

Chapter 2

Segmental Labeling to Study Multidomain Proteins

Jing Xue, David S. Burz, and Alexander Shekhtman

Abstract This chapter contains a review of methodologies and recent applications of segmental labeling for NMR structural studies of proteins and protein complexes. Segmental labeling is used to specifically label a segment of protein structure with NMR active nuclei, thus reducing NMR spectral complexity and greatly facilitating structural NMR studies of large multi-domain proteins. It can also be used to introduce a synthetic fragment into a protein structure to study post-translationally modified proteins. Detailed protocols describing segmental labeling techniques are also included.

2.1 Introduction

Multi-domain protein complexes are the eukaryotic solution to biogenetic pathways found in prokaryotes. Each domain is a fully functional unit whose activity is necessary for the processivity of the reaction pathway. Interactions between domains are a common means of regulating the overall activity of such a complex. Therefore it is useful to be able to study a single domain within a complex to understand regulation of biological activity.

NMR analyses of protein structure have been traditionally limited by the molecular weight or overall size of the molecule being examined; more residues contribute more complexity to the spectrum. To reduce spectral complexity, several methodological approaches have been used. Among them, partial labeling of the molecule in question by uniformly labeling the backbone or by selectively labeling specific amino acids. In both of these cases the information gleaned represents a subset of all structural information available. Segmental labeling is used to isotopically label one domain or segment within a multi-domain complex and offers the dual advantage of providing complete structural information while minimizing spectral complexity.

Segmental labeling of separate domains within a large multidomain protein became possible after extending methodologies of synthetic protein chemistry to the area of recombinant protein production. Since the 1980s, people have sought to increase the size of high purity, synthetic peptides or proteins available for structural, functional, or physiological study. Solid phase peptide synthesis (SPPS) combined with size-exclusion or ion-exchange chromatography was used to prepare large quantities of high purity peptides [1]. This synthetic technique led to native chemical ligation (NCL)

J. Xue • D.S. Burz • A. Shekhtman (✉)
Department of Chemistry, State University of New York at Albany,
Albany, NY 12222, USA
e-mail: ashekhtman@albany.edu

[2, 3] in which two synthetic peptides are ligated *in vitro* through a C-terminal cysteine on one peptide and an N-terminal thioester group on the other. NCL evolved to expressed protein ligation (EPL) [4] in which cloned, recombinant peptides and proteins, overexpressed in *E. coli*, are utilized to form a native peptide bond, resulting in a macromolecule that is functionally similar to the natural protein.

Segmental isotope labeling combined with modern NMR techniques can identify whether protein domains interact with one another and helps to define the precise interaction interface and orientation between them, or whether they do not interact and can therefore be structurally and biologically characterized independently. It has been used to study enzymatic reactions mechanisms and to structurally characterize proteins that are difficult to obtain in chemically pure form, such as glycosylated proteins.

2.2 Methods and Techniques

2.2.1 Stepwise Solid-Phase Peptide Synthesis (SPPS)

Stepwise solid-phase peptide synthesis (SPPS) was established in the 1980s [1, 5] as the most efficient way to prepare peptide fragments. The peptide is immobilized (covalently linked) to a stationary phase resin via a linker. Single amino acids are incorporated one at a time from the C-terminus to the N-terminus. The peptide is constructed through repeated cycles of deprotecting, washing, coupling and washing. The final product is cleaved from resin using HF or trifluoroacetic acid. This process allows chemical modifications to the peptide backbone and the incorporation of unnatural or isotopically-labeled amino acids.

The effective limit of SPPS is 60–70 amino acids, beyond this length the overall yield of product makes this technique impractical. To prevent unintended reactions, the N-terminal amine is protected. SPPS is defined by the nature of the chemical group used to protect the α -amino group during stepwise synthesis. The protecting groups typically employed are di-*tert*-butyl-dicarbonate (Boc) or 9-fluorenylmethyloxycarbonyl (Fmoc).

Boc is used to reduce aggregation during synthesis and during incorporation of base-sensitive peptide analogs (utilizing non-natural amino acids) [6]. The Boc group is removed using trifluoroacetic acid (TFA), which results in a positively-charged amino group that must be neutralized concomitant with coupling to the next activated amino acid. During the cleavage reaction cresol is added to scavenge *t*-butyl cations and prevent the formation of undesired products. Exposure to HF is harsh and may result in the degradation of the nascent peptides; this led to the use of a less harsh, base-labile reagent (Fmoc).

Fmoc uses piperidine in DMF to remove the protecting group and results in a neutral exposed amino group [7]. TFA is used to cleave the peptide from the resin. The lack of residual charge may lead to increased aggregation of the peptide. Nonetheless, the use of Fmoc is generally preferred over Boc because of the ease of cleavage, despite the increase in cost of synthesis. Finally, Boc SPPS generates fluoride salts, which are highly soluble, whereas Fmoc SPPS yields a TFA salt, which is less soluble.

SPPS had been optimized to yield high purity peptides by using reverse phase HPLC or ion exchange chromatography, for peptides less than 50 amino acids in length. Due to the limitations of SPPS, it was necessary to develop a better technique that not only produces a high yield and purity of protein, but abundantly synthesizes much larger proteins.

2.2.2 Native Chemical Ligation (NCL)

SPPS was combined with native chemical ligation (NCL) [2] to produce longer polypeptides. NCL is based on the reaction between two unprotected synthetic peptides, one of which contains a C-terminal

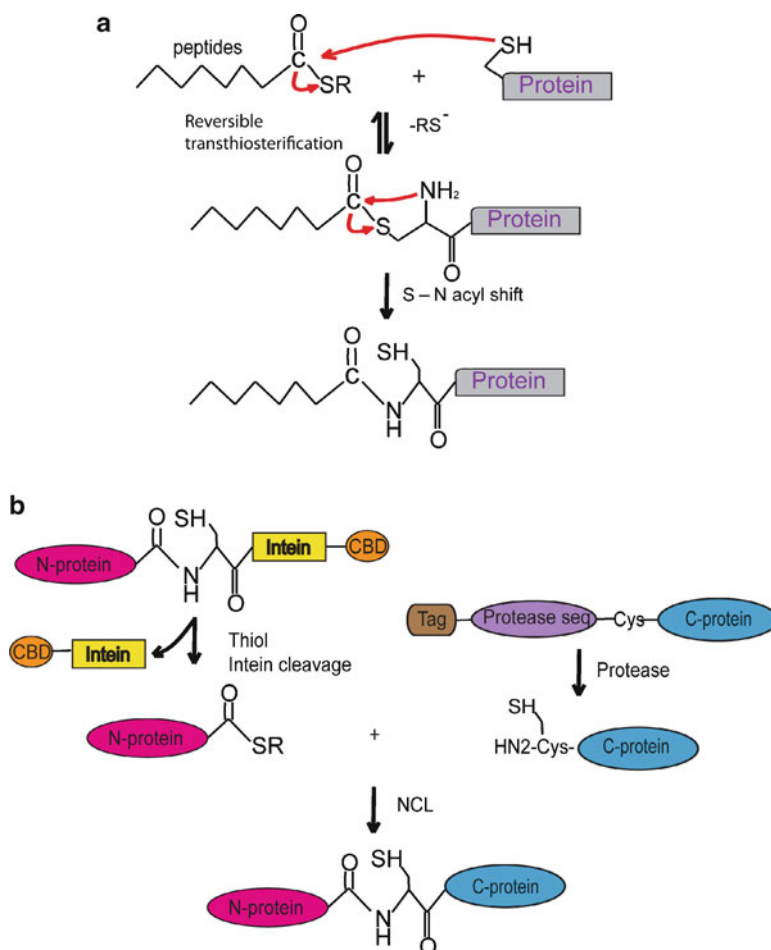


Fig. 2.1 (a) Peptide-protein ligation (semisynthesis). An N-terminal synthetic peptide containing a C-terminal thioester reacts with a C-terminal recombinant protein through an N-terminal cysteine via reversible trans-thioesterification and S-N acyl shift to form a native peptide bond. (b) Protein-protein ligation. A C-terminal thioester is created when intein-CBD is cleaved from the N-terminal protein fragment and an N-terminal cysteine is formed when protease cleaves the C-terminal fragment at the protease site. A native peptide bond is then formed via native chemical ligation. CBD is a chitin binding domain used for ease of purification

thioester (α -thioester) and the other an N-terminal cysteine residue (α -cysteine), to form a native peptide bond. The essence of NCL is the formation of an intermediate thioester-linked product. The thioester-linked product undergoes spontaneous rearrangement, via an intramolecular nucleophilic attack, to reach the final desired amide-linked product (Fig. 2.1). The result is a native polypeptide chain that functions *in vitro* or *in vivo*. NCL is widely used in conjunction with NMR [7, 8] and mass spectroscopy to observe single domains within a full-length protein by ligating together isotope-labeled and unlabeled peptides. In NMR, the technique alleviates the difficulty of obtaining clear and non-overlapping spectra by reducing spectral complexity [8].

