

Simple High-Resolution NMR Spectroscopy as A Tool in Molecular Biology

NMR and NMR-data analysis tools for molecular biology

Luca Mureddu and Geerten W. Vuister*

Leicester Institute of Structural and Chemical Biology

Department of Molecular and Cell Biology

University of Leicester

Henry Wellcome Building, Lancaster Road, Leicester LE1 7HN, United Kingdom

*Correspondence Address: Geerten W. Vuister, Department of Molecular and Cell Biology, Leicester Institute of Structural and Chemical Biology, University of Leicester, Henry Wellcome Building, Lancaster Road, Leicester, LE1 7HN, United Kingdom, Phone: +44 116 229 7076; Email: gv29@leicester.ac.uk

Keywords: NMR, Chemical Shift Mapping, Tstar, CcpNMR AnalysisAssign, ^{15}N -HSQC.

Abbreviations: 1D, one-dimensional; 2D, two-dimensional; 3D, three-dimensional; BMRB, Biological Magnetic Resonance Data Bank; CcpNmr, Collaborative Computing Project for NMR (software); CSP, chemical shift perturbation; eq, equivalent; HSQC, Heteronuclear Single Quantum Coherence Spectroscopy; NEF, NMR-exchange format; ppm, parts per million; NMR, Nuclear Magnetic Resonance; TROSY, Transverse Relaxation-Optimized Spectroscopy; Tstar, Testis-Signal Transduction and Activation of RNA.

Abstract

Nuclear Magnetic Resonance (NMR) is one of the major techniques for investigating the structure, dynamics and interactions between biomolecules. However, non-experts often experience NMR experimentation and data analysis as intimidating. We discuss a simple yet powerful NMR technique, the so-called chemical shift perturbation (CSP) analysis, as a tool to elucidate macromolecular interactions in small- and medium-sized complexes, including protein-protein, protein-drug, and protein-DNA/RNA interactions. We discuss current software packages for NMR data analysis and present a new interactive graphical tool implemented in CcpNmr AnalysisAssign version-3, which can drastically reduce the time required for the CSP analysis. Lastly, we illustrate the usefulness of a protein 3D structure for interpretation of the CSP data.

Introduction

It is the ultimate aim of the molecular biologist to understand cellular functioning in its molecular context. As such, it is imperative to know at which time and place specific biomolecules are active to exert their function. At the root of our understanding, however, is the realisation that the interactions between individual molecules, that together form active complexes of sometimes intricate complexity, constitute the underpinning basis of all the biological processes. Structural biology is the field of science which aims to describe such interactions between biologically relevant molecules at an atomic level. It is based on the notion that the interactions are facilitated by the specific molecular shapes and, as has nowadays become evident, also their dynamical changes. Together these are crucial in determining the affinities that drive the assembly of the macromolecular complexes [1].

Nuclear Magnetic Resonance (NMR) is one of the three major techniques that provides structural, dynamical and also interaction data [2]. In this minireview we will illustrate how a simple yet powerful experimental NMR technique, the so-called chemical shift perturbation (CSP) analysis, can be used to investigate interactions between biomolecules or biomolecules and small drug-like compounds. Since it was first proposed, the CSP analysis has become well-established, as illustrated by the increasing number of papers referring to the technique (Fig. 1A), with currently ~80 references annually. In this paper, we also discuss how current NMR software packages can facilitate the CSP data analysis. Particular focus will be given to the CcpNmr AnalysisAssign version-3, which provides several user-friendly tools for retrieving the relevant data thus, providing for invaluable biological information.

Chemical shift and exchange

NMR is a spectroscopic technique that employs an inherent property of many nuclei called 'spin' to yield spectra of various nuclei of biological interest, e.g. ^1H , ^{13}C , ^{15}N and ^{31}P . NMR's exquisite (bio-)chemical usefulness originates from its ability to discriminate the different nuclei in a biomolecule. The electronic environment of each nucleus slightly modifies its exact resonance frequency through a process called chemical shielding and consequently the positions of various peaks in the NMR spectrum are specific for each nucleus in the molecule. The position of a peak in an NMR spectrum is commonly referred to as the 'chemical shift' and denoted by the symbol δ . The chemical shift constitutes one of the most important parameters and in NMR provides a powerful tool for a biochemist; as it not only allows us to discriminate one nucleus from another, but also provides information about their conformation and nearby chemical environment. For example, an analysis using recorded chemical shifts from the Biological Magnetic Resonance Data Bank (BMRB) reveals distributions with median values of ~8.24, ~4.32 ppm and ~1.39 ppm for the Alanine H^{N} , H^{α} and H^{β} , respectively (Fig. 1B). The spread of each of the three distributions reflects the different conformations, i.e. the chemical environments, and dynamics, i.e. the change in these environments, of each of the nuclei in the various Alanine residues in their respective proteins.

