PERSPECTIVE

Recent advances in segmental isotope labeling of proteins: NMR applications to large proteins and glycoproteins

**Lenka Skrisovska · Mario Schubert ·
Frédéric H.-T. Allain**

Abstract In the last 15 years substantial advances have been made to place isotope labels in native and glycosylated proteins for NMR studies and structure determination. Key developments include segmental isotope labeling using Native Chemical Ligation, Expressed Protein Ligation and Protein Trans-Splicing. These advances are pushing the size limit of NMR spectroscopy further making larger proteins accessible for this technique. It is just emerging that segmental isotope labeling can be used to define inter-domain interactions in NMR structure determination. Labeling of post-translational modified proteins like glycoproteins remains difficult but some promising developments were recently achieved. Key achievements are segmental and site-specific labeling schemes that improve resonance assignment and structure determination of the glycan moiety. We adjusted the focus of this perspective article to concentrate on the NMR applications based on recent developments rather than on labeling methods themselves to illustrate the considerable potential for biomolecular NMR.

Keywords Segmental isotope labeling · Glycoproteins · Glycosylation · Expressed Protein Ligation · Protein Trans-Splicing

Lenka Skrisovska and Mario Schubert contributed equally to this work.

L. Skrisovska · M. Schubert · F. H.-T. Allain (✉)
Institute for Molecular Biology and Biophysics, ETH Zürich, 8093 Zurich, Switzerland
e-mail: allain@mol.biol.ethz.ch

Abbreviations

NCL	Native Chemical Ligation
EPL	Expressed Protein Ligation
PTS	Protein Trans-Splicing
aa	Amino acid
SPPS	Solid Phase Peptide Synthesis
IPL	Intein-mediated Protein Ligation

Introduction

NMR spectroscopy is after X-ray crystallography one of the two widely used methods to determine 3D structures of biomolecules. In addition it can complement 3D structure information by dynamic studies and binding interactions. However, the technique has limitations concerning the molecular size and the study of posttranslationally modified proteins.

The classical protein NMR techniques are limited to a molecular size typically in the range of 15–20 kDa. The size limit is caused by the slower tumbling time of larger molecules that give rise to increased transverse relaxation rates and thus severe line broadening. Both the resolution and sensitivity in ^{1}H, ^{13}C, ^{15}N dimensions are then decreased in spectra of larger molecules. In addition the number of atoms and thus NMR resonances increase with the molecular size. However, many biological problems involve larger molecules or molecular assemblies and since NMR spectroscopy can provide information complementary to crystallography it is highly desirable to extend the size limit.

Two major breakthroughs pushed the size limit ahead by downscaling the line broadening due to relaxation: incorporation of high levels of deuterium into the molecule eliminates proton-related relaxation pathways (Gardner and

Kay [1998]) and transverse relaxation-optimized spectroscopy (TROSY) uses a cancellation of relaxation contributions from dipole–dipole interaction and chemical shift anisotropy (CSA) (Pervushin et al. [1997]). The original TROSY reduces relaxation effects and thus line broadening of only ^{15}N–^{1}H correlations. Subsequently a TROSY for ^{13}C–^{1}H correlations of aromatic side chains (Pervushin et al. [1998]) was introduced and recently a methyl-TROSY (Tugarinov et al. [2003]). Combining deuteration with TROSY triple resonance experiments made the backbone assignment of proteins of 20–80 kDa possible. However, the chance that a vast amount of resonances in larger proteins overlap is high. Segmental isotope labeling is one approach to reduce the spectral complexity to a manageable level. Part I of this review is dedicated to the different approaches to obtain segmental isotope labeling of proteins and their NMR application to large or multidomain proteins. Part II of this review gives an overview of the different labeling schemes available for glycoproteins and in particular how segmental labeling of the proteins and the sugar moiety can considerably improve the structural definition of the attached sugar.

Part I: segmental isotope labeling of proteins

The development of approaches for the uniform incorporation of ^{15}N, ^{13}C and ^{2}H isotopes into protein sequences allowed NMR spectroscopy to achieve the resonance assignments and to determine structures of proteins of molecular sizes up to 20 kDa. However, above 20 kDa, in addition to line broadening, the increased spectral complexity makes the structure determination of such proteins very difficult. The ability to isotopically label only defined segments within intact proteins can certainly facilitate the NMR study of large proteins in the future. Although the segmental labeling has not yet been widely applied in NMR, we believe that the increased need to study larger proteins will render segmental labeling methods increasingly routine among NMR spectroscopists. We are reviewing here the current methods available for segmental labeling of proteins with a focus on the ones resulting in NMR application. For their detailed technical descriptions as well as applications outside of NMR we refer the reader to numerous excellent publications (Ayers et al. [1999]; Cotton and Muir [2000]; David et al. [2004]; Muralidharan and Muir [2006]). Moreover, we are focusing on how segmental labeling has been used in protein NMR: to investigate interdomain interactions within multidomain proteins, to study conformational changes and ligand binding, to help resonance assignment of large proteins and last but not least to facilitate protein structure determination.
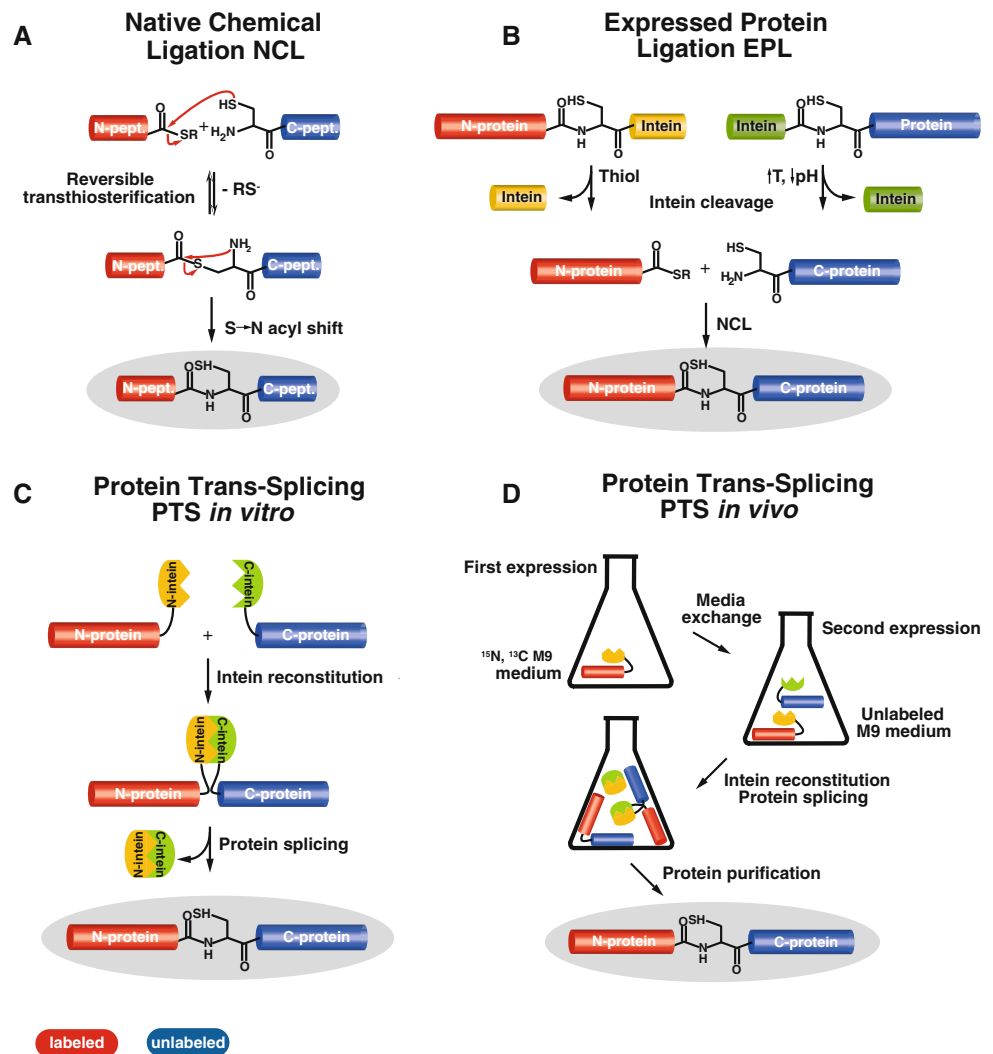
The different approaches

Native Chemical Ligation and Expressed Protein Ligation: the classical in vitro approaches

Several techniques have been successfully developed up to now to obtain segmentally isotopically labeled samples for NMR investigations (Fig. 1). The introduction of Native Chemical Ligation (NCL) has had an important role in the development of these techniques (Dawson and Kent [2000]; Dawson et al. [1994]; Hofmann and Muir [2002]). NCL is based on a reaction of two unprotected synthetic peptides, one containing a C-terminal thioester (α-thioester) and the other containing an N-terminal cysteine residue (α-cysteine), which results in formation of a native peptide bond under aqueous conditions (Fig. 1a). NCL can be used for segmental isotope labeling of proteins by ligating together isotopically labeled and unlabeled synthetic peptides (Balambika et al. [2007]; Kochendoerfer et al. [2004]; Rajagopal and Kent [2007]). Both peptides containing the appropriate reactive termini can be produced by solid-phase peptide synthesis (SPPS) (Dawson and Kent [2000]; Hofmann and Muir [2002]). However, SPPS is limited by the maximal possible length of synthesized peptides of accurate amino acid sequence, which is approximately 50 amino acids. Since most proteins of interest are composed of more than 100 amino acids, NCL of such proteins will require ligation of more than two synthetic peptides (Hackeng et al. [1999]). Additionally, the synthesis of isotopically labeled peptides is expensive, making this approach fairly unfavorable for most NMR applications. Nevertheless, total chemical protein synthesis gives not only the possibility of segmental isotope labeling of proteins but enables the incorporation of any type of modification (phosphorylation, methylation, glycosylation etc.) or site-specific label into the protein sequence.