NCL is performed in aqueous solution at neutral pH under denaturing conditions. This highly chemoselective reaction depends strongly on the amino acid present at the C-terminal thioester, with Gly increasing the reaction rate and β -branched amino acids, like Leu or Ile, reducing the reaction rate and producing lower yields [9]. Interestingly, cysteine and histidine, which are among the least sterically hindered amino acids, react at the same speed as glycine, while valine, isoleucine and proline, which are more sterically hindered, react less favorably [10].

2.2.3 Expressed Protein Ligation (EPL)

2.2.3.1 Protein-Peptide Ligation (Semisynthesis)

Because of the length limitation of synthetic peptides, chemical ligation of short synthetic peptides with recombinant proteins expressed in *E.coli* was developed to extend the size of ligated proteins from less than one hundred to several hundred amino acids. This method, called semisynthesis [11, 12], also allows unnatural amino acids [13], fluorescent probes [13], or posttranslational modifications [14] to be introduced into any size protein and can be used to attach a synthetic peptide to either the C-terminus or N-terminus of the recombinant expressed protein (Fig. 2.1a). To introduce a synthetic peptide at the N-terminus, the peptide is synthesized with a C-terminal α -thioester and the recombinant protein must have an N-terminal Cys. To position the peptide at the C-terminus, the protein is expressed as a fusion with an engineered intein. Self-cleavage of the intein results in a C-terminal α -thioester that can be used for the ligation with a synthetic peptide possessing an N-terminal Cys.

2.2.3.2 Protein-Protein Ligation

Protein-protein ligation requires overexpression of protein fragments in which the N-terminal fragment is fused to an intein and the C-terminal fragment has a Cys at the N-terminus. The engineered intein catalyzes its own excision to yield N-terminal protein fragments containing reactive termini for optimum ligation. There are three broad categories of inteins: (1) Maxi-inteins that contain a homing endonuclease domain within the core sequence, which is not required for splicing activity [15]; (2) Mini-inteins that lack a homing endonuclease domain [15–17]; (3) Trans-splicing inteins that have no peptide linkage between the N- and C-terminal halves of the intein resulting in two fragments that must come together for splicing activity to occur.

The basic protein splicing mechanism involves three steps (Fig. 2.1b): First, an N \rightarrow S (or N \rightarrow O) acyl shift in which the N-extein is transferred to the $-\text{SH}$ or $-\text{OH}$ group of a Cys or Ser at the N-terminus of the intein. Second, the entire N-extein is transferred, via thioesterification, to a second, conserved Cys/Ser/Thr at the +1 position within C-extein. Third, the resulting branched intermediate undergoes cyclization with a conserved asparagine at the C-terminus of the intein, and the intein is excised as a C-terminal succinimide derivative. Spontaneous chemical rearrangement leads to the formation of an amide bond between the two exteins in an intein-independent manner.

Intein-mediated ligation gene products are derived from *Synechocystis* sp. *dnaB* (*Ssp* DnaB) [18], *Mycobacterium xenopi gyrA* (*Mxe* GyrA) [19], and from *Methanobacterium thermoautotrophicum rir1* (*Mth* RIR1) [20]. The *Ssp* DnaB intein has been engineered to undergo pH or temperature dependent cleavage at the C-terminus to generate a fragment containing the desired N-terminal amino acid residue [21]. *Mxe* GyrA and *Mth* RIR1 inteins have been modified to undergo thio-induced cleavage at the N-termini to yield a C-terminal α -thioester on the resulting fragment [22].

To generate a C-terminal protein fragment for EPL, the protein is designed to contain a specific protease cleavage sequence that, after cleavage, will leave an N-terminal Cys. So far, three proteases have proved to be useful for this purpose: Factor Xa, which cleaves immediately after its recognition sequence, Ile-Glu-Gly-Arg; TeV protease, which cleaves within its recognition sequence, Glu-Asn-Leu-Tyr-Phe-Gln-Cys, between Gln and Cys; and thrombin, which cleaves within its recognition sequence, Leu-Val-Pro-Arg-Cys-Ser, between Arg and Cys.

It is important to choose less sterically hindered amino acids at the ligation site to improve the ligation reaction. It is also important to design a functional assay for the ligation product to ensure that the modifications introduced to facilitate EPL do not influence the structure and biological function of the protein [23]. For example, to construct SH32 protein [24], which consists of the Src homology type 3 (SH3) and type 2 (SH2) domains, the SH3 domain contained an α -thioester at the C-terminus, and the

SH2 domain contained a Cys at the N-terminus. The location of the ligation site was chosen to be within the short linker region between the SH3 and SH2 domains and involved two mutations, N120G and S121C. The S-C mutation is required to facilitate the ligation reaction, whereas the N-G mutation is expected to improve the kinetics of ligation reaction. As a result, the NMR spectra of the ligation product, SH32, and recombinant expressed SH32 were quite similar, which means that the ligation reaction did not affect protein folding, even though a few expected chemical shift changes are observed in the amino acids located spatially close to the ligation site.

2.3 Applications

2.3.1 Conformational Changes

The 400-kDa bacterial core RNA polymerase (RNAP) depends on the binding of σ factors for promoter recognition and specific transcription initiation in RNA polymerization. σ^{70} is responsible for the bulk of transcription during exponential growth. Structural studies confirm [25, 26] that the -35 elements of the binding site for RNAP are recognized by amino acid residues of σ^{70} region 4.2. The latent DNA binding activity of σ^{70} is inhibited by N-terminal region 1.1, which directly masks the DNA binding determinants of region 4.2 [27, 28]. This inhibition is relieved by a conformational change when σ^{70} factor binds to the RNAP core. The autoinhibition of σ^{70} was difficult to resolve by using X-ray crystallography due to the flexibility of region 1.1 [29]. Segmental labeling of region 4.2 and isotope edited NMR spectroscopy was used to observe interactions between regions 1.1 and 4.2 required for σ -factor autoinhibition. A thermostable variant of σ^{70} , σ_A from *Thermatoga maritima* [30] was used to facilitate NMR studies. Two constructs were created, σ -factor with [U - ^{15}N] region 4.2 and σ -factor with a deletion in region 1.1 and [U - ^{15}N] region 4.2.

Segmental isotopic labeling was accomplished by using expressed protein ligation (EPL) [31]. To facilitate the ligation reaction, a Cys was inserted between Gly348 and Lys349. Two different N-terminal fragments, full-length σ_A^* (1–348) and $\Delta 1.1$ - σ_A^* (137–348) which lacks region 1.1, were fused with an *Mxe* GyrA intein-CBD fragment. In this construct an α -thioester group is released by thiolysis to ligate with the C-terminal fragment, σ_A factor region 4.2, [U - ^2H , ^{13}C , ^{15}N]-CG- σ^A (349–399).

Experimentally, segmentally labeled σ^{A*} and $\Delta 1.1$ - σ^{A*} are both active *in vitro*. The σ^{A*} and $\Delta 1.1$ - σ^{A*} constructs are able to bind the -35 promoter DNA with similar affinities in low salt buffer. In high salt buffer, the affinity of σ^{A*} for the -35 promoter DNA is reduced by more than two orders of magnitude as compared to $\Delta 1.1$ - σ^{A*} . This result proves that deletion of region 1.1 allows the truncated σ^A factor, $\Delta 1.1$ - σ^{A*} , to make tight and specific interactions with the -35 promoter DNA, confirming that region 1.1 is involved in the previously observed autoinhibition of σ^A .

$^1\text{H}\{^{15}\text{N}\}$ HSQC-TROSY and $^1\text{H}\{^{13}\text{C}\}$ HSQC experiments were performed to present the spectra of region 4.2 in the context of $\Delta 1.1$ - σ^{A*} and in σ^{A*} (Fig. 2.2). The well-dispersed signals indicate that region 4.2 assumes a defined tertiary fold in $\Delta 1.1$ - σ^{A*} and in σ^{A*} . In contrast, the isolated region 4.2 lacks a defined fold in solution and many peaks are overlapping. Further experiments demonstrated that adding T4Asia, a known ligand of *E. coli* σ^{70} region 4.2, results in significant conformational changes in the NMR spectrum of $\Delta 1.1$ - σ^{A*} (Fig. 2.3), implying that the presence of region 1.1 inhibits binding of T4Asia to region 4.2. There are only minor differences between the NMR spectra of region 4.2 in the context of $\Delta 1.1$ - σ^{A*} and in σ^{A*} . It was concluded that region 1.1 indirectly inhibits σ_A binding to the promoter DNA, possibly by electrostatic interaction [32].

