Title

Sparse labeling of proteins: Structural characterization from long range constraints

Permalink

https://escholarship.org/uc/item/0nm5835v

Authors

Prestegard, James H
Agard, David A
Moremen, Kelley W
et al.

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], and **Kari Pederson**[a]

[a]Complex Carbohydrate Research Center, University of Georgia, Athens, GA 30602

[b]The Howard Hughes Medical Institute and the Dept. Biochem & Biophys, Univ. Calif. San Francisco, San Francisco, CA 94158

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.

Keywords

Protein NMR; Glycoprotein; Sparse Labeling; Paramagnetic constraints; Residual dipolar coupling; ligand docking; NMR assignments

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

*To whom correspondence should be addressed: J. H. Prestegard, Complex Carbohydrate Research Center, University of Georgia, 315 Riverbend Rd, Athens GA 30602, Phone: 706-542-6281, jpresteg@ccrc.uga.edu.

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 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.

Discussion

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 Figure 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 be determined from five or more non-degenerate RDCs. [44] When prior assignment of RDCs to specific bond vectors is not possible, parameters generated by calculation from an arbitrary assignment, or a search over parameter space, can be used to back-calculate RDCs for comparison to experiment. Scores generated from this comparison can be used to correctly associate RDCs with specific bond vectors, overcoming the assignment problem so often associated with sparse labeling..

For our illustration we have chosen to use ^{13}C-^{1}H alanine methyl labeling in a perdeuterated sample of human TRAP1 (hTRAP1). Methyl labeling has become a mainstay for structural investigation of large proteins that can be perdeuterated. Because of the unique relaxation properties of the methyl group, resolution is very high. It is usually implemented as isoleucine, leucine, valine (ILV) labeling[45], in which case methyl density is high enough to give a substantial number of NOEs.[9, 45] Some NOEs can be quite long range in an otherwise perdeuterated environment.[46] We chose alanine because the methyl is firmly anchored to the polypeptide backbone and ^{13}C-^{1}H RDCs, which have an effective vector along the Cα-Cβ bond due to rapid axial rotation of the methyl group, effectively constrain the Cα-Cβ angle providing backbone structural information.

Figure 1B illustrates the type of resolution that can be attained with the ~150 kDa TRAP1 system. The sample was perdeuterated by expression in E. coli using uniformly deuterated glucose in ^{2}H$_2$O. Supplementation with 200 mg ^{13}C-methyl, ^{2}H-α alanine per liter of minimal media shortly before induction of protein expression resulted in effective ^{13}C-^{1}H methyl labeling of alanines. The sample is 110μM in 70μL, pH 6.5, sodium phosphate buffer. Data were acquired at 900 MHz with a total acquisition time of 1hr. The spectrum

presented is a methyl-TROSY spectrum (essentially an HMQC experiment) in which the indirect dimension evolves as zero and two quantum coherences.[47] These coherences show coupling to just two of the three methyl protons resulting in a ^{13}C triplet. The central line of this triplet is extremely sharp due to interference of the two dipolar relaxation contributions. This narrowing of the central line, combined with the three protons contributing to intensity, results in very high sensitivity as well as resolution.

One expects 52 crosspeaks in the methyl-TROSY spectrum of hTRAP1, distributed as 19, 17, and 16 from the N-terminal, middle, and C-terminal domains, respectively; approximately 40 of these can be clearly resolved. Only an expanded central region of the spectrum is shown in Figure 1B. This contains ~30 identifiable crosspeaks. There are a few low intensity peaks here and outside the alanine region of the spectrum suggesting some scrambling of label to other amino acids. Scrambling to methyls of leucine, valine, and isoleucine at levels as high as 25% have been reported when using a higher level of supplementation (800 mg/L).[48] It appears that lower levels of supplementation result in a modest sacrifice in the extent of alanine labeling (65% at 200 mg),[48] but possibly a lower level of scrambling. It has also been shown that high levels of alanine labeling with minimal scrambling can be achieved by adding metabolic precursors of leucine, valine, and isoleucine to the medium.[48] While not available at this point, we expect that RDCs for methyl groups can be easily collected after selection of a suitable alignment medium. The number of RDCs expected in each domain, combined with connectivity constraints should be enough to define relative domain orientations and relative domain-domain motions, providing that the crosspeaks can be assigned to specific sites.

Assignment of a structurally characterized domain

Assignment strategies that exploit protein structural characteristics to predict NMR observables have a long history.[25, 37, 38] Here we illustrate an RDC-based process with a hypothetical example for a TRAP1 domain. The basic strategy is, of course, straightforward in that one back-calculates RDCs and matches these to an experimental set. Because of the necessity of evaluating five alignment parameters before any definitive structural or assignment information arises, one typically begins by making an arbitrary assignment of experimental RDCs to specific sites. First thoughts are to simply permute assignments and test all possibilities for agreement with observation. This computational task grows factorially with the number of RDCs. Even with the very efficient algorithms for order parameter determination[49] and back-calculation of RDCs (~100 ms/trial on a 2.26 GHz cpu) this quickly becomes impractical. For just 10 sites this corresponds to 4.2 days of computation. There have been a number of suggestions for more efficient search algorithms. [25, 39, 40] However, we present here a particularly simple approach that starts with a basic grid search.

Considering the five parameters that define domain orientation in the principal alignment frame to be three Euler angles (α, β, γ) and two order parameters (S_{ZZ}, S_{YY}), the scope of the search must initially be 0° to 180° for Euler angles and typically -10^{-3} to $+10^{-3}$ for S_{ZZ} and $-0.5S_{ZZ}$ to $-S_{ZZ}$ for S_{YY} (the latter is a consequence of the order tensor being traceless and the convention that $|S_{ZZ}| \geq |S_{YY}| \geq |S_{XX}|$). With large numbers of RDCs it may be possible to directly determine some of these parameters,[40] but then the number of assignment possibilities to test becomes prohibitive. A grid search over all parameters may not seem efficient either, however, the search can be performed iteratively beginning with a coarse grid and reducing the scope as well as grid size for one or more parameter at each iteration. An initial step size of no more than 5° for α and β proves necessary while γ can begin with 10° steps. Steps for S_{ZZ} and S_{YY} are initially 10^{-4} and $0.05S_{ZZ}$, respectively.

In the first stage of the procedure RDCs are calculated for each atom pair of interest (^{13}C and a ^{1}H pseudo-atom on the C_α-C_β axis projection for an alanine methyl group) at each combination of values over the ranges of the parameters. Atom positions can be taken from an NMR, X-ray or theoretical structure. The back-calculated sets of RDCs are then sorted into ascending numerical order and compared to the experimental RDCs sorted in the same manner. An RMSD between the experimental and back-calculated RDCs is calculated for each possible orientation. The data sets are then sorted by ascending RMSD. The lowest 1 to 10% (depending on number of parameter sets, > or < 10,000) are then separated in five RMSD-parameter pairs, normalized with respect to the maximum RMSD in each set, and a cluster with the low RMSDs and high solution density is used to redefine the search range and step size for the next iteration. Once no further improvement can be made in RMSD, the order and orientation parameters have been determined. Back calculated RDCs are then compared to experimental values. Any experimental RDC that has a single back-calculated RDC within experimental error of the back-calculated value is considered assigned. Assignments are typically not complete but unassigned RDCs can often be grouped into pairs or triplets of ambiguous assignments. The length of time required to reach this point for a 10 RDC set is typically less than 2 hrs.

Simulated data for the alanine methyl (^{13}C$^\beta$) labeled hTRAP1 C-terminus was used to illustrate this procedure. This domain contains 16 alanines and although no structure currently exists for the human protein it is highly homologous to other Hsp90s. Using the online server I-TASSER [50, 51] we obtained the structure shown in Figure 1A. A set of synthetic RDCs with a range of −8 to +5 HZ and random errors of up to ± 0.5 Hz were generated for this structure using REDCAT.[52] Representative results of the iterative grid search for one of the parameters (Euler angle α) can be seen in Figure 2. The black dots are the best results (lowest 1% of RMSDs) from the initial grid search. Red dots are the best results from the first refinement while green dots are from the second refinement. Best results from the final grid search are indicated by blue dots. The inset plot shows an expansion of the results and how well the system refined down to a single value as the ranges narrowed. Using the criteria described above for assessment of a match of back-calculated to experimental (simulated) RDC data, six residues (38% of the system under investigation) were unambiguously assigned and 14 of the 16 residues (88%) had an ambiguity of two or less.

Supplementing RDCs with other data can easily remove ambiguities in assignments. Perdeuterated proteins can, for example, allow measurement of NOE connectivities up to as much as 10Å.[46] Assuming that we could observe NOEs up to 6Å (Figure 3), pairwise association of sites and their NOEs would allow all but two of the residues to be unambiguously assigned. Calculating chemical shifts based on structure can also help to reduce ambiguities.[20] The use of simulated RDCs with random errors of +/− 0.5 Hz and a range of −8 to +5 Hz, as a substitute for actual experimental data in the above example, corresponds to an assumption that the domain structure is quite accurate and that data susceptible to conformational averaging have been excluded. Where experimental structures of individual domains are available, and spin relaxation or chemical shift data allow such exclusions, this assumption is likely to be valid. When structures are calculated from homology models validity may be in question. Some sense of the required accuracy can be achieved by examination of an ensemble of NMR structures. One with a 0.8 Å RMSD over backbone atoms of well defined secondary structure elements shows an RMSD over predicted RDCs for alanines in those elements of about 6% of the RDC range.

