# PROTEIN CHEMICAL LIGATION AS AN INVALUABLE TOOL FOR STRUCTURAL NMR

## A. Shekhtman*

## 1. Introduction

Structural analysis of proteins and their complexes comes to the forefront of the efforts to gain comprehensive knowledge of biological processes within the cell [1–3]. Protein NMR spectroscopy becomes the technique of choice for structural characterization of proteins and their complexes when the molecular weight does not exceed approximately 30 kDa [4, 5]. Though X-ray crystallography remains the gold standard for high-resolution determination of protein structures, the availability of NMR analysis expands the repertoire of structural biology to include flexible and otherwise non-crystallizable structural targets, such as multi-domain systems where separate domains are connected by flexible linkers [6–8]. The ability of NMR spectroscopy to characterize the dynamics of these systems on multiple timescales at atomic resolution is unique among spectroscopic methods and can provide necessary insights into the biological function of the corresponding molecules [4, 9].

The barrier for the application of NMR spectroscopy to characterize larger protein targets is the increased spectral complexity of these large proteins and to the significant line-width broadening associated with unfavorable relaxation properties. At the same time, there is a strong need to have a technique which adequately describes the molecular interactions of modular proteins consisting of multi-domain structures. Structural analysis of separate domains from these systems may not be adequate due to the inter-domain and intra-molecular regulatory interactions. Multi-domain proteins created through genetic shuttling are very common, especially, in eukaryotic systems [10]. Usually, the molecular weight of a single protein domain is about 10 kDa [7]. Thus, multi-domain proteins are difficult NMR targets due to their increasingly large molecular weight. Newly developed NMR techniques based on transverse relaxation optimized (TROSY) methodology significantly alleviates problems related to the increased line-width of the NMR signals and makes it possible to work with proteins and protein complexes of molecular weights beyond 100 kDa [11].

Although observable, NMR signals from the large proteins exhibit extreme spectral overlap, which cannot be resolved even in three- or four-dimensional NMR spectra [12]. The way to decrease structural complexity is to use samples where only few amino acids are labeled with NMR active nuclei, thus, editing out signals from the rest of the molecule [13]. Though powerful, this approach is limited because of the semi-uniform distribution of specific amino acids within the primary structure of proteins, which leads to the inability to obtain sequence specific information without highly non-trivial resonance assignments.

Expressed protein ligation presents a very valuable addition to the existing repertoire of protein over-expression technologies for NMR sample preparation [14–16].

---

* The State University of New York at Albany, USA.

It allows one to isotopically label part of the full-length protein leaving the rest of the protein cryptic. This approach dramatically decreases the spectral complexity of the NMR data since isotope edited NMR experiments leave only resonances originating from the isotopically labeled segment of the protein. Segmental labeling can be used both to expand structural NMR characterization of the proteins to much larger sizes and to obtain highly specific information about protein structure using the labeled segment as a chemical probe. Segmental labeling has successfully been used to study by NMR diverse modular systems such as the SH3-SH2 domains from abl kinase [17], the bacterial $\sigma^A$ factor [18] and the Gyrase intein system [19].
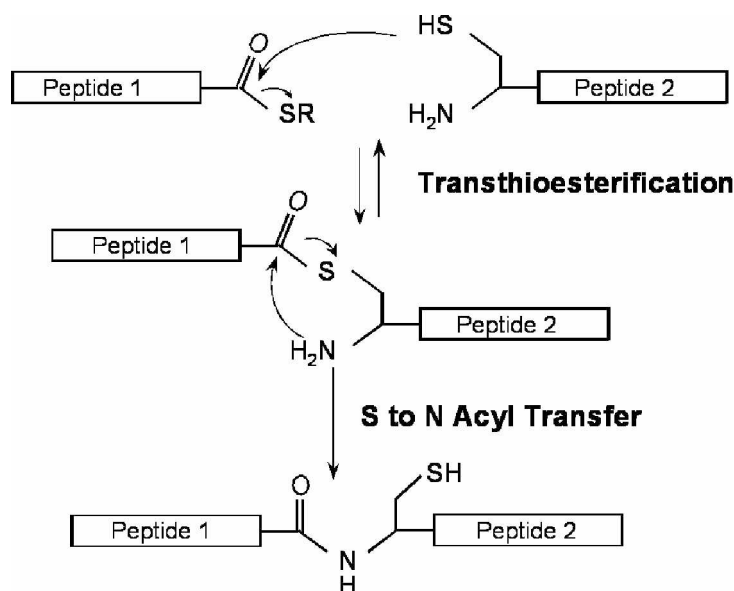


