Journal of Magnetic Resonance 241 (2014) 32–40

Contents lists available at ScienceDirect

Journal of Magnetic Resonance

journal homepage: www.elsevier.com/locate/jmr

Sparse labeling of proteins: Structural characterization from long range constraints

James H. Prestegard [a,*], David A. Agard [b], Kelley W. Moremen [a], Laura A. Lavery [b], Laura C. Morris [a], Kari Pederson [a]

[a] Complex Carbohydrate Research Center, University of Georgia, Athens, GA 30602, United States
[b] The Howard Hughes Medical Institute, Dept. Biochem. & Biophys., Univ. Calif. San Francisco, San Francisco, CA 94158, United States

ARTICLE INFO

Keywords:
Protein NMR
Glycoprotein
Sparse labeling
Paramagnetic constraints
Residual dipolar coupling
Ligand docking
NMR assignments

ABSTRACT

Structural characterization of biologically important proteins faces many challenges associated with degradation of resolution as molecular size increases and loss of resolution improving tools such as perdeuteration when non-bacterial hosts must be used for expression. In these cases, sparse isotopic labeling (single or small subsets of amino acids) combined with long range paramagnetic constraints and improved computational modeling offer an alternative. This perspective provides a brief overview of this approach and two discussions of potential applications; one involving a very large system (an Hsp90 homolog) in which perdeuteration is possible and methyl-TROSY sequences can potentially be used to improve resolution, and one involving ligand placement in a glycosylated protein where resolution is achieved by single amino acid labeling (the sialyltransferase, ST6Gal1). This is not intended as a comprehensive review, but as a discussion of future prospects that promise impact on important questions in the structural biology area.

1. Introduction

Protein structure determination by NMR has enjoyed many successful years based on a strategy that requires uniform labeling with ^{15}N and ^{13}C, depends on an ability to pass magnetic coherences between directly bonded pairs of nuclei in the polypeptide backbone, and results in extensive, if not complete, backbone assignments [1]. Extension to assignment of sidechain ^{13}C nuclei and directly bonded protons allows interpretation of NOEs as pairwise distance constraints and determination of structure. Refinements such as perdeuteration have improved resolution and allowed application to proteins many tens of kDa in size. Nevertheless this strategy has limitations for many important systems. Among them are biological systems involving proteins and protein complexes that are larger than sizes reached by this strategy and classes of proteins, glycoproteins for example, that are less amenable to expression in minimal media based on glucose and ammonium chloride, as commonly used with bacterial hosts. Eukaryotic hosts that produce proper glycosylation, including mammalian cells, also are less tolerant of high levels of deuterium and use of perdeuteration has been precluded in these cases. This is not a small issue as an estimated 50% of all human proteins are glycosylated [2], with glycosylation required for production of functional proteins in many cases [3]. Also, an estimated 40% of therapeutic proteins produced by the pharmaceutical industry are glycosylated [4]. These issues demand consideration of alternative ways for producing structures from NMR data.

This need has not gone unnoticed and there are clearly examples of systems characterized by utilization of smaller data sets, particularly sets that are more easily acquired, such as backbone-only NMR data or data coming from selectively labeled methyl groups [5–9]. However, many of these applications still depend on uniform labeling to accomplish resonance assignments. As an alternative to uniform labeling, NMR active isotopes can be introduced by supplying labeled versions of single or multiple amino acid sources in expression media. One loses the ability to use conventional assignment strategies, but there are advantages to these sparse labeling strategies. First, mammalian cells can utilize these amino acids as biosynthetic building blocks and certain isotopically labeled amino acids are relatively inexpensive. Second, the resulting reduction in numbers of labeled sites improves resolution even in the absence of perdeuteration. And third, while constraints can still come from NOEs, particularly when perdeuteration allows measurement of longer range interactions, one can also capitalize on complementary long range structural constraints, such as residual dipolar couplings (RDCs) [10–12], pseudo-contact shifts (PCSs) [13–15], and paramagnetic relaxation enhancements (PREs) [16,17]. By combining these long range measurements with local structural

* Corresponding author. Address: Complex Carbohydrate Research Center, University of Georgia, 315 Riverbend Rd., Athens, GA 30602, United States.
E-mail address: jpresteg@ccrc.uga.edu (J.H. Prestegard).

1090-7807/$ - see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmr.2013.12.012

information from chemical shifts [18–20] and cross-correlated relaxation experiments [21] on sparsely labeled sites one can easily pursue structural characterization. The pursuit of structure does become much more dependent on computational methods to produce structures with acceptable precision. But there are clearly continuing advances in this area [6,8,15,22,23]. In this perspective, we do not intend to provide a comprehensive review of contributions in all of these areas, but choose a few examples to illustrate the potential for solving problems of high biological interest and some of the problems that may be encountered in implementing new approaches.

