Published in final edited form as:

Methods Mol Biol. 2015 ; 1229: 325–333. doi:10.1007/978-1-4939-1714-3_26.

Characterizing protein-glycosaminoglycan interactions using solution NMR Spectroscopy

Prem Raj B. Joseph, Krishna Mohan Poluri, Krishna Mohan Sepuru, and Krishna Rajarathnam

Departments of Biochemistry and Molecular Biology, Sealy Center for Structural Biology and Molecular Biophysics, The University of Texas Medical Branch, 301 University Boulevard, Galveston, TX 77555-1055

Abstract

Solution nuclear magnetic resonance (NMR) spectroscopy and, in particular, chemical shift perturbation (CSP) titration experiments are ideally suited for characterizing the binding interface of macromolecular complexes. ^1H - ^{15}N -HSQC-based CSP studies have become the method of choice due to their simplicity, short time requirements, and not requiring high-level NMR expertise. Nevertheless, CSP studies for characterizing protein–glycosaminoglycan (GAG) interactions have been challenging due to binding-induced aggregation/precipitation and/or poor quality data. In this chapter, we discuss how optimizing experimental variables such as protein concentration, GAG size, and sensitivity of NMR instrumentation can overcome these roadblocks to obtain meaningful structural insights into protein–GAG interactions.

Keywords

Nuclear magnetic resonance (NMR); Chemical shift perturbation; Protein–ligand interactions; Glycosaminoglycan; Dissociation constant; Heparan sulfate; Heparin

1 Introduction

Crosstalk between macromolecules in both the intracellular and extracellular milieu orchestrates all cellular processes, and understanding the structural basis and molecular mechanisms that dictate the specificity and affinity requires a detailed atomistic description of the complexes. Nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography are both routinely used for structure determination of macromolecular complexes. However, NMR is the only avenue available when complexes fail to crystallize, which could be the case if one or both of the partners are dynamic and/or if the interactions are weak in nature.

Glycosaminoglycans, such as heparin and heparan sulfate, are highly negatively charged linear polysaccharides that are widely expressed by most cell types. They mediate a wide variety of crucial functions due to their ability to bind and regulate the activities of large

*Corresponding Author: Krishna Rajarathnam, Department of Biochemistry & Molecular Biology, 301 University Boulevard, University of Texas Medical Branch, Galveston, TX 77555, krrajara@utmb.edu.

classes of proteins [1, 2]. We refer the reader to other chapters for a more detailed description of GAG structures and properties. Remarkably very little is known regarding the structural basis of how GAGs interact with proteins and mediate function. The challenges that have plagued both NMR and X-ray studies can be attributed to multiple interrelated factors that include binding-induced protein precipitation, especially at the high concentrations used in structural studies, the flexibility and high negative charge of GAGs, intrinsic heterogeneity and GAG size, and the dynamic nature of the protein–GAG interface.

In this chapter, we outline how NMR chemical shift perturbation (CSP) methods can be effectively used to study protein–GAG interactions. The popularity of this method is due to the high sensitivity and robustness of 2D ^1H - ^{15}N HSQC experiments. Chemical shifts are exquisitely sensitive to binding-induced changes in the local electronic environment, and can provide residue-level details of the binding interface, binding affinities, and timescales of protein–ligand interactions, and together with molecular docking tools like HADDOCK can also provide structural models of the protein–ligand complexes [3].

CSP-based NMR experiments have been routinely used for studying protein binding to other proteins, peptides, DNA, and small molecule ligands, but have been challenging for characterizing protein–GAG interactions. In addition to the other challenges described above, one other reason could be working under nonoptimal conditions. In this chapter, we discuss different factors that must be taken into consideration in the experimental design, and in particular, the importance of protein concentration, GAG size, sensitivity of NMR instruments, and do's and don'ts for obtaining quality NMR data. Our guidelines are based on our experience of characterizing the binding of heparin oligosaccharides to neutrophil-activating chemokines ([4]; unpublished results). Structure-function and animal model studies have shown that gradients formed by chemokines bound to the cell surface and soluble GAGs direct and regulate neutrophil trafficking from the blood to the tissue [5].