*Fig. 1.* General mechanism of protein chemical ligation

## 2. Method

The general methodology of expressed protein ligation is based on the reaction between the C-terminal thioester of the N-terminal segment of the protein and the N-terminal cysteine residue of the C-terminal segment [20, 21] (Fig. 1). The critical issue for the sample construction is the position of the ligation site, which is usually chosen in a loop region not involved in the biological activity of the protein. This minimizes the introduction of any possible structural change at the ligation site and, more importantly, reduces the possibility to affect the biological function of the protein. The reaction is mild enough to be performed under native conditions when one or both precursor fragments are folded [15]. The C-terminal $\alpha$-thioester can be produced using either solid phase peptide synthesis or bacterially over-expressed protein-intein fusion precursor with subsequent intein cleavage [16].

354

The total synthesis of peptides is well developed and provides essentially complete control over the regio- and stereospecific placement of isotopic labels within the peptidicstructure. It is essential to use this method if information about selected bonds in a full-length protein is needed. It works well when the synthetic product does not exceed about 3 kDa. The economies and time-scales of such synthesis essentially preclude routine use of this technique for synthesis of larger peptidic fragments.
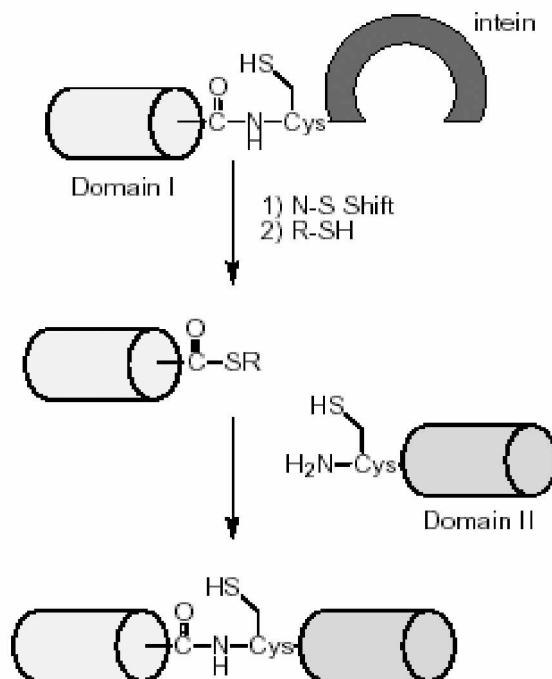


*Fig. 2.* General mechanism of expressed protein ligation
using intein fusion construct

The intein-based technique to generate polypeptide C-terminal $\alpha$-thioesters was proposed and shown to work on various systems [16]. It is based on genetically engineered inteins to release the N-terminal extein of the N-extein-intein fusion with a C-terminal $\alpha$-thioester group. The intein cleavage reaction is very robust and is accomplished using thiol containing compounds, such as ethanethiol and 2-mercaptoethanesulfonic acid (MESNA).

The N-terminal cysteine-containing fragment of the ligation reaction (Fig. 2) can be created using either factor Xa protease to release the final product from a specific fusion system or mutated inteins which release the N-terminal cysteine segment. Both methods are proven to be effective and can be interchangeably used to generate the desired product [15].