Despite the predicted success for RDC-based assignment for sets of RDCs, the size of those found in a TRAP1 domain, the problem of clustering RDCs domain by domain in large proteins remains. There are strategies based on a divide and conquer approach (expressing

domains individually and counting on overlap of crosspeaks in domain samples with crosspeaks from the intact protein). This has been shown to be effective for the cytosolic Hsp90β.[53] An alternate clustering strategy relies on the introduction of paramagnetic tags that can perturb crosspeaks in a distance dependent way (usually nitroxides on TEMPO groups linked covalently to exposed cysteines,[17, 54] but also lanthanide chelates[14, 55, 56], including some pioneering work from the Griesinger lab[57], and inserted lanthanide binding peptide sequences[58, 59]). The effects of PREs or PCSs can be seen out to 30Å or more, well within the dimensions of a typical domain. Strategically chosen paramagnetic sites should, in many cases, allow clustering of sites into those belonging to a single domain. The direct information provided on distances and angles within domains, of course, add additional information useful in removing assignment ambiguities as well as in structure refinement.

Sparse labeling of glycosylated proteins

Large proteins composed of structurally well defined domains are not the only systems benefitting from strategies involving sparse labeling combined with long range constraints. There are systems of moderate size that cannot benefit from the improved resolution provided by perdeuteration, simply because the required expression hosts used for isotope labeling do not tolerate growth in deuterated media. While there are classes where proper formation of disulfide bonds is an issue, the major class is represented by glycoproteins. These proteins are characterized by the addition of sugar chains (glycans) to polypeptides by covalent linkage to asparagines (N-glycans) or threonines/serines (O-glycans). Glycosylation is not a rare event for mammals and other Eukaryotes. Glycosylated proteins may even represent the majority of proteins in these systems. In many cases glycosylation is required for proper folding, stability, or function, and therefore is not dispensable.

Baculovirus systems have become a frequently used host for the production of glycosylated proteins.[60] If specific glycosylation is necessary, however, expression in mammalian hosts is required.[61] Isotopic labeling in mammalian cells, as well as the insect cells used in baculovirus systems, typically require supplementation with isotopically labeled amino acids. While a complete mix of amino acids can prove expensive, supplementation with one amino acid, or a select set of amino acids is not. Significant advances in this area have been made over the last few years.[62] We illustrate here one specific advance using HEK-293 cells in a suspension medium.[63] The details involving isotope labeling for NMR have been documented previously.[64] However, briefly, the protocol involves using a commercial drop-out medium supplemented with one or more of the amino acids carrying the desired isotope labels. Cells are transiently transfected with an expression vector as a polyamine complex that facilitates uptake and transport into the cytosol of the host cells. The vector includes a secretion signal sequence, a His-tag, an "AviTag" sequence for in vitro biotinylation, a "superfolder" GFP domain, a site cleavable with TEV protease and the sequence for the target protein.[64] Secreted protein is isolated from the medium by Ni^{2+}-NTA chromatography, followed by TEV protease cleavage, and Ni^{2+}-NTA column separation of the target protein from GFP. Yields in excess of 100 mg protein/L have been achieved for several proteins.

The type of data achievable is illustrated in Figure 4A. This is an HSQC spectrum of the glycosyltransferase, ST6Gal1 in which all phenylalanines have been labeled with ^{15}N. There are 16 phenylalanine sites in the protein and 16 crosspeaks are resolved in the figure, despite the rather large size (36 kDa not counting glycans). ST6Gal1 is the enzyme that adds sialic acid in an α2,6- linkage to galactose residues at the termini of complex glycans. Sialic acid termination of glycans is a hallmark of mammalian glycosylation.[65] High levels of sialylation correlate with many types of cancer and α2,6- versus α2,3-sialylation in the upper airways of humans is what distinguishes initial targets for infection by human versus

bird flu viruses. Hence, there is considerable interest in the way mammalian enzymes target specific glycans for α2,6-sialic acid addition.

Ligand placement using sparse labels

Our initial interest in sparse isotopic labeling of glycosylated proteins was motivated by the prospect of structure determination of previously uncharacterized proteins such as ST6Gal1. Despite decades of effort, crystallography had failed to produce a structure of this particular glycosylated protein. With successful sparse labeling, the prospect of structure determination using a combination of RDCs and long range distance constraints from paramagnetic sites was appealing. Introduction of paramagnetic sites through the design of protein constructs with single reactive cysteines proved challenging because of the structural and functional importance of native cysteines. Consequently, collecting data through the use paramagnetically tagged ligands, and using the data to screen and refine homology models, became the preferred option. Paramagnetically labeled (spin labeled) versions of both the sugar donor (CMP-sialic acid) and an acceptor (N-acetyllactosamine, LacNAc) were synthesized, replacing the carboxylated sialic acid with carboxy-TEMPO in the first case and adding a TEMPO ring to the reducing end of LacNAc in the second case[66] (see Figure 5).

Enhanced relaxation from bound ligands during INEPT transfer and refocusing periods of a standard HSQC experiment results in crosspeak intensity reduction in proportion to $\exp(-C\tau/r^6)$ where τ is the total time that proton magnetization is transverse in an HSQC experiment, C is a collection of constants dependent on spin properties, dynamics, and fractional occupation of ligand sites, and r is the distance between the unpaired electron position and various ^{15}N labeled phenylalanine sites. The resulting spectral perturbations are presented in Figure 4 and 6 for the two ligand analogs. For CMP-carboxyTEMPO the near disappearance of two crosspeaks suggest that two phenylalanines are very close to the donor nitroxide and the reduction in intensity of two additional crosspeaks indicates that two more are at intermediate distances. The addition of LacNAc-TEMPO produced a marginal reduction in intensity of the two most strongly CMP-carboxyTEMPO perturbed crosspeaks. Most striking, however, was a more visible reduction of a crosspeak that was not reduced at all in the presence of CMP-carboxyTEMPO.

At the time of data collection there were structures of bacterial α2,3- and α2,3/8 sialyltransferases (2IHZ, 1RO7) with one of the products of the sialylation reaction, CMP, in the donor binding site. While it was possible to make structures where two phenylalanines were close to a superimposed CMP-carboxyTEMPO, and even dock LacNAc-TEMPO so that a single additional phenylalanine was close to its TEMPO residue, the low sequence identity of the bacterial transferases prevented the selection of a definitive homology model. A few months ago, two structures of ST6Gal1 appeared, one on the rat protein by Meng et al. (4MPS) and one on the human protein by Kuhn et al.[67] (4JS1/2). This presents an opportunity to use the original paramagnetic perturbation data[68, 69] for a different purpose, namely placing ligands in the active site of ST6Gal1. In particular the structure of the rat protein does not contain either donor or acceptor ligands; the human protein contains a nucleotide product (CMP) and a glycan from a neighboring molecule appears to mimic acceptor placement. The paramagnetic data can now guide placement of donor and acceptor ligands in the rat ST6Gal1 structure and evaluate the placement of the acceptor in the human ST6Gal1 structure.

Extracting quantitative distance constraints from the original data is not straightforward. Unfortunately there are other potential ligand induced contributions to transverse relaxation times that can reduce crosspeak intensities (exchange broadening due to on and off processes for the ligands, for example). The existence of these contributions was verified in

the original work using a non-hydrolyzable donor analog in place of CMP-TEMPO and observing crosspeak reductions of the two most strongly CMP-carboxyTEMPO perturbed peaks. These reductions could come from exchange of ligands or from protein dynamics on an intermediate timescale (0.01–10.0ms). Interestingly, a loop near the binding pocket containing two phenylalanines is disordered and not observable in the recent rat ST6Gal1 crystal structure. To circumvent these complications, the original work also collected data using reduced and oxidized versions of CMP-carboxyTEMPO (reduction accomplished with ascorbic acid). Comparing reduced and oxidized spectra and assuming dynamics remained unchanged on reduction, one can use the distance dependence of spin-spin relaxation ($1/r^6$) and exponential loss during HSQC transfer and refocusing periods to estimate distances. Ideally the additional dependence of relaxation rates on dynamics and spin properties is removed using a calibration distance, possibly by observing interactions of the nitroxide with a nuclear spin site in the CMP molecule. However, in the absence of an isotopically enriched site and knowledge of the ligand conformation, a direct calculation using estimates for molecular tumbling times and fundamental nuclear and electron spin properties was used.[69] Three distances between the TEMPO nitroxide and amide protons of phenylalanine amide protons were determined (11 ± 2 Å, $\sim13\pm3$ Å, and 15 ± 3 Å). A fourth could not be quantitated because of the complete absence of the peak under oxidized conditions. However, it is likely the fourth distance is shorter than 11Å. A re-analysis of intensity reductions on adding LacNAc-TEMPO to ST6Gal1 already saturated with a non-hydrolyzable version of the donor suggests distances of 20 ± 5Å, 20 ± 5Å and 15 ± 4Å between the LacNAc-TEMPO nitroxide and the amide protons of the two previously perturbed and one newly perturbed phenylalanine residues.