An HSQC spectrum provides a fingerprint of the backbone amides of a protein. Each backbone amide is represented by a peak corresponding to the chemical shifts of the ^1H (x -axis) and ^{15}N (y -axis) nuclei. The total number of peaks corresponds to the number of amino acids in the protein chain excluding prolines (which do not have a backbone amide NH) plus those corresponding to side chain NH_2 of asparagine and glutamine residues. In general, chemical shifts are available from literature or the BMRB data bank (<http://www.bmrb.wisc.edu>); if not, the chemical shifts must be assigned using established NMR procedures.

An overlay of a series of HSQC spectra of a heparin oligosaccharide titration to a ^{15}N -labeled chemokine, and the 2D histogram plot of the weighted average of ^1H and ^{15}N chemical shift changes ($\delta = [\delta_{\text{H}}^2 + (\delta_{\text{N}}/5)^2]^{1/2}$) as a function of individual residues, are shown in Fig. 1a and c, respectively. The data show selective perturbation of a subset of protein peaks indicating specific binding. It is generally assumed that residues showing the largest CSP mediate the binding process. However, CSP changes are correlative and not necessarily causative of the binding process. Therefore, it is possible that not all of the interfacial residues will show significant CSP upon binding. Considering that binding is predominantly mediated via lysine NH_2 and arginine guanidinium groups, NMR chemical

shifts of some of the Lys/Arg amide backbone may not be sensitive enough due to the long intervening side chain. Similarly, changes in chemical shift may also occur due to indirect binding, arising from structural rearrangements or changes in packing interactions remote from the binding interface. Therefore it is important that the CSP data be interpreted cautiously.

If the structure of the protein is known, a residue-level description of the binding site can be obtained (Fig. 1d). Analysis of the protein structure also helps in teasing out direct interactions from indirect interactions. Binding affinities can be obtained from the CSP of the individual residues as described previously (Fig. 1b) [6, 7].

2 Materials

2.1 Protein Concentration and NMR Titrations

NMR experimental conditions that enable titrations at low protein concentrations would substantially increase the probability of acquiring quality data, considering high concentration could lead to aggregation/precipitation. Availability of high field NMR instruments and advances in cryoprobes and gradient accessories have significantly improved the scope and complexity of experiments including the ability to work at low protein concentrations. For instance, we have acquired an HSQC spectrum of a ~100 μM chemokine sample (MW ~16 kDa) in 10 min on an 800 MHz spectrometer equipped with cryoprobe and field gradient accessories. The amount of time required can be reduced by a factor of ~2 when using higher field instruments (800 vs. 600 or 500 MHz), and by a factor of ~8 when using an NMR instrument with and without a cryoprobe. However, the sensitivity of a cryoprobe is dependent on ionic strength, with low salt conditions providing the best sensitivity. In terms of protein concentration, reducing the concentration by a factor of 2 will increase the time requirement by a factor of 4. Therefore, time requirement for a 50 μM compared to a 200 μM sample will be 16 times higher to achieve a similar signal to noise (s/n).

For some proteins, binding studies may be feasible only at concentrations as low as ~10 μM . It is still possible to collect an HSQC spectrum of a ~10 μM sample in less than 12 hrs on an 800 MHz spectrometer with a cryoprobe. As titration experiments may not be practical and cost effective, we suggest collecting the spectrum of the unbound protein at high concentrations and collecting one or two spectra of the GAG-bound protein at the low concentrations. Ideally, these data could identify which residues mediate binding and allow describing the geometry of the binding interface. However, chemical shift assignments of all the residues in the GAG-bound form may not be possible, especially if the interface residues undergo significant binding-induced chemical shift changes. In this case, additional experiments must be performed to better define the binding-interface residues. We also suggest, if possible, characterizing the protein–GAG complexes using other biophysical techniques such as dynamic light scattering (DLS) and sedimentation velocity to gain insights into the relationship between protein concentration and aggregation state of the complexes before carrying out the NMR experiments.