Either one or two domains of the full-length proteins can be separately labeled with an NMR active nuclei using EPL. The general issues of isotopic labeling for NMR are well known. Molecular size consideration dictates extensive use of deuteration to alleviate line broadening associated with proton-proton relaxation in large molecular systems. The reduced proton (REDPRO) approach [22] of isotopic labeling provides a useful trade-off between the need for proton density to obtain structural information using NMR and requirement of isolated protons to eliminate proton-proton relaxation. The REDPRO procedure has the potential for applications which include structural determination, mapping of chemical shift changes, and for the study of higher molecular weight cases.

## 3. Applications

Expressed protein ligation has been successfully used to characterize diverse multi-domain systems such as the abl SH32 dual domain system [17] and bacterial sigma factor [18] using NMR spectroscopy. A semi-synthetic approach to EPL was also used to characterize an unusual scissile bond between the N-extein and intein [19].
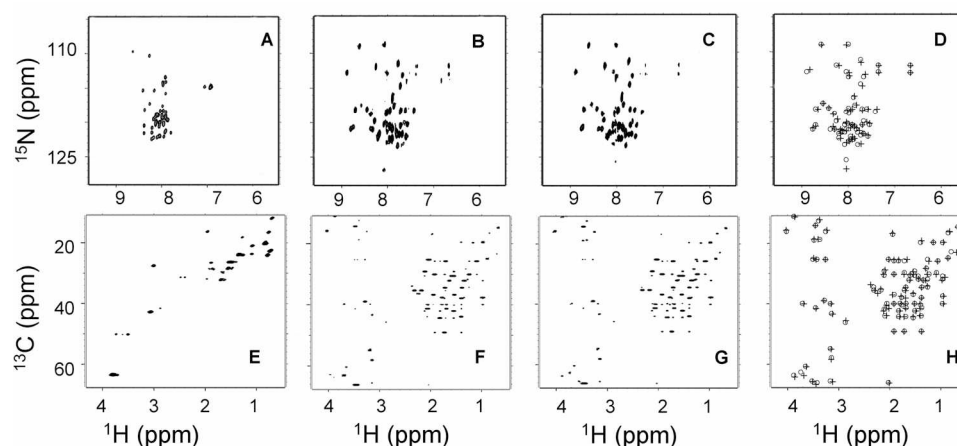


Fig. 3. Effect of the molecular context on the solution structure of bacterial sigma factor region 4.2. (A-C) $^1H\{^{15}N\}$HSQC-TROSY spectra of isolated region 4.2 (A), region 4.2 in the context of full-length bacterial sigma factor (B), region 4.2 in the context of bacterial sigma factor without region 1.1 (C); (E-G) $^1H\{^{13}C\}$ HSQC spectra of isolated region 4.2 (E), region 4.2 in the context of full-length bacterial sigma factor (F), region 4.2 in the context of bacterial sigma factor without region 1.1 (G); (D and H) Comparison of the reconstructed $^1H\{^{15}N\}$HSQC-TROSY and $^1H\{^{13}C\}$ HSQC spectra of the truncated bacterial sigma factor (circles) and full-length bacterial sigma factor (crosses), using chemical shifts extracted from individual spectra (B, C and F,G)

Bacterial sigma-factor binds to the RNA polymerase to form a holoenzyme competent for DNA transcription [23]. RNA polymerase itself does not bind to the promoter DNA region and completely relies on sigma-factor for DNA recognition. Para-