The derived distances were now used in a data restrained docking protocol using the rat crystal structure (residues 95–403, Uniprot P13721) and the software package HADDOCK. [70, 71] In our case assignments of crosspeaks to specific phenylalanine residues were not available, but HADDOCK allows paramagnetic constraints to be treated as ambiguous constraints involving small sets of possible assignments. Possible assignments, as well as some additional residue contact information for the CMP portion of the donor analog were derived by superimposing the rat structure on the structure of the ST3Gal1 pig protein that has CMP in its donor binding site and appeared shortly after collection of our perturbation data.[72] This protein had just two recognizable homologous segments (62% identity over 47 residues and 48% identity over 23 residues), but these segments are in the conserved Rossmann fold common to many nucleotide binding sites. We identified two residues in ST6Gal1 that are likely to be involved in cytosine contact as well as two that are likely to make contact with the ribose and phosphate of CMP respectively. These residues were entered as HADDOCK active residues as if they had been identified by chemical shift perturbation of assigned residue crosspeaks. The phenylalanines likely experiencing paramagnetic perturbations from the CMP-carboxyTEMPO were identified as PHE357 and PHE356 for the short distances and PHE340 and PHE208 were identified for the long distances. For LacNAc-TEMPO the long distances were assigned to PHE357 and PHE356. The only possibility for the additional phenylalanine perturbed by LacNAc-TEMPO was PHE371. The distance between the LacNAc O6 oxygen the Gal residue and the C1 carboxylated carbon of the TEMPO moiety of the CMP-carboxyTEMPO was constrained at 4 ± 1Å to mimic the attack geometry of the SN2 reaction leading to the α2,6-glycosidic bond between donor and acceptor. The missing loop in the rat structure was added to the ST6Gal1 structure using MODELLER[73] and this loop, as well as a segment of a helix near the entrance to the active site were designated as fully flexible. 1000 poses were generated in HADDOCK; the top 200 were subjected to initial simulated annealing refinement as well as a final water refinement step. Poses were clustered and the low energy members of the top clusters were evaluated.

The ligands as docked into the ST6Gal1 active site of the model with the fifth best score and the second lowest molecular energy are depicted in Figure 7. The model has substantial favorable electrostatic and van der Waals contributions. As expected, the CMP-donor mimic is buried at the bottom of a long channel in a place consistent with both the ST3Gal1 structure used to select constraints and probable phenylalanine contacts. It is also in good agreement with the positioning of the CMP product in the human ST6Gal1 structure (a near identical positioning of the phosphate and a 5.7 Å RMSD over cytosine, ribose and phosphate core atoms when the nucleotide binding elements of the proteins are overlaid). The LacNAc portion of our acceptor mimic is at the top of the channel with the O6 oxygen of the Gal residue positioned for attack on the carboxylated ring carbon of carboxyTEMPO (the mimic of the C2 atom in sialic acid). There is less consistency with the positioning of the other atoms of the GlcNAc and Gal residues of LacNAc (7.9Å RMSD). However these residues may be constrained in the human structure by crystal packing and their covalent attachment to glycosylation sites in adjacent molecules. It is also possible that the hydrophobic TEMPO moiety influences placement. The TEMPO moiety can be seen to occupy a well-defined cavity in an area where a mannose residue may normally reside. The GlcNAc also occupies a well-defined cavity showing contacts with Tyr119, Asp271, Ala365, and Tyr272. These latter contacts could explain the preference for the N-acetylated glucose of LacNAc as opposed to the free glucose of lactose.

The general strategies described above for placement of ligands are of course applicable to other sparsely labeled glycosylated enzymes. The strategies can clearly be improved, possibly by reversing strategy, paramagnetically tagging the proteins with some of the novel reagents being developed, and observing perturbations on NMR active isotopes in the ligands. Applications of either strategy would take an important step toward explaining, and possibly manipulating, the donor specificities of this important class of enzymes.

Prospects for structure determination of sparsely labeled proteins

Not all systems of interest will have experimental domain structures available or even good homology models for species of interest. Assignment of crosspeaks without the aid of a preexisting structure then becomes an issue. We have been working on a strategy that uses amide exchange rates to correlate crosspeaks in native protein spectra with crosspeaks in denatured spectra of whole proteins or isolated peptides where assignments are more straightforward.[74] This has proven successful for small proteins, but major hurdles remain for larger proteins and proteins that contain disulfide linkages that must be disrupted in the denaturation process. There is also the prospect of simultaneously carrying out assignments and structure determination, something that may be feasible with sufficient structural constraints.[39, 75, 76]

Whether crosspeaks are assigned independent of structure or based on reliable models for whole proteins or individual domains, subsequent structural characterization is going to be increasingly dependent on computational methods for protein structure prediction. Fortunately there have been significant strides in the structure prediction area. At the moment, the data used in most cases is more extensive than the very sparse RDCs, PCSs and PREs discussed above. In general some backbone NOEs, NOEs from methyl labeling, and torsion angles from backbone chemical shifts are used along with RDC, PCS, or PRE data. There have also been tremendous advances in using fragment assembly strategies and incorporating at least some experimental data in the fragment selection process. Limits to the size of molecules amenable to analysis have improved from around 25 kDa a few years ago[77] to over 40 kDa at this time.[8] It is clear at this point that for moderate resolution structures (~3Å backbone structures), extensive side chain assignments will not be essential. Placement of sidechain atoms, at least in the well packed center of proteins, can now be done with reasonable accuracy allowing a focus on more easily acquired backbone centered

NMR data.[77] Thus, the prospects for contributions of NMR to structural characterization of larger proteins and proteins difficult to express in bacterial hosts remain high. We can expect both high quality structures of whole proteins and efficient means for investigation bound ligand geometries to emerge. This will lay the groundwork for structural and functional analysis of many biologically important proteins.

Acknowledgments

References

1. Kay LE. NMR studies of protein structure and dynamics. Journal of Magnetic Resonance. 2005; 173:193–207. [PubMed: 15780912]

2. Apweiler R, Hermjakob H, Sharon N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. Biochimica Et Biophysica Acta-General Subjects. 1999; 1473:4–8.

3. Moremen KW, Tiemeyer M, Nairn AV. Vertebrate protein glycosylation: diversity, synthesis and function. Nature Reviews Molecular Cell Biology. 2012; 13:448–462.

4. Walsh G. Post-translational modifications of protein biopharmaceuticals. Drug Discovery Today. 2010; 15:773–780. [PubMed: 20599624]

5. Campos M, Francetic O, Nilges M. Modeling pilus structures from sparse data. Journal of Structural Biology. 2011; 173:436–444. [PubMed: 21115127]

6. Ganguly S, Weiner BE, Meiler J. Membrane Protein Structure Determination using Paramagnetic Tags. Structure. 2011; 19:441–443. [PubMed: 21481766]

7. Karaca E, Bonvin A. Advances in integrative modeling of biomolecular complexes. Methods. 2013; 59:372–381. [PubMed: 23267861]

8. Lange OF, Rossi P, Sgourakis NG, Song YF, Lee HW, Aramini JM, Ertekin A, Xiao R, Acton TB, Montelione GT, Baker D. Determination of solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109:10873–10878. [PubMed: 22733734]

9. Ruschak AM, Kay LE. Methyl groups as probes of supra-molecular structure, dynamics and function. Journal of Biomolecular NMR. 2010; 46:75–87. [PubMed: 19784810]

10. Chen, K.; Tjandra, N. The Use of Residual Dipolar Coupling in Studying Proteins by NMR. In: Zhu, G., editor. Nmr of Proteins and Small Biomolecules. 2012. p. 47-67.

11. Prestegard, JH.; Mayer, KL.; Valafar, H.; Benison, GC. Determination of protein backbone structures from residual dipolar couplings. In: James, TL., editor. Nuclear Magnetic Resonance of Biological Macromolecules, Part C. 2005. p. 175-+

12. Fitzkee NC, Bax A. Facile measurement of H-1-N-15 residual dipolar couplings in larger perdeuterated proteins. Journal of Biomolecular NMR. 2010; 48:65–70. [PubMed: 20694505]

13. Allegrozzi M, Bertini I, Janik MBL, Lee YM, Lin GH, Luchinat C. Lanthanide-induced pseudocontact shifts for solution structure refinements of macromolecules in shells up to 40 angstrom from the metal ion. Journal of the American Chemical Society. 2000; 122:4154–4161.

