Methods xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

Methods

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

Structure determination of protein-ligand complexes by NMR in solution

Julien Orts [a], Alvar D. Gossert [b,c,*]

[a] Laboratory of Physical Chemistry, ETH Zürich, HCI F217, Vladimir-Prelog-Weg 2, 8093 Zürich, Switzerland
[b] Institute for Molecular Biology and Biophysics, ETH Zürich, HPP L25.1, Hönggerbergring 24, 8093 Zürich, Switzerland
[c] Biomolecular NMR Spectroscopy Platform, ETH Zürich, HPP L25.1, Hönggerbergring 24, 8093 Zürich, Switzerland

ARTICLE INFO

Article history:
Received 6 November 2017
Received in revised form 24 January 2018
Accepted 29 January 2018
Available online xxxx

ABSTRACT

In this paper, we discuss methods for determining structures of protein-ligand complexes by NMR in solution. Our discussion is based on small ligands (<2 kDa) as for example drugs, metabolites or oligopeptides, but most of the considerations also apply to more general cases. In NMR in solution, the kinetics of association and dissociation of the complex – the exchange rate – determines the optimal sample preparation and the NMR experimental approach. Additionally, depending on the part of the complex that will be studied (only the bound ligand, the protein, the protein-ligand interface or the entire protein-ligand complex structure), different types of NMR experiments are needed. Therefore, the choice of a combination of the appropriate experiment and a suitable sample preparation in terms of ligand to protein ratios are discussed in detail. Also, considerations for practically preparing samples of protein-ligand complexes and carrying out experiments including trouble shooting are described. For structure determination, the scope of this paper is limited to NOE-based methods and some of the most recent approaches will be covered.

© 2018 Elsevier Inc. All rights reserved.

Contents

1. Introduction: structure determination of protein-ligand complexes then and now . . . 00
   1.1. Sample preparation . . . 00
   1.2. Collection of NMR data . . . 00
   1.3. Deriving distance restraints . . . 00
   1.4. Structure calculation . . . 00
2. Fundamentals of protein-ligand interactions . . . 00
   2.1. Thermodynamics: dissociation constant, K_D, and the observable fraction bound, p_B . . . 00
   2.2. Significance of K_D for NMR experiments . . . 00
   2.3. Kinetics: on- and off-rates, k_{on}, k_{off}, and the observable exchange rate, k_{ex} . . . 00
   2.4. Significance of exchange kinetics for NMR spectroscopy . . . 00
3. Requirements for attempting a structure determination of protein-ligand complexes . . . 00
   3.1. Parameters for characterizing a protein-ligand complex . . . 00
   3.2. Assessing feasibility of a complex structure determination . . . 00
4. Sample preparation . . . 00
   4.1. Preparation of complexes with fast exchange kinetics . . . 00
   4.2. Preparation of complexes with slow exchange kinetics . . . 00
   4.3. Considerations for cases where ligand solubility is limiting . . . 00
      4.3.1. Maximizing fraction of bound protein at limited solubility of the ligand . . . 00
      4.3.2. Maximizing the concentration of complex in situations of limited ligand solubility . . . 00
      4.3.3. Potential additives for increasing ligand solubility . . . 00
   4.4. Minimizing content of protonated small molecules to avoid t_1-noise and baseline irregularities . . . 00
   4.5. Changing temperature, viscosity and concentrations of protein and ligand to avoid the intermediate exchange regime . . . 00
5. NMR experiments for recording and identifying intra-protein, intra-ligand and inter-molecular NOE cross peaks . . . 00

* Corresponding author.
E-mail addresses: julien.orts@phys.chem.ethz.ch (J. Orts), alvar.gossert@mol.biol.ethz.ch (A.D. Gossert).

https://doi.org/10.1016/j.ymeth.2018.01.019
1046-2023/© 2018 Elsevier Inc. All rights reserved.

5.1. Information content of conventional 3D NOESY experiments . . . 00
5.2. NMR experimental elements for isotope editing and filtering . . . 00
5.2.1. Half-filter experiments . . . 00
5.2.2. Practical considerations when recording filtered experiments . . . 00
5.2.3. Improvements on half-filter experiments: purging schemes and matched adiabatic pulses . . . 00
5.3. Variants of isotope editing experiments . . . 00
5.3.1. Filtered-edited or edited-filtered NOESY experiments for recording inter-molecular NOEs? . . . 00
5.3.2. All-inclusive NOESY . . . 00
5.3.3. Filtering based on fast exchange: Transferred NOEs yield intra-ligand and inter-molecular protein-to-ligand NOEs . . . 00
5.4. Selecting the appropriate experiments for recording intra-ligand, inter-molecular and intra-protein NOEs . . . 00
5.4.1. Complexes with fast exchanging ligands . . . 00
5.4.2. Complexes with slow exchanging ligands . . . 00
6. Experimental optimizations . . . 00
6.1. Optimizing NOE mixing time . . . 00
6.2. Simplified description of relaxation processes during mixing time to assess useful duration of mixing time . . . 00
6.3. Maximizing NOE cross peaks . . . 00
6.4. NOE mixing time for quantification of inter-nuclear distances . . . 00
6.5. Special considerations for inter-molecular NOEs . . . 00
6.6. Changing temperature and viscosity to increase the NOE . . . 00
7. Deriving distance restraints . . . 00
7.1. Correcting for incomplete occupancy . . . 00
7.2. Accounting for differential relaxation of different nuclei . . . 00
7.3. Calibration of NOEs . . . 00
8. Structure calculation . . . 00
9. Conclusion . . . 00
Acknowledgements . . . 00
Appendix A. Supplementary data . . . 00
References . . . 00

1. Introduction: structure determination of protein-ligand complexes then and now

NMR has been very successful in determining three-dimensional structures of proteins in the last thirty years [1–4]. An important part of this success is uniform isotope labelling of the target protein with ^{13}C and ^{15}N [5,6], which enables recording 3D NOESY spectra with little signal overlap. However, for solving the structure of molecular complexes of proteins with ligands, there is the complication that ligands in most cases cannot be isotope labelled. Therefore, several methods had to be devised, which allow recording separate sets of signals for the unlabelled ligand and the isotope labelled protein. The methods for solving 3D solution structures of such protein-ligand complexes were developed in the early 1990s, and the general approach is still followed nowadays. As an example to illustrate the general approach, we will use here the work on the complex of cyclosporine A (CsA) and cyclophillin A (Cyp), which was described in detail in the publication of 1994 [7]. In the main part of this text we will then turn to the details of contemporary approaches.

1.1. Sample preparation

In order to study the Cyclophillin-Cyclosporin complex, Cyclophillin was produced recombinantly in *E. coli* with uniform ^{13}C, ^{15}N isotope labelling and an equimolar complex with unlabelled Cyclosporin A was prepared at 1.1 mM concentration.

The Cyclophillin-Cyclosporin complex represents one extreme of possible complexes, it is a stable complex in slow exchange, for which a equimolar (1:1) complex could be prepared, which is an ideal situation. For weaker and more transient interactions further considerations need to be taken into account, which will be addressed in more detail in the section about sample preparation.