The second important step in the development of techniques for segmental isotope labeling of proteins has been the exploitation of a naturally occurring process called protein splicing (Evans and Xu [2002]; Paulus [2000]). Protein splicing is a posttranslational process in which internal segments (inteins) catalyze their own excision from the precursor proteins with consequent formation of a native peptide bond between two flanking external regions (exteins). Up to now more than three hundred inteins have been identified (see www.neb.com/neb/inteins.html) and many of them were extensively characterized (Derbyshire et al. [1997]; Mathys et al. [1999]; Mills et al. [1998]; Telenti et al. [1997]; Wu et al. [1998]). Their self-splicing properties were used to develop very convenient tools for protein engineering. There are two methods based on intein properties that have been used for segmental isotope

Fig. 1 Overview of the methods for segmental isotope labeling of proteins. **a** Mechanism of the Native Chemical Ligation. **b** Principle of Expressed Protein Ligation. **c** Principle of Protein Trans-Splicing by split inteins **d** Schematic representation of the protocol for in vivo Protein Trans-Splicing

labeling of proteins: Expressed Protein Ligation (EPL) and Protein Trans-Splicing (PTS).

Expressed Protein Ligation is based on Native Chemical Ligation (Fig. 1a) except that both or at least one of the protein fragments for protein synthesis is produced by bacterial expression (David et al. 2004; Muir 2003; Severinov and Muir 1998). This approach has also been called intein-mediated protein ligation (IPL) (Evans et al. 1998). Since the reaction involves protein fragments containing an α-thioester and an α-cysteine, a cysteine is required at the ligation site. This criterion is not always easily fulfilled within natural protein sequences, therefore often a non native cysteine residue needs to be introduced. Such a mutation should be as conservative as possible, in most studies serine or alanine were chosen. For the recombinant production of protein fragments with reactive termini EPL uses engineered inteins designed to cleave only on one of their termini (Fig. 1b). As an example, New

England Biolabs developed the IMPACT[TM] system, a very convenient set of bacterial vectors that allow easy recombinant production of protein fragments with α-thioesters and α-cysteines (Xu and Evans 2003; 2001). While the production of recombinant protein fragments with an α-thioesther can only be achieved through intein cleavage, other methods have been used for production of recombinant protein fragments with an α-cysteine. Using proteolytic leader sequences for specific proteases like factor Xa protease (Camarero et al. 2002; Xu et al. 1999; Zhang et al. 2007), TEV (Tobacco Etch Virus) protease (Tolbert and Wong 2002), thrombin (Busche et al. 2009) or enterokinase one can as well generate protein fragments with N-terminal cysteine residues. Furthermore, endogenous methionyl aminopeptidase treatment of bacterially expressed constructs starting with Met-Cys can also generate protein fragments with N-terminal cysteine residues (Camarero et al. 2001; Iwai and Pluckthun 1999). In order to reach

good yields, EPL requires high concentrations of the intein-protein precursors. The efficiency of the ligation step for EPL (as well as for NCL) strongly depends on the concentration of the ligating fragments, thus high concentrations in the mM range are often needed. We have recently reported the significant improvement of the protocol used for EPL, by increasing the concentration of the ligating fragments as well as by introducing a refolding step before intein cleavage in case of insoluble intein-protein precursors (Skrisovska and Allain 2008).

EPL is frequently used for segmental isotope labeling of proteins (Camarero et al. 2002; Skrisovska and Allain 2008; Vitali et al. 2006; Xu et al. 1999; Zhang et al. 2007). In most studies, only two protein fragments where ligated. However, a very convenient way to study large proteins would potentially be to ligate three or even more protein fragments in order to isotopically label internal segments of a protein. Such a three-piece protein ligation was first realized by Cotton et al. (1999) where a synthetic peptide containing both a α-cysteine and α-thioester was inserted between two recombinant protein fragments containing the appropriate reactive termini. Subsequently, a very elegant sequential ligation strategy of a multidomain protein was reported by Blaschke et al. (2000). In this work the three recombinant protein fragments containing proper reactive termini (α-thioester and α-cysteine termini generated through intein and Xa cleavage, respectively) were ligated together by a two-step ligation.

In addition to segmental isotope labeling application, EPL of a recombinant protein fragment with synthetic peptide gives the possibility to incorporate amino acid modifications and labels specifically into native protein sequences (Ayers et al. 1999; Cotton and Muir 2000; Muir et al. 1998). This allows studying proteins with posttranslational modifications relevant to their biological activity and structure. Furthermore, EPL can be used as well as a tool for production of cyclic proteins (Camarero et al. 2001; Iwai and Pluckthun 1999) and toxic proteins (Evans et al. 1998).

Protein Trans-Splicing: a convenient in vitro and in vivo expression approach

Inteins can be fragmented into two parts which do not have activity on their own. After their association they reconstitute into an active intein which performs a splicing reaction resulting in ligation of their fusion protein fragments (Fig. 1c). Such a process is known as Protein Trans-Splicing (PTS) (Muralidharan and Muir 2006; Xu and Evans 2005). Fragmented inteins are called split inteins and they occur naturally (Wu et al. 1998) or they can be designed artificially as demonstrated in the first reports of

PTS (Shingledecker et al. 1998; Southworth et al. 1998; Yamazaki et al. 1998). Both precursors containing the split inteins fused with protein fragments can be produced separately by bacterial expression. If one fragment is isotopically labeled while the second one is unlabeled, after their purification and reconstitution, the splicing reaction will result in a segmentally labeled protein. PTS has a few sequence requirements in order to obtain efficient splicing activity. Since the first transesterification step of PTS requires a thiol or hydroxyl group, the N-terminal residue of the C-extein must be a cysteine, serine or threonine residue. Additionally, many split inteins require several natural extein amino acid residues at the intein–extein junction, which after the splicing reaction will be included in the ligated protein.

Artificially split inteins require denaturation and renaturation in order to restore their splicing activity, which may not always be achievable. An elegant solution to this problem is the usage of naturally occurring split inteins that have the ability to spontaneously reconstitute into a functional intein. This also offers the possibility of performing PTS in vivo, which significantly simplifies the procedure (Zuger and Iwai 2005) (Fig. 1d). Both precursor proteins are expressed in the same culture using two plasmids containing different inducible promoters. Segmental isotope labeling is achieved by expressing the first precursor in labeled medium which is followed by transfer of the cell culture to a non-labeled medium and induction of expression of the second precursor. The protein splicing reaction occurs directly after both precursors are present in the cell, thus there is no need for any intermediate isolation and purification steps. The ligated product can be directly purified from the cell culture. Another benefit of in vivo PTS compared to other methods such as NCL and EPL is that it does not require high concentrations of the precursor proteins and efficient protein ligation can be achieved at micromolar concentrations. Although this protocol seems to be very simple and straightforward, there are some disadvantages. This method depends on the solubility and stability of at least one of the precursor proteins in *E. coli*, which may limit the application of the method. In addition, non-native amino acid residues are introduced into the sequence of the ligated product depending on the split intein that is used. Due to the change of media, isotope scrambling might take place during the expression of precursor proteins. However, in recent work from Muona and coworkers, the authors managed to suppress the effect of isotope scrambling by optimizing the expression protocol (Muona et al. 2008). Similarly to EPL, PTS can be used for isotope labeling of the central segment of a large or a multidomain protein (Otomo et al. 1999). This can be achieved either by combining two split inteins which are divided at different positions in order to prevent their

misassociation (Busche et al. 2009; Otomo et al. 1999) or by using a combination of a naturally occurring split intein and of a designed inducible split intein (Shi and Muir 2005).

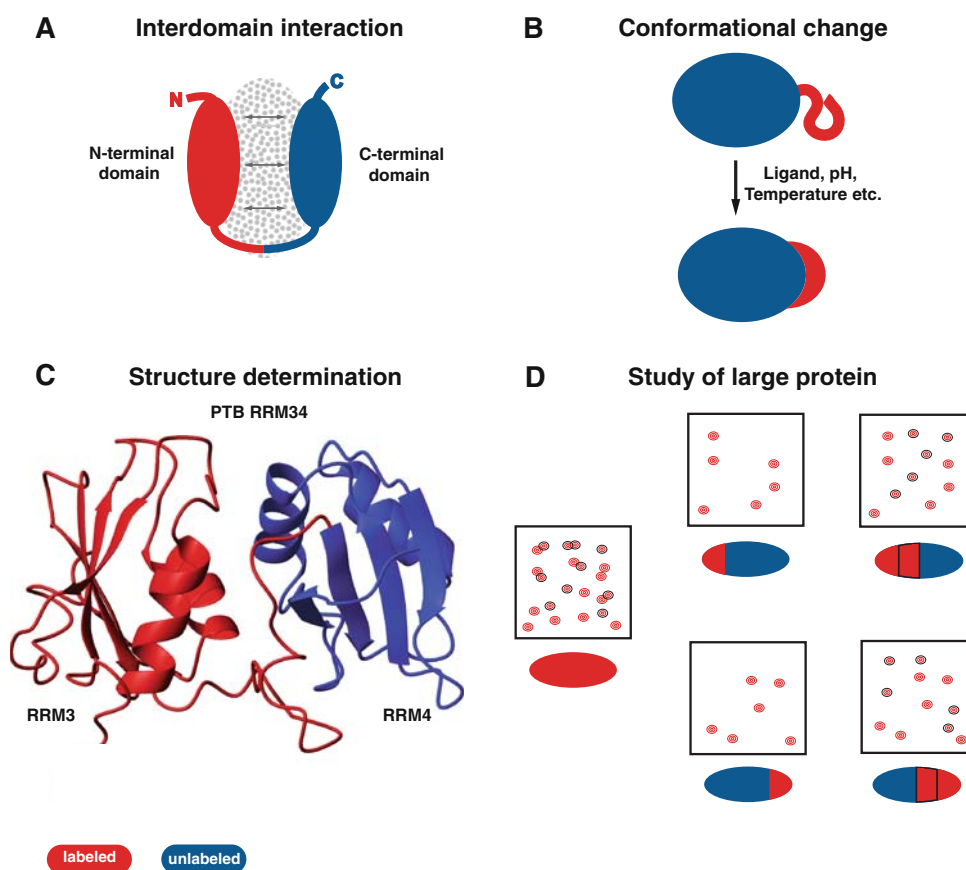### Enzymatic protein ligation: an unexplored approach

An alternative method for segmental isotope labeling of proteins is based on enzyme mediated protein ligation. Enzymes successfully used for protein ligation are for example subtiligase and sortase A. Subtiligase is an engineered variant of the serine protease subtilisin BPN′, which catalyzes peptide bond formation in aqueous solutions between the esterified carboxyl group of one peptide and the amino group of the other peptide. The use of subtiligase for protein synthesis was demonstrated by total synthesis of Ribonuclease A from six synthetic peptides (Jackson et al. 1994). Despite the promising use of subtiligase for protein synthesis, this method has not yet been used for segmental isotope labeling of proteins. Sortase A is an enzyme which in gram-positive bacteria is attaching surface proteins to the peptidoglycan cross bridge of the cell wall. The enzyme

catalyzes the peptide bond cleavage between threonine and glycine in the recognition sequence LPETG of the surface protein and subsequently ligates the carboxyl group of threonine with an amino group of glycine from the peptidoglycan. Sortase A requires the recognition motif LPXTG, which is included in the amino acid sequence of the ligated protein. Recently, this approach has successfully been used to introduce an unlabeled solubility enhancement tag to an isotopically labeled protein for the NMR study (Kobashigawa et al. 2009). We believe that this approach opens the possibility for more applications of segmental labeling in the future.