To further understand the nature of autoinhibition, region 1.1 was segmentally labeled and characterized by NMR spectroscopy in the context of full-length σ^A [33]. Region 1.1 (residues 25–120) was expressed in and purified from *E. coli* [32]. The standard set of double and triple resonance 2D and 3D experiments and restraints generated from a series of multi-dimensional NMR experiments were

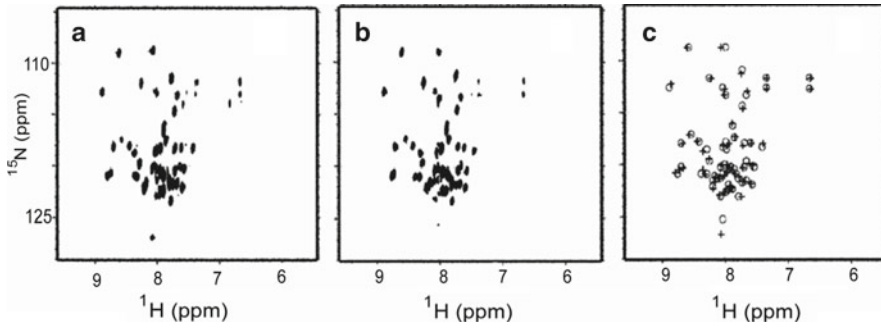


Fig. 2.2 Effect of context on the solution structure of σ^A region 4.2. (a) $^1\text{H}\{^{15}\text{N}\}$ HSQC-TROSY spectrum of [$U\text{-}^{15}\text{N}$] region 4.2 in context of σ^A . (b) $^1\text{H}\{^{15}\text{N}\}$ HSQC-TROSY spectrum of [$U\text{-}^{15}\text{N}$] region 4.2 in context of $\Delta 1.1\text{-}\sigma^A$. (c) The overlay of *panel (a)* and *panel (b)*. *Panel (a)* peaks are represented by *circles*, and *panel (b)* peaks are shown as *crosses* (This figure is reproduced from Camarero et al. [32])

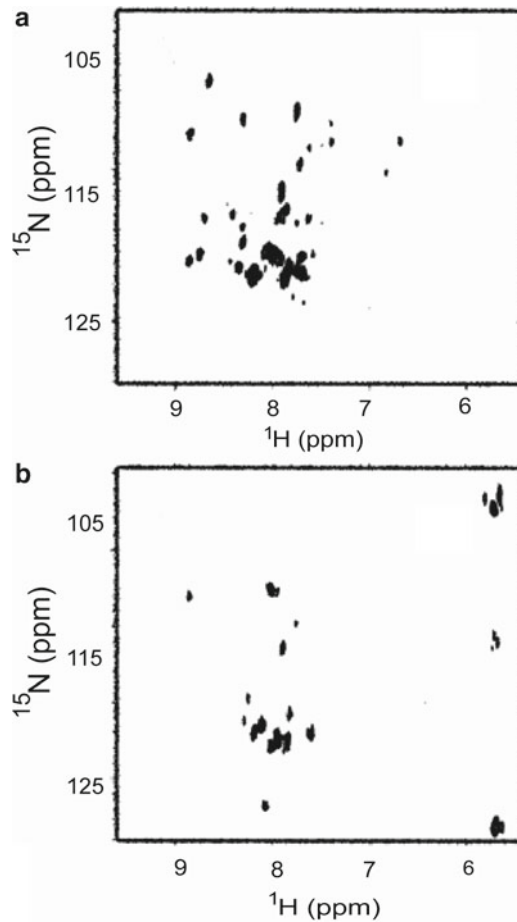


Fig. 2.3 Binding of T4 AsiA and promoter DNA to $\sigma^{1.1}\text{-}\sigma^A$. $^1\text{H}\{^{15}\text{N}\}$ HSQC-TROSY spectra of [$U\text{-}^{15}\text{N}$] region 4.2 in context of $\Delta 1.1\text{-}\sigma^A$ with 1.2 M equivalents of purified AsiA (a) and promoter DNA (b) (This figure is reproduced from Camarero et al. [32])

performed for the chemical shifts assignments and the solution structure determination of region 1.1 [27, 28, 32]. The $^1\text{H}\{^{15}\text{N}\}$ -HSQC NMR spectrum of region 1.1 in the context of full-length σ^A was sufficiently broadened to allow the interaction surface of σ^A to be mapped. These results, in combination with cross-linking experiments [34], clearly indicate that region 1.1 interacts with regions 3 and 4.1 of the full length σ^A .

Importantly, this work shows that segmental isotopic labeling does not interfere with the folding of full-length proteins and allows us to observe domain-domain interactions without the absolute requirement for assignments. It can be used to observe the effects of ligands on a segmentally labeled domain in the context of full-length protein by using isotope edited NMR experiments.

2.3.2 Interdomain Interactions

Advances in the study of multidomain proteins by using solution NMR have been made possible in recent years by the development of new segmental isotope labeling methods that identify and map interdomain interactions and allow structural characterization in the absence and presence of such interactions. Skrisovska and Allain [35] developed a technique to segmentally isotope label multidomain proteins and to provide a high yield recovery of ligated product. The protocol employs an on-column expressed protein ligation (EPL) step and permits ligation of insoluble, non-interacting and improperly folded domains. The technique was successfully demonstrated by using two multidomain proteins, heterogenous nuclear ribonucleoprotein L (hnRNP L) and Npl3p, each of which contain RNA recognition motifs (RRMs).

hnRNP L is an abundant RNA-binding protein involved in alternative splicing and mRNA degradation [36]. hnRNP L contains four RRM3 and RRM4, which are connected by a long linker region (417–461), may interact with one another. Two constructs were prepared, the first, RRM3-Mxe GyrA-CBD, consists of RRM3 fused at the C-terminus to the *Mxe* GyrA intein and a chitin binding domain (CBD), the second, CBD-Ssp DnaB-RRM4, consists of RRM4 fused at the N-terminus to CBD and the *Ssp* DnaB intein. To ligate the constructs following cleavage of the intein-CBD moieties requires a cysteine residue at the N-terminus of RRM4 that can react with a thioester at the C-terminus of RRM3. The ligation site lies in the linker region at position 452. Since the linker does not contain any cysteine residues, serine 452 was mutated to cysteine (S452C) at the N-terminus of RRM4.

Each construct was overexpressed in *E. coli* under non-labeling and [U - ^{15}N] or [U - ^{15}N , ^{13}C] labeling growth conditions and resulted in the formation of inclusion bodies. Inclusion bodies from labeling and non-labeling overexpression were combined, solubilized in 8 M urea, refolded by fast dilution and bound to a chitin column. Intein cleavage and subsequent ligation was induced by the addition of sodium 2-mercaptoethanesulfonate (MESNA); the reaction was allowed to proceed for 24 h at 37°C on the column before eluting the final product, hnRNP RRM34. Ligation efficiency was ~90%, hnRNP L RRM34 ran as a single band on SDS-PAGE and its molecular weight was confirmed by mass spectrometry. The cleavage reaction, estimated to be 80–90% complete for *Mxe* GyrA [19] and 60–70% for *Ssp* DnaB [18], proved to be the limiting step in preventing an even higher ligation efficiency.

$^1\text{H}\{^{15}\text{N}\}$ -HSQC and $^1\text{H}\{^{13}\text{C}\}$ -HSQC spectra were collected for hnRNP L RRM34 containing [U - ^{13}C , ^{15}N] RRM3 and [U - ^{15}N] RRM4 and compared to $^1\text{H}\{^{15}\text{N}\}$ -HSQC and $^1\text{H}\{^{13}\text{C}\}$ -HSQC spectra collected for [U - ^{13}C , ^{15}N]- hnRNP L RRM34 to confirm that the ligated product was properly folded. 2D homonuclear and 3D ^{13}C -edited NOESY half-filter spectra acquired using segmentally labeled hnRNP L RRM34 identified 101 NOE crosspeaks between the two domains, confirming that the domains interact. The interaction interface was defined and the structure of the ligated structure is currently being characterized.

Npl3p (nuclear protein localization) is a yeast RNA binding protein. It is a member of the serine/arginine-rich (SR) protein family that selects and regulates splice sites in eukaryotic mRNA. Npl3p contains two RRM s (RRM1 and RRM2) connected by a short linker and a C-terminal glycine/arginine-rich domain. Two constructs were prepared: RRM1-*Mxe* GyrA-CBD and CBD-*Ssp* DnaB-RRM2 to yield Npl3p RRM12. A cysteine, introduced into the short linker by mutating serine 193 (S193C) was used as the N-terminal residue of RRM2.