1.2. Collection of NMR data

After obtaining the complete resonance assignment, a set of different NOESY spectra was recorded in order to obtain three sets of NOE-based distance restraints:

1. Intra-protein NOEs for determining the protein structure in the complex
2. Intra-ligand NOEs for determining the ligand conformation in the complex
3. Inter-molecular NOEs for determining the structure of the binding interface

For obtaining intra-protein NOEs (1), 3D ^{13}C-resolved $^{1}H,^{1}H$ NOESY and 3D ^{15}N-resolved $^{1}H,^{1}H$ NOESY spectra were recorded with mixing times of 80–100 ms. For intra-ligand NOEs (2), so-called double half-filter experiments with isotope-filtering elements in both dimensions were recorded on samples containing unlabelled CsA and labelled Cyp to just record NOEs within CsA. In the special case of CsA, in previous work, ^{13}C-labelled CsA was prepared and the conformation of the bound state was determined on a sample containing unlabelled Cyp and ^{13}C-labelled CsA using 2D NOESY spectra with ^{13}C-editing elements in both dimensions [8,9]. Additionally, $^{3}J_{HH}$ couplings were used to restrain values of dihedral angles. Inter-molecular NOEs (3) were recorded on samples of isotope labelled Cyp in complex with unlabelled CsA using double half-filter experiments with isotope-filtering in one dimension and isotope-editing in the other dimension. Again, due to the special circumstance that ^{13}C-labelled CsA was available, the same experiments were also recorded on a sample containing isotope labelled CsA and unlabelled Cyp.

NOESY data are not collected much differently nowadays. There are improvements such as purging elements and matched adiabatic sweep pulses which are more efficient than half-filters. However, the approach of recording several spectra for individual types

Please cite this article in press as: J. Orts, A.D. Gossert, Methods (2018), https://doi.org/10.1016/j.ymeth.2018.01.019

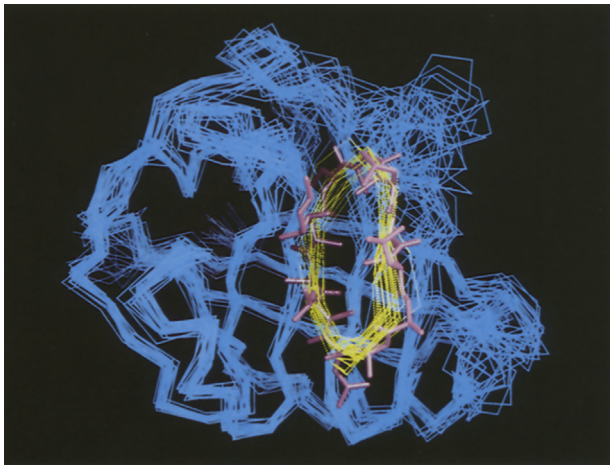

Fig. 1. Solution model of cyclophilin A in complex with cyclosporine A. Cα trace of the polypeptide backbone of the 22 selected conformers representing the NMR solution structure of the Cyp-CsA complex (blue, Cyp; yellow, CsA; magenta: all-heavy-atom molecular model of CsA in the conformer with the lowest residual Lennard-Jones potential.) Reproduced from [7] with permission from Springer. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of intra- and inter-molecular NOEs is still followed. New strategies like 4D spectra and the All-inclusive NOESY may change the data acquisition strategy but this still needs to be established.

1.3. Deriving distance restraints

For deriving distance restraints, the intensities of the recorded NOE cross peaks were translated into distance restraints without further correction factors, since this complex was stable and fully populated in the sample.

Nowadays, inter-molecular NOEs are calibrated differently, accounting for the generally longer distances between protons at interfaces compared to intramolecular NOEs, where short distance NOEs are also present, like for example those between geminal protons. Additionally, for weaker complexes that are only partially populated, correction factors accounting for the bound population are used.

1.4. Structure calculation

Using these distance restraints, structure calculations were performed in torsion angle space. This was crucial at that time, because computers weren't powerful enough to perform these calculations in Cartesian space. In torsion angle space, however, only one molecule can be represented. Thus, Cyp and CsA were linked to yield a single molecular chain, by means of linker residues without van der Waals potential (without attraction and repulsion), which were thus able to penetrate other atoms. The linker length was set to 2.6-fold the diameter of Cyp, in order to allow CsA to freely access the entire surface of Cyp.

For preparing the structure calculation, a number of unnatural amino acids that are present in CsA had to be added to the residue library of the program DIANA [10].

The structure calculation was then run in a standard way. All distance restraints were used from the beginning of the calculation. Alternative protocols, using first only intra-molecular restraints and only later adding inter-molecular restraints lead to the same results as including all restraints from the beginning.

Although computers are much more powerful today, calculations can be carried out in Cartesian space but probably still in most cases in torsion angle space [10–13]. While the original structures were calculated from a set of manually assigned and calibrated upper distance restraints, NOE cross peak assignment is performed automatically by highly refined routines that are based on multiple cycles of iterative structure determination [10,11,14]. Therefore, during a contemporary structure calculation many hundreds of structures are calculated and the process thus still profits highly form running it in torsion angle space. There are a few minor inconveniences with this approach in torsion angle space: Ligands need to be virtually connected to the protein chain and if ligands are not made up of natural amino acids or nucleic acids, the ligands need to be described in a library file. Fortunately, there are automated routines that automatically determine such library entries [15,16]. A second inconvenience is the handling of cyclic molecules. In torsion angle space, only linear and branched structures can be represented, but not circular ones. Therefore, one needs to choose between representing ring-structures as fixed ring systems, or open chains with artificial ring-closure constraints that reduce the speed of the calculation.

The resulting structure is shown in Fig. 1.

The NMR field has come a long way since these early results on structure determination of protein-ligand complexes. There has been progress on all aspects of structure determination of complexes: protein isotope labelling and preparation of samples, acquisition of the necessary isotope filtered experiments, alternative sources of structural restraints, computational methods and structure determination, which have been covered in several high-quality reviews [17–21]. (Interestingly, little progress has been made in preparing uniformly isotope labelled ligands.) In spite of the impressive progress of the NMR field, the hallmarks of the determination of the Cyp-CsA complex are still valid nowadays, and thus serve as suitable introductory example to the topic of NOE-based structure determination of protein-ligand complexes.

2. Fundamentals of protein-ligand interactions

Before addressing specific details of structure determination of protein-ligand complexes by NMR, we want to set the scene with a minimal description of the thermodynamic and kinetic properties of protein-ligand interactions, and their consequences for NMR experiments.

2.1. Thermodynamics: dissociation constant, K_D, and the observable fraction bound, p_B

Proteins (P) and ligands (L) can interact to form a complex (PL) (Eq. (1)). Thermodynamically complex formation is governed by a multitude of atomic interactions and entropic factors that determine the free energies of the free (P and L) and bound (PL) states. If the free energy of the bound state more favorable, complex formation will occur. Proteins and ligands constantly associate and dissociate with rate constants k_{on} and k_{off}, respectively, and establish equilibrium populations of free and bound states.

$$P + L \underset{k_{off}}{\overset{k_{on}}{\rightleftarrows}} PL \tag{1}$$

NMR measurements of protein-ligand complexes are taken at equilibrium conditions. The binding equilibrium is described accurately by the thermodynamic equilibrium constant, the dissociation constant K_D for protein-ligand interactions (Eq. (2)).

$$K_D = \frac{[P]_{free}[L]_{free}}{[PL]} \tag{2}$$

where $[L]_{free}$, $[P]_{free}$ and $[PL]$ are the concentrations of the free protein, the free ligand and the complex, respectively. K_D values