In practice, the analysis of one dimensional biomolecular spectra is prohibitive because of spectral overlap. To overcome this problem, it is possible to correlate one nucleus with another, generating two-dimensional (2D) or even higher-dimensional (3D, 4D, nD) NMR spectra. For proteins in particular, a simple and very informative example is the 2D heteronuclear ^{15}N -HSQC [3] (or alternatively for larger proteins the ^{15}N -TROSY-HSQC [4]) experiment. The resulting 2D ^{15}N -HSQC spectrum affords greater resolution and valuable information as peaks can be used as a 'fingerprint' of a protein. In practice, nearly each peak represents a backbone amide group of an individual residue, with the exception of peaks originating from the HN containing side chains of amino acids Asn, Gln, His, Trp, Lys and Arg. However, except for His the signals from the sidechain moieties are easily recognisable. Furthermore, proline residues are absent in 2D heteronuclear ^{15}N -HSQC spectra due to the lack of an amide group. The chemical shift values of the various nuclei can be extremely useful when used as a proxy to monitor protein-ligand interactions, a process called chemical shift mapping or chemical shift perturbation (CSP).

When a protein is titrated with a ligand, e.g. with a drug or another biomolecule, the chemical shift of the nitrogen and proton nuclei of the residues that are in close proximity to the binding site will be most affected. Thus, the binding of the ligand results in changes in the chemical shifts of these nuclei, causing the resulting peaks to alter their position in the NMR spectrum. By recording a series of NMR experiments at varying stoichiometries of protein and ligand, the resulting series of spectra conveys information regarding the affinity of the ligand, as well as identifying the important residues involved in the interaction.

To briefly explain the theoretical aspect of this phenomenon, consider for simplicity a peak for a nucleus i in a protein P at 7.0δ ppm in a one-dimensional NMR spectrum (Fig. 1C). For a simple two-state binding process with ligand L :



the $[P]:[L] = 1:0$ equivalent condition (Fig. 1C, 1:0 eq) represents the left side of equation 1, where the protein is in its unbound state. Upon addition of a high-affinity ligand, under the condition that the k_{off} is small, the equilibrium lies fully towards the bound state. Consequently, addition of the ligand causes the i peak in the unbound state to decrease in intensity, whereas a new peak at a different position appears; this new peak represents the bound state of the same protein nucleus i . At a $[P]:[L] = 1:0.5$ stoichiometry (Fig. 1C, 1:0.5 eq) peaks for both the unbound and bound states will be present with equal intensities (neglecting dynamic effects). The difference between the bound- and unbound peak position is called $\Delta\delta_i$. At a $[P]:[L] = 1:1$ stoichiometry, only the peak of the bound state will be present, as the peak for the unbound state will have disappeared (Fig. 1C, 1:1 eq).

In NMR, the above situation when the k_{off} is much smaller than $\Delta\delta_i$ is called the “slow exchange” regime. In contrast, in a “fast exchange” regime, when k_{off} is much greater than $\Delta\delta_i$, in each of the spectra recorded at different [P]:[L] stoichiometries the peak position for nucleus i represents the population weighted average of its free and bound positions. Consequently, the peak appears to be “moving” from its original position (Fig. 1D 1:0 eq.) towards a its bound position as the ligand concentration increases (Fig. 1D, 1:0.5 eq to 1:2 eq). In cases where $k_{off} \approx \Delta\delta_i$, the peak typically disappears due to line broadening effects and this situation is called “intermediate-exchange” (not shown).

The real power of the CSP method lies in the identification of protein residues most affected by the ligand, i.e. those residues with nuclei that display large $\Delta\delta_i$ values. In case of multi-dimensional spectra, such as the ^{15}N -HSQC spectrum (Fig. 2A), the total chemical shift change involving all dimensions is usually taken (*vide infra*, equation 2) and used as a proxy for the importance of the specific residue in the interaction. Furthermore, for the fast exchange regime K_d can be determined from the observed $\Delta\delta_i$ values as a function of ligand concentration [5].