The applications for NMR studies: how can this help?

Segmental isotope labeling has a large potential for application in NMR spectroscopy (Fig. 2). Different isotope labeling schemes within one protein can be used to resolve the spectral complexity of large proteins thus simplifying the assignment and making their structure determination achievable. Besides providing a solution to the size problem, segmental isotope labeling has been also

Fig. 2 Schematic illustration of the NMR applications of segmental isotope labeling for proteins **a** Interdomain interaction studies **b** Studies of induced conformational changes of protein segments **c** Structure of the PTB RRM34 determined using data from a segmentally labeled protein **d** Example of the modular application of segmental isotope labeling for the study of large proteins, showing schematically how segmental labeling of different protein fragments can help their resonance assignments

used to study interdomain interactions, the relative orientation of domains or conformational changes that occur upon ligand binding. In the next sections we review the different types of NMR applications of segmental isotopic labeling in proteins.

Introducing an invisible solubility enhancement tag

In certain cases expression yield and solubility of a protein can be significantly increased by adding a solubility tag, like for example the small tag GB1 (B1 domain of Streptococcal protein G) (Zhou et al. 2001). However, if the presence of this tag is crucial to maintain the solubility of the protein construct, additional NMR signals from the solubility domain might complicate NMR assignment and automated structure calculations. To circumvent this problem Kobashigawa and coworkers recently expressed an uniformly labeled Vav C-terminal SH3 domain (VcSH3) in fusion with an N-terminal GB1 tag. Subsequently, a non-labeled C-terminal GB1 tag was added to the fusion protein using sortase A mediated protein ligation and the isotopically labeled N-terminal GB1 tag was removed by proteolytical cleavage. The result of this strategy was a VcSH3-GB1 fusion protein with only the VcSH3 domain isotopically labeled (Kobashigawa et al. 2009). Alternatively the in vivo PTS method can be used to incorporate a solubility enhancing tag to a protein, as it was demonstrated for the prion-inducing domain of yeast Sup35 by Zuger and Iwai (2005).

Segmental isotope labeling to study interdomain interactions

Clearly the most frequent application of segmental isotope labeling is the study of interdomain interactions within multidomain proteins (Fig. 2a). A convenient way of determining whether two or more domains interact or not is to isotopically label one of the domains while leaving the other domain(s) unlabeled. There are several approaches which can reveal whether labeled and unlabeled protein segments are in contact or not. Comparison of 2D ^{15}N–^{1}H HSQC spectra can already provide significant structural information about protein segments present in different environments. Camarero et al. (2002) used EPL for segmental isotope labeling of the σ^A factor from *Thermotoga maritima* to investigate the proposed direct interaction between the N- and C-terminal regions of the protein and their effect on DNA binding. Two segmentally labeled samples were prepared with only the C-terminal region isotopically labeled (aa 348–399), one containing the full length sequence (399 aa) and one with a shorter N-terminal

region (aa 137–399). The superimposed ^{15}N–^{1}H HSQC-TROSY and ^{13}C–^{1}H HSQC spectra of both samples showed high similarities, indicating a similar fold of the C-terminal region in both constructs, thus disproving the expected strong interaction between the N-terminal and C-terminal regions. The addition of DNA caused significant chemical shift changes in the spectra of the construct with the smaller N-terminal region but not in those of the full length construct. Based on these data, the authors were able to draw the conclusion that the presence of the N-terminal region (called 1.1) does inhibit the binding of DNA of the C-terminal region (called 4.2), but not via a direct interaction between the two regions.

Similarly, EPL has been applied for segmental isotopic labeling of the monomeric apolipoprotein E3 (apoE3) in order to study the interaction between the N- and C-terminal domains (Zhang et al. 2007). The C-terminal domain within the full length apoE3 appeared to be more structured than the isolated construct, suggesting a weak interdomain interaction. Comparison of the ^{15}N–^{1}H HSQC spectra of the full length construct containing a ^{2}H-labeled N-terminal domain (aa 1–214) and a ^{13}C, ^{15}N-labeled C-terminal domain (aa 215–299) with the ^{13}C, ^{15}N labeled isolated C-terminal domain (aa 215–299) showed several spectral differences. From a careful analysis, the authors concluded that there are weak interactions between N- and C-terminal domains in apoE3 that stabilize the C-terminal domain fold without affecting the fold of the N-terminal domain.

Walters et al. (2003) used EPL for segmental isotopic labeling of the 40 kDa hHR23a protein with the purpose of studying the interaction between the N-terminal UBL domain and the C-terminal UBA and XPC domains. NMR relaxation experiments (^{15}N longitudinal and transverse relaxation measurements as well as heteronuclear dipole–dipole cross relaxation measurements) supported the presence of interdomain interactions. In order to detect NOEs between the UBL domain (aa 1–117) and UBA and XPC domains (aa 118–363), a 3D ^{13}C filtered edited NOESY (Zwahlen et al. 1997) was recorded on a construct where only the C-terminal domains were ^{13}C labeled. Additionally a ^{15}N-edited NOESY experiment was recorded on a construct with ^{2}H, ^{15}N-labeled C-terminal domains (Walters et al. 1997). However, no interdomain NOEs could be detected, in support of the interdomain interactions. To investigate this further, a chemical shift perturbation analysis of individually expressed domains was performed and confirmed the interaction of the UBL domain with the UBA domain. Based on these data the authors concluded that the UBL domain does indeed interact with the UBA domain within the hHR23a protein and suggested that its structure is not rigidly locked into one conformation, which could explain the lack of interdomain NOEs.

Segmental isotope labeling to study conformational changes

Segmental isotope labeling of proteins is a very practical tool to study conformational changes of protein segments induced for example by ligand binding (Fig. 2b). As demonstrated in the following study, EPL of a recombinant protein fragment (aa 1–345) with a synthetic nine amino acids peptide containing isotopically labeled amino acids at specific sites was used to study the conformational change that occur at the carboxy terminus of the G protein α subunit upon G protein activation (Anderson et al. 2005). ^{1}H–^{13}C HSQC spectra of the labeled amino acids of the semi-synthetic segmentally labeled Gα in the free protein and in the GDP-bound state were very similar and showed little dispersion indicating that the C-terminus is highly mobile. In order to study the conformational changes of the carboxyl terminus between the GDP-bound and the activated states, AlF$_4^-$, an analog of GTP, was titrated into Gα to mimic the active state of the G protein. Each addition of AlF$_4^-$ resulted in loss of intensities in all of the resonances observed in the ^{1}H–^{13}C HSQC spectrum suggesting that the carboxyl terminus adopts an ordered conformation upon addition of AlF$_4^-$.

Segmental isotope labeling for determining structures of multidomain proteins

The full potential of segmental isotope labeling for structure determination was demonstrated in a study carried out in our laboratory. Using the pTWIN^{TM} system, we performed EPL in order to segmentally isotopically label the last two C-terminal RNA Recognition Motifs (RRM) (RRM3 aa 324–442 and RRM4 aa 443–531) of polypyrimidine tract binding protein (PTB) (Vitali et al. 2006). In a previous study, the structure of the same protein construct was determined by NMR using uniform labeling and it was concluded that no interdomain interactions could be detected between the RRMs (Conte et al. 2000). However, when we studied these same RRMs in the complex with RNA, we could observe a large interdomain interface (Oberstrass et al. 2005). In order to determine if this interdomain interface was induced by RNA binding, we decided to re-evaluate the structure of the free RRM34 using EPL. Two segmentally labeled samples were generated containing either RRM3 or RRM4 ^{15}N, ^{13}C-labeled. The overlay of ^{15}N–^{1}H HSQC spectra of the uniformly and of the two segmentally labeled RRM34 samples showed no differences, indicating that the Ser to Cys mutation at the ligation site did not affect the protein fold. 2D ^{13}C filtered edited NOESY (Peterson et al. 2004) and 3D ^{13}C edited filtered NOESY (Lee et al. 1994) experiments revealed a high number (130) of interdomain NOEs which could be

used as long-range interproton distance constraints in the structure determination of PTB1 RRM34 (Fig. 2c). Evidently, without segmental labeling only a small fraction of these interdomain NOEs could have been observed in the uniformly labeled sample due to spectral overlap. The PTB RRM34 structure revealed a large interdomain interface with a fixed relative orientation of both RRMs which is very similar to the interaction found in the structure of the complex with RNA. This study still represents to our knowledge the first and only application of segmental isotope labeling for structure determination of a protein by NMR.