doxically, sigma-factor alone does not bind to DNA due to autoinhibition [24]. It was shown that the small acidic N-terminal domain Region 1.1 regulates binding of the C-terminal Region 4.2 of sigma factor to the DNA promoter region. It was hypothesized that Region 1.1 obstructs the DNA binding surface of region 4.2, thus preventing its interaction with DNA [25]. Using EPL segmental labeling, two constructs were created [18]. One construct was a full-length sigma factor with REDPRO labeled domain 4.2 and domains 1 through 4.1 unlabeled. The second construct was truncated form of sigma factor consisting of unlabeled domains 1.2 through 4.1 and REDPRO labeled 4.2. NMR spectra of the backbone amide protons and side-chains showed no significant change in chemical shifts between the two constructs (Fig. 3). Due to the exquisite sensitivity of the NMR chemical shift to the changes in the chemical environment, it was concluded that there is no direct interaction between domains 1.1 and 4.2. A new model of indirect, electrostatic interaction between domains 1.1 and 4.2 was proposed. Essentially, this study used segmental labeling to create a sequence specific probe of the tertiary structure of the protein without absolute need for laborious NMR resonance assignments. Semi-synthesis of a segmental isotopically labeled protein and NMR was used to decipher the mechanism of intein splicing autocatalysis [19].

The critical issue of the intein splicing reaction is the conformation of the scissile protein bond between the N-extein and the intein. Crystallographic studies found significant conformational heterogeneity in the scissile (-1) peptide bond ranging from normal *trans* [26, 27], distorted trans [28] and *cis* [29] conformation. NMR chemical shifts and the scalar coupling constant $^1J_{NC}$ are exquisitely sensitive to peptide bond conformation [30] and have been used to probe the conformation of the scissile (-1) peptide bond in the small intein *Mxe* GyrA. A N-extein containing a $^{13}C$ labeled C-terminal carbonyl was synthesized using Boc-SPPS [31]. It was chemically ligated to the [U- $^{15}N$] wild type intein, *Mxe* GyrA as well as the [U- $^{15}N$] inactive mutant GyrA(H75A), thus, creating an unique $^{13}C$-$^{15}N$ bond between the extein and intein. Using isotope edited NMR experiments [32], the scalar coupling constant $^1J_{NC'}$ of the labeled scissile bond was extracted from wild type and mutant GyrA. The significant difference in the coupling constant between the wild type N-extein-GyrA and inactive mutant N-extein-GyrA(H75A) suggests a distorted *trans* bond at the extein-intein junction existing in solution. Since twisted amide bonds are known to be significantly more susceptible to alkaline hydrolysis, this study supports the "ground-state destabilization" model as a part of the mechanism of autocatalysis [19].

## 4. Conclusions

Expressed protein ligation provides a unique opportunity to obtain structural high-resolution information about chemically defined segments of large proteins and protein complexes using NMR spectroscopy. It complements recently developed transverse relaxation optimized (TROSY) based NMR techniques to provide NMR windows into molecular masses of 100 kDa and more. Combination of semisynthesis and expressed protein ligation creates new possibilities of analyzing unusual electron configurations within full-length proteins using NMR observables.