Each construct was overexpressed in *E. coli* under non-labeling and [U - ^{15}N] or [U - ^{15}N , ^{13}C] labeling growth conditions and was soluble. The on-column cleavage efficiency of both proteins was 80–90%, but the ligation efficiency was only ~10% as estimated from SDS-PAGE. The chitin column elution containing ligated and non-ligated product was concentrated to ~1 mM and further incubated at 42°C for 24 h; this additional step improved the ligation efficiency to 80–90%. RRM12 was separated from the non-ligated species by using gel filtration chromatography. To show that ligation is independent of the folded state of the proteins, segmentally labeled protein was successfully prepared by concentrating and ligating protein domains in the presence of 6 M guanidinium chloride. Ligation using a naturally occurring cysteine (C211) located within RRM2 left the domain unstructured but capable of undergoing ligation, albeit at a lower level than obtained when the domains were properly folded.

$^1\text{H}\{^{15}\text{N}\}$ -HSQC spectra acquired for each domain indicates that both fold properly after cleavage and during the ligation reaction. There were no significant differences between the $^1\text{H}\{^{15}\text{N}\}$ -HSQC spectra of segmentally and uniformly labeled Npl3p RRM12 indicating that the S193C mutation has no effect on the protein fold of ligated RRM12. An overlay of the spectra of each domain is very similar to the $^1\text{H}\{^{15}\text{N}\}$ -HSQC spectrum of Npl3p RRM12, suggesting that there is no interaction between the two domains. No NOE crosspeaks were observed between RRM1 and RRM2 in 3D ^{13}C -edited NOESY half-filtered data collected using a segmentally labeled sample in which only one of the RRM s was ^{13}C -labeled, further indicating that RRM1 and RRM2 do not interact. Minor changes in chemical shifts indicate that there may be small changes in the conformation and dynamics of the domains or weak interactions for which NOEs were not detected. Because RRM1 and RRM2 do not interact, the structure of each domain was determined individually, greatly simplifying the analysis.

In sum, the on-column ligation technique is very robust, interacting domains ligate more efficiently than non-interacting domains, but this is not necessary for successful ligation. Low level ligation efficiency is improved by concentrating the eluted fragments, thereby increasing the concentration of protein termini available for the ligation reaction. Successful ligation of two protein fragments is largely independent of their solubility and folding state. This technique will be broadly applicable in future solution NMR studies on large, multidomain proteins.

2.3.3 Glycoproteins

Glycosylation is a common post translational modification (PTM) that facilitates a variety of biological processes involving primarily inter- and intra-cellular communication. Understanding the molecular basis for these processes is limited by the dearth of structural information available for glycoproteins. The use of NMR to acquire structural information is hampered by the inability to generate sufficient quantities of uniformly glycosylated protein and the spectral complexity arising from overlapping signals attributed to carbohydrate and protein moieties. Segmental labeling can help overcome these problems.

Slynko et al. [37] used *in vitro* glycosylation to attach an unlabeled glycan to [U - ^{13}C , ^{15}N]-labeled protein and NMR spectroscopy to deduce the structures of the N-linked oligosaccharide PTM and the corresponding modified protein (Fig. 2.4). The protein, AcrA^{61–210 $\Delta\Delta$} from *Campylobacter jejuni*, is a drug efflux pump protein with broad substrate specificity that is easily glycosylated *in vivo* and *in vitro*. The *in vitro* glycosylation method requires three components, each isolated from separate, dedicated strains of *E. coli*: Oligosaccharyltransferase, PglB, which is purified from solubilized membrane fractions [38]; a lipid-linked oligosaccharide (LLO) prepared from cells containing an inactive

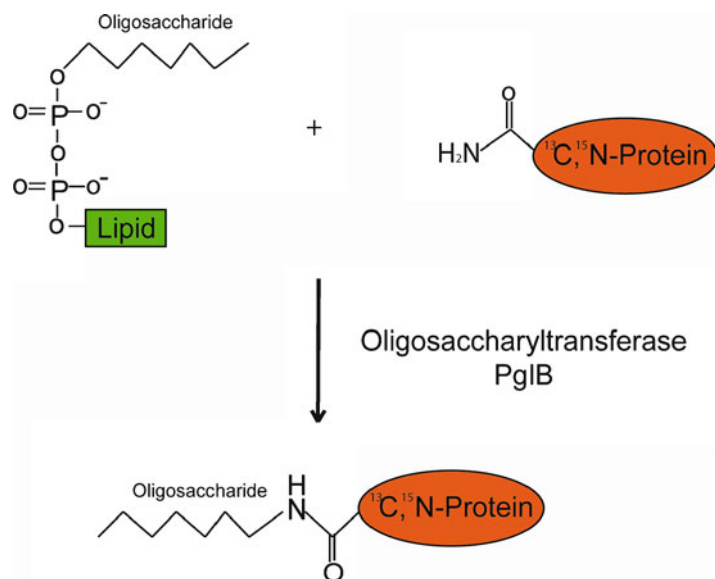


Fig. 2.4 *In vitro* synthesis of glycosylated protein. Oligosaccharyltransferase PglB mediates the reaction between an oligosaccharide attached to a lipid and an Asn of a labeled protein

pgl ORF, and the [U - ^{13}C , ^{15}N] target protein, AcrA^{61–210 $\Delta\Delta$} . The small (13 kDa) protein contains a lipoyl domain and an extended loop that includes the glycosylation site. In addition, the model protein contains two deletions ($\Delta\text{F97-N117}$ and $\Delta\text{F146-D166}$) and two point mutations (K96Q and K131Q). Purified AcrA^{61–210 $\Delta\Delta$} was ~90% glycosylated *in vitro* and further purified by Ni-NTA chromatography before using in NMR experiments. The protein yields good dispersion and line widths in NMR.

By using [U - ^{13}C , ^{15}N] protein, chemical shift overlap between the protein and the carbohydrate is resolved using specific filtered/edited 2D NOESY-type experiments that suppress resonances from the labeled protein and allow the non-labeled glycan to be assigned. In these experiments NOE transfer becomes very efficient due to increased overall tumbling rate. The use of a very high magnetic field (900 MHz) also maximizes the sensitivity and resolution of the NOESY spectra, enabling long-range distance restraints to be extended to 6 Å. Amino acid type-selective ^1H - ^{15}N correlation and standard triple resonance experiments were used to identify N42 as the glycosylation site [39]. The observed ^1H and ^{15}N resonances are consistent with those previously observed for glycosylated asparagines [40]. The segmentally labeled complex contains a lipoyl domain fold with four N-terminal and four C-terminal intertwined β -strands forming a β -sandwich. The glycosylation site is contained within a flexible loop that lies between the two half-lipoyl domains motifs. Flexibility was assessed by acquiring steady-state heteronuclear $^1\text{H}\{^{15}\text{N}\}$ NOE spectra. The NH of the glycosylated N42 side chain is less flexible than the backbone of the surrounding amino acid residues most likely due to the reduced degree of conformational freedom imparted by the attached glycan.

By acquiring 2D ^{13}C -filtered-filtered NOESY and natural abundance ^{13}C -HSQC spectra, assignment of all residues in the attached glycan was accomplished. 2D ^{13}C -filtered-filtered NOESY suppresses all protein resonances, resulting in a NOESY spectrum of the unlabeled glycan. 2D ^{15}N -filtered-filtered NOESY, containing NOEs to the carbohydrate amides while suppressing protein amide signals, confirmed the assignment of the linked polysaccharide structure. In total, 125 interproton distance constraints within the attached glycan were obtained. Half of which are inter-residue NOEs (~11 per glycosidic linkage). The final structure had an RMSD of 0.56 ± 0.10 Å. The glycan structure was calculated exclusively by using torsion angle dynamics from experimentally derived upper limit NOE distance restraints.

Segmental labeling of a glycosylated target protein is a new strategy for studying glycoproteins by NMR [41]. Problems associated with the spectral overlap between protein and oligosaccharide resonances are avoided. This technique also generates chemically pure glycosylated protein, which is difficult to achieve by enzymatic glycosylation.

2.3.4 Reaction Mechanisms

To gain insight into the mechanistic details of autocatalysis, Romanelli et al. (2004) [12] used a combination of NMR spectroscopy and segmental isotopic labeling to study the structure of an active protein splicing precursor. In the first step of the splicing reaction an N→S (or N→O) acyl shift occurs in which the N-extein is transferred to the –SH or –OH group of a Cys or Ser at the N-terminus of the intein. In the ground state destabilization model, it is believed that the scissile (–1) peptide bond is distorted from its most favorable conformation rendering it susceptible to nucleophilic attack. To examine the validity of this model in the autocatalytic splicing reaction mechanism, semisynthesis was employed to prepare an N-extein fusion of the *Mxe* GyrA intein in which the protein was uniformly labeled with ¹⁵N, and the peptide ¹³C labeled at the (–1) phenylalanine residue. This ligation scheme allowed the scissile (–1) peptide bond to be doubly labeled and examined in detail.