Segmental isotope labeling to study large proteins

As outlined in the introduction, the use of NMR spectroscopy to study proteins of high molecular weights is one of the major potential benefits of segmental isotope labeling. One way of doing this is to simplify the complexity of spectra and their resonance assignments by recording NMR experiments for proteins in which only subfragments are isotopically labeled. By labeling each subfragment at a time, the resonance assignment and the structure calculation of the whole protein is facilitated (Fig. 2d). Such strategy was used by Yagi et al. (2004) to obtain the backbone assignments of the β subunit monomer of F0F1-ATP synthase (52 kDa). Protein Trans Splicing was used in this case in order to obtain the backbone resonance assignments and to investigate the conformational change of the protein upon nucleotide binding. Four constructs, each with a different isotopically labeled subfragment of the protein were produced (aa 1–271, 1–124, 272–473, and 391–473). Despite inserting five additional residues (GGGTG) required for the splicing, a comparison of the ^{15}N–^{1}H TROSY-HSQC spectra with those of the uniformly labeled protein confirmed the intact structure of all four ligated proteins and showed significantly improved signal resolution due to the reduced resonance overlap. To obtain the sequential backbone assignment of such a large protein, 3D triple resonance experiments and a 3D ^{15}N edited NOESY were recorded for all four segmentally labeled proteins. Furthermore, to investigate the relative orientation of the N-terminal and C-terminal domains residual dipolar couplings (RDC) were measured for the constructs with appropriate ^{15}N, ^{13}C and ^{2}H labeled segments (aa 1–124 and 391–473). Based on the collected data, the secondary structure of the β subunit could be predicted showing that the structure adopts an open form in the absence of bound nucleotide. Subsequently, the authors showed that the β subunit undergoes a conformational change from an open to a closed form upon nucleotide binding based on observed chemical shift perturbation and changes of the RDCs. This elegant example clearly

demonstrates the strength of segmental isotope labeling for NMR studies of large proteins.

Part II: segmental isotope labeling of glycoproteins

Glycoproteins play crucial roles in a variety of biological processes such as cell growth and differentiation, development, cell–cell interactions, modulation of the immune system, inflammation and cancer, protein folding, quality control and turnover or pathogenicity and host invasion of bacteria (Varki 1993). Despite the fact that more than 50% of all proteins are predicted to be glycosylated (Apweiler et al. 1999) only a small fraction of the known protein structures are glycoproteins with an intact glycan. Out of those only few 3D structures were solved by NMR spectroscopy (Erbel et al. 2000; Fletcher et al. 1994; Hashimoto et al. 1999; Metzler et al. 1997; Slynko et al. 2009; Wyss et al. 1995). The study of glycoproteins by NMR spectroscopy face a variety of problems: in vivo expression or extractions from natural sources give usually poor yields and glycosylation is often inhomogeneous and incomplete. Even if a homogeneous sample is obtained the ^{1}H chemical shift overlap between carbohydrate and protein signals is usually severe making it difficult to be analyzed by NMR. Often the signals of the free ends of the glycan cannot be unambiguously assigned due to overlap and multiple signal sets resulting from inhomogeneities (De Beer et al. 1996; Wyss et al. 1995). Trimming the inhomogeneous ends enzymatically can lead to more homogeneous samples (Metzler et al. 1997) but may alter the glycan conformation. To overcome ^{1}H frequency degeneracies labeling strategies with ^{13}C and ^{15}N are crucial to obtain a sufficient

amount of unambiguously assigned chemical shifts and NOEs for structure determination. The different methods and labeling schemes that have been or could be applied are illustrated in Fig. 3 and discussed in the following sections in detail.

The different isotopic labeling approaches in glycoproteins

Uniform/metabolic labeling using in vivo eukaryotic expression systems

In order to do uniform labeling of glycoproteins (Fig. 3a) only eukaryotic expression systems have been used, e.g. CHO (Chinese hamster ovary) cells (Lustbader et al. 1996; Metzler et al. 1997; Wyss et al. 1993), yeast (Blanchard et al. 2008; Pickford and O'Leary 2004; Wood and Komives 1999; Wood et al. 2000), plants (Ippel et al. 2004), insect cells (Walton et al. 2006), slime mold (Cubeddu et al. 2000) and hybridoma cells designed for the production of antibodies (Yamaguchi et al. 2006).

CHO cells contain a mammalian N-glycosylation system and are therefore able to produce "authentic" mammalian glycoproteins. However, the cells grow very slowly, the protein yields are low making ^{13}C/^{15}N labeling very expensive considering the complexity of media used. However, after time consuming optimizations of the expression parameters good yields can be reached (>10 mg/L) (Wyss et al. 1993). If only ^{15}N labeling is desired an alternative could be the addition of ^{15}NH$_4$Cl to the unlabeled medium for CHO cells resulting in 50–75% incorporation into the N-linked oligosaccharides (Gawlitzek et al. 1999). The yeast *Pichia pastoris* is currently the
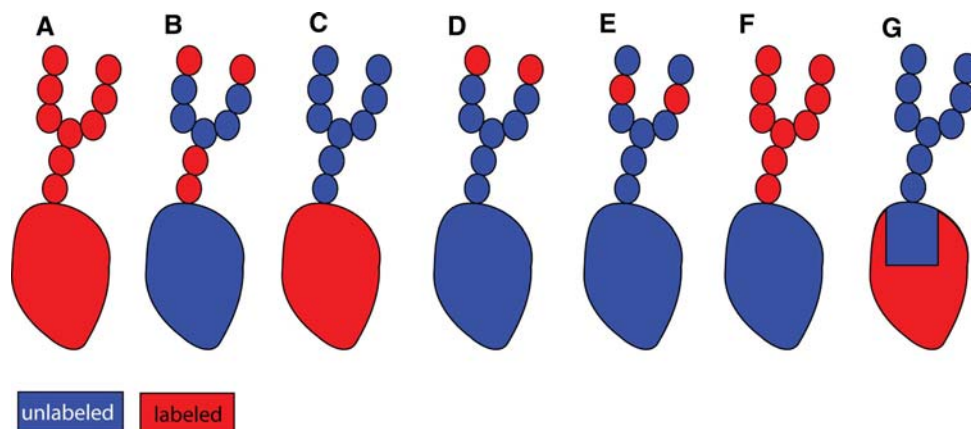

Fig. 3 Classification of the different possible labeling schemes of glycoproteins: **a** uniform labeling, **b** residue specific labeling, **c** segmental labeling in which the glycan is unlabeled and the protein labeled, **d** residue specific labeling of the non-reducing terminal residues of the glycan, **e** residue specific labeling of non-terminal residues of the glycan, **f** segmental labeling in which the glycan is labeled and the protein unlabeled and **g** ligated protein consisting of a synthesized unlabeled glycopeptide and one or two labeled peptide chains

second mostly used expression system for NMR studies behind *E. coli* (Pickford and O'Leary 2004). *P. pastoris* has the advantage that a sole carbon source can be used: ^{13}C glycerol or ^{13}C glucose (using special strains). In the case of glycoprotein expression, moderate to good protein yields could be reached ($\sim$5 mg from 10 g ^{13}C glucose and only slightly higher with the more expensive ^{13}C glycerol) (Wood and Komives 1999). Since the cells grow slowly, proteolytic degradation during the long expression times is a problem although certain mutant strains with reduced proteolytic activity could be envisaged. One clear limitation of this expression system for general use is that the N- and O-linked glycosylation pattern in *Pichia pastoris* are different from higher eukaryotes. Labeling glycoproteins in plants and insect cells presents severe limitations. The former is limited to naturally occurring plant proteins (Ippel et al. 2004) and low yields, slow growth and expensive media make a wide use of expression in Sf9 insect cells at the moment economically unfeasible (Walton et al. 2006).

The slime mold *Dictyostelium discoideum* has emerged as promising eukaryotic expression system (Arya et al. 2008) but has not been widely used. The advantages are rapid cell growth, simple media and good yields, e.g. $\sim$9 mg protein from 10 g ^{13}C glycerol (Cubeddu et al. 2000). The second promising expression system is hybridoma cells which is limited to the production of uniformly ^{15}N/^{13}C labeled antibodies using serum free media (Yabe et al. 1986) containing labeled glucose, sodium pyruvate, succinic acid and a mixture of amino acids (Yamaguchi et al. 2006). Yields in the range of 20–40 mg/L cell culture were reached (Kato et al.). Another advantage of this system is that one can obtain a residue-type selective labeling within the glycan (e.g. GlcNAc; schematically depicted in Fig. 3b) (Yamaguchi et al. 1998) or an amino-acid type selective labeling (Kato et al. 1993; 1991a; 1991b; Kim et al. 1994) by using labeling of certain components (e.g. ^{13}C-GlcN or certain amino acids). Labeling of the entire glycan (Fig. 3f) with >95% incorporation can be achieved by growth on ^{13}C glucose (Kato and Yamaguchi 2008; Yamaguchi et al. 2000). However, a small fraction of isotope scrambling can occur.

Segmental labeling by in vitro methods

An alternative to in vivo uniform labeling are in vitro approaches that allow a segmental labeling of the glycoprotein. Very recently, taking advantage of a new in vitro strategy to produce bacterial glycoproteins (Kowarik et al. 2006a), developed in the group of Markus Aebi (ETH Zurich), we could transfer a complete glycan unit to a ^{13}C/^{15}N labeled recombinantly expressed protein using an oligosaccharyltransferase in vitro (Fig. 3c). The separate expression of protein and the carbohydrate precursor

enables differential labeling of the two components resulting in a segmentally labeled N-linked glycoprotein. In this first application the N-glycan of *Campylobacter jejuni* was transferred to a small model protein using the oligo-saccharyltransferase pglB from *C. jejuni* (Slynko et al. 2009). The N-glycan was synthesized by an engineered *E. coli* strain containing the N-glycosylation locus of *C. jejuni* (Wacker et al. 2002) except the oligosaccharyl-transferase. A clear advantage of this method is to produce a glycoprotein with a homogeneous glycan. The yield of the glycosylated protein AcrA was $\sim$90% after 15 h of incubation. An additional advantage is that isotope scrambling cannot occur resulting in >99% ^{13}C labeling of the protein segment and 98.9% ^{12}C (natural abundance ^{13}C) occurrence on the glycan. However, the method is restricted so far to proteins with an accessible bacterial N-glycosylation sequence D/E-X-N-Z-S/T (X, Z: any amino acid except Pro) (Chen et al. 2007; Kowarik et al. 2006b) and a limited variation in the accepted glycan structure.