14. Yagi H, Pilla KB, Maleckis A, Graham B, Huber T, Otting G. Three-Dimensional Protein Fold Determination from Backbone Amide Pseudocontact Shifts Generated by Lanthanide Tags at Multiple Sites. Structure. 2013; 21:883–890. [PubMed: 23643949]

15. Schmitz C, Vernon R, Otting G, Baker D, Huber T. Protein Structure Determination from Pseudocontact Shifts Using ROSETTA. Journal of Molecular Biology. 2012; 416:668–677. [PubMed: 22285518]

16. Clore GM. Exploring sparsely populated states of macromolecules by diamagnetic and paramagnetic NMR relaxation. Protein Science. 2011; 20:229–246. [PubMed: 21280116]

17. Battiste JL, Wagner G. Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. Biochemistry. 2000; 39:5355–5365. [PubMed: 10820006]

18. van der Schot G, Zhang ZY, Vernon R, Shen Y, Vranken WF, Baker D, Bonvin A, Lange OF. Improving 3D structure prediction from chemical shift data. Journal of Biomolecular NMR. 2013; 57:27–35. [PubMed: 23912841]

19. Han B, Liu YF, Ginzinger SW, Wishart DS. SHIFTX2: significantly improved protein chemical shift prediction. Journal of Biomolecular NMR. 2011; 50:43–57. [PubMed: 21448735]

20. Sahakyan AB, Vranken WF, Cavalli A, Vendruscolo M. Structure-based prediction of methyl chemical shifts in proteins. Journal of Biomolecular NMR. 2011; 50:331–346. [PubMed: 21748266]

21. Reif B, Hennig M, Griesinger C. Direct measurement of angles between bond vectors in high-resolution NMR. Science. 1997; 276:1230–1233. [PubMed: 9157875]

22. Hirst SJ, Alexander N, McHaourab HS, Meiler J. RosettaEPR: An integrated tool for protein structure determination from sparse EPR data. Journal of Structural Biology. 2011; 173:506–514. [PubMed: 21029778]

23. Steiner D, van Gunsteren WF. An improved structural characterisation of reduced French bean plastocyanin based on NMR data and local-elevation molecular dynamics simulation. European Biophysics Journal with Biophysics Letters. 2012; 41:579–595. [PubMed: 22706892]

24. Mayer MP. Gymnastics of Molecular Chaperones. Molecular Cell. 2010; 39:321–331. [PubMed: 20705236]

25. Pintacuda G, Keniry MA, Huber T, Park AY, Dixon NE, Otting G. Fast structure-based assignment of N-15 HSQC spectra of selectively N-15-labeled paramagnetic proteins. Journal of the American Chemical Society. 2004; 126:2963–2970. [PubMed: 14995214]

26. Shiau AK, Harris SF, Southworth DR, Agard DA. Structural Analysis of E. coli hsp90 Reveals Dramatic Nucleotide-Dependent Conformational Rearrangements. Cell. 2006; 127:329–340. [PubMed: 17055434]

27. Ali MMU, Roe SM, Vaughan CK, Meyer P, Panaretou B, Piper PW, Prodromou C, Pearl LH. Crystal structure of an Hsp90–nucleotide–p23/Sba1 closed chaperone complex. Nature. 2006; 440:1013–1017. [PubMed: 16625188]

28. Dollins DE, Warren JJ, Immormino RM, Gewirth DT. Structures of GRP94-Nucleotide Complexes Reveal Mechanistic Differences between the hsp90 Chaperones. Molecular Cell. 2007; 28:41–56. [PubMed: 17936703]

29. Fogliatto G, Gianellini LM, Brasca MG, Casale E, Ballinari D, Ciomei M, Degrassi A, De Ponti A, Germani M, Guanci M, Paolucci M, Polucci P, Russo M, Sola F, Valsasina B, Visco C, Zuccotto F, Donati D, Felder E, Pesenti E, Mantegani S, Galvani A, Isacchi A. NMS-E973, a Novel Synthetic Inhibitor of Hsp90 with Activity against Multiple Models of Drug Resistance to Targeted Agents, Including Intracranial Metastases. Clinical Cancer Research. 2013; 19:3520–3532. [PubMed: 23674492]

30. Davies NG, Browne H, Davis B, Drysdale MJ, Foloppe N, Geoffrey S, Gibbons B, Hart T, Hubbard R, Jensen MR, Mansell H, Massey A, Matassova N, Moore JD, Murray J, Pratt R, Ray S, Robertson A, Roughley SD, Schoepfer J, Scriven K, Simmonite H, Stokes S, Surgenor A, Webb P, Wood M, Wright L, Brough P. Targeting conserved water molecules: design of 4-aryl-5-cyanopyrrolo[2,3-d]pyrimidine Hsp90 inhibitors using fragment-based screening and structure-based optimization. Bioorganic & Medicinal Chemistry. 2012; 20:6770–6789. [PubMed: 23018093]

31. Bussenius J, Blazey CM, Aay N, Anand NK, Arcalas A, Baik T, Bowles OJ, Buhr CA, Costanzo S, Curtis JK, Defina SC, Dubenko L, Heuer TS, Huang P, Jaeger C, Joshi A, Kennedy AR, Kim AI, Lara K, Lee J, Li J, Lougheed JC, Ma S, Malek S, Manalo JC, Martini JF, Mcgrath G, Nicoll M, Nuss JM, Pack M, Peto CJ, Tsang TH, Wang L, Womble SW, Yakes M, Zhang W, Rice KD. Discovery of XL888: a novel tropane-derived small molecule inhibitor of HSP90. Bioorganic & Medicinal Chemistry Letters. 2012; 22:5396–5405. [PubMed: 22877636]

32. Yun TJ, Harning EK, Giza K, Rabah D, Li P, Arndt JW, Luchetti D, Biamonte MA, Shi J, Lundgren K, Manning A, Kehry MR. EC144, a synthetic inhibitor of heat shock protein 90, blocks

innate and adaptive immune responses in models of inflammation and autoimmunity. Journal of Immunology. 2011; 186:563–575.

33. Didenko T, Duarte AMS, Karagöz GE, Rüdiger SGD. Hsp90 structure and function studied by NMR spectro, Biochimica et Biophysica Acta. Molecular Cell Research. 2012; 1823:636–647.

34. Pearl LH, Prodromou C. Structure and Mechanism of the Hsp90 Molecular Chaperone Machinery. Annual Review of Biochemistry. 2006; 75:271–294.

35. Venditti V, Fawzi NL, Clore GM. Automated sequence- and stereo-specific assignment of methyl-labeled proteins by paramagnetic relaxation and methyl-methyl nuclear overhauser enhancement spectroscopy. Journal of Biomolecular NMR. 2011; 51:319–328. [PubMed: 21935714]

36. Skinner SP, Moshev M, Hass MAS, Ubbink M. PARAssign-paramagnetic NMR assignments of protein nuclei on the basis of pseudocontact shifts. Journal of Biomolecular NMR. 2013; 55:379–389. [PubMed: 23526169]

37. Xu YQ, Matthews S. MAP-XSII: an improved program for the automatic assignment of methyl resonances in large proteins. Journal of Biomolecular Nmr. 2013; 55:179–187. [PubMed: 23292498]

38. Hus JC, Prompers JJ, Bruschweiler R. Assignment strategy for proteins with known structure. Journal of Magnetic Resonance. 2002; 157:119–123. [PubMed: 12202140]

39. Wang XS, Tash B, Flanagan JM, Tian F. RDC derived protein backbone resonance assignment using fragment assembly. Journal of Biomolecular NMR. 2011; 49:85–98. [PubMed: 21191805]

40. Shealy P, Liu YZ, Simin M, Valafar H. Backbone resonance assignment and order tensor estimation using residual dipolar couplings. Journal of Biomolecular NMR. 2011; 50:357–369. [PubMed: 21667298]

41. Fischer MWF, Losonczi JA, Weaver JL, Prestegard JH. Domain orientation and dynamics in multidomain proteins from residual dipolar couplings. Biochemistry. 1999; 38:9013–9022. [PubMed: 10413474]

42. Jensen MR, Ortega-Roldan JL, Salmon L, van Nuland N, Blackledge M. Characterizing weak protein-protein complexes by NMR residual dipolar couplings. European Biophysics Journal with Biophysics Letters. 2011; 40:1371–1381. [PubMed: 21710303]

43. Tossavainen H, Koskela O, Jiang PJ, Ylanne J, Campbell ID, Kilpelainen I, Permi P. Model of a Six Immunoglobulin-Like Domain Fragment of Filamin A (16–21) Built Using Residual Dipolar Couplings. Journal of the American Chemical Society. 2012; 134:6660–6672. [PubMed: 22452512]

44. Prestegard JH, Bougault CM, Kishore AI. Residual dipolar couplings in structure determination of biomolecules. Chemical Reviews. 2004; 104:3519–3540. [PubMed: 15303825]

45. Goto NK, Gardner KH, Mueller GA, Willis RC, Kay LE. A robust and cost-effective method for the production of Val, Leu, Ile (delta 1) methyl-protonated N-15-, C-13-, H-2-labeled proteins. Journal of Biomolecular NMR. 1999; 13:369–374. [PubMed: 10383198]

46. Godoy-Ruiz R, Guo CY, Tugarinov V. Alanine Methyl Groups as NMR Probes of Molecular Structure and Dynamics in High-Molecular-Weight Proteins. Journal of the American Chemical Society. 2010; 132:18340–18350. [PubMed: 21138300]

47. Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE. Cross-correlated relaxation enhanced H-1-C-13 NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. Journal of the American Chemical Society. 2003; 125:10420–10428. [PubMed: 12926967]

48. Ayala I, Sounier R, Use N, Gans P, Boisbouvier J. An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein. Journal of Biomolecular Nmr. 2009; 43:111–119. [PubMed: 19115043]

49. Losonczi JA, Andrec M, Fischer MWF, Prestegard JH. Order matrix analysis of residual dipolar couplings using singular value decomposition. Journal of Magnetic Resonance. 1999; 138:334–342. [PubMed: 10341140]

50. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols. 2010; 5:725–738.

