, *270*, 4846-4858.

(24) Revington, M.; Zuiderweg, E. R. *J Biomol NMR* **2004**, *30*, 113-114.

(25) Tugarinov, V.; Muhandiram, R.; Ayed, A.; Kay, L. E. *J Am Chem Soc* **2002**, *124*, 10025-10035.

(26) Gotte, G.; Libonati, M.; Laurents, D. V. *J Biol Chem* **2003**, *278*, 46241-46251.

(27) Guttler, B. H.; Cynis, H.; Seifert, F.; Ludwig, H. H.; Porzel, A.; Schilling, S. *Amino Acids* **2013**, *44*, 1205-1214.

(28) Flora, D.; Mo, H.; Mayer, J. P.; Khan, M. A.; Yan, L. Z. *Bioorg Med Chem Lett* **2005**, *15*, 1065-1068.

(29) Xie, M. L.; VanderVelde, D.; Morton, M.; Borchardt, R. T.; Schowen, R. L. *J Am Chem Soc* **1996**, *118*, 8955-8956.

(30) Fesik, S. W.; Gampe, R. T.; Zuiderweg, E. R. P. *J Am Chem Soc* **1989**, *111*, 770-772.

(31) Bruckner, C.; Fahr, A.; Imhof, D.; Scriba, G. K. *J Pharm Sci* **2012**, *101*, 4178-4190.

(32) Schon, I.; Kisfaludy, L. *Int J Pept Prot Res* **1979**, *14*, 485-494.

(33) Jornvall, H. *FEBS Lett* **1974**, *38*, 329-333.

(34) Tomizawa, H.; Yamada, H.; Ueda, T.; Imoto, T. *Biochemistry* **1994**, *33*, 8770-8774.

(35) Vicar, J.; Budesins.M; Blaha, K. *Collect Czech Chem C* **1973**, *38*, 1940-1956.

(36) Schwalbe, H.; Fiebig, K. M.; Buck, M.; Jones, J. A.; Grimshaw, S. B.; Spencer, A.; Glaser, S. J.; Smith, L. J.; Dobson, C. M. *Biochemistry* **1997**, *36*, 8977-8991.

(37) Hennig, M.; Bermel, W.; Spencer, A.; Dobson, C. M.; Smith, L. J.; Schwalbe, H. *J Mol Biol* **1999**, *288*, 705-723.

(38) Schwarzinger, S.; Kroon, G. J.; Foss, T. R.; Wright, P. E.; Dyson, H. J. *J Biomol NMR* **2000**, *18*, 43-48.

(39) Wishart, D. S.; Bigam, C. G.; Holm, A.; Hodges, R. S.; Sykes, B. D. *J Biomol NMR* **1995**, *5*, 67-81.

(40) Schwarzinger, S.; Kroon, G. J.; Foss, T. R.; Chung, J.; Wright, P. E.; Dyson, H. J. *J Am Chem Soc* **2001**, *123*, 2970-2978.

(41) Parker, D. *Chem Rev* **1991**, *91*, 1441-1457.

(42) Wang, Y.; Bjorndahl, T. C.; Wishart, D. S. *J Biomol NMR* **2000**, *17*, 83-84.

(43) Chelius, D.; Rehder, D. S.; Bondarenko, P. V. *Anal Chem* **2005**, *77*, 6004-6011.

(44) Kumar, M.; Chatterjee, A.; Khedkar, A. P.; Kusumanchi, M.; Adhikary, L. *J Am Soc Mass Spectr* **2013**, *24*, 202-212.

(45) Yang, H. Q.; Zubarev, R. A. *Electrophoresis* **2010**, *31*, 1764-1772.

(46) Alexandrescu, A. T. *PLoS One* **2013**, *8*, e56467.

(47) Luhrs, T.; Ritter, C.; Adrian, M.; Riek-Loher, D.; Bohrmann, B.; Dobeli, H.; Schubert, D.; Riek, R. *Proc Natl Acad Sci U S A* **2005**, *102*, 17342-17347.

Table of Contents artwork