Instead of uniformly labeling the complete glycan particular sugar units can be labeled using specific glycosyl-transferases (Fig. 3d) (Gilhespy-Muskett et al. 1994; Goux et al. 1982; Macnaughtan et al. 2008; Miyazaki et al. 2000; Yamaguchi et al. 1998). The method is usually applied following a selective trimming of inhomogeneous non-reducing terminal residues of the glycoproteins. If the non-reducing terminal residues of the glycan are labeled, unlabeled carbohydrate units can be subsequently added enzymatically (Fig. 3e) (Gilhespy-Muskett et al. 1994). So far ^{13}C galactose (Gilhespy-Muskett et al. 1994) and ^{13}C sialic acid (Macnaughtan et al. 2008) have been attached in vitro by a galactosyltransferase and α-2,6 sialyltransferase, respectively. The main disadvantages of those methods are the long incubation times in the range of days increasing the risk of proteolytic cleavage. Typical yields are $\sim$90% with a ^{13}C incorporation of $\sim$90% (Gilhespy-Muskett et al. 1994).

Perspective on future labeling schemes

Although labeling of the glycan moiety in an otherwise unlabeled glycoprotein (Fig. 3f) can be achieved using metabolic labeling (Yamaguchi et al. 2000), it would be appealing to attach a labeled glycan to an unlabeled protein using an oligosaccharyltransferase (Fig. 3f) but this remains to be done. For such an approach isotopically labeled carbohydrates would be necessary. Isotope labeling of carbohydrates can be achieved by in vivo expression methods, or by chemical or chemoenzymatic synthesis (Kato et al. 2008; Live et al. 2001). Recent advancements in genetic engineering of *E. coli* for the production of oligosaccharides (Dumon et al. 2006; Fierfort and Samain 2008; Hancock et al. 2006) offer a very promising avenue

for producing isotopically labeled carbohydrates. So far this approach has been exclusively directed toward expression of $^{15}N/^{13}C$ labeled polysaccharides and their cleavage products (Azurmendi et al. 2007; Blundell et al. 2004; Gitti et al. 1994; Kern et al. 2008; Yu et al. 1993). Lewis[x] and sialyl Lewis[x] have been successfully chemo-enzymatically synthesized with ^{13}C labeling on a milligram scale (Ichikawa et al. 1992; Probert et al. 1997). Strategies to ligate a protein with a labeled glycan efficiently still remain to be developed.

Ligation between a chemically or chemoenzymatically synthesized glycopeptide and another peptide chain has been used in order to obtain a homogeneous glycoprotein (Brik et al. 2006; Piontek et al. 2009; Tolbert et al. 2005; Yamamoto et al. 2008). Two ligations on either side of the glycopeptide have also been used (Piontek et al. 2009). However, isotope labeling of the non-modified peptide chains would be possible (Fig. 3g) but has not yet been realized.

Alternative methods to synthesize glycoproteins (Brik et al. 2006; Gamblin et al. 2009) like glycoprotein remodeling, in vivo suppressor tRNA technology or chemical methods could also open possible avenues to introduce isotopes into glycoproteins.

Applications of uniform and segmental isotope labeling of glycoproteins
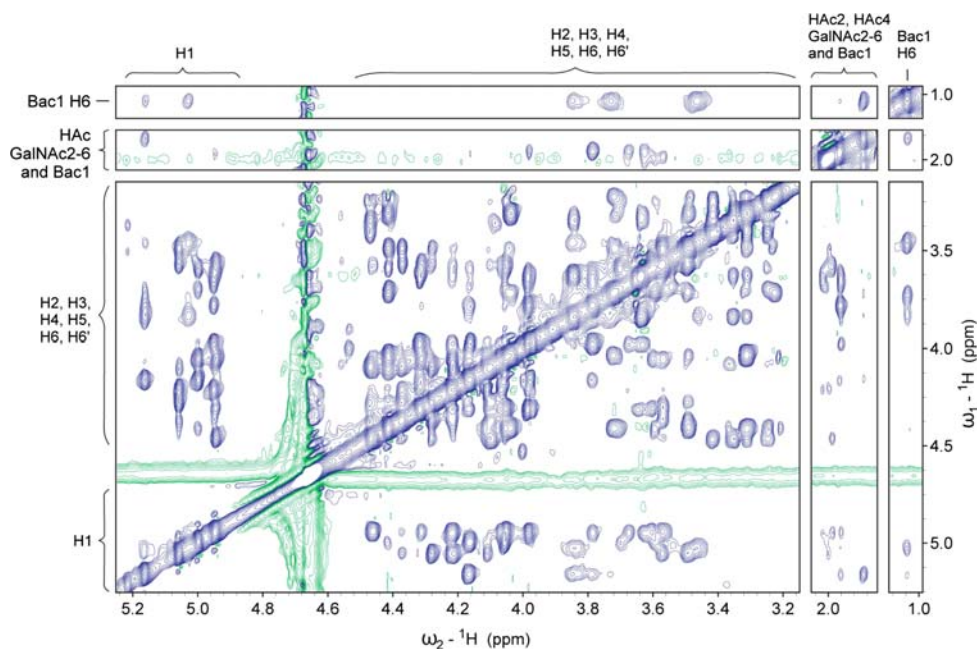
Using uniform labeling to study glycoprotein structures

Uniform $^{13}C/^{15}N$ labeling of a glycoprotein has been mainly used to apply standard methodology of multidimensional heteronuclear NMR spectroscopy for resonance assignment and structure calculation of the polypeptide chain of the glycoprotein (Cubeddu et al. 2000; Metzler et al. 1997; Wyss et al. 1995; Yamaguchi et al. 2006). Uniform ^{15}N labeling enables the resonance and NOE assignment of amide groups within the carbohydrate, the glycosylated asparagines and protein backbone, e.g. using ^{15}N edited TOCSY, ^{15}N edited NOESY, HNHA and HNHB experiments (Wood et al. 2000; Wyss et al. 1995). ^{13}C labeling helps resonance assignment of the glycan by separating degenerate ^{1}H resonances in an additional, better-dispersed ^{13}C dimension, e.g. using an HCCH-TOCSY (Weller et al. 1996) or an HCCH-COSY (Yamaguchi et al. 2000). NOE assignment of the glycan is facilitated using a ^{13}C-edited NOESY. Overlapped regions in the 2D NOESY can often be resolved (Weller et al. 1996). Obtaining more unambiguous restraints helps to better define the structure of the protein and carbohydrate.

Using segmental protein-carbohydrate labeling to study carbohydrate structures

The segmental labeling scheme in which the protein is uniformly $^{13}C/^{15}N$ labeled and the glycan is unlabeled (Fig. 3c) facilitates not only to assign and distinguish overlapped ^{1}H signals between the glycan and the protein but most importantly enables the separation of NOEs within the glycan, NOEs between the protein and the glycan and NOEs within the protein using filter and editing NMR techniques. 2D ^{13}C filtered filtered NOESY spectra (Peterson et al. 2004) contain solely NOEs within the glycan while all NOEs within the protein and the protein-glycan NOEs are suppressed (Fig. 4). Segmental labeling is

Fig. 4 Structural ^{13}C filtered filtered 2D NOESY of the glycoprotein AcrA from *C. jejuni* ($^{13}C/^{15}N$-labeled protein). Only NOEs within the unlabeled carbohydrate moiety are observed
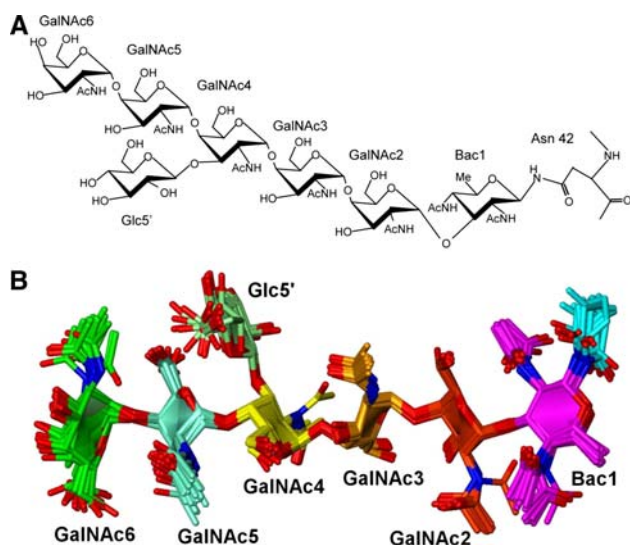
Fig. 5 **a** Scheme of N-glycan of *C. jejuni* and **b** structural ensemble of the same sugar superimposed on the heavy atoms of all saccharides. The sugar was attached to an isotopically labeled protein which resulted in the measurement of a high number of NOEs within the sugar moiety

the only method so far that helped to resolve the severe chemical shift overlap between the glycan and protein resonances.

The attachment of the oligosaccharide to a globular protein has the advantage that the NOE transfer within the carbohydrate becomes very efficient due to the increased overall tumbling time. The NOE build up and efficiency is then comparable to those observed in proteins of a similar size. High magnetic field NMR spectroscopy at 900 MHz is beneficial to obtain sufficient dispersion in this homonuclear spectrum for unambiguous assignment. Such a NOESY spectrum, recorded at mixing times in the linear range, contains all the distance information necessary for structure calculations of the glycan moiety. In the N-glycan of *C. jejuni* attached to the protein AcrA, 125 distance constraints could be collected within this bacterial heptasaccharide (Slynko et al. 2009). On average 11 inter-residue NOEs were observed per glycosidic linkage, a number that significantly exceeds the amount of NOEs observed for isolated carbohydrates (Wormald et al. 2002). Structure calculations revealed a well-defined glycan ensemble (Fig. 5) demonstrating that the improved NOE transfer results in a crucially enhanced quality of oligosaccharide 3D structures.