51. Zhang Y. I-TASSER server for protein 3D structure prediction. Bmc Bioinformatics. 2008; 9:8. [PubMed: 18179724]

52. Valafar H, Prestegard JH. REDCAT: a residual dipolar coupling analysis tool. Journal of Magnetic Resonance. 2004; 167:228–241. [PubMed: 15040978]

53. Karagöz GE, Duarte AMS, Ippel H, Uetrecht C, Sinnige T, Rosmalen Mv, Hausmann J, Heck AJR, Boelens R, Rüdiger SGD. N-terminal domain of human Hsp90 triggers binding to the cochaperone p23. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108:580–585. [PubMed: 21183720]

54. Liu YZ, Kahn RA, Prestegard JH. Dynamic structure of membrane-anchored Arf center dot GTP. Nature Structural & Molecular Biology. 2010; 17 876-U128.

55. Peters F, Maestre-Martinez M, Leonov A, Kovacic L, Becker S, Boelens R, Griesinger C. Cys-Ph-TAHA: a lanthanide binding tag for RDC and PCS enhanced protein NMR. Journal of Biomolecular Nmr. 2011; 51:329–337. [PubMed: 21892794]

56. Haussinger D, Huang JR, Grzesiek S. DOTA-M8: An Extremely Rigid, High-Affinity Lanthanide Chelating Tag for PCS NMR Spectroscopy. Journal of the American Chemical Society. 2009; 131:14761–14767. [PubMed: 19785413]

57. Ikegami T, Verdier L, Sakhaii P, Grimme S, Pescatore B, Saxena K, Fiebig KM, Griesinger C. Novel techniques for weak alignment of proteins in solution using chemical tags coordinating lanthanide ions. Journal of Biomolecular Nmr. 2004; 29:339–349. [PubMed: 15213432]

58. Allen KN, Imperiali B. Lanthanide-tagged proteins - an illuminating partnership. Current Opinion in Chemical Biology. 2010; 14:247–254. [PubMed: 20102793]

59. Barb AW, Ho TG, Flanagan-Steet H, Prestegard JH. Lanthanide binding and IgG affinity construct: potential applications in solution NMR, MRI and luminescence microscopy. Protein Science. 2012; 21:1456–1466. [PubMed: 22851279]

60. Saxena, K.; Dutta, A.; Klein-Seetharaman, J.; Schwalbe, H. Isotope Labeling in Insect Cells. In: Shekhtman, A.; Burz, DS., editors. Protein NMR Techniques. Third Edition. 2012. p. 37-54.

61. Dutta, A.; Saxena, K.; Schwalbe, H.; Klein-Seetharaman, J. Isotope Labeling in Mammalian Cells. In: Shekhtman, A.; Burz, DS., editors. Protein NMR Techniques. 2012. p. 55-69.

62. Gossert AD, Hinniger A, Gutmann S, Jahnke W, Strauss A, Fernandez C. A simple protocol for amino acid type selective isotope labeling in insect cells with improved yields and high reproducibility. Journal of Biomolecular NMR. 2011; 51:449–456. [PubMed: 21964698]

63. Backliwal G, Hildinger M, Chenuet S, Wulhfard S, De Jesus M, Wurm FM. Rational vector design and multi-pathway modulation of HEK 293E cells yield recombinant antibody titers exceeding 1 g/l by transient transfection under serum-free conditions. Nucleic acids research. 2008; 36:e96. [PubMed: 18617574]

64. Barb AW, Meng L, Gao Z, Johnson RW, Moremen KW, Prestegard JH. NMR Characterization of Immunoglobulin G Fc Glycan Motion on Enzymatic Sialylation. Biochemistry. 2012; 51:4618–4626. [PubMed: 22574931]

65. Varki, NM.; Strobert, E.; Dick, EJ.; Benirschke, K.; Varki, A. Biomedical Differences Between Human and Nonhuman Hominids: Potential Roles for Uniquely Human Aspects of Sialic Acid Biology. In: Abbas, AK.; Galli, SJ.; Howley, PM., editors. Annual Review of Pathology: Mechanisms of Disease. Vol. Vol 6. 2011. p. 365-393.

66. Jain NU, Venot A, Umemoto K, Leffler H, Prestegard JH. Distance mapping of protein-binding sites using spin-labeled oligosaccharide ligands. Protein Science. 2001; 10:2393–2400. [PubMed: 11604544]

67. Kuhn B, Benz J, Greif M, Engel AM, Sobek H, Rudolph MG. S.D.-. Journal: (2013) Acta Crystallogr., The structure of human [alpha]-2,6-sialyltransferase reveals the binding mode of complex glycans Acta Crystallogr., Sect.D. 2013; 69:1826–1838.

68. Liu S, Meng L, Moremen KW, Prestegard JH. Nuclear Magnetic Resonance Structural Characterization of Substrates Bound to the alpha-2,6-Sialyltransferase, ST6Gal-I. Biochemistry. 2009; 48:11211–11219. [PubMed: 19845399]

69. Liu S, Venot A, Meng L, Tian F, Moremen KW, Boons GJ, Prestegard JH. Spin-labeled analogs of CMP-NeuAc as NMR probes of the alpha-2,6-sialyltransferase ST6Gal I. Chemistry & Biology. 2007; 14:409–418. [PubMed: 17462576]

70. de Vries SJ, van Dijk M, Bonvin AMJJ. The HADDOCK web server for data-driven biomolecular docking. Nat. Protoc. 2010; 5:883–897. [PubMed: 20431534]

71. Dominguez C, Boelens R, Bonvin AMJJ. HADDOCK: A Protein–Protein Docking Approach Based on Biochemical or Biophysical Information. Journal of the American Chemical Society. 2003; 125:1731–1737. [PubMed: 12580598]

72. Rao FV, Rich JR, Rakic B, Buddai S, Schwartz MF, Johnson K, Bowe C, Wakarchuk WW, DeFrees S, Withers SG, Strynadka NCJ. Structural insight into mammalian sialyltransferases. Nature Structural & Molecular Biology. 2009; 16:1186–1188.

73. Yang Z, Lasker K, Schneidman-Duhovny D, Webb B, Huang CC, Pettersen EF, Goddard TD, Meng EC, Sali A, Ferrin TE. UCSF Chimera, MODELLER, and IMP: An integrated modeling system. Journal of Structural Biology. 2012; 179:269–278. [PubMed: 21963794]

74. Prestegard JH, Sahu SC, Nkari WK, Morris LC, Live D. Christian Chemical shift prediction for denatured proteins. Journal of Biomolecular NMR. 2013; 55:201–209. [PubMed: 23297019]

75. Bermejo GA, Llinas M. Deuterated protein folds obtained directly from unassigned nuclear overhauser effect data. Journal of the American Chemical Society. 2008; 130:3797–3805. [PubMed: 18318535]

76. Tian Y, Schwieters CD, Opella SJ, Marassi FM. AssignFit: A program for simultaneous assignment and structure refinement from solid-state NMR spectra. Journal of Magnetic Resonance. 2012; 214:42–50. [PubMed: 22036904]

77. Raman S, Lange OF, Rossi P, Tyka M, Wang X, Aramini J, Liu GH, Ramelot TA, Eletsky A, Szyperski T, Kennedy MA, Prestegard J, Montelione GT, Baker D. NMR Structure Determination for Larger Proteins Using Backbone-Only Data. Science. 2010; 327:1014–1018. [PubMed: 20133520]

78. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF chimera - A visualization system for exploratory research and analysis. Journal of Computational Chemistry. 2004; 25:1605–1612. [PubMed: 15264254]

Highlights

- Sparse isotopic labeling improves resolution in large and glycosylated proteins.

- Paramagnetic perturbations provide needed long range structural information.