The peptide, H-AAMR[¹³C']F-SR, was synthesized by Boc-N^a-SPPS using the *in situ* neutralization/2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) activation protocol [6] on 3-mercaptopropionamide-4-methylbenzhydramine (MNHA) resin. The peptide was cleaved from the resin using anhydrous HF containing 4% (v/v) *p*-cresol, purified by reverse phase chromatography (RP-HPLC) and characterized by using electrospray mass spectrometry (ESMS). DNA coding for GyrA protein was amplified by PCR off plasmid pTXB1 (NEBiolabs), which encodes a GyrA intein with a single mutation (N198A) that prevents cleavage of the intein-C-extein bond, but does not affect N-terminal splicing [42]. The PCR product was cloned into pTrcHisA (Invitrogen) to generate pTrcHis-Xa-GyrA_{WT}, in which a factor Xa cleavage site is inserted between the His-tag and the GyrA coding region. pTrcHis-Xa-GyrA_{WT} was used as a template to prepare a host of mutant GyrA proteins, including GyrA(H75A), which contains a H75A mutation that abolishes N-terminal splicing activity. GyrA_{WT} was overexpressed from plasmid pTrcHis-SH3-GyrA_{WT}, which contains a Src homology domain 3 between the His-tag and the intein.

To uniformly label the inteins, pTrcHis-Xa-GyrA(H75A) and pTrcHis-SH3-GyrA_{WT} were transformed into *E. coli* strain BL21(DE3), grown in minimal (M9) medium containing 0.2% (w/v) ¹⁵N-NH₄Cl as the sole nitrogen source. The cells were induced with IPTG and the fusion proteins purified by using Ni²⁺-NTA affinity chromatography. To prepare [*U*-¹⁵N]GyrA, His-SH3-GyrA was incubated overnight in DTT cleavage buffer (50 mM Tris, pH 8, 100 mM DTT). To prepare [*U*-¹⁵N]GyrA(H75A), His-Xa-GyrA(H75A) was incubated in proteolysis buffer (50 mM Tris, pH 8, 100 mM NaCl, 1 mM CaCl₂) in the presence of factor Xa for 10 h at room temperature. Both proteins were further purified by using RP-HPLC and characterized by ESMS.

Two segmentally labeled constructs, AAMR[¹³C']F-[*U*-¹⁵N]GyrA and AAMR[¹³C']F-[*U*-¹⁵N]GyrA(H75A), were prepared by chemically ligating H-AAMR[¹³C']F-SR with either [*U*-¹⁵N]GyrA or [*U*-¹⁵N]GyrA(H75A). Reactions were carried out using ~1 mM protein and ~10 mM peptide in 0.1 M NaP_i buffer, pH 8, containing 6 M guanidinium chloride, 3% MESNA and 2% ethanethiol, for 5 h at room temperature. The ligation products were purified by using RP-HPLC and characterized by using ESMS. Purified protein was re-folded by stepwise dialysis into NMR buffer (20 mM KP_i, pH 6.6, 100 mM NaCl) for spectroscopy.

The uniformly labeled N-exteins have similar ¹H{¹⁵N}HSQC spectra, only a small number of signals are unique to one construct or other. The spectral similarity factor [32] between the constructs is 16 Hz, indicating that the H75A mutation does not result in a major structural rearrangement.

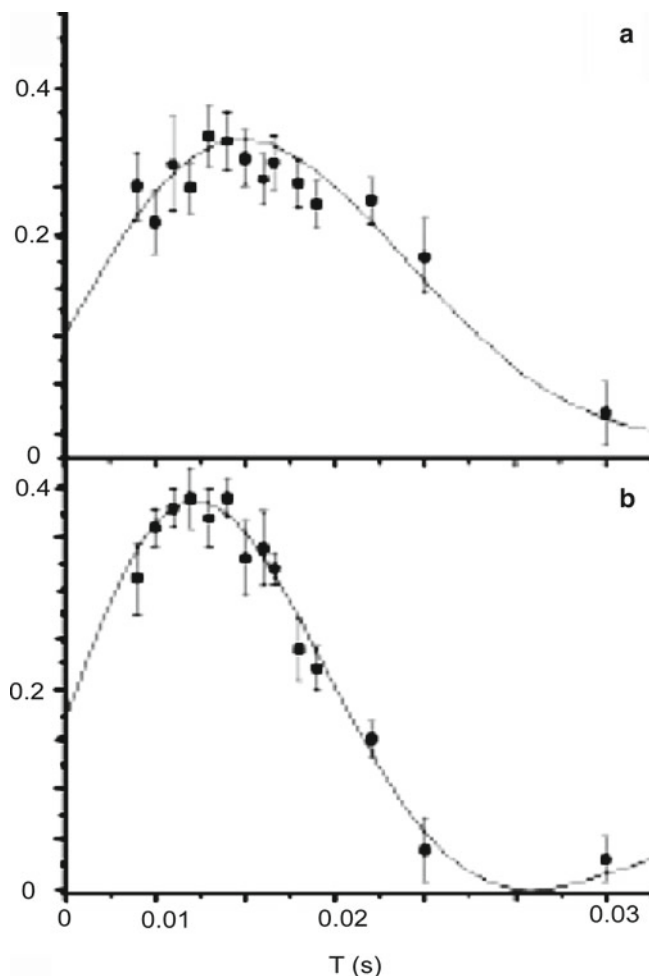


Fig. 2.5 Determination of the $^1J_{\text{NC}'}$ coupling constant for the scissile (-1)amide in AAMR[$^{13}\text{C}'$]F-[$U\text{-}^{15}\text{N}$]GyrA and AAMR[$^{13}\text{C}'$]F-[$U\text{-}^{15}\text{N}$]GyrA(H75A), respectively. The time evolution of the normalized peak intensities extracted from a series of H(N)CO experiments was nonlinearly fit to Eq. 2.1 to give the single-bond coupling constants. (a) Fit obtained for AAMR[$^{13}\text{C}'$]F-[$U\text{-}^{15}\text{N}$]GyrA, $^1J_{\text{NC}'} = 12.3 \pm 0.3$ Hz and $R_2 = 17.3 \pm 0.3$ s $^{-1}$. (b) Fit obtained for AAMR[$^{13}\text{C}'$]F-[$U\text{-}^{15}\text{N}$]GyrA(H75A), $^1J_{\text{NC}'} = 16.2 \pm 0.2$ Hz and $R_2 = 17.5 \pm 0.4$ s $^{-1}$ (This figure is reproduced from Romanelli et al. [12])

The small differences are attributed to localized changes in structure and/or a magnetic shielding effect due to the loss of the imidazole ring in the H75A construct. Since the H75A mutant renders the intein inactive, these results imply that H75 plays an essential, catalytic role in the splicing reaction.

The dual-isotopic labeling pattern allowed unequivocal assignment of the scissile (-1) amide resonances. The most dramatic effect is on the amide proton, which shifts from 6.61 ppm in AAMR[$^{13}\text{C}'$]F-[$U\text{-}^{15}\text{N}$]GyrA to 10.01 ppm in AAMR[$^{13}\text{C}'$]F-[$U\text{-}^{15}\text{N}$]GyrA(H75A). The chemical shift of the (-1) amide proton in AAMR[$^{13}\text{C}'$]F-[$U\text{-}^{15}\text{N}$]GyrA lies upfield of the typical value observed for cysteine within a random coil [8], however, neither the ^{13}C nor the ^{15}N chemical shifts of the (-1) amide are unusually shifted. For AAMR[$^{13}\text{C}'$]F-[$U\text{-}^{15}\text{N}$]GyrA(H75A), all of the resonances are shifted downfield.

One-bond dipolar couplings were evaluated from the time evolution of normalized peak intensities derived from HNCO-type experiments [43]. The $^1J_{\text{NC}'}$ values obtained for the scissile (-1) amide were 12.3 ± 0.3 and 16.2 ± 0.2 Hz, for AAMR[$^{13}\text{C}'$]F-[$U\text{-}^{15}\text{N}$]GyrA and AAMR[$^{13}\text{C}'$]F-[$U\text{-}^{15}\text{N}$]GyrA(H75A), respectively (Fig. 2.5). Typical $^1J_{\text{NC}'}$ values reported for proteins are 13–17 Hz [44, 45].

Amide $^1J_{\text{NC}'}$ values correlate with hydrogen bonding: Hydrogen bonding to the amide carbonyl increases the coupling constant while hydrogen bonding to the amide hydrogen decreases the coupling constant [45–47]. The authors speculate that the low $^1J_{\text{NC}'}$ coupling observed for the scissile amide in the active construct is evidence of a backbone distortion primarily due to the fact that the H75A mutant has a significantly higher $^1J_{\text{NC}'}$ value similar to what is typically observed in proteins. The conclusion is that the first step in protein splicing is facilitated in part by destabilizing the scissile amide bond, in agreement with the proposed destabilization theory.