Chemical shift perturbation in practice

The first step in a CSP procedure is the assignment of individual peaks to a specific residue in the protein. Different software packages and algorithms have been developed to establish the assignments of the backbone nuclei, i.e. H^N , N, C^α , C' , either by manual or automated approaches [6]. According to Lee and Markley, based on BMRB statistics in 2014 Sparky was still the most widely used NMR data analysis tool for backbone assignment [7]. A much more recent programme, not included in their list, is CcpNmr AnalysisAssign version-3 [8], which provides tools for simple and semi-automated backbone assignments and a dedicated interactive CSP analysis module, from here on referred to as the CSP module (Fig. 2C). An *ab-initio* protein backbone assignment can be a time-limiting factor, as it requires some effort in terms of sample preparation, NMR spectra recording and data analysis. However, assignments can potentially be retrieved from databases, such as the BMRB, and serve as the starting point, adjusting them as required by the user’s experimental conditions.

Table 1 displays an overview of functionalities from the various NMR data analysis programmes that are relevant for a CSP analysis. After the initial setup, either Sparky [7], CcpNmr Analysis Version 2 [9], NmrView [10] and CARA [11] can automatically calculate the $\Delta\delta$ from chemical shifts. However, only AnalysisAssign version-3 has capabilities for interactive inspection of the $\Delta\delta$ data in relation to the underpinning spectra, interactive thresholding and easy adaptation of various relevant parameters. Moreover, apart from AnalysisAssign version-3 users will have to manually plot and/or export to other software packages for further analysis. This typically will involve multiple manual steps before retrieving the biological relevance from their initial NMR dataset. In addition, parameter adjustments will necessitate repeating these various steps, thus increasing the time required for the whole analysis and inevitably increasing the risk of introducing human errors. The AnalysisAssign version-3 CSP module, contains all required functionalities for a full CSP analysis, including automatically generated bar charts

of CSP values as function of residue number which are linked to the underpinning spectra, peak tables and a binding plot. Additionally, if the protein molecular structure information is available, the annotations and selections can be mirrored to a graphical molecular visualisation, such as PyMOL [12].

We tested the AnalysisAssign CSP module by exploring the binding of Clip2 RNA, AAAUAA, to the Tstar-KH domain [13] [14]. We imported both the assignments of the free Tstar-KH domain in addition to a series of five ^{15}N -HSQC spectra of uniformly ^{15}N -labelled Tstar-KH at 0.0, 0.5, 1.0, 1.5, and 2.0 equivalent of Clip2 ligand, directly from the original Sparky data using the inherent data conversion routines of AnalysisAssign. The assignments of the KH domain were propagated from the spectrum at 0.0 equivalent of ligand to all other spectra using the simple drag and drop feature of AnalysisAssign for copying peak lists. As expected, some of the peaks had changed their positions upon titration with the ligand. Using the interactive tables and spectrum displays it was possible to easily identify shifted peaks and correct their position in the target spectra, either individually, or as a group of peaks across multiple spectra, or automatically. Fig. 2B shows the overlay of the five spectra, with trajectories of shifted peaks indicated for selected residues.

Fig. 2C shows the overlay of the five ^{15}N -HSQC spectra at the position of G78. The gradual change in peak position upon increasing ligand concentration is evident. Importantly, to ensure a valid analysis all spectra should be properly referenced as changes in peak positions could otherwise be misinterpreted. Fortunately, AnalysisAssign has routines to establish, and where needed report, on the spectral alignment that functions even in the case of non-fully identical spectra [8]. Using an automated analysis, in which all the peaks were accurately matched to their extrema, CSP_i values are calculated for each Tstar-KH residue i using equation 2 and displayed automatically as a bar plot in the CSP module interface.

$$\text{CSP}_i = \sqrt{(\Delta\delta_{\text{H}_i})^2 + \alpha(\Delta\delta_{\text{N}_i})^2} \quad (2)$$

where α denotes the relative weighting of chemical shift changes of the ^{15}N nuclei relative to the ^1H nuclei, by convention set to 0.14 [15], and $\Delta\delta_{\text{H}_i}$ and $\Delta\delta_{\text{N}_i}$ denote the observed changes of the proton and nitrogen chemical shifts for residue i , respectively. Crucially, the threshold below which the CSP_i values are deemed not significant needs to be established. A value of 1σ derived from the distribution of all CSP values is set by default as the first estimate without a need of other filters [5]. Rapid manual inspection of the affected residues establishes if this threshold needs upward or downward adjustment. In the case of G78, the peak follows a consistent trajectory upon ligand titration with well-defined changes in peak positions, rendering its CSP an appropriate threshold value. In contrast, another residue with a similar CSP value, A138, is located in a crowded region of the spectrum (cf. Fig. 2B) and thus its peak movements in such areas are potentially compromised by mis-assignment of the peaks. Such residues should be flagged and we recommend that, barring further NMR data confirming their proper assignment, they first be excluded from the analysis. In our view it is wise to reduce any possible false positives, until the moment that their inclusion appears warranted, e.g. after careful inspection of