2.2 Choice of GAG for NMR Titrations

NMR studies require homogeneous GAG samples, and as only heparin is commercially available in different sizes, we discuss our experimental strategy on the basis of our experience using heparin oligosaccharides. Nevertheless, our discussion is most likely applicable to heparan sulfate and all other GAGs. Various studies have shown that the structural and functional properties of heparin mimic heparan sulfate, especially protein binding to the highly sulfated regions of the heparan sulfate. Most biophysical and structural studies reported in literature have also used unfractionated or size-fractionated heparin oligosaccharides. Heparin oligosaccharides are available in various sizes from a disaccharide to a 26 mer from specialized vendors like Neoparin, Iduron, and V-labs.

2.3 NMR Sample Preparation

The protein must be isotopically labeled with ^{15}N for characterizing binding interactions using ^1H - ^{15}N HSQC titrations. Isotopically labeled proteins are overexpressed in *E. coli* grown in minimal media using $^{15}\text{NH}_4\text{Cl}$ as the sole nitrogen source. Since the growth characteristics of *E. coli* in minimal media are somewhat compromised compared to growth in rich media (LB media), growing larger cultures is necessary to produce sufficient quantities of protein. In recent years, adding labeled growth supplements (available from vendors such as Cambridge Isotopes and Spectra Isotopes) circumvents this problem by increasing cell growth rates and overall protein expression. Once the protein is purified, the purity and molecular weight are confirmed using mass spectrometry. Sample purity is important because contaminations could complicate interpreting real signals from the background spurious noise peaks.

Protein samples for NMR studies are typically ~500 μl in volume, containing ~5–10 % D_2O for spectrometer frequency lock, 1 mM 2,2-dimethyl-2-silapentanesulfonic acid (DSS) for spectral referencing, and 1 mM sodium azide (NaN_3) to prevent microbial growth. Sample temperature for NMR data collection can vary depending on the behavior of the protein, and typically is between 20 and 40 $^\circ\text{C}$.

The choice of buffer could influence the quality of the GAG-bound spectra. If the initial choice of buffer results in poor quality spectra, we suggest collecting spectra at different buffers, pH, and ionic strength. In our experience, significant line broadening could occur under nonoptimal pH conditions (*see* Note 1).

It is also important to ensure that the protein and GAG samples are prepared in the same buffer to minimize chemical shift changes due to pH changes which could complicate and in worst case scenario lead to wrong interpretation on the binding interactions. If necessary, dialyze the protein and ligand in the same buffer. Alternatively, the lyophilized powders can be dissolved in the NMR buffer of interest, but the pH of the samples needs to be checked and adjusted before proceeding with the experiments.

¹Choice of pH could be a critical factor in obtaining quality NMR CSP data. One can collect trial HSQC spectra of a protein–GAG sample by varying pH conditions (say pH 5.0–8.0). This can help in arriving at the right pH condition for the titration experiments. In the case of some chemokines, we observe severe line broadening below pH 7.0 for some heparin oligosaccharides but not all.

3 Methods

3.1 Experimental Design

Prior knowledge of the protein, including behavior of the protein in solution, dimerization and oligomerization properties (including K_d), its GAG binding properties including binding-induced oligomerization and precipitation issues would be useful. Using the right protein concentration is a critical parameter for successful titration. A major problem is binding-induced precipitation, especially at high protein concentrations typically used in NMR studies. We propose an initial concentration of ~200 μ M, and in the event of precipitation, continuing to reduce the concentration until there is no evidence of precipitation. For chemokines, we carried out titrations on samples anywhere between 50 and 150 μ M. In principle, lower concentrations will require longer data acquisition times, and therefore whenever possible, we strongly recommend using spectrometers with cryoprobes.

We propose that the initial experiments are carried out in low ionic strength buffers, which could lead to stronger binding, resulting in lower GAG requirements. If the binding data suggest nonspecific interactions, spectra collected at varying ionic strengths could allow teasing out specific vs. nonspecific interactions (*see* Note 2). If starting protein concentrations is dependent on the oligosaccharide length, then titrations have to be carried out with shorter oligosaccharides or the protein concentration must be reduced sufficiently where the complex does not aggregate or precipitate (*see* Notes 3 and 4).