Another study [48] investigated the next step in the autocatalytic splicing reaction: the excision of the intein. They showed that intern-succinimide formation, which follows branched intermediate formation, is the rate-limiting step in protein splicing, and that this helps regulate the overall fidelity of the reaction. To examine the hypothesis that structural changes during the splicing reaction reflects a re-organization of the catalytic apparatus to accelerate succinimide formation at the C-terminal splice junction, branched intermediates of the *Mxe* GyrA intein were prepared using semi-synthesis.

Branched peptides, corresponding to residues 185–198 of the *Mxe* GyrA intein, were synthesized by using Fmoc/tBu SPPS using the HBTU activation protocol for linear chain assembly on Rink-amide ChemMatrix resin. Branched chain assembly utilized HOBt and (N,N'-diisopropylcarbodiimide) DICP activation. The peptides were cleaved from the resin using a cocktail comprised of TFA, triisopropylsilane (TIS) ethandiol and water, purified by RP-HPLC and characterized by using ESI-MS. DNA coding for *Mxe* GyrA intein residues 1–184 was amplified from pTXB1 to incorporate a factor Xa site and His-tag and cloned back into pTXB1, which also contains a chitin binding domain (CBD). The first intein residue, Cys1, was mutated to Ser (C1S) to generate the final fusion product, His-Xa-Ser1-(2–184)-GyrA-CBD, used to prepare branched constructs.

To uniformly [U -, ^{15}N] label the fusion protein, the appropriate plasmid was transformed into *E. coli* strain BL21(DE3), grown in minimal (M9) medium containing 0.2% (w/v) $^{15}\text{N-NH}_4\text{Cl}$ as the sole nitrogen source. Overexpressed protein was purified from inclusion bodies by using Ni-NTA affinity chromatography under denaturing conditions and renatured by stepwise dialysis. Thiolysis was performed in 50 mM Tris-HCl, pH 7.6, 100 mM NaCl, 1 mM EDTA and 200 mM MESNA for 2 days at 25°C to yield the protein α -thioester. Products were purified by using RP-HPLC on a C4 column. Purified thioester was lyophilized and refolded by stepwise dialysis into 50 mM Tris-HCl, pH 7.6, 100 mM NaCl and 100 mM MESNA. The N-terminal tag was removed with factor Xa and the final products purified by C4 RP-HPLC and characterized by ESI-MS.

EPL was performed using a 3:1 equivalent ratio of peptide to protein α -thioester in ligation buffer (100 mM NaPi, pH 7.8, 6 M guanidinium chloride and 100 mM NaCl) containing 100 mM MESNA and 10 mM Tris[2-carboxyethyl] phosphine (TCEP) for 5 days at 4 °C. Semisynthetic protein was separated from unreacted material on a Ni-NTA column and the ligated product purified by C4 RP-HPLC and characterized by using ESI-MS. Purified constructs were refolded by stepwise dialysis into NMR buffer (50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM TCEP) for analysis.

NMR spectroscopy was used to compare the local structure around the scissile +1 peptide bond in the context of the linear precursor and the branched intermediate. The linear construct showed a signal around 8 ppm at pH 7.5 and 4°C; in contrast, no signal was obtained for the branched construct under the same condition, however, when the branched construct was denatured, a clear HNC0 signal was obtained. The lack of signals in the HNC0 spectrum of the branched construct reflects an exchange process, either chemical and/or conformational, around the labeled amide. A new construct was prepared in which two peptide bonds, the scissile +1 amide and the amide connecting Phe 194 and Val 195, were labeled with ^{13}C and ^{15}N . A single resonance was observed at 7.89 ppm at pH 7.5, while two peaks were detected, at 7.89 and 8.29 ppm, at pH 4.5. This observation was reversible, demonstrating that the signal from the scissile +1 amide is highly sensitive to pH, whereas that of the Phe 194-Val195 amide is not. The $^1J_{\text{NC}'}$ coupling constants for the scissile +1 and Phe 194-Val 195 amides at pH 4.5 were found to be 15.4 ± 0.5 and 15.5 ± 0.2 Hz, respectively, which indicates a normal trans-planar conformation.

Segmental labeling proved to be a viable technique to analyze the mechanism of intein mediated protein splicing reaction at atomic resolution and defined key conformations of the protein backbone preceding the enzymatic catalysis.

2.4 Protocols

2.4.1 Semisynthesis of a Segmental Isotopically Labeled Protein Splicing Precursor

To determine the structure of an active N-extein-intein splicing precursor, [12] used semisynthesis to prepare segmentally isotopic labeled constructs in which a short N-extein peptide α -thioester, H-AAMR[$^{13}\text{C}'$]F-SR, is ligated to an intein sequence derived from the *Mycobacterium xenopi* DNA gyrase A (*Mxe* GyrA) intein. The peptide is prepared by Boc-Na-SPPS and contains a single ^{13}C isotope at the C' position of the phenylalanine. The intein protein is overexpressed as a uniformly ^{15}N labeled [U -, ^{15}N] polyhistidine-cleavage site-intein fusion product. This approach results in only the scissile (-1) amide being dual labeled with ^{13}C and ^{15}N .

2.4.1.1 Peptide Synthesis

1. The peptide, H-AAMR[$^{13}\text{C}'$]F-SR, is synthesized on 3-mercaptopropionamide-4-methylbenzhydrylamine (MBHA) resin by using the *in situ* neutralization/2-(H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) activation protocol for Boc-SPPS.
2. The peptide is cleaved off the resin by using anhydrous HF containing 4% (v/v) *p*-cresol for 1 h at 4°C.
3. The peptide is purified by preparative reverse phase HPLC (Vydac C18 resin) using a linear gradient of 13.5–31.5% solution B (9:1 MeCN:water, 0.1% trifluoroacetic acid) over 60 min at a flow rate of 3 mL/min. The final product, ~40 mg of purified peptide, is characterized by using electrospray mass spectrometry (ESMS).

2.4.1.2 Cloning and Protein Expression

1. Plasmid pTXB1 (NEBiolabs) encodes the GyrA intein with a single mutation, N198A, which prevents cleavage of the intein–C-extein peptide bond but does not affect the N-terminal splicing reaction. DNA encoding the *Mxe* GyrA intein (residues 1–198) is PCR amplified using the pTXB1 vector as a template.
2. To construct the plasmid pTrc-His-Xa-GyrA_{WT} which encodes a factor Xa cleavage sequence between a poly(His) tag and the wild type intein, the PCR product is cloned into the *Bam*HI and *Hind*III restriction sites of pTrcHisA (Invitrogen).
3. pTrc-His-Xa-GyrA_{WT} is used as a template for site-directed mutagenesis using the QuikChange site-directed mutagenesis kit (Stratagene) to generate single site mutations in the GyrA intein, in particular H75A, which abolishes N-terminal splicing activity.
4. To overexpress the fusion proteins for segmental labeling, *E. coli* strain BL21 (DE3), is transformed with either pTrc-His-Xa-GyrA_{H75A} or pTrc-His-SH3-GyrA_{WT} and grown to mid-log phase at 37°C in LB medium. pTrc-His-SH3-GyrA_{WT} (V. Muralidharan, Rockefeller University), which encodes

the Src homology 3 (SH3) domain of murine c-Crk-II between the poly(His) tag and the intein, is used to overexpress the wild type fusion protein.

5. For [U - ^{15}N]-labeling, LB medium is replaced by M9 minimal medium containing 0.2% (w/v) $^{15}\text{NH}_4\text{Cl}$ as the sole nitrogen source.
6. Overexpression is induced with 0.4 mM isopropyl β -D-thiogalactoside (IPTG) at 37°C for 5 h, after which cells are harvested and lysed.
7. Fusion protein is purified by affinity chromatography using a Ni^{2+} high-trap column (Amersham) and dialyzed into 50 mM Tris-HCl, pH 8.0, 1 mM EDTA.

2.4.1.3 Generation of [U - ^{15}N]-GyrA_{H75A} and [U - ^{15}N]-GyrA_{WT}

1. [U - ^{15}N]-GyrA_{WT} is generated by incubating purified, labeled His-SH3-GyrA_{WT} overnight in DTT cleavage buffer (50 mM Tris-HCl, pH 8.0, 100 mM DTT).
2. [U - ^{15}N]-GyrA_{H75A} is generated by incubating purified, labeled His-Xa-GyrA_{H75A} in proteolysis buffer (50 mM Tris-HCl, pH 8.0, 0.1 M NaCl, 1 mM CaCl_2) with factor Xa for 10 h at room temperature.
3. Both proteins are purified to >95% homogeneity by preparative RP-HPLC and characterized by ESMS.