the structure (*vide infra*) or on the basis of other data that confirms that they can be considered relevant for the binding event. Depending on k_{off} and the residue-specific $\Delta\delta_i$, peaks can disappear in any of the spectra with ligand concentrations > 0 . Such situation still conveys useful information about the involvement of the specific residue, albeit that the exact magnitude of its CSP_i value cannot be established. In general, the disappearance of a peak for a specific residue is assumed to imply a $\Delta\delta_i$ (CSP_i) value resulting in exchange broadening due to an intermediate exchange regime and hence signifies importance. Consequently, these residues are automatically flagged by the CSP analysis module, e.g. residue I97 in Tstar-KH.

The true power of the CSP analysis is revealed when mapping the CSP results onto a molecular structure. The CSP module can automatically map the annotations and residue selections onto a molecular structure of the biomolecule under investigation using an external molecular visualisation programme, such as PyMOL [12]. The CSP analysis of the interaction of Clip2 RNA with the Tstar-KH domain identified residues F72, V73, G74, K75, L77, G78, G81, S83 T93, I97, R104, K106, K108, E110, R113, Y120 and L130 with significant CSP values (Fig. 2C). When mapped onto the structure of Tstar-KH, a clustering of several affected residues is observed across an interface formed by one α -helix, two β -strands and two loops, corresponding to the KH hydrophobic groove (Fig. 2D). In fact, according to Feracci *et al.* [13] residues G78 and I97, (the latter flagged as ‘missing’ by the CSP analysis), belong to a set of crucial residues in the groove that stabilise the interaction with the Clip2 RNA.

In cases where no 3D structure is available, often the structures of homologous proteins can be used as a reference. Protein structure is more highly conserved than primary sequence; therefore, small changes in protein composition usually do not significantly alter the structure of the macromolecule. Alternatively, in many cases protein structures can be fairly accurately obtained using homology modelling, where an existing structure of a homologous protein is used as a template to generate the structure of the protein of interest. Such models can be reliably used for mapping of protein-ligand interactions, assuming the binding interface has not been affected by the mutations. A description of contemporary available homology modelling software packages and servers are discussed in the review by Vyas *et al* [16].

Currently freely available NMR software suites, such as Sparky, allow retrieving CSP values from their tables, but users are required to manually export to third party software to carry on the analysis or are limited to single and static plotting, like in the case of CcpNmr Analysis Version 2. The possibility to graphically and interactively inspect the NMR data that identify the residues involved in a protein-ligand interaction, makes the new AnalysisAssign CSP module extremely useful for non-experts and drastically reduces the time required for a CSP analysis. The CSP module is also not limited to ^1H - ^{15}N as it can accommodate any combination of nuclei, e.g. ^1H - ^{13}C in case of methyl residues. AnalysisAssign is implemented in a flexible fashion that will facilitate easy adaptation to insert specific calculation modes for $\Delta\delta_i$ values, automatic pre- and user-defined K_d fittings as well as direct links to external auto-docking software such as HADDOCK [17]. For more advanced users, it is also possible

to use the AnalysisAssign libraries to create specific, but simple macros to extrapolate further information from the dataset (Table 1). For example, a macro can be used to plot the minimal shift changes resulting from the mutation of a specific residue in a protein, including in this calculation only residues in which the CSP is above the defined threshold value. Several settings and data exporters based on the NMR-exchange format (NEF, [18]) or tabular .xls format have been implemented, thus providing tools to easily export information to other programs for further analysis if so required.

A series of other programmes have also been developed to address specific tasks using peaks in NMR spectra and NMR assignments, such as the programme Farseer, which performs analyses on large and multivariable datasets, including CSP [19]. Other programmes include auto-FACE, which facilitates the identification of binding mechanisms from CSP data [20] and TITAN, which uses peak perturbations trajectories to help the identification of interaction mechanisms [21].