Applications for labeled glycans

Isotope labeling of glycans with ^{15}N and ^{13}C (Fig. 3a, b, d–f) enables a variety of NMR applications. First, through-bond experiments can be used for the assignment of the carbohydrate, e.g. HCCH-TOCSY (Weller et al. 1996), HCCH-COSY (Yamaguchi et al. 2000), HNCO and HNCA for acetamido group assignment (Weller et al. 1996) or experiments designed especially for carbohydrates (Colebrooke et al. 2005; Macnaughtan et al. 2008). Second, chemical shift degeneracies especially in NOESY spectra can be resolved to obtain more distance restraints as reported in a protein–carbohydrate complex (Harris et al. 1999). Third, carbohydrate dynamics could be studied using ^{15}N (Yamaguchi 2008) and ^{13}C relaxation (Miyazaki et al. 2000; Yamaguchi et al. 1998).

Sugar-type specific labeling (Fig. 3b, d, e) simplifies the resonance assignment of the glycan by simplifying the spectra (Gilhespy-Muskett et al. 1994; Yamaguchi et al. 1998). But most importantly, such labeling scheme can potentially be used to selectively detect NOEs between the non-labeled and the labeled sugar units using 2D and 3D filtered-editing NOESY spectra. Finally, in using ^{2}H/^{13}C labeled glucose precursors for the expression of the glycoprotein, certain types of sugar units could be distinguished by ^{1}H/^{2}H ratios at specific hydrogen positions. This way GlcNac, Man and Fuc can be distinguished which helps the assignment process (Yamaguchi et al. 2000).

Conclusion and perspectives

As shown in this review, structure determination of segmentally labeled biomolecules strongly depends on inter-segmental NOEs obtained by filtered and edited NOESY spectra between the two fragments. However, for larger proteins the filter elements will lead to a significant reduction in NOE intensities due to relaxation losses. Two alternative differential labeling schemes originally developed for protein–protein complexes should probably be used to overcome these losses. One is based on a highly deuterated (>98%) and ^{15}N labeled protein together with an unlabeled molecule (Walters et al. 2001). A 3D ^{15}N edited NOESY spectrum will reveal then inter-segmental NOEs between the amides of the deuterated segment and aliphatic protons from the unlabeled segment. The experiment is very sensitive due to the absence of filter delays and the deuterated environment of the NH. However, only inter-segmental NOEs involving amides are observable. This labeling scheme has been used once for a segmentally labeled multi-domain protein but the system did not reveal interdomain NOEs probably due to dynamics (Walters et al. 2003). A reverse approach uses a ^{1}H/^{13}C-I,L,V-methyl/^{2}H labeled segment with an unlabeled segment to enable the observation of methyl to aliphatic inter-segmental NOEs. This approach was demonstrated in a protein–protein complex (Gross et al. 2003a; Gross et al. 2003b).

We believe that the in vitro glycosylation technique has large potential for the 3D structure determination of glycoproteins. However, further development is needed to adapt the in vitro glycosylation system to the transfer of a larger range of glycans and to circumvent limitations in regard to the glycosylation sequence requirements. The in vitro glycosylation technique has also potential for the study of carbohydrate conformations by attaching them to a model protein taking advantage of the slower tumbling regime and the differential labeling. A variety of methods exist to label carbohydrates and it now remains to discover methods to attach those to an unlabeled protein. The methodology consisting of ligating synthesized glycopeptides with recombinantly expressed labeled proteins is in principle available but has not been used yet. We expect that it will be only a matter of time until the first segmental labeled glycoprotein will be obtained based on this method.

As discussed here, segmental isotope labeling methods have already proven to be very useful for the study of biomolecules using NMR spectroscopy. Although these methods have not been widely used for NMR structure determination, it is likely considering the increased interest for solving structures of large proteins or membrane proteins by solid and liquid state NMR that there will be an increased demand for preparing proteins with segmental isotopic labeling to resolve resonance overlap. For such applications, the recent methodological development allowing ligation of three pieces is crucial as this will be the only approach allowing the central fragment of a protein to be labeled separately. Moreover, methods developed for segmental labeling of proteins could also be used for studying protein–protein complexes, a second major area in biological NMR spectroscopy. Several NMR studies (Mal et al. 2007; Volkman et al. 2002) showed a dramatic increase of the spectral quality of protein–protein complexes when the two proteins are expressed together covalently attached by a Glycine–Serine-rich linker. One could now easily envisage to segmentally label such "complex fusions" to better detect intermolecular contacts and to help solving such structures.

References

Anderson LL, Marshall GR, Crocker E, Smith SO, Baranski TJ (2005) Motion of carboxyl terminus of Galpha is restricted upon G protein activation. A solution NMR study using semisynthetic Galpha subunits. J Biol Chem 280:31019–31026

Apweiler R, Hermjakob H, Sharon N (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. Biochim Biophys Acta 1473:4–8

Arya R, Bhattacharya A, Saini KS (2008) Dictyostelium discoideum—a promising expression system for the production of eukaryotic proteins. FASEB J 22:4055–4066

Ayers B, Blaschke UK, Camarero JA, Cotton GJ, Holford M, Muir TW (1999) Introduction of unnatural amino acids into proteins using expressed protein ligation. Biopolymers 51:343–354

Azurmendi HF, Vionnet J, Wrightson L, Trinh LB, Shiloach J, Freedberg DI (2007) Extracellular structure of polysialic acid explored by on cell solution NMR. Proc Natl Acad Sci USA 104:11557–11561

Balambika R, Inui T, Sargsyan H, Arshava B, Cohen L, Ding F, Becker J, Naider F (2007) Synthesis of a double transmembrane domain fragment of Ste2p by native chemical ligation. Int J Peptide Res & Ther 13:251–263

Blanchard V, Gadkari RA, George AV, Roy S, Gerwig GJ, Leeflang BR, Dighe RR, Boelens R, Kamerling JP (2008) High-level expression of biologically active glycoprotein hormones in Pichia pastoris strains–selection of strain GS115, and not X-33, for the production of biologically active N-glycosylated ^{15}N-labeled phCG. Glycoconj J 25:245–257

Blaschke UK, Cotton GJ, Muir TW (2000) Synthesis of multi-domain proteins using expressed protein ligation: strategies for segmental isotopic labeling of internal regions. Tetrahedron 56:9461

Blundell CD, DeAngelis PL, Day AJ, Almond A (2004) Use of ^{15}N-NMR to resolve molecular details in isotopically-enriched carbohydrates: sequence-specific observations in hyaluronan oligomers up to decasaccharides. Glycobiology 14:999–1009

Brik A, Ficht S, Wong CH (2006) Strategies for the preparation of homogenous glycoproteins. Curr Opin Chem Biol 10:638–644

Busche AEL, Aranko AS, Talebzadeh-Farooji M, Bernhard F, Dötsch V, Iwaï H (2009) Segmental isotopic labeling of a central domain in a multidomain protein by protein trans-splicing using only one robust DnaE intein. Angew Chem Int Ed 48:6128–6131

Camarero JA, Fushman D, Cowburn D, Muir TW (2001) Peptide chemical ligation inside living cells: in vivo generation of a circular protein domain. Bioorg Med Chem 9:2479–2484

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

Chen MM, Glover KJ, Imperiali B (2007) From peptide to protein: comparative analysis of the substrate specificity of N-linked glycosylation in C. jejuni. Biochemistry 46:5579–5585

Colebrooke SA, Blundell CD, DeAngelis PL, Campbell ID, Almond A (2005) Exploiting the carboxylate chemical shift to resolve degenerate resonances in spectra of ^{13}C-labelled glycosaminoglycans. Magn Reson Chem 43:805–815

Conte MR, Grune T, Ghuman J, Kelly G, Ladas A, Matthews S, Curry S (2000) Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals novel features of the RRM fold. EMBO J 19:3132–3141

Cotton GJ, Muir TW (2000) Generation of a dual-labeled fluorescence biosensor for Crk-II phosphorylation using solid-phase expressed protein ligation. Chem Biol 7:253

Cotton GJ, Ayers B, Xu R, Muir TW (1999) Insertion of a synthetic peptide into a recombinant protein framework: a protein biosensor. J Am Chem Soc 121:1100–1101

Cubeddu L, Moss CX, Swarbrick JD, Gooley AA, Williams KL, Curmi PM, Slade MB, Mabbutt BC (2000) Dictyostelium discoideum as expression host: isotopic labeling of a recombinant glycoprotein for NMR studies. Protein Expr Purif 19:335–342

David R, Richter MP, Beck-Sickinger AG (2004) Expressed protein ligation. Method and applications. Eur J Biochem 271:663–677

Dawson PE, Kent SBH (2000) Synthesis of native proteins by chemical ligation. Annu Rev Biochem 69:923–960

Dawson PE, Muir TW, Clark-Lewis I, Kent SB (1994) Synthesis of proteins by native chemical ligation. Science 266:776–779

De Beer T, Van Zuylen CW, Leeflang BR, Hard K, Boelens R, Kaptein R, Kamerling JP, Vliegenthart JF (1996) NMR studies of the free alpha subunit of human chorionic gonadotropin. Structural influences of N-glycosylation and the beta subunit on the conformation of the alpha subunit. Eur J Biochem 241:229–242

Derbyshire V, Wood DW, Wu W, Dansereau JT, Dalgaard JZ, Belfort M (1997) Genetic definition of a protein-splicing domain: functional mini-inteins support structure predictions and a model for intein evolution. Proc Natl Acad Sci USA 94:11466–11471

Dumon C, Bosso C, Utille JP, Heyraud A, Samain E (2006) Production of Lewis x tetrasaccharides by metabolically engineered Escherichia coli. Chembiochem 7:359–365

Erbel PJ, Karimi-Nejad Y, van Kuik JA, Boelens R, Kamerling JP, Vliegenthart JF (2000) Effects of the N-linked glycans on the 3D structure of the free alpha-subunit of human chorionic gonadotropin. Biochemistry 39:6012–6021

Evans TC, Xu MQ (2002) Mechanistic and kinetic considerations of protein splicing. Chem Rev 102:4869–4884