# References

1. *Burley S. K., Bonanno J. B.* // Methods Biochem. Anal. 2003. Vol. 44. P. 591.
2. *Christendat D. Yee A.* et al. // Prog. Biophys. Mol. Biol. 2000. Vol. 73. P. 339.
3. *Koonin E. V., Wolf Y. I., Karev G. P.* // Nature. 2002. Vol. 420. P. 218.
4. *Cavanagh J., Fairbrother W. J.* et al. // Protein NMR Spectroscopy: Principles and practice. San Diego, 1996. P. 1.
5. *Montelione G. T., Zheng D.* et al. // Nat. Struct. Biol. 2000. Vol. 7. P. 982.
6. *Bax A., Kontaxis G., Tjandra N.* // Methods Enzymol. 2001. Vol. 339. P. 127.
7. *Cowburn D.* // J. of Cellular Biochemistry. 1995. P. 24.
8. *Fushman D., Cowburn D.* // Biological Magnetic Resonance. New York. 2003. P. 20, 53.
9. *Palmer A. G., Grey M. J., Wang C.* // Methods Enzymol. 2004.
10. *Gerstein M., Lin J., Hegyi H.* // Pac. Symp. Biocomput. 2000. P. 30.
11. *Riek R., Pervushin K., Wuthrich K.* // Trends Biochem. Sci. 2000. Vol. 25. P. 462.
12. *Salzmann M., Pervushin K.* et al. // Proc. Natl. Acad. Sci. U.S.A. 1998. Vol. 95. P. 13585.
13. *Matthews S.* // Methods Mol. Biol. 2004. Vol. 278. P. 35.
14. *Cowburn D., Muir T. W.* // Nuclear Magnetic Resonance of Biological Macromolecules. 2001. Vol. 339. P. 41.
15. *Cowburn D., Shekhtman A.* et al. // Methods Mol. Biol. 2004. Vol. 278. P. 47.
16. *Muir T. W.* // Annu. Rev. Biochem. 2003. Vol. 72. P. 249.
17. *Xu R., Ayers B., Cowburn D., Muir T. W.* // Proc. Nat. Acad. Sci. U.S.A. 1999. Vol. 96. P. 388.
18. *Camarero J. A., Shekhtman A.* et al. // Proc. Nat. Acad. Sci. U.S.A. 2002. Vol. 99. P. 8536.
19. *Romanelli A., Shekhtman A.* et al. // Proc. Nat. Acad. Sci. U.S.A. 2004. Vol. 101. P. 6397.
20. *Dawson P. E., Kent S. B.* // Annu. Rev. Biochem. 2000. Vol. 69. P. 923.
21. *Dawson P. E., Muir T. W.* et al. // Science. 1994. Vol. 266. P. 776.
22. *Shekhtman A., Ghose R.* et al. // FEBS Lett. 2002. Vol. 524. P. 177.
23. *Severinova E., Severinov K.* et al. // J. Mol. Biol. 1996. Vol. 263. P. 637.
24. *Gross C. A., Chan C.* et al. // Cold Spring Harb Symp. Quant. Biol. 1998. Vol. 63. P. 141.
25. *Dombroski A. J., Walter W. A., Gross C. A.* // Genes Dev. 1993. Vol. 7. P. 2446.
26. *Ding Y. Xu M. Q., Ghosh I.* et al. // J. Biol. Chem. 2003. Vol. 278. P. 39133.
27. *Mizutani R., Nogami S.* et al. // J. Mol. Biol. 2002. Vol. 316. P. 919.
28. *Poland B. W., Xu M. Q., Quiocho F. A.* // J. Biol. Chem. 2000. Vol. 275. P. 16408.
29. *Klabunde T., Sharma S.* et al. // Nat. Struct. Biol. 1998. Vol. 5. P. 31.
30. *Juranic N., Moncrieffe M. C.* et al. // J. Am. Chem. Soc. 2002. Vol. 124. P. 14221.
31. *Schnolzer M., Alewood P.* et al. // Int. J. Pept. Protein Res. 1992. Vol. 40. P. 180.
32. *Konrat R., Muhandiram D. R.* et al. // J. Biomol. NMR. 1997. Vol. 9. P. 409.

## БЕЛКОВОЕ ХИМИЧЕСКОЕ ЛЕГИРОВАНИЕ КАК ЭФФЕКТИВНЫЙ ИНСТРУМЕНТ ДЛЯ СТРУКТУРНОГО ЯМР

### А. Шехтман

В работе сделан обзор последних приложений метода белкового легирования для ЯМР исследований структуры протеинов и протеиновых комплексов. Эта уникальная методология позволяет создавать химерные белки, которые имеют определенные химические сегменты, помеченные ЯМР активными ядрами ($^{15}$N и $^{13}$C), оставляющие без изменений белковые свойства. Этот метод приводит к существенному сокращению спектральной сложности ЯМР и открывает новые возможности для анализа структур и структурных взаимодействий очень больших ( >100 к а. е. м.) биологических молекул и биологических механизмов.

358