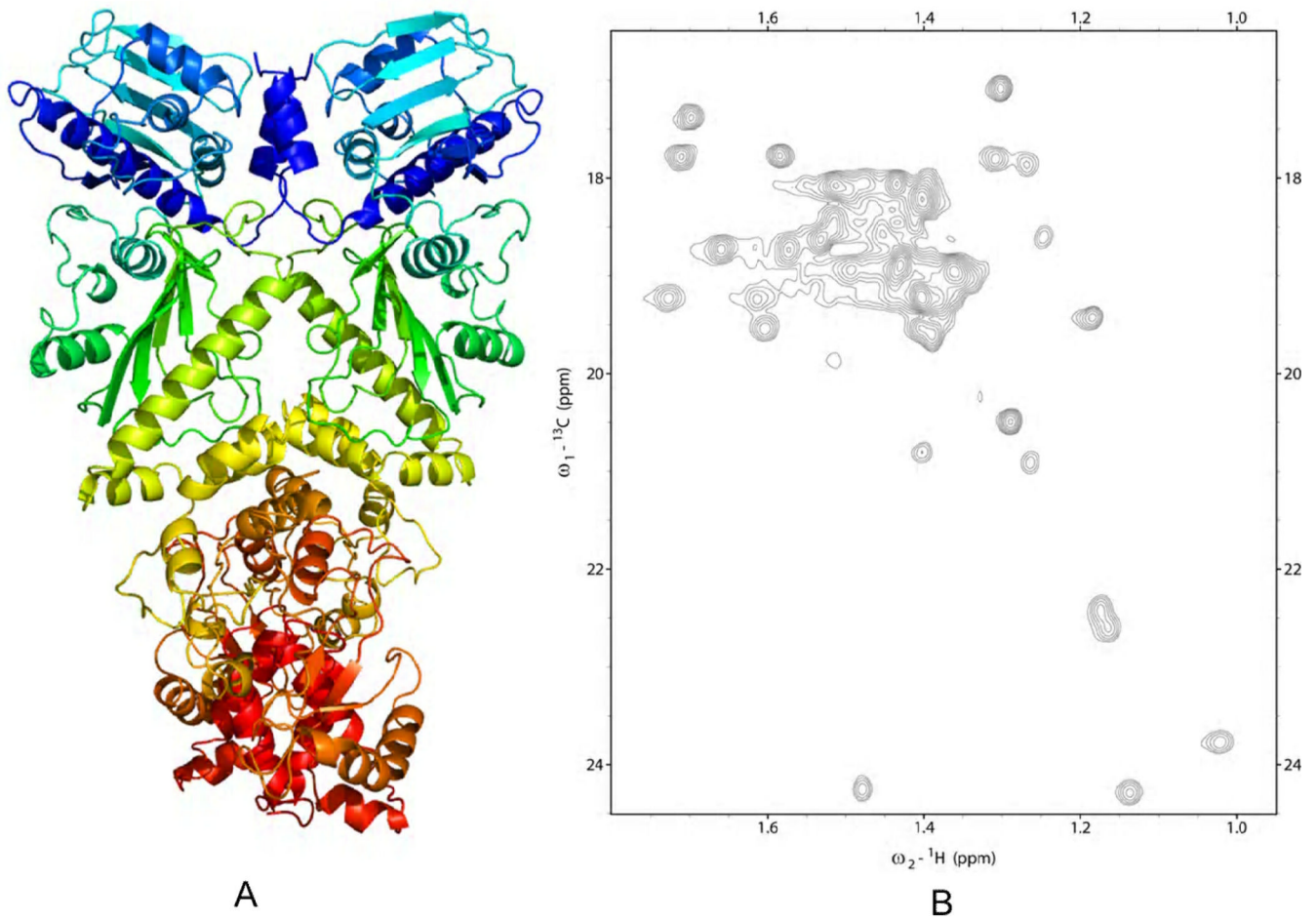- Developments in computational modeling complement sparse labeling approaches.