2.4.1.4 Preparation of AAMR[^{13}C]F-[U - ^{15}N]-GyrA_{WT} and AAMR[^{13}C]F-[U - ^{15}N]GyrA_{H75A}

1. Ligation reactions between H-AAMR[^{13}C]F-SR and either -[U - ^{15}N]-GyrA_{WT} or -[U - ^{15}N]-GyrA_{H75A}, are initiated by dissolving purified, lyophilized peptide (10 eq) and protein (1 eq) in ligation buffer (6 M Gdm-HCl, pH 8.0, 0.1 M NaPi containing 3% MESNA and 2% ethanethiol). The final concentration of peptide is ~10 mM and that of protein is ~1 mM. The reaction is complete after 5 h at room temperature.
2. The ligation products are purified by preparative RP-HPLC and characterized by ESMS.
3. Purified proteins (1 mg/mL) are folded by stepwise dialysis at 4 °C from denaturing buffer (6 M Gdm-HCl, pH 6.6, 0.1 M NaCl, 20 mM KPi, 1 mM DTT) into NMR sample buffer (20 mM KPi, pH 6.6, 0.1 M NaCl).

2.4.1.5 NMR Spectroscopy

1. NMR samples are prepared by concentrating labeled proteins to 100 μM in NMR sample buffer containing 10% D_2O and 0.01% NaN_3 .
2. $^1\text{H}\{^{15}\text{N}\}$ HSQC (heteronuclear single quantum correlation) and 2D planes of HNC0 spectra are collected at 4 °C on a spectrometer equipped with a cryoprobe.
3. For the HSQC experiments, 512 complex points are collected in the ^1H and ^{15}N dimensions. In the 2D H{N}CO experiments, 512 complex points are collected in the ^1H dimension and 40 complex points in the ^{13}C dimension. Data sets are multiplied by a cosine-bell window function and zero-filled to 1,000 points using XWINNMR (Bruker) before Fourier transformation. The corresponding sweep widths are 12.5, 12, and 30 ppm in the ^1H , ^{13}C , and ^{15}N dimensions, respectively.
4. Experimental amide $^1\text{J}_{\text{NC}}$ coupling constants are obtained by fitting the time evolution of the normalized peak intensities extracted from a series of HNC0-type experiments using the expression:

$$I_k = \exp(-4t_1 R_{2k}) \sin^2(2\pi {}^1\text{J}_{\text{NC}} t_1) \quad (2.1)$$

where I_k is the normalized peak intensity for peak k , R_{2k} is the transverse relaxation time for peak k , and t_1 is the indirect dimension delay.

2.4.2 Expressed Protein Ligation of a Segmentally Labeled Bacterial σ Factor

To examine the effect of σ factor region 1.1 interactions with promoter DNA, and, in particular, intermolecular interactions between regions 1.1 and 4.2, segmentally labeled σ factor was prepared containing $\sigma^A(1-348)$ -GyrA or $\Delta 1.1$ - $\sigma^A(137-348)$ -GyrA ligated to [U - 2H , ^{13}C , ^{15}N]-CG- $\sigma^A(349-399)$ [12].

2.4.2.1 Cloning and Protein Expression

1. Intein proteins $\sigma^A[1-348]$ -GyrA-CBD and $\Delta 1.1$ - $\sigma^A[137-348]$ -GyrA-CBD are expressed as a chitin binding domain fusions (CBD) in LB medium off pTXB1 (NEBiolabs) in *E. coli* strain BL21(DE3)pLysS.
2. The proteins are purified on chitin-agarose beads (NEBiolabs).
3. Region 4.2 of σ^A is expressed off pET28 (Novagen) in *E. coli* strain BL21(DE3) as a His-tagged fusion containing a factor Xa cleavage site between the poly(His) tag and region 4.2. [U - 2H , ^{13}C , ^{15}N]-CG- $\sigma^A(349-399)$ protein is prepared by growing the cells in M9 minimal medium in 2H_2O containing 0.2% [U - ^{13}C]glucose and 0.1% $^{15}NH_4Cl$. Introducing a Gly residue immediately after the Cys was found to greatly improve the yield of the cleavage reaction.
4. His-tagged protein is purified by affinity chromatography on Ni-nitrilotriacetate (NTA) beads (Qiagen) followed by preparative RP-HPLC on a Vydac C18 column.

2.4.2.2 Generation of Labeled Species

1. [U - 2H , ^{13}C , ^{15}N]-CG- $\sigma^A(349-399)$ is generated by incubating purified, triple-labeled His-Xa-CG- $\sigma^A(349-399)$ in proteolysis buffer (50 mM Tris-HCl, pH 8.0, 0.1 M NaCl, 1 mM $CaCl_2$) with factor Xa for 10 h at room temperature.
2. The protein is purified to >95% homogeneity by preparative RP-HPLC.

2.4.2.3 Preparation of Segmentally Labeled Species

1. Ethyl α -thioester derivatives of $\sigma^A[1-348]$ -GyrA-CBD and $\Delta 1.1$ - $\sigma^A[137-348]$ -GyrA-CBD are generated *in situ* in the ligation mixture by thiolysis of chitin beads. Equal molar amounts of each of the two fragments ([U - 2H , ^{13}C , ^{15}N]-CG- $\sigma^A(349-399)$ and $\sigma^A[1-348]$ -GyrA-CBD or $\Delta 1.1$ - $\sigma^A[137-348]$ -GyrA-CBD) are used at a concentration of ~ 50 μM each. The ligation reaction is carried out in ligation buffer (25 mM KPi, pH 7.2, 200 mM Gdm-HCl, 250 mM NaCl, 1 mM EDTA containing 0.2% octyl glucoside and 3% (v/v) ethanethiol) overnight at room temperature.
2. The slurry is filtered and the beads washed several times with ligation buffer. All washes are combined with the supernatant.
3. The ligated protein is concentrated and exchanged into storage buffer (30 mM Tris-HCl, pH 7.6, 100 mM NaCl, 20 mM CHAPSO, 20 mM DTT). *N.b.* CHAPSO is 3-[(3-cholamidopropyl)dimethylammonio]-2-hydroxy-1-propanesulfonate.

2.4.2.4 NMR Spectroscopy

1. NMR samples are prepared by exchanging pure protein (final concentration 100–200 μM) into 30 mM Tris-HCl, pH 7.6, containing 100 mM NaCl, 20 mM CHAPSO, 20 mM [$^2H_{10}$]DTT, 0.1% NaN_3 , and 10% (v/v) 2H_2O .

- $^1\text{H}\{^{15}\text{N}\}$ HSQC-TROSY and $^1\text{H}\{^{13}\text{C}\}$ constant-time HSQC spectra are collected at 35°C with 1,000 scans per transient. Five hundred and twelve complex points are collected in the ^1H , ^{13}C and ^{15}N dimensions and multiplied by a cosine-bell window function and zero-filled to 1,000 points using XWINNMR (Bruker) before Fourier transformation. The corresponding sweep widths are 12.5, 70, and 30 ppm in the ^1H , ^{13}C , and ^{15}N dimensions, respectively. The triple-labeling procedure affords a low proton density in region 4.2.
- Unlabeled -35 element promoter DNA or T4 AsiA, which binds to region 4 to inhibit gene expression, is added to the NMR samples to a final molar ratio of 1.2:1.