Conclusions

Chemical shift perturbation is very useful as a simple tool to elucidate macromolecular interactions in small and medium-sized complexes, including protein-protein, protein-drugs, and protein-DNA/RNA interactions [5], either in solution, or in solid state NMR [22]. The CSP method works best when recording heteronuclear NMR spectra on samples in which the biomolecule, e.g. the protein, is isotopically labelled as to allow for selective detection. Fortunately, protein overexpression and isotope labelling, e.g. by ^{15}N or ^{13}C , has now become routine in *E.coli* and *Pichia Pastoris*, with new and promising developments for expression in higher eukaryotic systems, which guarantees the presence of a more complex folding machinery and post-translational modifications [23]. Together with new developments in NMR technology, e.g. direct ^{15}N -observation [24], the CSP method will find even more widespread application.

In comparing the various programmes for CSP data analysis, we find that the CSP module of AnalysisAssign version-3 provides for a simple and interactive graphical tool that allows users to significantly reduce the time required for a CSP analysis.

Downloads

AnalysisAssign can be freely downloaded for non-commercial academic usage from <http://www.ccpn.ac.uk/v3-software/downloads>. The CSP module and tutorials will be included in its upcoming version 3.0 release.

Acknowledgements

We thank Drs Cyril Dominguez and Mikael Feracci for the usage of their Tstar-KH/Clip2 data. We also thank Drs Ed Brooksbank, Wayne Boucher and Rasmus Fogh for their expert contributions to the CcpNmr AnalysisAssign programme. We thank the editors for helpful comments. LM acknowledges his stipend provided by MRC-IMPACT PhD programme (grant MR/NO13913/1) and GWV acknowledges funding of the CCPN project by MRC (grants MR/L000555/1 and MR/P00038X/1).

Author Contributions

LM designed and developed the CSP analysis module for CcpNmr AnalysisAssign Version 3.0. LM & GWV analysed the data. LM & GWV wrote the manuscript.

Conflict of interest

The authors declare no conflict of interest.

References

- 1 Breukels V, Konijnenberg A, Nabuurs SM, Doreleijers JF, Kovalevskaya N V & Vuister GW (2011) Overview on the use of NMR to examine protein structure. In *Curr Protoc Protein Sci* p. 17.5.2-17.5.44.
- 2 Kwan AH, Mobli M, Gooley PR, King GF & MacKay JP (2011) Macromolecular NMR spectroscopy for the non-spectroscopist. *FEBS J.* **278**, 687–703.
- 3 Bodenhausen G & Ruben DJ (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* **69**, 185–189.
- 4 Pervushin K, Riek R, Wider G & Wuthrich K (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci.* **94**, 12366–12371.
- 5 Williamson MP (2013) Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73**, 1–16.
- 6 Würz JM, Kazemi S, Schmidt E, Bagaria A & Güntert P (2017) NMR-based automated protein structure determination. *Arch. Biochem. Biophys.* **628**, 24–32.
- 7 Lee W, Tonelli M & Markley JL (2015) NMRFAM-SPARKY: Enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–7.
- 8 Skinner SP, Fogh RH, Boucher W, Ragan TJ, Mureddu LG & Vuister GW (2016) CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis. *J. Biomol. NMR* **66**, 111–124.
- 9 Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J & Laue ED (2005) The CCPN data model for NMR spectroscopy: Development of a software