Figure 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.
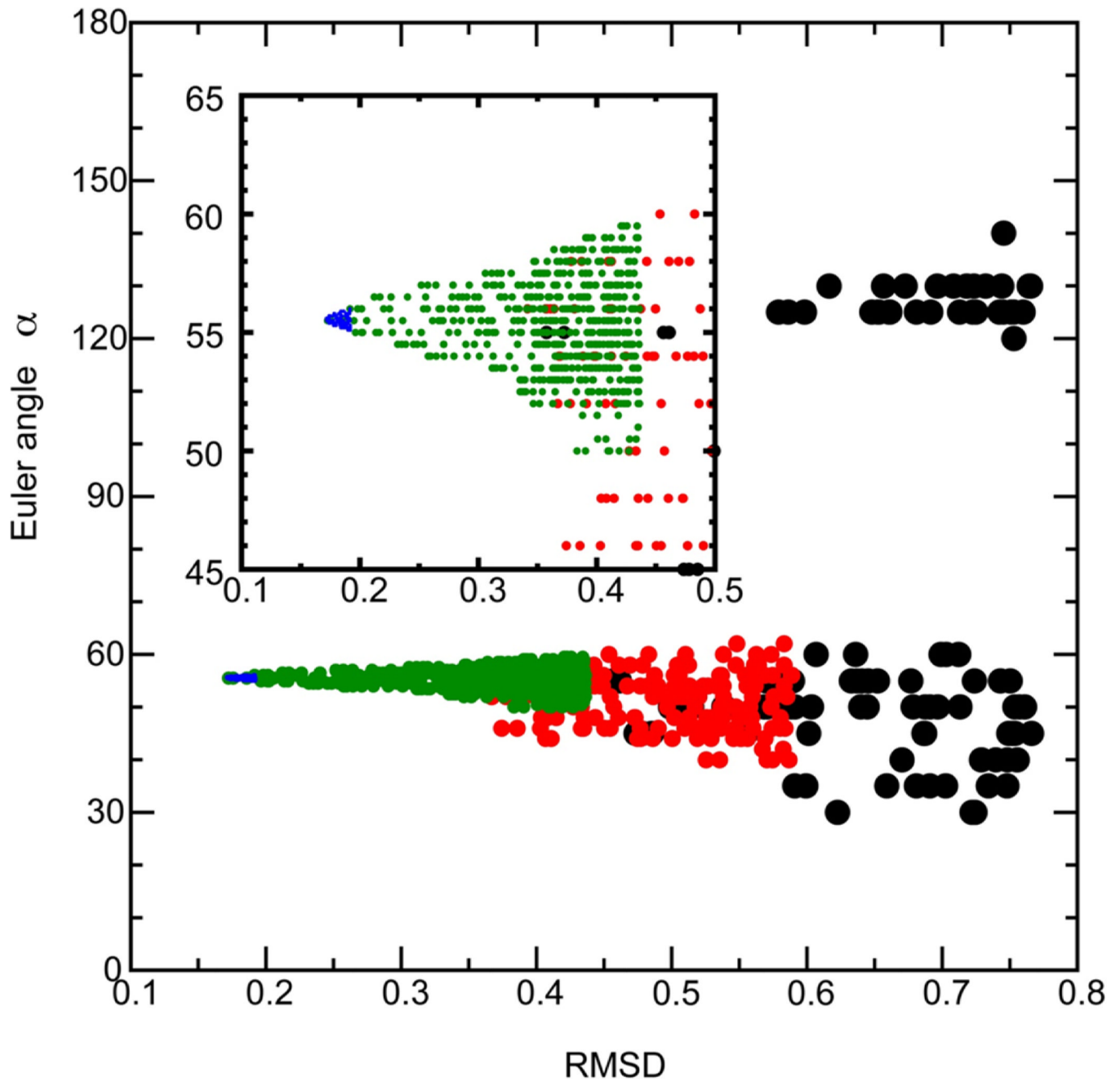
Figure 2.
Iterative grid search optimization of the Euler angle α. The lowest 1 to 10% of RMSDs from experimental values are plotted in an ordered array. Each new iteration begins with a reduced range based on analysis of the prior 1 to 10%. The black dots represent the initial grid search. Red dots are for the first refinement. Green dots are for the second refinement and blue dots are for the final refinement.
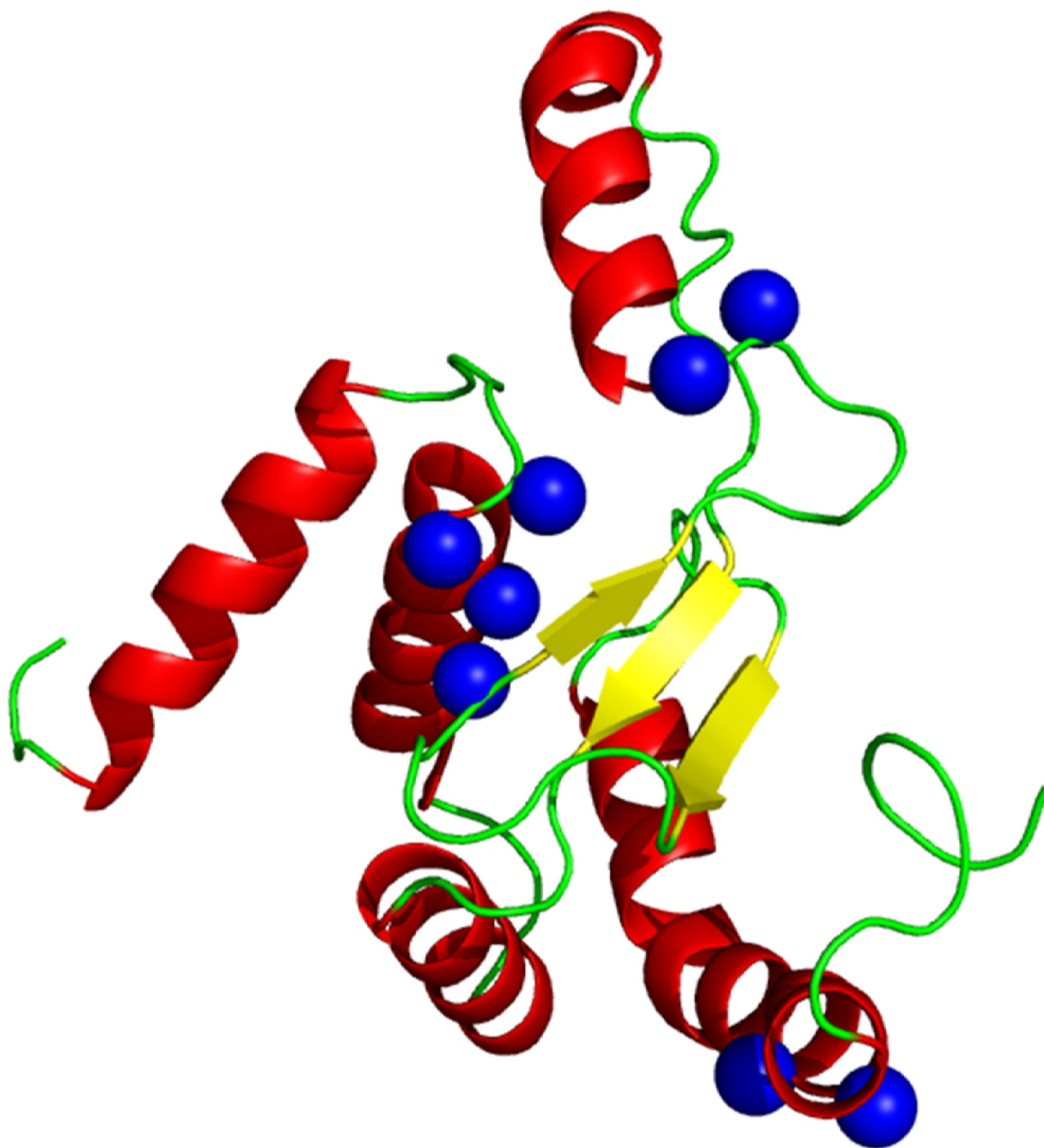
Figure 3.
Alanine methyl groups potentially associated through NOE contacts. The model represents the C-terminal domain of TRAP1. The blue spheres represent the alanine methyl groups that are within 6 Å of another alanine methyl.
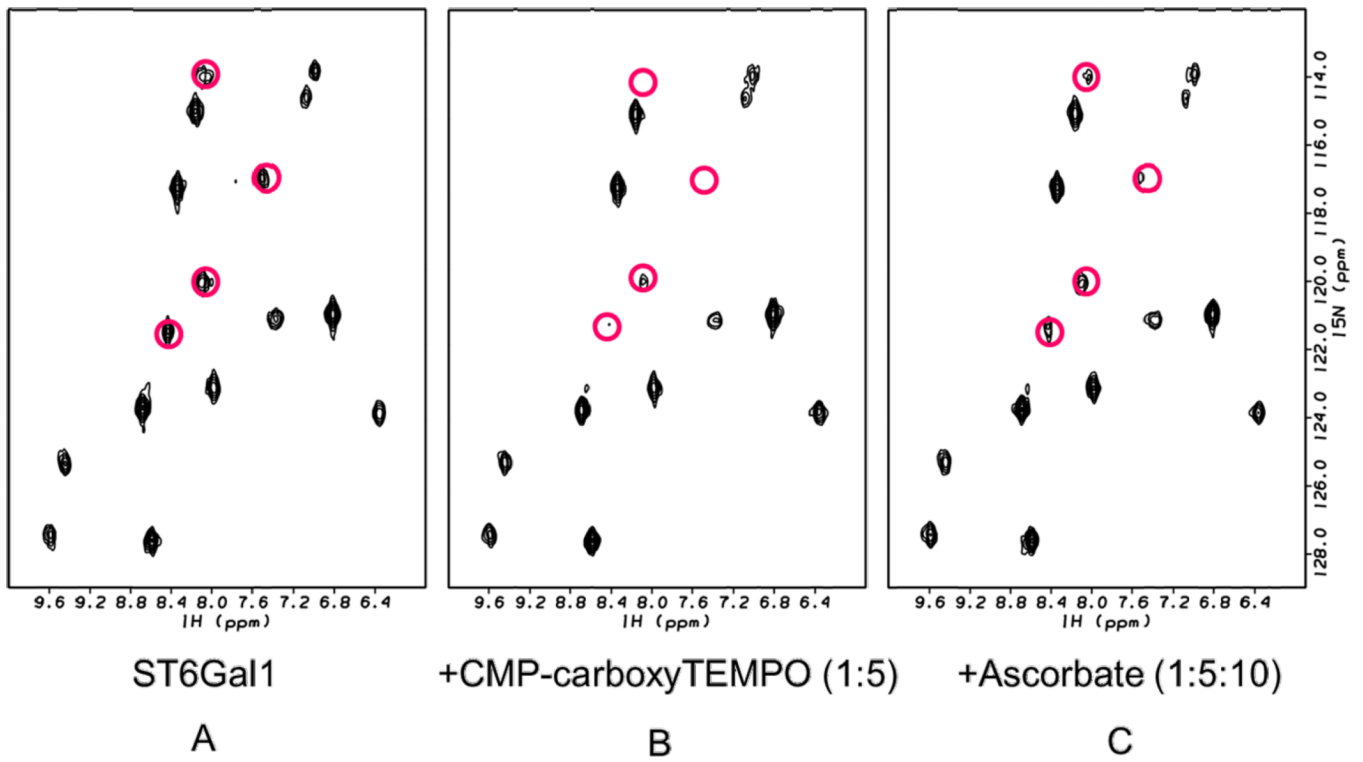
Figure 4.
HSQC Spectra of ^{15}N-phenylalanine labeled ST6Gal1 perturbed by CMP-carboxyTEMPO.
(A) Unperturbed spectrum. (B) Spectrum perturbed by CMP-carboxyTEMPO. (C)
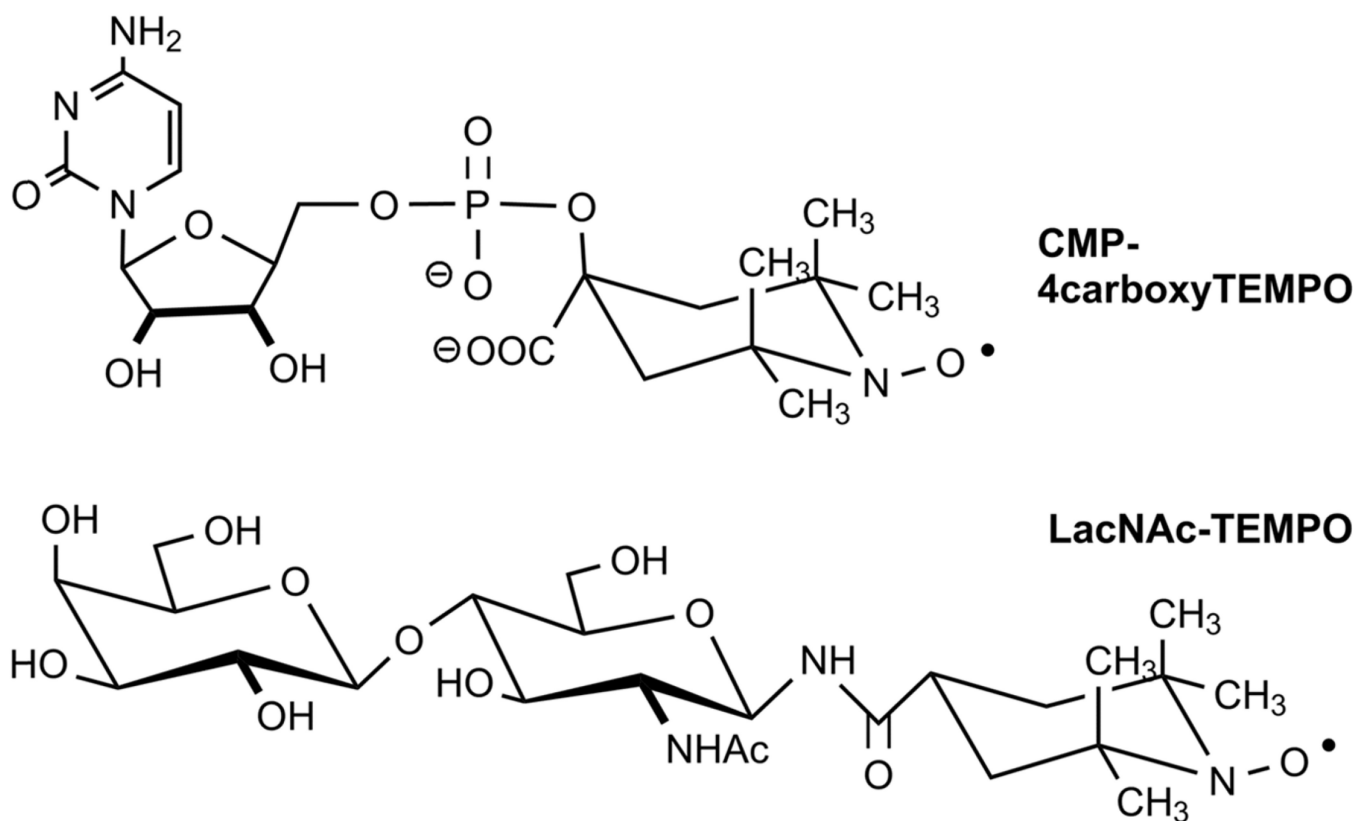Perturbations partially reversed by reduction of the spin-label.