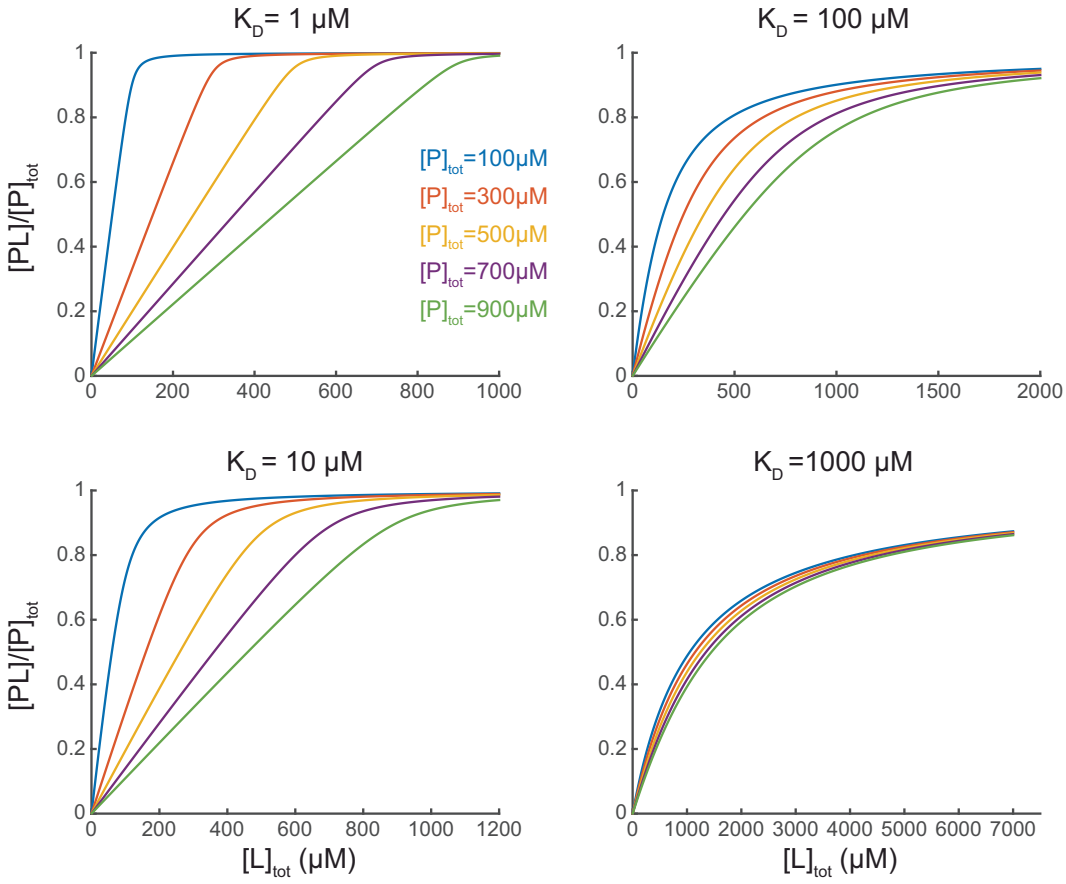

Fig. 2. Bound population of protein with respect to the total ligand concentration at different complex affinities, K_D. Curves were plotted using Eqs. (3b) and (4) for typical concentrations employed in NMR samples for structure determination. The same curves apply if ligand and protein are exchanged (Eq. (2) is symmetric with respect to P and L). In this case, the vertical axis would read $[PL]/[L]_{tot}$ and the horizontal axis $[P]_{tot}$.

typically range from 1 mM for weak binders to nM or pM for tight binders.

The K_D is not directly observable in NMR experiments. However, depending on the sample, all three species L, P and PL can be observed, but they may be difficult to quantify. For quantification, typically the ratio of concentrations of complex $[PL]$ to total ligand $[L]_{tot}$ or protein $[P]_{tot}$ is used, that is, the fraction of bound ligand (p_B^L) or protein (p_B^P), respectively.

$$p^L_{\text{bound}} = \frac{[PL]}{[L]_{\text{tot}}} = 1 - p^P_{\text{free}} \tag{3a}$$

$$p^P_{\text{bound}} = \frac{[PL]}{[P]_{\text{tot}}} = 1 - p^L_{\text{free}} \tag{3b}$$

The equation for the concentration of complex $[PL]$ can be derived easily form Eq. (2) by setting $[L]_{free} = [L]_{tot} - PL$ and $[P]_{free} = [P]_{tot} - [PL]$, and solving the resulting quadratic equation.

$$[PL] = \frac{[L]_{\text{tot}} + [P]_{\text{tot}} + K_D - \sqrt{([L]_{\text{tot}} + [P]_{\text{tot}} + K_D)^2 - 4[L]_{\text{tot}} \cdot [P]_{\text{tot}}}}{2} \tag{4}$$

2.2. Significance of K_D for NMR experiments

For solving structures of a protein-ligand complex, the concentration of complex $[PL]$ should be maximized in the sample. As we will see in the section on sample preparation, the ratio of complex over free protein is also very important. If there is too much free protein in the sample, the NOEs recorded on the sample will be contaminated with those from the free protein, and a structure determination will be very difficult if not impossible. However, by adjusting the concentration of protein and ligand in the sample, the fraction of bound – or occupied – protein can be adjusted. Typically, a value larger than 0.8 is aimed at. Fig. 2 gives an overview of multiple combinations of protein and ligand concentrations typical for NMR experiments and the resulting population of the bound state, depending on the affinity of the complex.

2.3. Kinetics: on- and off-rates, k_{on}, k_{off}, and the observable exchange rate, k_{ex}

As mentioned previously, complex association and dissociation occur at the rates termed k_{on} and k_{off}, respectively. On-rates can be as high as $10^9\ s^{-1}\ M^{-1}$, which describes the diffusion limit of molecular collisions with the correct orientation for ligands hitting the binding pocket[1]. On-rates can also be considerably slower, for instance if slow conformational changes of the protein are a

[1] We use here the value of $10^9\ s^{-1}\ M^{-1}$ as the maximal on-rate of small molecules, which was originally determined for diffusion controlled enzymatic reactions [22]. For drugs, measured values, or values inferred from k_{off} and K_D measurements, range up to $10^8\ s^{-1}\ M^{-1}$, and may be higher for fragments [23]. Higher on-rates ($10^{10}\ s^{-1}\ M^{-1}$) have been shown for enzyme-substrate pairs where electrostatic attraction increases the probability of productive encounters. For protein-protein interactions maximal on-rates are in the order of $10^6\ s^{-1}\ M^{-1}$ [24].

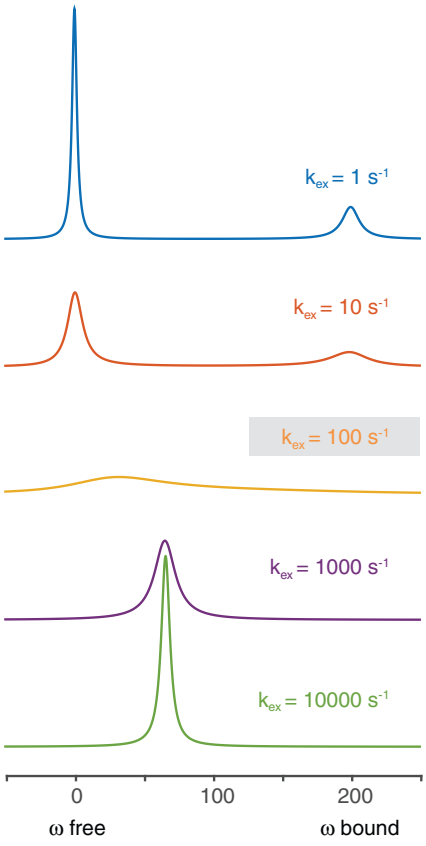

Fig. 3. Illustration of line broadening effect on ligand resonance induced by protein binding with different exchange kinetics. Spectra are calculated with a ligand concentration of 1500 μM, a protein concentration of 500 μM and a K_D value of 10 μM. The resonance frequencies of the free and the bound ligand protons (single proton) are 0 Hz and 200 Hz with linewidths of 5 Hz and 20 Hz, respectively (line-widths of free small molecules are typically below 1 Hz, but for this illustration, this would have resulted in overly high ligand signals). The signal of the free ligand is larger than the signal of the bound ligand, because of the excess of ligand over protein and the smaller linewidth of free ligand. The resonances are progressively broadened when the exchange rate k_{ex} increases, almost disappear for intermediate exchange (k_{ex} underlined in grey) and are finally merged in a single averaged sharp line in the fast exchange regime.

prerequisite for binding. The off-rates depend on the affinity of the ligand in the following way:

$$K_D = \frac{k_{off}}{k_{on}}, \tag{5}$$

which can be derived by using the equilibrium condition $k_{off}[PL] = k_{on}[P]_{free}[L]_{free}$, and rearranging to $k_{off}/k_{on} = [P]_{free}[L]_{free}/[PL]$ and using the relation shown in Eq. (2).