References

- Heath WF, Merrifield RB (1986) A synthetic approach to structure-function relationships in the murine epidermal growth factor molecule. *Proc Natl Acad Sci USA* 83:6367–6371
- Dawson PE, Muir TW, Clark-Lewis I, Kent SB (1994) Synthesis of proteins by native chemical ligation. *Science* 266:776–779
- Tam JP, Lu YA, Liu CF, Shao J (1995) Peptide synthesis using unprotected peptides through orthogonal coupling methods. *Proc Natl Acad Sci USA* 92:12485–12489
- Dawson PE, Kent SB (2000) Synthesis of native proteins by chemical ligation. *Annu Rev Biochem* 69:923–960
- Kent SB (1988) Chemical synthesis of peptides and proteins. *Annu Rev Biochem* 57:957–989
- Schnolzer M, Alewood P, Jones A, Alewood D, Kent SB (1992) In situ neutralization in Boc-chemistry solid phase peptide synthesis. Rapid, high yield assembly of difficult sequences. *Int J Pept Protein Res* 40:180–193
- Camarero JA, Hackel BJ, de Yoreo JJ, Mitchell AR (2004) Fmoc-based synthesis of peptide alpha-thioesters using an aryl hydrazine support. *J Org Chem* 69:4145–4151
- Wishart DS, Sykes BD, Richards FM (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J Mol Biol* 222:311–333
- Hackeng TM, Mounier CM, Bon C, Dawson PE, Griffin JH, Kent SB (1997) Total chemical synthesis of enzymatically active human type II secretory phospholipase A2. *Proc Natl Acad Sci USA* 94:7845–7850
- Hackeng TM, Griffin JH, Dawson PE (1999) Protein synthesis by native chemical ligation: expanded scope by using straightforward methodology. *Proc Natl Acad Sci USA* 96:10068–10073
- Severinov K, Muir TW (1998) Expressed protein ligation, a novel method for studying protein-protein interactions in transcription. *J Biol Chem* 273:16205–16209
- Romanelli A, Shekhtman A, Cowburn D, Muir TW (2004) Semisynthesis of a segmental isotopically labeled protein splicing precursor: NMR evidence for an unusual peptide bond at the N-extein-intein junction. *Proc Natl Acad Sci USA* 101:6397–6402
- Muir TW, Sondhi D, Cole PA (1998) Expressed protein ligation: a general method for protein engineering. *Proc Natl Acad Sci USA* 95:6705–6710
- Tolbert TJ, Franke D, Wong CH (2005) A new strategy for glycoprotein synthesis: ligation of synthetic glycopeptides with truncated proteins expressed in *E. coli* as TEV protease cleavable fusion protein. *Bioorg Med Chem* 13:909–915
- Chong S, Xu MQ (1997) Protein splicing of the *Saccharomyces cerevisiae* VMA intein without the endonuclease motifs. *J Biol Chem* 272:15587–15590
- Derbyshire V, Belfort M (1998) Lightning strikes twice: intron-intein coincidence. *Proc Natl Acad Sci USA* 95:1356–1357
- Shingledecker K, Jiang SQ, Paulus H (1998) Molecular dissection of the *Mycobacterium tuberculosis* RecA intein: design of a minimal intein and of a trans-splicing system involving two intein fragments. *Gene* 207:187–195
- Wu H, Xu MQ, Liu XQ (1998) Protein trans-splicing and functional mini-inteins of a cyanobacterial dnaB intein. *Biochim Biophys Acta* 1387:422–432
- Telenti A, Southworth M, Alcaide F, Daugelat S, Jacobs WR Jr, Perler FB (1997) The *Mycobacterium xenopi* GyrA protein splicing element: characterization of a minimal intein. *J Bacteriol* 179:6378–6382
- Smith DR, Doucette-Stamm LA, Deloughery C, Lee H, Dubois J, Aldredge T, Bashirzadeh R, Blakely D, Cook R, Gilbert K, Harrison D, Hoang L, Keagle P, Lumm W, Pothier B, Qiu D, Spadafora R, Vicaire R, Wang Y, Wierzbowski J, Gibson R, Jiwani N, Caruso A, Bush D, Reeve JN et al (1997) Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. *J Bacteriol* 179: 7135–7155
- Muthy S, Evans TC, Chute IC, Wu H, Chong S, Benner J, Liu XQ, Xu MQ (1999) Characterization of a self-splicing mini-intein and its conversion into autocatalytic N- and C-terminal cleavage elements: facile production of protein building blocks for protein ligation. *Gene* 231:1–13

22. Evans TC Jr, Benner J, Xu MQ (1999) The in vitro ligation of bacterially expressed proteins using an intein from *Methanobacterium thermoautotrophicum*. *J Biol Chem* 274:3923–3926
23. Hauser PS, Ryan RO (2007) Expressed protein ligation using an N-terminal cysteine containing fragment generated in vivo from a pelB fusion protein. *Protein Expr Purif* 54:227–233
24. Xu R, Ayers B, Cowburn D, Muir TW (1999) Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies. *Proc Natl Acad Sci USA* 96:388–393
25. Hawley DK, McClure WR (1983) Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acids Res* 11:2237–2255
26. Harley CB, Reynolds RP (1987) Analysis of *E. coli* promoter sequences. *Nucleic Acids Res* 15:2343–2361
27. Dombroski AJ, Walter WA, Record MT Jr, Siegele DA, Gross CA (1992) Polypeptides containing highly conserved regions of transcription initiation factor sigma 70 exhibit specificity of binding to promoter DNA. *Cell* 70:501–512
28. Dombroski AJ, Walter WA, Gross CA (1993) Amino-terminal amino acids modulate sigma-factor DNA-binding activity. *Genes Dev* 7:2446–2455
29. Patikoglou GA, Westblade LF, Campbell EA, Lamour V, Lane WJ, Darst SA (2007) Crystal structure of the *Escherichia coli* regulator of sigma70, Rsd, in complex with sigma70 domain 4. *J Mol Biol* 372:649–659
30. Gruber TM, Bryant DA (1997) Molecular systematic studies of eubacteria, using sigma70-type sigma factors of group 1 and group 2. *J Bacteriol* 179:1734–1747
31. Camarero JA, Muir TW (2001) Native chemical ligation of polypeptides. *Curr Protoc Protein Sci* Chapter 18: Unit18.14
32. Camarero JA, Shekhtman A, Campbell EA, Chlenov M, Gruber TM, Bryant DA, Darst SA, Cowburn D, Muir TW (2002) Autoregulation of a bacterial sigma factor explored by using segmental isotopic labeling and NMR. *Proc Natl Acad Sci USA* 99:8536–8541
33. Vuthoori S, Bowers CW, McCracken A, Dombroski AJ, Hinton DM (2001) Domain 1.1 of the sigma(70) subunit of *Escherichia coli* RNA polymerase modulates the formation of stable polymerase/promoter complexes. *J Mol Biol* 309:561–572
34. Sorenson MK, Darst SA (2006) Disulfide cross-linking indicates that FlgM-bound and free sigma28 adopt similar conformations. *Proc Natl Acad Sci USA* 103:16722–16727
35. Skrisovska L, Allain FH (2008) Improved segmental isotope labeling methods for the NMR study of multidomain or large proteins: application to the RRM of Npl3p and hnRNP L. *J Mol Biol* 375:151–164
36. Hui J, Bindereif A (2005) Alternative pre-mRNA splicing in the human system: unexpected role of repetitive sequences as regulatory elements. *Biol Chem* 386:1265–1271
37. Slynko V, Schubert M, Numao S, Kowarik M, Aebi M, Allain FH (2009) NMR structure determination of a segmentally labeled glycoprotein using in vitro glycosylation. *J Am Chem Soc* 131:1274–1281
38. Kowarik M, Numao S, Feldman MF, Schulz BL, Callewaert N, Kiermaier E, Catrein I, Aebi M (2006) N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science* 314:1148–1150
39. Schubert M, Oschkinat H, Schmieder P (2001) MUSIC, selective pulses, and tuned delays: amino acid type-selective (1)H-(15)N correlations, II. *J Magn Reson* 148:61–72
40. Wood MJ, Sampoli Benitez BA, Komives EA (2000) Solution structure of the smallest cofactor-active fragment of thrombomodulin. *Nat Struct Biol* 7:200–204
41. Skrisovska L, Schubert M, Allain FH (2010) Recent advances in segmental isotope labeling of proteins: NMR applications to large proteins and glycoproteins. *J Biomol NMR* 46:51–65
42. Southworth MW, Amaya K, Evans TC, Xu MQ, Perler FB (1999) Purification of proteins fused to either the amino or carboxy terminus of the *Mycobacterium xenopi* gyrase A intein. *Biotechniques* 27:110–114, 116, 118–120
43. Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* 13:289–302
44. Ritt S, Boschitz ET, Meier R, Tacic R, Wessler M, Junker K, Konter JA, Mango S, Renker D, van den Brandt B, Efimovych VV, Kovaljov A, Prokofiev A, Mach R, Chaumette P, Deregul J, Durand G, Fabre J, Thiel W (1991) Measurement of the vector analyzing power iT_{11} in $\pi^+ \rightarrow e^+ \mu^-$. *Phys Rev C Nucl Phys* 43:745–760
45. Juranic N, Moncrieffe MC, Likic VA, Prendergast FG, Macura S (2002) Structural dependencies of $h^3J_{NC'}$ scalar coupling in protein H-bond chains. *J Am Chem Soc* 124:14221–14226
46. Juranic N, Vuk-Pavlovic S, Nikolic AT, Chen TB, Macura S (1996) Nitrogen-15 NMR chemical shifts in oligopeptides coordinated to cobalt(III). *J Inorg Biochem* 62:117–126
47. Juranic N, Macura S (2001) Correlations among (1)J(NC') and (h3)J(NC') coupling constants in the hydrogen-bonding network of human ubiquitin. *J Am Chem Soc* 123:4099–4100
48. Frutos S, Goger M, Giovani B, Cowburn D, Muir TW (2010) Branched intermediate formation stimulates peptide bond cleavage in protein splicing. *Nat Chem Biol* 6:527–533