- pipeline. *Proteins Struct. Funct. Genet.* **59**, 687–696.
- 10 Johnson BA & Blevins RA (1994) NMR View: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603–614.
- 11 Keller R (2004) *The computer aided resonance assignment tutorial*.
- 12 DeLano WL (2014) The PyMOL Molecular Graphics System, Version 1.8. *Schrödinger LLC*, <http://www.pymol.org>.
- 13 Feracci M, Foot JN, Grellscheid SN, Danilenko M, Stehle R, Gonchar O, Kang HS, Dalglish C, Meyer NH, Liu Y, Lahat A, Sattler M, Eperon IC, Elliott DJ & Dominguez C (2016) Structural basis of RNA recognition and dimerization by the STAR proteins T-STAR and Sam68. *Nat. Commun.* **7**.
- 14 Foot JN, Feracci M & Dominguez C (2014) Screening protein - Single stranded RNA complexes by NMR spectroscopy for structure determination. *Methods* **65**, 288–301.
- 15 Ayed A, Mulder FAA, Yi GS, Lu Y, Kay LE & Arrowsmith CH (2001) Latent and active p53 are identical in conformation. *Nat. Struct. Biol.* **8**, 756–60.
- 16 Vyas VK, Ukawala RD, Ghate M & Chintla C (2012) Homology Modeling a Fast Tool for Drug Discovery: Current Perspectives. *Indian J. Pharm. Sci.* **1**, 1–17.
- 17 Dominguez C, Boelens R & Bonvin AMJJ (2003) HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–7.
- 18 Gutmanas A, Adams PD, Bardiaux B, Berman HM, Case DA, Fogh RH, Güntert P, Hendrickx PMS, Herrmann T, Kleywegt GJ, Kobayashi N, Lange OF, Markley JL, Montelione GT, Nilges M, Ragan TJ, Schwieters CD, Tejero R, Ulrich EL, Velankar S, Vranken WF, Wedell JR, Westbrook J, Wishart DS & Vuister GW (2015) NMR Exchange Format: A unified and open standard for representation of NMR restraint data. *Nat. Struct. Mol. Biol.* **22**, 433–434.
- 19 Teixeira JMC, Skinner SP, Arbesú M, Breeze AL & Pons M (2018) Farseer-NMR: automatic treatment, analysis and plotting of large, multi-variable NMR data. *J. Biomol. NMR* **71**, 1–9.
- 20 Krishnamoorthy J, Yu VCK & Mok YK (2010) Auto-FACE: An NMR based binding site mapping program for fast chemical exchange protein-ligand systems. *PLoS One* **5**, e8943.
- 21 Waudby CA, Ramos A, Cabrita LD & Christodoulou J (2016) Two-Dimensional NMR Lineshape Analysis. *Sci. Rep.* **6**, 24826.
- 22 Gardiennet C, Wiegand T, Bazin A, Cadalbert R, Kunert B, Lacabanne D, Gutsche I, Terradot L, Meier BH & Böckmann A (2016) Solid-state NMR chemical-shift perturbations indicate domain reorientation of the DnaG primase in the primosome of *Helicobacter pylori*. *J. Biomol. NMR* **64**, 189–195.
- 23 Verardi R, Traaseth NJ, Masterson LR, Vostrikov V V. & Veglia G (2012) Isotope labeling for solution and solid-state NMR spectroscopy of membrane proteins. *Adv. Exp. Med. Biol.* **992**, 35–62.
- 24 Takeuchi K, Arthanari H, Shimada I & Wagner G (2015) Nitrogen detected TROSY at high field yields high resolution and sensitivity for protein NMR. *J. Biomol. NMR* **63**, 323–331.