The on- and off-rates are not directly observable in NMR experiments. The observable quantity is the exchange rate k_{ex}, which describes the sum of total exchange events for the ligand or the protein, and can actually be different for the two.

$$k_{ex}^{L} = k_{on}[P]_{free} + k_{off} \tag{6a}$$

$$k_{ex}^{P} = k_{on}[L]_{free} + k_{off} \tag{6b}$$

Eq. (6) can alternatively be expressed in the form $k_{ex}^{L} = k_{on}([P]_{free} + K_D)$ using the relation shown in Eq. (5).

2.4. Significance of exchange kinetics for NMR spectroscopy

The binding kinetics play also a major role in the protein-ligand NMR experiments. We define three types of kinetics regime or exchange regime: slow exchange, intermediate exchange and the fast exchange (Fig. 3). These three exchange regimes are defined with respect to the difference of an NMR parameter (A) between the bound and free state (Fig. 4).

$$\text{slow exchange regime : if } |A_{bound} - A_{free}| \ll k_{ex} \tag{7a}$$

$$\text{intermediate exchange regime : if } |A_{bound} - A_{free}| \approx k_{ex} \tag{7b}$$

$$\text{fast exchange regime : if } |A_{bound} - A_{free}| \gg k_{ex} \tag{7c}$$

where the parameter A can be for example a relaxation rate or the chemical shift. On the chemical shift time scale, this therefore concerns the change in chemical shift in Hz between the free and bound states.[2] If k_{ex} matches the chemical shift difference between free and bound state, which often happens in the range of 100 s^{-1} (corresponding to ΔCS = 0.2 ppm on a 500 MHz spectrometer) this defines the mid-point of this time scale, the intermediate exchange regime. Two orders of magnitude above (10,000 s^{-1}) and below (1 s^{-1}) is then well within the fast and slow exchange regime.

In the slow exchange regime, NMR spectra exhibit two sets of signals with the individual properties of each state. The spectroscopic properties, chemical shifts and relaxation times, of these species are practically independent from each other. Regarding the linewidths, the protein peaks are broad (>10 Hz), the same is true of signals of the complex, while the free ligand peaks are narrow (<2 Hz).

If the exchange kinetics are faster, as for a naturally weak binder or by changes in the temperature or viscosity, the spectroscopic properties of the different species start to average. Before reaching complete averaging, the so-called intermediate exchange regime is crossed. In the intermediate exchange regime, the spectroscopic properties cannot be easily predicted and the resulting NMR signals are difficult to interpret (Fig. 3). We will therefore always try to avoid this blinding exchange regime by modifying the experimental conditions such as the temperature, the viscosity, the protein-ligand ratio, the spectrometer field, etc. Increasing for example the temperature will therefore shift the kinetics towards the fast exchange regime, while increasing the viscosity of the sample will shift the kinetics towards the slow exchange regime.

In the fast exchange regime, which is typical for weak binders, the spectroscopic properties of the ligand and the protein are averaged and depend on the population-weighted average of the bound and the free properties of the ligand and the protein.

$$A_{avg} = A_{free} \cdot p_{free} + A_{bound} \cdot p_{bound} \tag{8}$$

where A can be the chemical shift, but also any type of relaxation rate. For example, under fast exchange conditions, the chemical shift (CS) of the ligand is population-averaged to CS_{avg}. The auto-relaxation and cross-relaxation rates of the protein-ligand, which are typically in the range of 0.5–10 s^{-1} also follow a population averaging. These rates are important for the structure determination of the complex since they define the NOEs, and R_1.

It is therefore important to characterize a protein ligand interaction by determining its affinity, K_D, and assessing the kinetic exchange regime, which will be discussed in the following section.

3. Requirements for attempting a structure determination of protein-ligand complexes

Before attempting a work-intense and expensive protein-ligand structure determination, a number of parameters should be

[2] Note that we use Hz as the unit for the chemical shift and s^{-1} for rate constants, to emphasize the distinction between a periodic modulation and stochastic events, respectively.

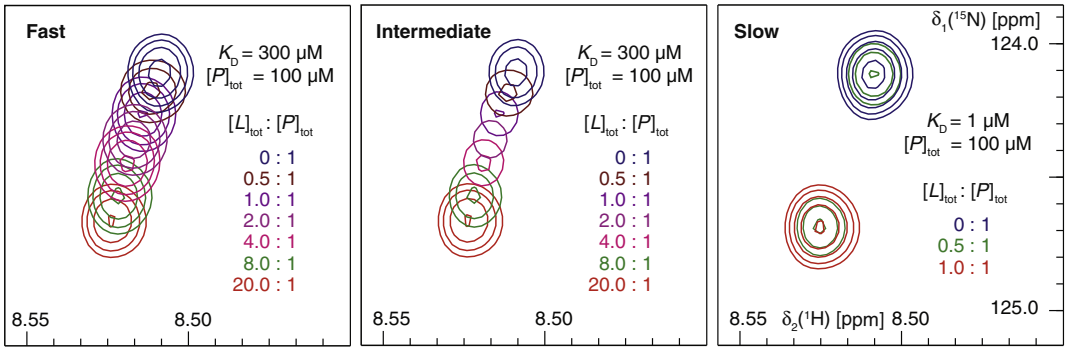

Fig. 4. Qualitative determination of exchange regime. [^{15}N,^{1}H]-HSQC titration spectra of a protein with a ligand, in fast exchange (left), intermediate exchange (middle) and in slow exchange (right). Experimental details are written on the panels. Note that for different signals in the protein, different exchange regimes may apply, depending on the chemical shift difference in Hertz between the bound in free signals in the ^{1}H and ^{15}N dimension.

determined, which will allow preparation of an optimal sample, and assessing the feasibility of the structure determination.

3.1. Parameters for characterizing a protein-ligand complex

The most important parameters characterizing a protein-ligand complex are its affinity (K_D), the exchange kinetics of the complex (k_{ex}) and the concentration of protein $[P]_{tot}$ and ligand $[L]_{tot}$ in the sample. For understanding the limitations for preparation of a sample of a complex and for enabling or the potentially best preparation of such a complex, maximal achievable concentrations of protein and ligand should be known, i.e. their solubility, and their long-term stability in solution. Therefore, methods for accurate concentration determination will now be summarized, and then K_D determination and k_{ex} assessment will be discussed.

Protein concentration determination is most precisely achieved in practice with HPLC-UV$_{215}$ (high-pressure liquid chromatography with UV detection at 215 nm), where a protein preparation is separated on a reversed-phase column under denaturing conditions (80 °C, water-acetonitrile mixtures, spiked with TFA) and then quantified based on the UV light absorption of the peptide bonds at 215 nm. This method has the advantage that small contaminants in the protein preparation are separated and the signal which is used for quantification is essentially only stemming from the protein of interest. The peptide bond has a constant extinction coefficient across all proteins, and therefore a HPLC-UV$_{215}$ system can be calibrated with a standard protein solution (typically serum albumin at 1 mg/mL). A less expensive alternative is measuring UV light absorption at 276 nm in bulk samples. This will always result in an overestimation of the concentration due to non-100% purity of the sample and additional absorption of contaminating proteins. Absorption at 276 nm is dominated by aromatic side-chains of the protein and can be calculated. The calculated extinction coefficients are however only valid for denatured proteins, a fact that needs to be kept in mind if measuring non-fully denatured proteins.