3.2 NMR Titrations and Data Analysis

CSP experiments involve collecting a series of HSQC spectra by adding GAG aliquots until essentially there are no binding-induced changes in protein backbone amide shifts. We suggest collecting a minimum of 6 to 8 spectra, which includes those of the free protein, around 50 % fraction bound population, and at saturation (*see* Note 5). More data points collected around 50 % bound population help in better defining the binding isotherms for accurate calculation of the dissociation constants. We advise using stock solutions of ~10–15 mM heparin oligosaccharides in order to minimize errors in protein concentration due to

²Initial experiments must be carried out in low ionic strength buffers. In our experience, we observed no significant differences in the perturbation profile or the binding affinities between low and high salt conditions for various oligosaccharides, though the overall extent of perturbation was higher in low salt buffers. Using low ionic strength buffers can result in non-native interactions leading to aggregation and precipitation. On the other hand, use of high salt buffers can lead to screening of electrostatic interactions and weak binding.

³Binding-induced oligomerization and precipitation effects are highly sensitive to heparin oligosaccharide chain length and the protein of interest. We suggest starting with shorter oligosaccharides such as a disaccharide and then progressing to higher oligosaccharides. For some proteins, titrations even with tetrasaccharide have led to precipitation.

⁴Working with shorter oligosaccharides would be a compromise but nonetheless can provide useful information on binding. Some limitations include smaller chemical shift perturbations, lower binding affinity, multiple binding modes, and nonspecific interactions. Therefore it may not be possible to come up with a unique binding model, and one needs to be cautious and not over interpret the data. On the other hand, selective perturbations would suggest specific binding and can provide the binding geometry. Further, studying shorter oligosaccharides can also be exploited for designing GAG decoys that could function as therapeutics in a clinical setting.

⁵When performing titration experiments, it is important that the sample is thoroughly mixed. After collecting the first HSQC spectrum of the free protein, the sample is transferred back to an Eppendorf tube using a glass pipette before addition of the ligand. After adding an aliquot of GAG (~2–10 μ l), the sample is mixed well by pipetting the solution a few times. Look for any cloudiness or precipitation. Precipitate sticking to the sides of the NMR tube will interfere with the shimming of the sample. Switch sample to a new NMR tube if necessary. Once the sample is in the NMR machine it is shimmed before starting the next experiment. It is not uncommon for the initial cloudiness to disappear on subsequent addition of GAG aliquots. We advise not to add GAG directly into the NMR tube, which can cause uneven mixing leading to errors in the estimation of the protein–GAG molar ratios.

dilution. Prior knowledge of an estimate of binding affinities could be useful in selecting the starting protein concentration and amount of GAG to be added.

NMR data processing, analysis, and spectra viewing can be carried out using NMRpipe, NMRview, Sparky, or instrument-specific Bruker and Varian software [8–10]. Most processing and analysis script and programs for data fitting are available in the respective software websites. Binding affinities of protein–GAG interactions can vary by orders of magnitude (nM to mM), and accordingly kinetics (especially off constant, k_{off}) can vary by many orders. The kinetics of binding are classified as slow, intermediate, and fast exchange on the NMR time scale, which can influence the nature and quality of the spectra [7]. If binding occurs in the slow exchange regime, spectra will contain separate peaks at the chemical shifts of the free (δ_{P}) and GAG-bound (δ_{PL}) protein. As the ligand is titrated into the protein, the peak intensity at δ_{P} will decrease and of the peak at δ_{PL} will increase. If binding occurs in the fast exchange regime, spectra will contain a single set of peaks at the population-weighted average chemical shifts. As the ligand is titrated into the protein, the peak position will move from δ_{P} to δ_{PL} . If binding is in the intermediate exchange regime, peaks are exchange-broadened resulting in poor quality spectra (*see* Note 6).