Evans TC Jr, Benner J, Xu MQ (1998) Semisynthesis of cytotoxic proteins using a modified protein splicing element. Protein Sci 7:2256–2264

Fierfort N, Samain E (2008) Genetic engineering of Escherichia coli for the economical production of sialylated oligosaccharides. J Biotechnol 134:261–265

Fletcher CM, Harrison RA, Lachmann PJ, Neuhaus D (1994) Structure of a soluble, glycosylated form of the human complement regulatory protein CD59. Structure 2:185–199

Gamblin DP, Scanlan EM, Davis BG (2009) Glycoprotein synthesis: an update. Chem Rev 109:131–163

Gardner KH, Kay LE (1998) The use of ^{2}H, ^{13}C, ^{15}N multidimensional NMR to study the structure and dynamics of proteins. Annu Rev Biophys Biomol Struct 27:357–406

Gawlitzek M, Papac DI, Sliwkowski MB, Ryll T (1999) Incorporation of ^{15}N from ammonium into the N-linked oligosaccharides of an immunoadhesin glycoprotein expressed in Chinese hamster ovary cells. Glycobiology 9:125–131

Gilhespy-Muskett AM, Partridge J, Jefferis R, Homans SW (1994) A novel ^{13}C isotopic labelling strategy for probing the structure and dynamics of glycan chains in situ on glycoproteins. Glycobiology 4:485–489

Gitti R, Long G, Bush CA (1994) Measurement of long-range2 ^{13}C–^{1}H coupling constants of 95% uniformly ^{13}C-labeled polysaccharide from Streptococcus mitis J22. Biopolymers 34:1327–1338

Goux WJ, Perry C, James TL (1982) An NMR study of ^{13}C-enriched galactose attached to the single carbohydrate chain of hen ovalbumin. Motions of the carbohydrates of glycoproteins. J Biol Chem 257:1829–1835

Gross JD, Gelev VM, Wagner G (2003a) A sensitive and robust method for obtaining intermolecular NOEs between side chains in large protein complexes. J Biomol NMR 25:235–242

Gross JD, Moerke NJ, Von der Haar T, Lugovskoy AA, Sachs AB, McCarthy JE, Wagner G (2003b) Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E. Cell 115:739–750

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

Hancock SM, Vaughan MD, Withers SG (2006) Engineering of glycosidases and glycosyltransferases. Curr Opin Chem Biol 10:509–519

Harris R, Kiddle GR, Field RA, Milton MJ, Ernst B, Magnani JL, Homans SW (1999) Stable-isotope-assisted NMR studies on C-13-enriched sialyl Lewis(x) in solution and bound to E-selectin. J Am Chem Soc 121:2546–2551

Hashimoto Y, Toma K, Nishikido J, Yamamoto K, Haneda K, Inazu T, Valentine KG, Opella SJ (1999) Effects of glycosylation on the structure and dynamics of eel calcitonin in micelles and lipid bilayers determined by nuclear magnetic resonance spectroscopy. Biochemistry 38:8377–8384

Hofmann RM, Muir TW (2002) Recent advances in the application of expressed protein ligation to protein engineering. Curr Opin Biotechnol 13:297

Ichikawa Y, Lin YC, Dumas DP, Shen GJ, Garciajunceda E, Williams MA, Bayer R, Ketcham C, Walker LE, Paulson JC, Wong CH (1992) Chemical-enzymatic synthesis and conformational-analysis of Sialyl Lewis-X and derivatives. J Am Chem Soc 114:9283–9298

Ippel JH, Pouvreau L, Kroef T, Gruppen H, Versteeg G, van den Putten P, Struik PC, van Mierlo CP (2004) In vivo uniform (15)N-isotope labelling of plants: using the greenhouse for structural proteomics. Proteomics 4:226–234

Iwai H, Pluckthun A (1999) Circular beta-lactamase: stability enhancement by cyclizing the backbone. FEBS Lett 459:166–172

Jackson DY, Burnier J, Quan C, Stanley M, Tom J, Wells JA (1994) A designed peptide ligase for total synthesis of ribonuclease A with unnatural catalytic residues. Science 266:243–247

Kato K, Yamaguchi Y (2008) Structural analyses of glycoconjugates by NMR. In: Taniguchi N et al (eds) Experimental glycoscience: glycochemistry. Springer, Tokyo, pp 45–50

Kato K, Matsunaga C, Igarashi T, Kim H, Odaka A, Shimada I, Arata Y (1991a) Complete assignment of the methionyl carbonyl carbon resonances in switch variant anti-dansyl antibodies labeled with [1–13C]methionine. Biochemistry 30:270–278

Kato K, Matsunaga C, Odaka A, Yamato S, Takaha W, Shimada I, Arata Y (1991b) Carbon-13 NMR study of switch variant anti-dansyl antibodies: antigen binding and domain-domain interactions. Biochemistry 30:6604–6610

Kato K, Gouda H, Takaha W, Yoshino A, Matsunaga C, Arata Y (1993) ^{13}C NMR study of the mode of interaction in solution of the B fragment of staphylococcal protein A and the Fc fragments of mouse immunoglobulin G. FEBS Lett 328:49–54

Kato K, Sasakawa H, Kamiya Y, Utsumi M, Nakano M, Takahashi N, Yamaguchi Y (2008) 920 MHz ultra-high field NMR approaches to structural glycobiology. Biochim Biophys Acta 1780:619–625

Kern T, Hediger S, Muller P, Giustini C, Joris B, Bougault C, Vollmer W, Simorre JP (2008) Toward the characterization of peptidoglycan structure and protein-peptidoglycan interactions by solid-state NMR spectroscopy. J Am Chem Soc 130:5618–5619

Kim H, Matsunaga C, Yoshino A, Kato K, Arata Y (1994) Dynamical structure of the hinge region of immunoglobulin G as studied by ^{13}C nuclear magnetic resonance spectroscopy. J Mol Biol 236:300–309

Kobashigawa Y, Kumeta H, Ogura K, Inagaki F (2009) Attachment of an NMR-invisible solubility enhancement tag using a sortase-mediated protein ligation method. J Biomol NMR 43:145–150

Kochendoerfer GG, Jones DH, Lee S, Oblatt-Montal M, Opella SJ, Montal M (2004) Functional characterization and NMR spectroscopy on full-length Vpu from HIV-1 prepared by total chemical synthesis. J Am Chem Soc 126:2439–2446

Kowarik M, Numao S, Feldman MF, Schulz BL, Callewaert N, Kiermaier E, Catrein I, Aebi M (2006a) N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. Science 314:1148–1150

Kowarik M, Young NM, Numao S, Schulz BL, Hug I, Callewaert N, Mills DC, Watson DC, Hernandez M, Kelly JF, Wacker M, Aebi

M (2006b) Definition of the bacterial N-glycosylation site consensus sequence. EMBO J 25:1957–1966

Lee W, Revington MJ, Arrowsmith C, Kay LE (1994) A pulsed field gradient isotope-filtered 3D ^{13}C HMQC-NOESY experiment for extracting intermolecular NOE contacts in molecular complexes. FEBS Lett 350:87–90

Live D, Silks LA 3rd, Schmidt J (2001) ^{13}C isotopic enrichment for nuclear magnetic resonance studies of carbohydrates and glyco-conjugates. Methods Enzymol 338:305–319

Lustbader JW, Birken S, Pollak S, Pound A, Chait BT, Mirza UA, Ramnarain S, Canfield RE, Brown JM (1996) Expression of human chorionic gonadotropin uniformly labeled with NMR isotopes in Chinese hamster ovary cells: an advance toward rapid determination of glycoprotein structures. J Biomol NMR 7:295–304

Macnaughtan MA, Tian F, Liu S, Meng L, Park S, Azadi P, Moremen KW, Prestegard JH (2008) ^{13}C-sialic acid labeling of glycans on glycoproteins using ST6Gal-I. J Am Chem Soc 130:11864–11865

Mal TK, Takahata S, Ki S, Zheng L, Kokubo T, Ikura M (2007) Functional silencing of TATA-binding protein (TBP) by a covalent linkage of the N-terminal domain of TBP-associated factor 1. J Biol Chem 282:22228–22238

Mathys S, EvansJr TC, Chute IC, Wu H, Chong S, Benner J, Liu X-Q, Xu M-Q (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

Metzler WJ, Bajorath J, Fenderson W, Shaw SY, Constantine KL, Naemura J, Leytze G, Peach RJ, Lavoie TB, Mueller L, Linsley PS (1997) Solution structure of human CTLA-4 and delineation of a CD80/CD86 binding site conserved in CD28. Nat Struct Biol 4:527–531

Mills KV, Lew BM, Jiang S, Paulus H (1998) Protein splicing in trans by purified N- and C-terminal fragments of the Mycobacterium tuberculosis RecA intein. Proc Natl Acad Sci USA 95:3543–3548

Miyazaki T, Sato H, Sakakibara T, Kajihara Y (2000) An approach to the precise chemoenzymatic synthesis of C-13-labeled sialylol-igosaccharide on an intact glycoprotein: a novel one-pot [3-C-13]-labeling method for sialic acid analogues by control of the reversible aldolase reaction, enzymatic synthesis of [3-C-13]-NeuAc-alpha-(2 → 3)-[U-C-13]-Gal-beta-(1 → 4)-GlcNAc-beta-sequence onto glycoprotein, and its conformational analysis by developed NMR techniques. J Am Chem Soc 122:5678–5694

Muir TW (2003) Semisynthesis of proteins by expressed protein ligation. Annu Rev Biochem 72:249–289

Muir TW, Sondhi D, Cole PA (1998) Expressed protein ligation: a general method for protein engineering. Proc Natl Acad Sci USA 95:6705–6710

Muona M, Aranko AS, Iwai H (2008) Segmental isotopic labelling of a multidomain protein by protein ligation by protein trans-splicing. Chembiochem 9:2958–2961

Muralidharan V, Muir TW (2006) Protein ligation: an enabling technology for the biophysical analysis of proteins. Nat Methods 3:429–438