For assessing the maximal solubility and stability of a protein, the protein is concentrated until first signs of precipitation occur. The solution is centrifuged and the supernatant concentration is measured. The supernatant is then incubated at 25 °C (or another temperature if NMR experiments for structure determination are indented to be carried out at another temperature) for a week, and its concentration is measured every two days to determine long-term stability. Since the NOESY experiments typically take several days to record, it must be ensured at this point that the protein is long-term stable under the chosen buffer conditions.

Ligand concentration is most conveniently measured by NMR, because typically, just weighing in an accurate amount of ligand is often not precise, due to limited solubility of the ligand or due to contaminants in the powder. To determine the concentration of a ligand in solution by NMR, a reference substance – usually DSS – at precisely known concentration is mixed with the ligand and the ^{1}H NMR signals are compared, taking the multiplicity of the hydrogen signals into account (e.g. a 0.11 mM solution of DSS will yield a 1 mM proton signal, due to the 9-fold multiplicity of the signal at 0 ppm). For accurate concentration measurements, it is important to use long relaxation delays (>5 s). For determining the solubility, two solutions with nominal concentrations above the expected solubility are prepared (e.g. 1 mM and 2 mM, the range depends on the expected solubility) and the concentration is determined by the method above. There are three possible outcomes of this experiment, which are illustrated with example numbers. (i) If the respective concentrations determined for both samples are 1 and 2 mM, the compound is pure and solubility is $\geq$2 mM. (ii) If the concentrations are 0.7 and 1.4 mM, the compound is 70% pure and solubility is $\geq$1.4 mM. (iii) If both concentration measurements yield the same value (e.g. 0.7 mM) then this corresponds to the solubility of the ligand in this buffer.[3]

The next step in characterizing a protein-ligand complex is measuring its dissociation constant (K_D). This is ideally directly performed by NMR, because the HSQC spectra that are recorded to this end also contain information about the exchange kinetics of the complex. There are several methods for determining K_D values by NMR [26], in the context of structure determination, however, HSQC-based methods are chosen, since labelled protein must anyway be available. Ideally, twenty point titrations would be carried out with a protein concentration similar to K_D and a ligand concentration spanning two orders of magnitude around the K_D ($K_D/10 < [L] < 10 \times K_D$).[3] In practice, a good estimate can be derived from a four- to six-point titration, carried out with constant protein concentration and logarithmically increasing ligand concentrations (e.g. 0, 50, 100, 200, 400, 800, 1600 µM, or 0, 30, 100, 300, 1000, 3000 µM). The magnitude of changes in the protein spectrum are fitted to the equation describing the bound fraction of protein in a binding equilibrium (Eqs. (3b) and (4)). Potential pitfalls are that changes in the protein spectrum can arise as well from pH, temperature, contaminants of the ligand (e.g. counter ion, degradation products) as well as the solvent of the ligand stock solution (e.g. DMSO). Appropriate controls should be in place in order to account for these effects [27].

The main limitation of this protein-observed method for K_D determination is that relatively high concentrations of protein (>20 µM on a modern cryoprobe [28]) are needed in order to be able to record an ^{15}N or ^{13}C-HSQC spectrum. This sets the minimum for

[3] As a comment: It would be desirable to determine concentrations of protein and ligand with the same method. For NMR, PULCON would be the method of choice, it however depends on having access to enough measurement time during all the preparatory experiments [25].

the K_D that can be determined to about 10 μM. This is not an important limitation, as most complexes with K_D < 10 μM can be prepared as nearly saturated complexes for structural studies, and exact knowledge of the value of the K_D is therefore not essential. There are other biophysical methods for determination of dissociation constants below 10 μM, like surface plasmon resonance or isothermal titration calorimetry [29]. Alternatively, K_D values be calculated from IC_{50} values of biochemical assays if exact concentrations and affinities of the reagents in the assay are known [30]. Since the precise value of K_Ds below 10 μM are not so important for a structure determination, it depends on the work involved in establishing such an additional assay, whether it should be carried out.

The final parameter important for characterizing a protein-ligand complex is the exchange rate (k_{ex}) between the bound and free forms of protein and ligand. This rate does not need to be quantified exactly, but it is important to know whether k_{ex} is slow, intermediate or fast in regard of the chemical shift time scale, i.e. k_{ex} is smaller, similar or larger, respectively, than the chemical shift difference of the free and bound state resonances in Hz. Values for slow, intermediate and fast are approximately $k_{ex} < 10\ s^{-1}$, $10\ s^{-1} < k_{ex} < 1000\ s^{-1}$ or $k_{ex} > 1000\ s^{-1}$, respectively. Qualitatively, fast and slow exchange regimes are easily identified from HSQC spectra of titrations [31]. If protein signals shift in response to increasing ligand concentrations, this indicates fast exchange kinetics. If protein signals gradually disappear and new signals appear in response to increasing ligand concentrations, this indicates slow exchange kinetics. For complexes with intermediate exchange kinetics, signals can be severely broadened due to exchange and are often not detectable, therefore, in a titration, signals progressively broaden in response to increasing ligand concentrations, but no new signals appear.

Pragmatically, signals of the molecular interface need to be visible in order to determine a structure. Therefore, precise quantification of exchange kinetics is not required. However, it is important to know whether the complex exhibits fast of slow exchange kinetics, because this has consequences for sample preparation.

Parameters for characterization of a protein-ligand complex

- Protein concentration and stability in sample
- Ligand concentration and stability in sample
- Dissociation constant (K_D) of complex
- Exchange kinetics of complex (k_{ex}: slow, intermediate, fast)

3.2. Assessing feasibility of a complex structure determination

Feasibility of a structure determination of a protein-ligand complex depends on several factors. The protein solubility must be high enough to enable recording NOESY spectra with sufficient signal to noise ratio. On modern spectrometers, concentrations as low as 200–300 μM may be sufficient for a small protein. Most NMR methods rely on having obtained the resonance assignment of the protein, which therefore stands as a stringent pre-requisite defining feasibility.

Ligand solubility is another key factor determining feasibility. The ligand must be soluble enough in order to highly saturate the protein. Ligand solubility is linked to the maximal population of bound protein that can be achieved for a protein-ligand pair with a given K_D. In the slow exchange case, if the bound fraction of protein is much lower than 80%, the spectrum is contaminated with signals of the free protein, which leads to heavy signal overlap and makes it very difficult if not impossible to pick the correct signals. In the fast exchange case, the bound fraction can be rather low, as only one set of signals is observed. In particular, for the intra-molecular NOEs on the ligand and inter-molecular NOEs detected on the ligand a low population of bound ligand is tolerable and sometimes even advantageous. However, for protein observed experiments, averaged NOEs of bound and free populations hamper analysis significantly, therefore, it is also advisable to maximize the population of bound protein and 80% is a good minimal target value.

As already mentioned in the preceding paragraph, slow and fast exchange kinetics have different influences on spectra and can impact feasibility in different ways. Intermediate exchange kinetics, however, will almost always render a structure determination impossible. Intermediate exchange kinetics lead to broadening of signals often to below the detection limit and hinder collection of NOE data for the binding interface. It is therefore important to examine protein HSQC spectra and 1D filtered ligand spectra for signs of severe line broadening due to intermediate exchange. If a complex is in the intermediate exchange regime, the kinetics may be influenced by changing temperature, changing the viscosity of the solution, changing the protein/ligand ratio, or ultimately by chemically modifying the ligand.