Figure 5.
Spin-labeled donor and acceptor ligands for ST6Gal1. The carboxyl group on the TEMPO moiety of CMP-carboxyTEMPO is positioned to mimic the carboxyl group on the sialic acid of the native donor. The TEMPO group on the reducing end of LacNAc would be a mannose residue in most native N-linked glycans.
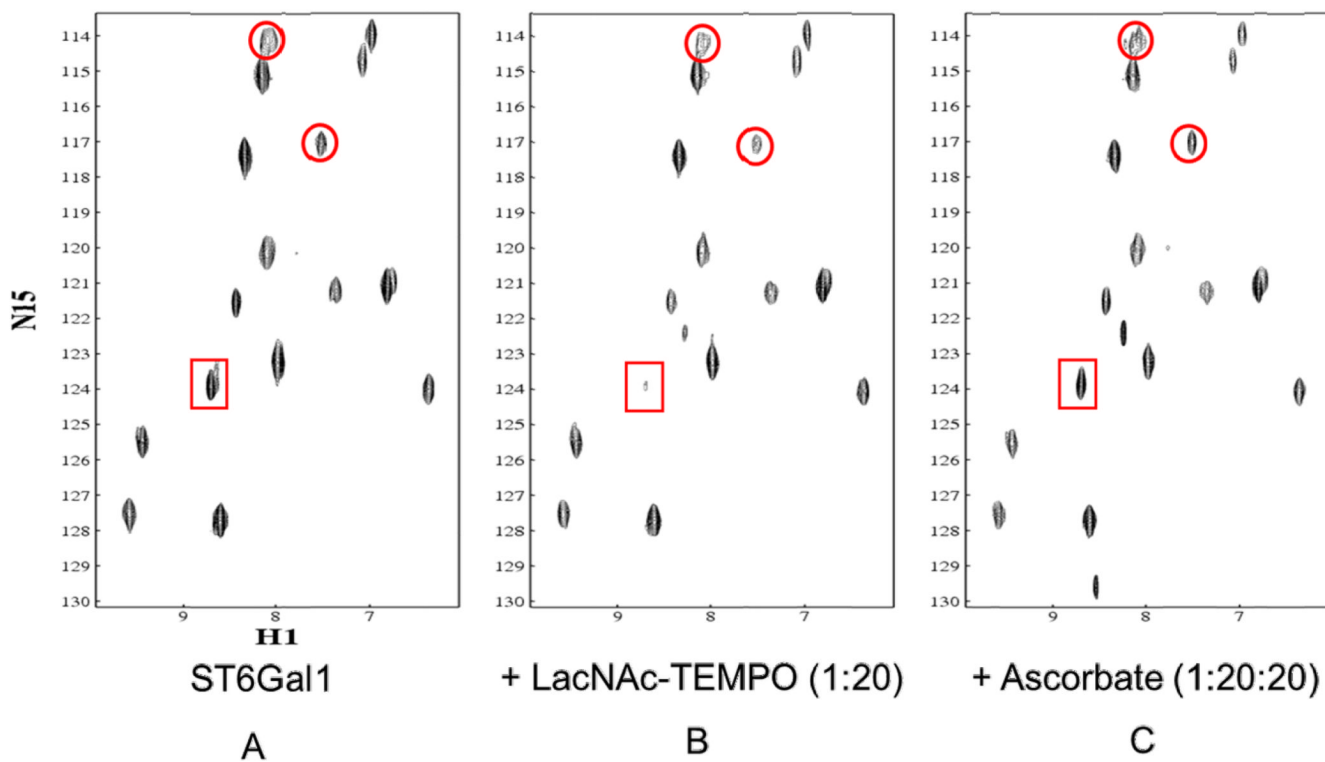
Figure 6.
HSQC Spectra of ^{15}N-phenylalanine labeled ST6Gal1 perturbed by LacNAc-TEMPO. (A) Unperturbed spectrum. (B) Spectrum perturbed by LacNAc-TEMPO. (C) Perturbations partially reversed by reduction of the spin-label.
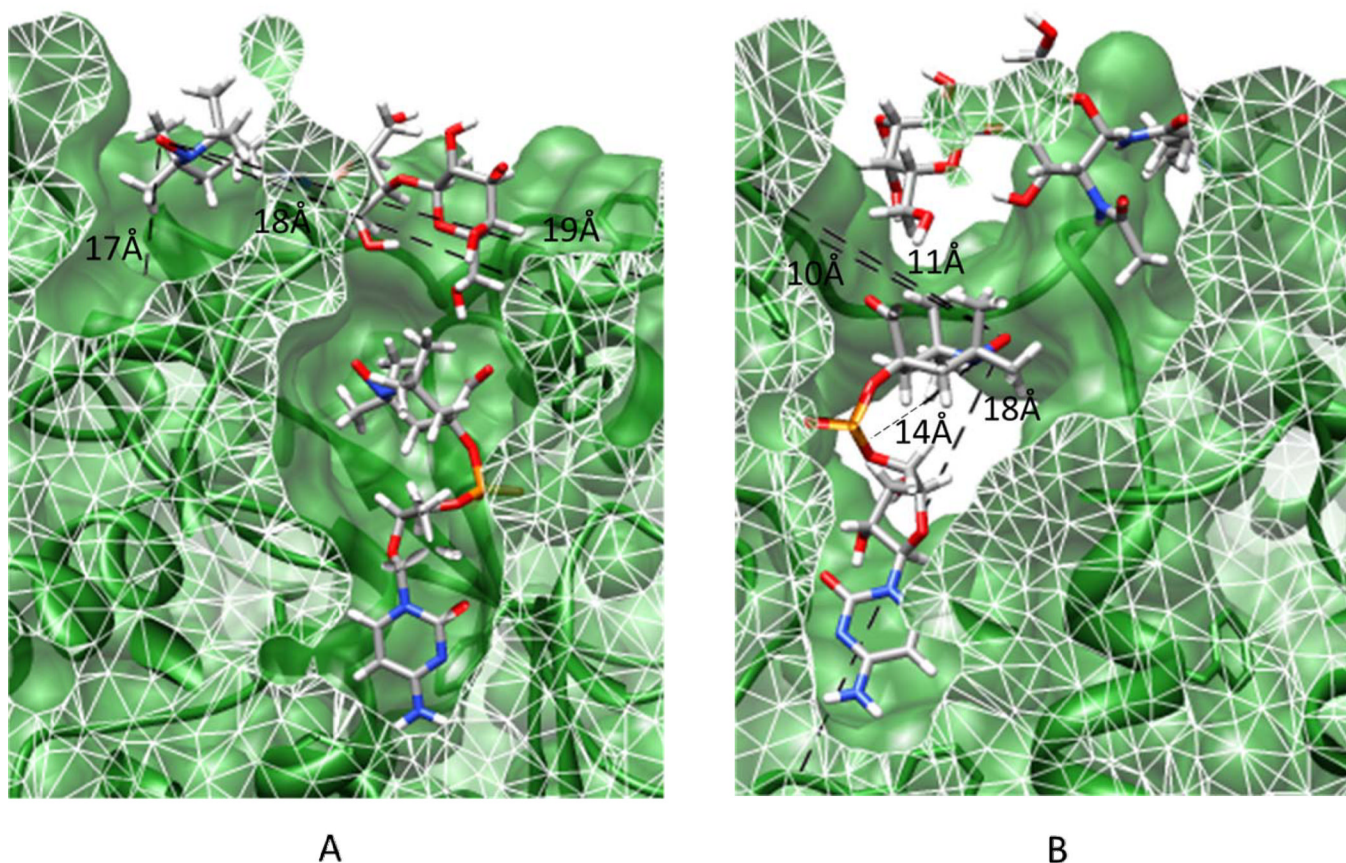
Figure 7.
Ligand placement in ST6Gal1 as guided by PRE constraints from spin labeled ligands. The protein structure has been cut away (white mesh) to show the active site cavity. A) and B) show views from the front and back respectively. Black lines depict distances in the final model to phenylalanines perturbed by CMP-carboxyTEMPO and/or LacNAcTEMPO. Molecular graphics and analyses were performed with the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311).[78]