Oberstrass FC, Auweter SD, Erat M, Hargous Y, Henning A, Wenter P, Reymond L, Amir-Ahmady B, Pitsch S, Black DL, Allain FH (2005) Structure of PTB bound to RNA: specific binding and implications for splicing regulation. Science 309:2054–2057

Otomo T, Ito N, Kyogoku Y, Yamazaki T (1999) NMR observation of selected segments in a larger protein: central-segment isotope labeling through intein-mediated ligation. Biochemistry 38:16040–16044

Paulus H (2000) Protein splicing and related forms of protein autoprocessing. Annu Rev Biochem 69:447–496

Pervushin K, Riek R, Wider G, Wuthrich K (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proc Natl Acad Sci USA 94:12366–12371

Pervushin K, Riek R, Wider G, Wuthrich K (1998) Transverse relaxation-optimized spectroscopy (TROSY) for NMR studies of aromatic spin systems in C-13-labeled proteins. J Am Chem Soc 120:6394–6400

Peterson RD, Theimer CA, Wu H, Feigon J (2004) New applications of 2D filtered/edited NOESY for assignment and structure elucidation of RNA and RNA-protein complexes. J Biomol NMR 28:59–67

Pickford AR, O'Leary JM (2004) Isotopic labeling of recombinant proteins from the methylotrophic yeast Pichia pastoris. Methods Mol Biol 278:17–33

Piontek C, Ring P, Harjes O, Heinlein C, Mezzato S, Lombana N, Pohner C, Puttner M, Varon Silva D, Martin A, Schmid FX, Unverzagt C (2009) Semisynthesis of a homogeneous glycoprotein enzyme: ribonuclease C: part 1. Angew Chem Int Ed Engl 48:1936–1940

Probert MA, Milton MJ, Harris R, Schenkman S, Brown JM, Homans SW, Field RA (1997) Chemoenzymatic synthesis of GM(3), Lewis x and sialyl Lewis x oligosaccharides in C-13-enriched form. Tetrahedron Lett 38:5861–5864

Rajagopal S, Kent SB (2007) Total chemical synthesis and biophysical characterization of the minimal isoform of the KChIP2 potassium channel regulatory subunit. Protein Sci 16:2056–2064

Severinov K, Muir TW (1998) Expressed protein ligation, a novel method for studying protein-protein interactions in transcription. J Biol Chem 273:16205–16209

Shi J, Muir TW (2005) Development of a tandem protein trans-splicing system based on native and engineered split inteins. J Am Chem Soc 127:6198–6206

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

Skrisovska L, Allain FH (2008) Improved segmental isotope labeling methods for the NMR study of multidomain or large proteins: application to the RRMs of Npl3p and hnRNP L. J Mol Biol 375:151–164

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

Southworth MW, Adam E, Panne D, Byer R, Kautz R, Perler FB (1998) Control of protein splicing by intein fragment reassembly. EMBO J 17:918–926

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

Tolbert TJ, Wong C-H (2002) New methods for proteomic research: preparation of proteins with N-terminal cysteines for labeling and conjugation13. Angew Chem Int Ed 41:2171–2174

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

Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE (2003) Cross-correlated relaxation enhanced ^{1}H[bond]^{13}C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. J Am Chem Soc 125:10420–10428

Varki A (1993) Biological roles of oligosaccharides: all of the theories are correct. Glycobiology 3:97–130

Vitali F, Henning A, Oberstrass FC, Hargous Y, Auweter SD, Erat M, Allain FH (2006) Structure of the two most C-terminal RNA recognition motifs of PTB using segmental isotope labeling. EMBO J 25:150–162

Volkman BF, Prehoda KE, Scott JA, Peterson FC, Lim WA (2002) Structure of the N-WASP EVH1 domain-WIP complex: insight into the molecular basis of Wiskott-Aldrich syndrome. Cell 111:565–576

Wacker M, Linton D, Hitchen PG, Nita-Lazar M, Haslam SM, North SJ, Panico M, Morris HR, Dell A, Wren BW, Aebi M (2002) N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. Science 298:1790–1793

Walters KJ, Matsuo H, Wagner G (1997) A simple method to distinguish intermonomer nuclear overhauser effects in homodimeric proteins with C2 symmetry. J Am Chem Soc 119: 5958–5959

Walters KJ, Ferentz AE, Hare BJ, Hidalgo P, Jasanoff A, Matsuo H, Wagner G (2001) Characterizing protein-protein complexes and oligomers by nuclear magnetic resonance spectroscopy. Methods Enzymol 339:238–258

Walters KJ, Lech PJ, Goh AM, Wang Q, Howley PM (2003) DNA-repair protein hHR23a alters its protein structure upon binding proteasomal subunit S5a. Proc Natl Acad Sci USA 100: 12694–12699

Walton WJ, Kasprzak AJ, Hare JT, Logan TM (2006) An economic approach to isotopic enrichment of glycoproteins expressed from Sf9 insect cells. J Biomol NMR 36:225–233

Weller CT, Lustbader J, Seshadri K, Brown JM, Chadwick CA, Kolthoff CE, Ramnarain S, Pollak S, Canfield R, Homans SW (1996) Structural and conformational analysis of glycan moieties in situ on isotopically ^{13}C, ^{15}N-enriched recombinant human chorionic gonadotropin. Biochemistry 35:8815–8823

Wood MJ, Komives EA (1999) Production of large quantities of isotopically labeled protein in *Pichia pastoris* by fermentation. J Biomol NMR 13:149–159

Wood MJ, Sampoli Benitez BA, Komives EA (2000) Solution structure of the smallest cofactor-active fragment of thrombomodulin. Nat Struct Biol 7:200–204

Wormald MR, Petrescu AJ, Pao YL, Glithero A, Elliott T, Dwek RA (2002) Conformational studies of oligosaccharides and glycopeptides: complementarity of NMR, X-ray crystallography, and molecular modelling. Chem Rev 102:371–386

Wu H, Hu Z, Liu XQ (1998) Protein trans-splicing by a split intein encoded in a split DnaE gene of Synechocystis sp. PCC6803. Proc Natl Acad Sci USA 95:9226–9231

Wyss DF, Withka JM, Knoppers MH, Sterne KA, Recny MA, Wagner G (1993) ^{1}H resonance assignments and secondary structure of the 13.6 kDa glycosylated adhesion domain of human CD2. Biochemistry 32:10995–11006

Wyss DF, Choi JS, Li J, Knoppers MH, Willis KJ, Arulanandam AR, Smolyar A, Reinherz EL, Wagner G (1995) Conformation and function of the N-linked glycan in the adhesion domain of human CD2. Science 269:1273–1278

Xu MQ, Evans TC Jr (2001) Intein-mediated ligation and cyclization of expressed proteins. Methods 24:257–277

Xu M-Q, Evans TC (2003) Purification of recombinant proteins from *E. coli* by engineered inteins. Methods Mol Biol 205:43–68

Xu MQ, Evans TC Jr (2005) Recent advances in protein splicing: manipulating proteins in vitro and in vivo. Curr Opin Biotechnol 16:440–446

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

Yabe N, Matsuya Y, Yamane I, Takada M (1986) Enhanced formation of mouse hybridomas without hat treatment in a serum-free medium. In Vitro Cell Dev Biol 22:363–368

Yagi H, Tsujimoto T, Yamazaki T, Yoshida M, Akutsu H (2004) Conformational change of H+ -ATPase beta monomer revealed on segmental isotope labeling NMR spectroscopy. J Am Chem Soc 126:16632–16638

Yamaguchi Y (2008) Development and applications of stable-isotope-labeling methods oriented to structural glycobiology. Trends Glycosci Glycotechnol 20:117–130

Yamaguchi Y, Kato K, Shindo M, Aoki S, Furusho K, Koga K, Takahashi N, Arata Y, Shimada I (1998) Dynamics of the carbohydrate chains attached to the Fc portion of immunoglobulin G as studied by NMR spectroscopy assisted by selective ^{13}C labeling of the glycans. J Biomol NMR 12:385–394

Yamaguchi Y, Takizawa T, Kato K, Arata Y, Shimada I (2000) ^{1}H and ^{13}C NMR assignments for the glycans in glycoproteins by using ^{2}H/^{13}C-labeled glucose as a metabolic precursor. J Biomol NMR 18:357–360

Yamaguchi Y, Nishimura M, Nagano M, Yagi H, Sasakawa H, Uchida K, Shitara K, Kato K (2006) Glycoform-dependent conformational alteration of the Fc region of human immunoglobulin G1 as revealed by NMR spectroscopy. Biochim Biophys Acta 1760:693–700

Yamamoto N, Tanabe Y, Okamoto R, Dawson PE, Kajihara Y (2008) Chemical synthesis of a glycoprotein having an intact human complex-type sialyloligosaccharide under the Boc and Fmoc synthetic strategies. J Am Chem Soc 130:501–510

Yamazaki T, Otomo T, Oda N, Kyogoku Y, Uegaki K, Ito N, Ishino Y, Nakamura H (1998) Segmental isotope labeling for protein NMR using peptide splicing. J Am Chem Soc 120:5591–5592

Yu L, Goldman R, Sullivan P, Walker GF, Fesik SW (1993) Heteronuclear NMR studies of ^{13}C-labeled yeast cell wall beta-glucan oligosaccharides. J Biomol NMR 3:429–441

Zhang Y, Vasudevan S, Sojitrawala R, Zhao W, Cui C, Xu C, Fan D, Newhouse Y, Balestra R, Jerome WG, Weisgraber K, Li Q, Wang J (2007) A monomeric, biologically active, full-length human apolipoprotein E. Biochemistry 46:10722–10732

Zhou P, Lugovskoy AA, Wagner G (2001) A solubility-enhancement tag (SET) for NMR studies of poorly behaving proteins. J Biomol NMR 20:11–14

Zuger S, Iwai H (2005) Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies. Nat Biotechnol 23:736–740

Zwahlen C, Legault P, Vincent SJF, Greenblatt J, Konrat R, Kay LE (1997) Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: application to a bacteriophage lambda N-peptide/boxB RNA complex. J Am Chem Soc 119:6711–6721