Finally, the quality of the ligand and its NMR spectrum are important. If a ligand contains only very few hydrogen atoms and if these are all located at one end of the ligand, it will not be possible to determine the exact orientation of the ligand in the binding pocket. Ideal ligands have several hydrogen atoms, which are equally spread over the entire ligand, therefore allowing obtaining distance restraints between all parts of the ligand and the protein. An additional pre-requisite for large coverage of the ligand is that the ligand signals need to be well-dispersed and that the resonance assignment is possible. If ligand signals all cluster, only ambiguous distance restraints can be obtained. Therefore, it is advisable to examine the spectra of several potential ligands, and if the resources are available, synthesizing suitable ligands with more hydrogens or with groups that increase the chemical shift dispersion of the hydrogens in the molecule.[4]

Parameters determining feasibility of a structure determination of a protein-ligand complex

- Protein solubility (>300μM)
- Ligand solubility (>K_D)
- Minimal population of bound protein (>0.8 achievable)*
- Exchange kinetics (slow or fast for signals to be visible, e.g. $k_{ex}<1s^{-1}$ or $k_{ex}>1000s^{-1}$)**
- Quality of protein spectrum (dispersion, resonance assignments available)
- Quality of ligand and its spectrum (number and distribution of hydrogens on ligand, dispersion of NMR signals, resonance assignment)

* The minimal population does not apply for ligand-observed trNOE experiments

** Fast and slow exchange limits depend on individual chemical shift differences of free and bound states and the external magnetic field. The numbers given here are rough boundaries.

[4] As a side note: interestingly, the feasibility of a NMR structure determination of a complex can be predicted rather well from the parameters discussed above. For X-ray crystallography, it is not possible to predict, whether a certain compound will co-crystallize or soak into a crystal.

Table 1
Suggested relative concentrations of protein and ligand depending on exchange kinetics and on the parts of a protein-ligand complex that will be characterized. In all cases a fraction of bound species larger than 0.8 should be aimed at, except where stated otherwise.

		Part of complex to be studied			
		Bound ligand	Inter-molecular	Bound protein	Full complex
Exchange rate	k_{ex} fast	$[P]_{tot} << [L]_{tot}$ 1:5–50 $[PL]/[L]$ = 0.2–0.02	$[P]_{tot} < [L]_{tot}$ (max $[PL]$)	$[P]_{tot} < [L]_{tot}$ (max $[PL]$ and min $[P]_{free}$)	$[P]_{tot} < [L]_{tot}$ (max $[PL]$ and min $[P]_{free}$)
	k_{ex} slow	$[P]_{tot} \geq [L]_{tot}$ (min $[L]_{free}$)	$[P]_{tot} \approx [L]_{tot}$ (max $[PL]$)	$[P]_{tot} \leq [L]_{tot}$ (min $[P]_{free}$)	$[P]_{tot} \leq [L]_{tot}$ (min $[P]_{free}$)

4. Sample preparation

Sample preparation depends on the exchange kinetics of the complex and on the experimental aim, e.g. determining the structure of the entire complex or just of the bound conformation of the ligand. Therefore, in the following discussion complexes exhibiting slow exchange kinetics are treated separately from complexes in fast exchange, and then different experimental aims are discussed. Our suggestions for optimal sample conditions for these different cases are summarized in Table 1.

A pre-requisite for optimal sample preparation is the precise knowledge on concentration of the stock solutions and the K_D, and a qualitative assessment of the exchange kinetics (see previous section).

For practical preparation of samples, ligands are usually added as highly concentrated d_6-DMSO stock solutions (50–100 mM), and the protein is previously checked for tolerance towards the solvent.

4.1. Preparation of complexes with fast exchange kinetics

As discussed in the introduction, in the case of fast kinetics, averaged signals of the bound and free states will be recorded. The suggested preparations are summarized in Table 1.

- Sample preparation for determination of structure of the ligand in the bound state
  For calculating the bioactive conformation of the small molecule bound to the protein, under fast exchange, the measurement of transferred-NOEs (tr-NOE) is the method of choice. In this case, the priority is to have a protein: ligand concentration ratio that is heavily biased towards the ligand, with 5–50-fold excess of ligand over the protein, depending on the size of the protein and the affinity of the ligand. This leads to very sensitive tr-NOE spectra. The calculated curves published by Campbell and Sykes [32] are a good theoretical basis for optimal sample preparation. However, one should not aim at the fraction of bound ligand yielding the maximal tr-NOE intensity. IN order to be able to interpret tr-NOE intensities to intramolecular distances, a large excess of ligand is necessary to avoid protein driven spin diffusion. tr-NOEs do not show a linear behavior with respect to the ligand bound fraction [32,33]. For a given mixing time, the tr-NOE will increase and then sharply decrease as the fraction of the bound ligand increases. In other words, depending on the fraction of the bound ligand the tr-NOEs intensities will be translated to different distances. Only at low ligand bound fraction the tr-NOEs magnetization intensity build-up has a linear behavior and can be correctly translated to distances. We advise to work at low ligand bound fraction, with 10–20 times less protein than ligand, and use an appropriately short mixing time (50–100 ms). One advantage to work with large ligand concentration and reduced protein concentration is that the spectra will be simpler to analyze as the protein signal vanishes. [32,33]
- Sample preparation for determination of the structure of the interface
  Here, the concentration of complex should be maximized. This can be achieved by using an excess of protein or of ligand. A large excess of ligand may lead to t_1-noise artefacts, but employing an excess of protein may be expensive.
- Sample preparation for determination of the structure or the protein in the bound state
  For the protein, and population weighted mixtures of NOEs of the free and bound state will be recorded that may be contradictory and the two states cannot be separated as in the case of slow exchange. It is therefore highly important to maximize the bound population of the protein to above 0.8. This often requires very high concentrations of ligand (>1 mM, Fig. 2). The difficulties often lie in the limited solubility of the ligand (treated below) and high ligand concentrations lead to t_1-noise ridges in spectra.
- Sample preparation for determination of the structure of the entire complex
  Here the same considerations as for determination of the protein structure apply. Excess of ligand is not an issue for obtaining intra-ligand NOEs and also not for inter-molecular NOEs. Therefore, determining a complex structure using a single sample doesn't require major compromises.

4.2. Preparation of complexes with slow exchange kinetics

- Sample preparation for determination of structure of the ligand in the bound state
  In this case a slight excess of protein is advisable in order to maximize the concentration of the bound ligand. It is however not that critical to minimize the concentration of the free ligand, because it will only yield very weak NOEs (if any at all) of opposite sign than the signals of interest, and usually don't represent a problem for spectral analysis. However, care should be taken not to have too high concentrations of free ligand, which could produce t_1-noise artefacts.
- Sample preparation for determination of the structure of the interface
  Here, the concentration of complex should be maximized, based on knowledge of the K_D value and using Eq. (3) and Fig. 2. For strong binders (K_D < 10 mM), unnecessary excess of ligand should be avoided, due to potential problems with t_1-noise. A convenient way of preparing an equimolar sample, is by adding a slight excess of the ligand to the protein and washing away the unbound ligand for example during a desalting step. The desalting step typically comes with a dilution of the sample and is thus usually followed by a concentrating protocol. In this step, great care should be taken to not introduce unlabeled contaminants coming from the concentrator devices; contaminants can be avoided by extensively washing the tools before use.