2. Discussion

2.1. Sparse labeling in large perdeuterated proteins

The first example relates to large proteins and protein complexes, one that stems from a collaboration between the Agard lab at UCSF and the Northeast Structural Genomics Consortium. TRAP1, the mitochondrial Hsp90 homolog, is a ~75 kDa molecular chaperone involved in protein unfolding/folding [24]. The chaperone has three major domains and functions as a dimer. There are crystal structures of full length bacterial homologs [25,26], yeast homologs [27], a canine ER homolog [28], and a recently determined structure of the zebrafish TRAP1 (Lavery et al., 2013, under revision), as well as a number of individual domain structures including those of the human cytosolic homologs [29–32]. However, structures of full length proteins remain rare and structural characterization of domain reorganization on client binding remains a challenge [33,34] (see Fig. 1A for a homology model of TRAP1). Expression in a bacterial host with perdeuteration is possible, but the dimer, even without a client present, pushes the resolution limits of uniform labeling approaches. Sparse labeling presents a reasonable option for problems like this, partly because one expects the structure of individual domains to be largely conserved in various states along a functional pathway, and one of the major hurdles for sparse labeling approaches, namely assignment of NMR resonances, can be overcome by assuming conservation of domain structure in the full length protein.

Assignment of sparsely labeled proteins can be approached by using a conserved structural motif to calculate NMR observables that are structure-dependent and measurable in a sparsely labeled system, and then matching these to back-calculated values. This approach can be taken with certain types of chemical shifts [20], PREs [35], PCSs [36], NOEs [37], and RDCs [38–40]. We choose to illustrate the viability of such approaches in the context of RDCs. RDCs are the residual of dipolar couplings, typically between pairs of directly bonded spin ½ nuclei (^{15}N–^{1}H or ^{13}C–^{1}H), that remain when the isotropy of motional averaging is disrupted by partial orientation of a molecule of interest. They are easily measured as additions to normal scalar couplings and they are highly dependent on structure. When distances between bonded pairs are assumed, a combination of structure and extent of order is represented in five independent parameters (alignment or order parameters), often defined in an arbitrarily selected molecular frame. These can be transformed to a principal order frame in which three parameters now describe the relationship of the original molecular frame to the principal order frame (Euler angles) and the remaining two parameters define molecular order (principal order and asymmetry). The alignment of principle order frames as determined for different domains can be a powerful constraint on the relative orientation of domains in multi-domain proteins or proteins in multi-protein complexes. An early demonstration [41] has expanded into numerous applications to the study of structure and motion in proteins over the years [42]. A particularly nice example appearing recently has used RDC data to help construct a model of how six Ig-like domains in actin binding filimin molecules fold into a three dimensional structure [43]. Most of these applications have used RDCs measured from uniformly labeled proteins, but large numbers of RDCs are not absolutely required, and there is great potential for coupling this type of application with sparse labeling strategies.

When the molecular structure of a domain is known and RDCs are assigned to specific bond vectors, order parameters can easily

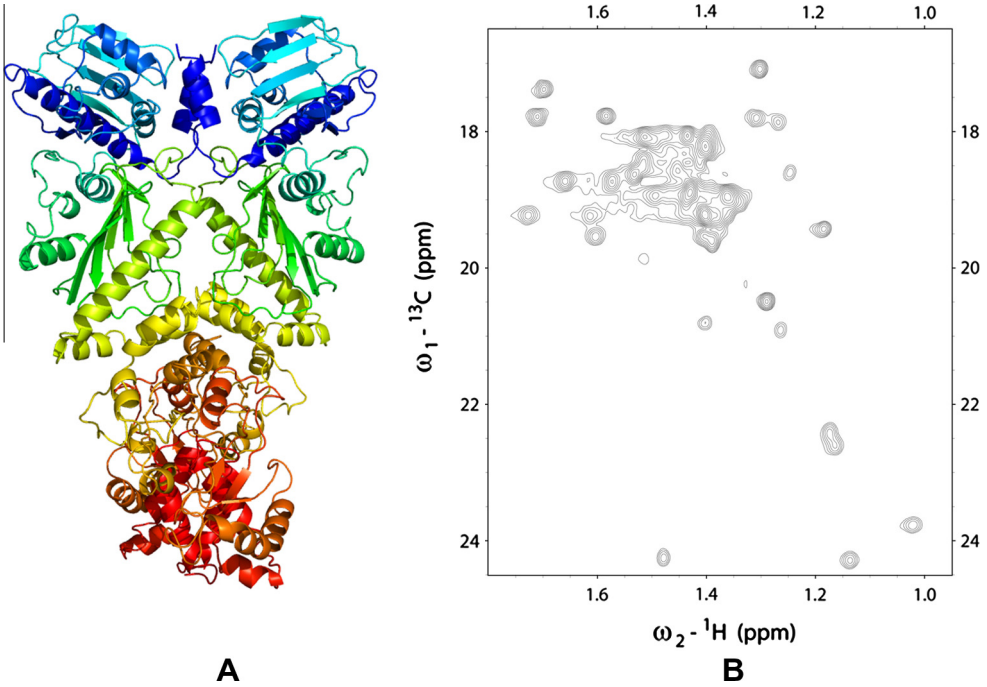

Fig. 1. Methyl-TROSY spectrum of TRAP1. (A) Model of the TRAP1 dimer. The structure was created as a monomer then aligned to the HSP90 dimer structure (2CG9) to mimic the dimer structure. (B) Methyl-TROSY spectrum of $^{13}C_{\beta}$–$^{1}H_{\beta}$ alanine labeled TRAP1.