The CSP experiments are best performed under the fast exchange regime as assigning chemical shifts of the GAG-bound form is straightforward as shown in Fig. 1a, b. However, in the intermediate and slow exchange regimes, assigning chemical shifts of all the residues in the GAG-bound form may not be possible as the information on the direction and magnitude of the individual peak movements is missing (*see* Note 6). In particular, this will be the case if the binding residues are in the crowded region of the spectrum and/or undergo large chemical shift changes. Therefore, it may be necessary to change experimental conditions so that the binding occurs in the fast exchange regime by using smaller GAGs, increasing ionic strength, and/or other experimental parameters.

In the fast exchange regime, the CSP follows a hyperbolic dependence as a function of ligand concentration. Proper choice of protein and ligand concentrations combined with nonlinear least squares analyses using two independent variables (ligand and protein concentrations) and two parameters (K_{d} and maximum CSP) can increase the accuracy of measured dissociation constant as shown in Fig. 1b [6].

References

1. Lindahl U, Kjellén L. Pathophysiology of heparan sulphate: many diseases, few drugs. *J Intern Med.* 2013; 273:555–571. [PubMed: 23432337]
2. Kamhi E, Joo EJ, Dordick JS, Linhardt RJ. Glycosaminoglycans in infectious disease. *Biol Rev Camb Philos Soc.* 2013; 88:928–943. [PubMed: 23551941]
3. Dominguez C, Boelens R, Bonvin AM. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc.* 2003; 125:1731–1737. [PubMed: 12580598]

⁶Line broadening could occur due to aggregation or due to intermediate exchange binding kinetics. In the latter case, the peaks would reappear on excess GAG titration (fractional bound population >90 %). Therefore, we suggest not to abort the experiment, but to collect a few more spectra in the presence of excess GAG.

4. Poluri KM, Joseph PR, Sawant KV, Rajarathnam K. Molecular basis of glycosaminoglycan heparin binding to the chemokine CXCL1 dimer. *J Biol Chem.* 2013; 288:25143–25153. [PubMed: 23864653]
5. Gangavarapu P, Rajagopalan L, Kolli D, Guerrero-Plata A, Garofalo RP, Rajarathnam K. The monomer-dimer equilibrium and glycosaminoglycan interactions of chemokine CXCL8 regulate tissue-specific neutrophil recruitment. *J Leukoc Biol.* 2011; 91:259–265. [PubMed: 22140266]
6. Markin CJ, Spyropoulos L. Increased precision for analysis of protein-ligand dissociation constants determined from chemical shift titrations. *J Biomol NMR.* 2012; 53:125–138. [PubMed: 22534787]
7. Mittermaier A, Meneses E. Analyzing protein-ligand interactions by dynamic NMR spectroscopy. *Methods Mol Biol.* 2013; 1008:243–266. [PubMed: 23729255]
8. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR.* 1995; 6:277–293. [PubMed: 8520220]
9. Johnson BA, Blevins RA. NMR View: a computer program for the visualization and analysis of NMR data. *J Biomol NMR.* 1994; 4:603–614. [PubMed: 22911360]
10. Goddard, TD.; Kneller, DG. SPARKY 3. University of California; San Francisco:

Fig. 1. Mapping of chemokine-heparin oligosaccharide binding interface using NMR chemical shift perturbation data. **(a)** A section of a ^1H - ^{15}N HSQC spectrum showing heparin oligosaccharide binding-induced chemical shift changes in a chemokine. The unbound and final bound peaks are in *black* and *red* and the intermediate peaks are shown in *cyan*, *blue*, and *green*. Note the selective perturbation of a subset of amino acids indicated by an *arrow*. **(b)** A representative plot of CSP vs. GAG/chemokine concentration (in molar ratio), which allows calculation of the dissociation constant (K_d). **(c)** A histogram of the CSP as a function of chemokine sequence. *Dotted line* indicates the cutoff for residues to be considered perturbed. **(d)** A surface plot of the chemokine showing residues (*blue*) that are significantly perturbed on GAG binding