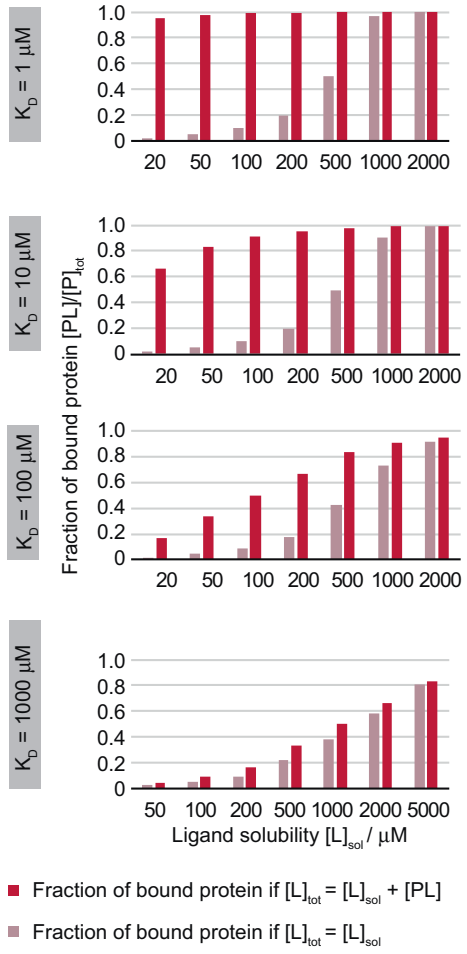

Fig. 5. Maximizing the fraction of bound protein ($p_B = [PL]/[P]_{tot}$) in situations of limited ligand solubility. The fraction of bound protein is plotted for a constant protein concentration of 1 mM and variable solubility values of ligands. If the ligand is added at its nominal solubility, only low fractions of the protein are populated (light red bars). Adding the ligand at the optimal concentration fully exploiting the ligand solubility ($[L]_{tot} = [L]_{sol} + [PL]$, see text and Eq. (9)), much higher values for the bound fraction can be reached (dark red bars), especially for low K_D values. Therefore, as an example a 1 mM protein sample can nearly be saturated with a ligand that has only 50 µM solubility if the K_D is smaller than 10 µM. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

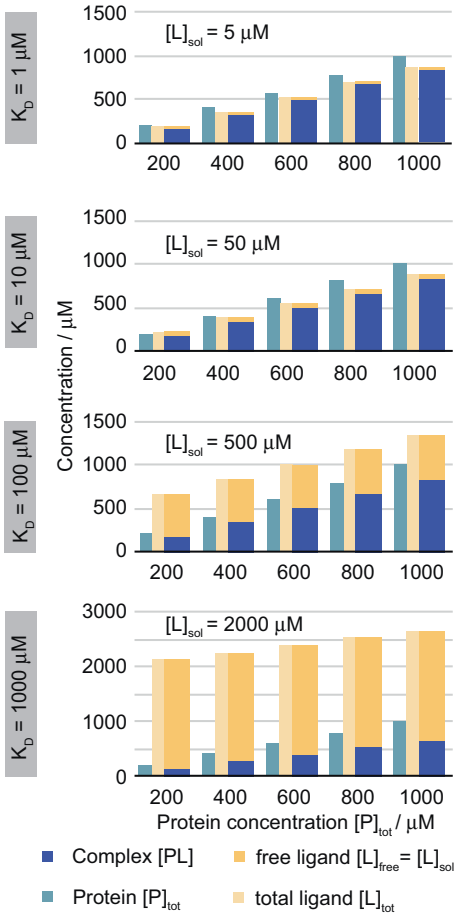

Fig. 6. Maximizing the concentration of protein-ligand complex [*PL*] in situations of limited ligand solubility The absolute concentration of complex (dark blue bars) is maximized in situations with limited ligand solubility. Total ligand and protein concentrations are shown in light yellow and light blue, respectively. The concentration of free ligand (dark yellow) is constant at its maximum solubility value. In this situation, the bound fraction of protein is constant for a given ligand (Eq. (9), compare light blue and dark blue bars). For the upper three examples the combination of ligand solubility and K_D yields a constant maximal p_B of 83%, for the lowest case 67%. It is evident that the absolute concentration of complex is linearly increased by increasing the concentration of protein, if of course ligand is added at the same time to fully exploit its solubility. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

- Sample preparation for determination of the structure or the protein in the bound state
  Here it is important to minimize the concentration of free protein in the sample by all means. A second set of signals from the free state of the protein will strongly impact subsequent analysis. Therefore, the fraction of bound protein should be pushed above 0.9.
- Sample preparation for determination of the structure of the entire complex
  Here the same considerations apply as for determination of the protein structure alone. Excess of free ligand is tolerable, since its free state doesn't contribute strong NOE cross peaks. Again, mind that unnecessary excess of ligand may introduce t_1-noise to the spectrum.

4.3. Considerations for cases where ligand solubility is limiting

Often ligand solubility is a limiting factor for preparation of highly populated complexes. However, it should be considered that ligands bound to the protein can be treated as if they had been removed from the solution in a first approximation. Therefore, the total ligand concentration in a protein-ligand mixture can be higher than the nominal solubility of the ligand in a given buffer ($[L]_{tot} = [L]_{free} + [PL]$). In terms of solubility only the concentration of free ligand ($[L]_{free}$) is limited.

For the limiting case, where $[L]_{free}$ is maximized to the solubility value $[L]_{sol}$, the K_D equation (Eq. (2)) can be solved using $[L]_{free} = [L]_{sol}$, which leads to the following expression of $[PL]/[P]_{tot}$ (i.e. the fraction of bound protein):

$$p_B^P = \frac{[PL]}{[P]_{tot}} = \frac{[L]_{sol}}{[L]_{sol} + K_D} \tag{9a}$$

$$[L]_{tot} = [L]_{sol} + [PL] = [L]_{sol}\left(1 + \frac{[P]_{tot}}{[L]_{sol} + K_D}\right) \tag{9b}$$

where $[L]_{sol}$ is the maximal solubility of the ligand. This is a much simpler expression than Eq. (4), and represents the asymptotic limit of Eq. (4) if $[L]_{free}$ approximates solubility. From the assumption that $[L]_{tot} = [L]_{sol} + [PL]$ and Eq. (9), optimal sample preparation in cases of limited ligand solubility can be derived.

4.3.1. Maximizing fraction of bound protein at limited solubility of the ligand

For example, let us consider a ligand with maximum solubility of 1 mM. If a sample is prepared with 1 mM of protein, for a 100 μM K_D about 0.75 mM of complex ([*PL*]) will result, if the ligand is added at its solubility limit (see Eqs. (3) and (4), and Fig. 2). However, in this situation only 0.25 mM of ligand is in the free state, because 0.75 mM are bound in the complex. The sample can therefore be topped-up with additional 0.75 mM of ligand before it starts precipitating. By increasing the total ligand concentration to 1.75 mM, the population of bound protein is raised to 0.9. This is a significant increase in the fraction of bound protein, which will simplify all subsequent steps in structure determination in important ways. (To finish the argument, the process of topping up the ligand concentration would need to be repeated again and again: In the second iteration, 0.9 mM of ligand is bound, and 0.85 is free. This allows adding yet another 0.15 mM of ligand staying within the solubility limit. This will again raise the concentration of bound ligand allowing for an additional small amount before reaching the solubility limit. This iterative approach yields the resulting value that can directly be calculated from Eq. (8))

For practical sample preparation, the optimal total ligand concentration is calculated from the previously determined K_D and $[L]_{sol}$ values using Eq. (9b) and added to the sample (Fig. 5). Alternatively, if co-precipitation of protein and ligand is not feared, a high excess of ligand can be added to the protein sample, and the resulting precipitate can be removed by centrifugation. We however prefer the more diligent way described above.

4.3.2. Maximizing the concentration of complex in situations of limited ligand solubility

In cases of limited ligand solubility, we have often encountered the following misconception about increasing the fraction of bound protein: intuitively the protein concentration is lowered on purpose (e.g. to 0.5 mM in the upper example) in order to have a larger excess of ligand over protein concentration because a higher fraction of bound protein is expected. However, Eq. (8) which applies to these cases is independent of the protein concentration. The fraction of protein occupied by ligand is therefore only depending on the affinity of the interaction and the concentration of free ligand in the solution, which under solubility-limiting conditions is constant if samples are prepared in the manner discussed above.