Figures

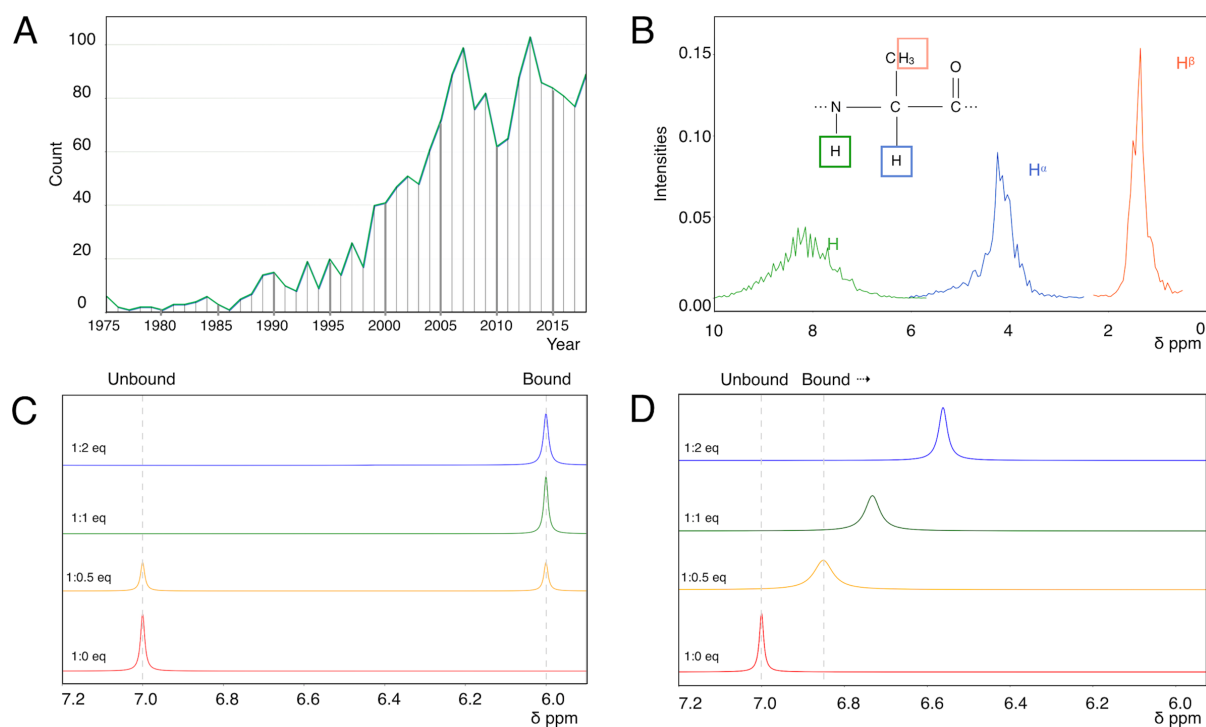


Figure 1. Chemical shift and exchange.

A) Number of CSP publications as function of year of publication. The plot shows the number of journal articles in the PubMed database by querying for “chemical shift (perturbation or mapping)”.

B) Distribution of deposited chemical shifts for the H^N , H^α and H^β nuclei of Alanine as derived from the Biological Magnetic Resonance Data Bank (BMRB). For historical reasons, the scale in NMR is expressed in relative terms, the so-called ppm scale, which runs from high positive values on the left to low, or negative values, on the right of the scale.

C) Simulated 1D $^1H^N$ NMR spectra under the slow chemical exchange regime. Spectra are shown at 0.0 (red), 0.5 (orange), 1.0 (green) and 2.0 (blue) equivalent (eq) of (NMR-invisible) ligand. $\Delta\delta$ was assumed to be -1 ppm. D) Simulated 1D $^1H^N$ NMR spectra under the fast chemical exchange regime. Spectra are coloured as previously. Peak positions were calculated using equation 6 in reference [5], using $\Delta\delta = -1$ ppm, $[protein] = 100 \mu M$, $K_d = 200 \mu M$. Note the gradual shift of the peak as function of ligand concentration. The exchange-induced broadening of peaks at 0.5 equivalent and 1 equivalent are slightly exaggerated for illustrative purposes.

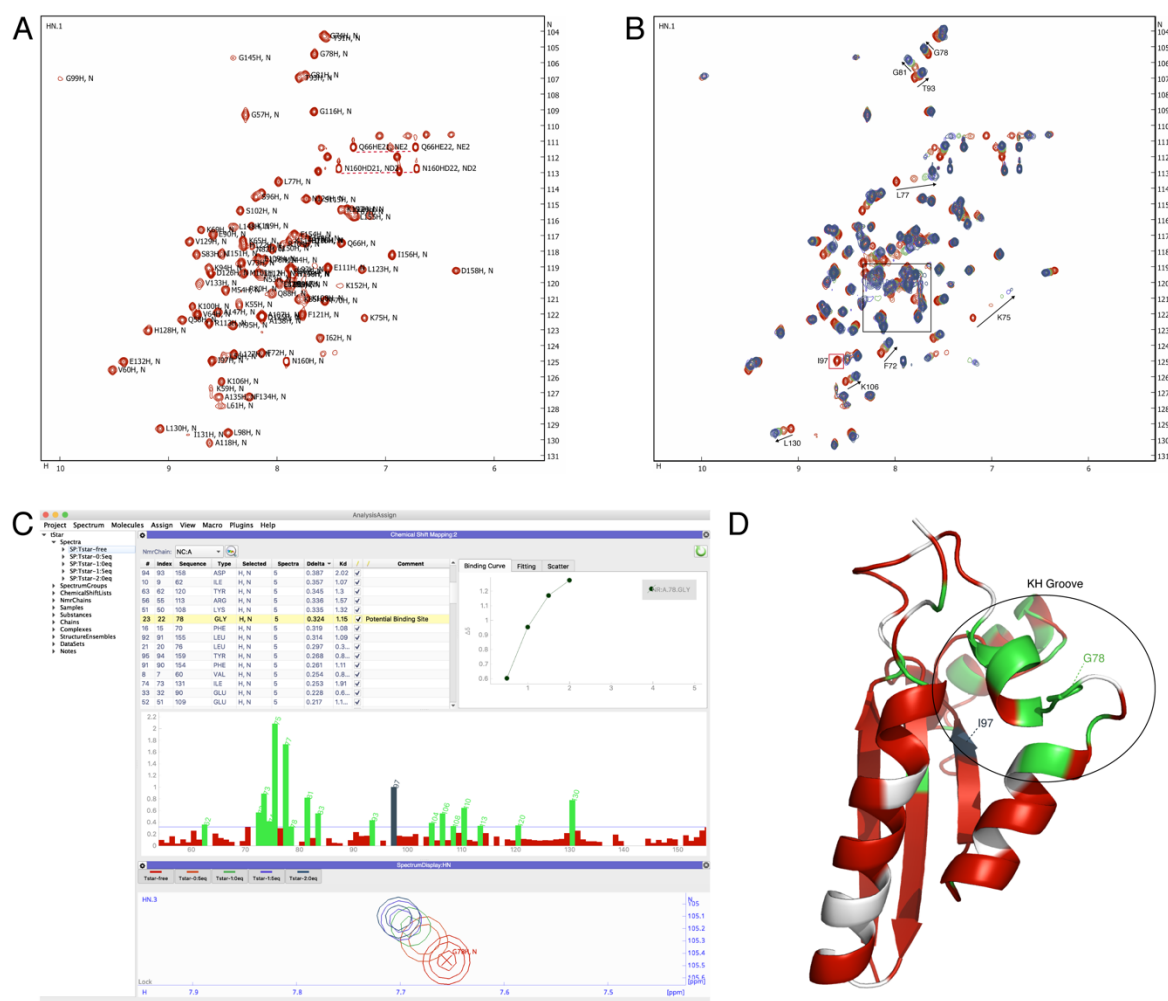


Figure 2. Chemical shift perturbation analysis (CSP) of the binding of Clip2 RNA (AAAUAA) to the Tstar-KH domain [13].

A) ¹⁵N-HSQC spectrum of 200 μM uniformly ¹⁵N-labelled Tstar KH domain; selected assignments are indicated.

B) Five overlaid spectra of Tstar-KH ¹⁵N-HSQC domain. Spectra are shown at 0.0 (red), 0.5 (orange), 1.0 (green), 1.5 (purple) and 2.0 (dark blue) equivalent of Clip2 ligand. The black square box indicates the crowded region. Arrows indicate different peak perturbation trajectories. The I97 peak (red box) is present only at 0.0 equivalent but disappears upon addition of Clip2.

C) The CSP analysis module in AnalysisAssign version-3. Clicking any bar in the bar chart (middle; included residues in green, excluded residues in red) or residue in the residue table (top) will navigate to the peaks of the corresponding residue in the spectra (bottom). The binding curve can be automatically displayed on the right side of the table. Multiple selection overlays related binding curves. All adjustments of parameters and settings of the CSP analysis module, such as setting the threshold line (i.e. horizontal line in middle panel) or excluding a residue from the analysis (checkboxes in the residue table), will result in a real-time update of all plots without the need for any tedious or error-prone manual actions.

D) Ribbon representation of the Tsar-KH domain (PDB code **5EL3**) with residues coloured according to their CSP values resulting from interaction of Clip2 RNA. Residues flagged with missing peaks in the spectra containing ligand, e.g. I97, are highlighted in dark blue. The black circle highlights the KH domain clip2 binding groove. Unassigned residues or residues removed from the analysis are indicated in light grey.

Features	CcpNmr V3	CcpNmr V2	Sparky	NmrView/CARA
Peak Automation ¹	Simultaneous peak selections, copying and re-fitting	Simultaneous peak selections, copying and re-fitting	Simultaneous peak selections, copying and re-fitting	Single peak re-fitting
Interactivity ²	Selectable plot items and tables with live updates	None or static plots	None	None
Settings ³	Multiple dimensionality Multiple Atoms Multiple $\Delta\delta$ calculation modes Several GUI parameters	Limited dimensionality	Limited dimensionality	Limited dimensionality
Extras ⁴	Link to Molecular Visualisation IPython Console Macro editor	None	User's macros	User's macros
Exports ⁵	Images: Various formats Texts: Various formats Software readable: Json	Text: Various formats	Text: Various formats	Text: Various formats

Table 1. Comparison of common freely available NMR software packages with built-in backbone assignment (as shown on the BMRB statistics in 2014, [7]) and CSP analysis capabilities.

¹ Graphical peak selection, copying assignments between spectra, peak adjustment and refitting, provisions to follow peaks across titration series.

² Live updates of results, interactive adjustment of parameters.

³ Adaptable to different experiment types, ability to handle different dimensionalities, ability to handle different peak parameters.

⁴ Interaction graphical visualisation tools, ability to link to other software packages.

⁵ Exports to external formats