To understand the constant fraction of bound protein the following analogy may be helpful: In an analogy to partial pressures in ideal gas laws, one can think of the free ligand concentration as a partial pressure that a ligand can exert on the protein. If solubility is limiting the concentration of free ligand to a constant value, then the pressure exerted on the protein will always be the same, regardless of the protein concentration, and therefore p_B of the protein will be constant. Therefore, in order to increase the concentration of [*PL*], the species we are mostly interested in, the protein concentration must be increased and the ligand concentration accordingly (Fig. 6)

4.3.3. Potential additives for increasing ligand solubility

In some cases, additives may help to increase the solubility of the ligand. But here a distinction needs to be made. Additives like glycerol which are homogeneously distributed in the solution, may lead to truly higher solubility of the ligand, which will help achieving higher fractions of bound protein. In contrast, adding detergents may strongly increase the amount of ligand that can be added to a solution without precipitation. However, in most cases the ligand will be partitioned into the detergent micelle and not lead to higher occupation of the protein. The active species of ligand, is the free ligand in solution. Therefore, detergents, especially if added above their critical micelle concentration, are often not aiding in increasing the fraction of bound protein.

4.4. Minimizing content of protonated small molecules to avoid t_1-noise and baseline irregularities

For regular protein structure determination, one usually doesn't pay particular attention to buffer composition, as potential signals of protonated buffer are well suppressed by the HSQC element in the pulse sequence. When recording NOESY spectra of unlabelled ligands, however, no editing step suppresses the intense signals of unlabelled buffer components. Therefore, the buffer should be optimized by removing protonated substances as far as possible. This is relevant for spectral quality: Intense signals of small molecules are very difficult to subtract with phase cycling and typically serious t_1-noise bands result. Additionally, large diagonal signals may have extensive ridges with wiggles, which cover part of the signals of interest.

Due to these reasons, protonated buffer substances (like Tris, glycerol, DTT)[5] should be avoided or if they are essential for protein stability, they must be replaced by deuterated ones. Because this is expensive, only the final sample is transferred to the deuterated NMR buffer. We typically run the protein over two subsequent desalting columns equilibrated with deuterated NMR buffer, which yields 98–99% buffer exchange at little protein loss. In detail, 0.5 mL of NMR sample is run over a first NAP-5 column (GE Healthcare) and the resulting 1 mL of eluate is split in two. These two samples are run in parallel over another two NAP-5 columns equilibrated with deuterated NMR buffer. The resulting 2 mL of eluate are concentrated to the final desired volume (care must be taken to flush the concentrator with ddH_2O in order to remove unlabelled contaminants from its membranes), and topped up with ligand to the targeted ligand concentration.

4.5. Changing temperature, viscosity and concentrations of protein and ligand to avoid the intermediate exchange regime

Depending on the properties of the complex, one may be caught up in the intermediate exchange regime, where signals are so strongly broadened that it is nearly impossible to extract any information (Fig. 3). The parameters that affect the exchange regime with respect to the chemical shift difference are: (i) the external magnetic field (changing the chemical shift difference in Hz, Eq. (7)), (ii) the temperature and viscosity of the solvent (affecting the on and off-rates Eq. (6b)) and (iii) the concentrations of protein and ligand (affecting the concentration of free ligand – Eq. (6b)).

(ii) One can try to reach the slow or fast exchange condition, by lowering or raising the temperature, respectively. As a rule of thumb, by raising temperature by 10 K the rate constants are accelerated by about a factor of two, because the population of the transition state is increased according to Boltzmann's law (exponential term in Arrhenius equation). Additionally, increased temperature leads to more inter-molecular collisions due to faster molecular motions and lower viscosity of the solvent (first term in Arrhenius equation), moving the system further to the fast exchange regime. In order to move to the slow exchange condition, the temperature can be lowered and the viscosity of the solution can be increased, by e.g. adding glycerol up to 30% to the sample. This however, will lead to broader lines and potentially to a reduction of observable signals.

(iii) Besides the modifications of the physico-chemical parameters of the solvent, which affect k_{on} and k_{off}, alternatively,

[5] The ligand stock solution should be prepared in deuterated DMSO (typically as 50–100 mM solution). To prevent freezing of d_6-DMSO at 4 °C storage temperature, we add 10% D_2O.

it is possible to modulate k_{ex} by manipulating the protein and ligand concentrations. For example, by adding a large excess of ligand the concentration of free ligand can be increased so that the off-rate becomes less significant (Eq. (6b)). That is, if the ligand dissociates from the protein, it is immediately (e.g. within μs) replaced by a new ligand and the slow dissociation process is not manifested in the signal anymore.

Generally, it is difficult to calculate the effect of these changes on resonance lines of a given complex. Before running long NOESY experiments, the influence of temperature, viscosity and change of the external field should be empirically assessed by recording HSQC spectra under these modified conditions.

Often, these measures are often not sufficient to change the kinetics enough in order to obtain well-interpretable signals. Then it is often advisable to try to obtain a slightly modified ligand, which has weaker or stronger affinity and different exchange kinetics.

Preparation of protein-ligand samples

- Use Table 1 for suggested protein to ligand ratios, try to reach p_B of >0.8 (Eqs. (3) and (4), Fig. 2) except where stated differently
- Avoid the intermediate exchange regime (change magnetic field, temperature, viscosity and concentrations of protein and ligand. Eqs. (6) and (7), Fig. 3)
- If possible, avoid large excess of free ligand to minimize spectral artefacts

In cases of limited solubility of the ligand

- Use Eq. (9b) to calculate maximal possible ligand concentration
- Add ligand to protein solution, not vice versa, in order to exploit the full concentration of bound ligand (Fig. 5)

5. NMR experiments for recording and identifying intra-protein, intra-ligand and inter-molecular NOE cross peaks

Ligands can typically not be isotope labelled, therefore NMR experiments with so-called isotope-filtered or -edited dimensions are employed for collecting NOEs. That is, because in standard 3D NOESY experiments signals from the unlabeled ligand are either suppressed or not distinguishable from protein signals, special types of experiments are employed the allow identifying inter- and intra-molecular NOE signals involving an unlabeled ligand.

In this paper, we follow a modified nomenclature based on A. Breeze [34]: isotope-*filtered* means that hydrogen nuclei bound to an isotope labelled heteroatom (X = ^{13}C or ^{15}N) are *rejected*. While isotope-*edited* means that hydrogen nuclei bound to an isotope labelled heteroatom are *selected*. The information content of isotope-edited experiments can be increased by recording the chemical shift of the heteronucleus: in this case, the dimension is called isotope-*resolved* (resolving versus editing reduces sensitivity of the experiment by factor $\sqrt{2}$, due to quadrature detection on the X-nucleus).

In an isotope-filtered ^{1}H dimension therefore, only ^{1}H not bound to ^{13}C and ^{15}N are visible, termed $^1H^{!X}$ ($^1H^{!X}$ = *not* bound to X), i.e. typically resonances of the ligand. In X-edited dimensions only ^{1}H resonances bound to ^{13}C or ^{15}N are visible ($^1H^X$), i.e. typically from the protein.

5.1. Information content of conventional 3D NOESY experiments

First, the information content of a conventional NOESY will be discussed, which demonstrates the necessity for the filtered and edited experiments described below. A conventional NOESY would be called in our nomenclature an 3D F_3-resolved NOESY (often called "edited" NOESY, when the distinction between editing and resolving is not made). Here, only signals of the isotope labelled protein are recorded in the heteronuclear resolved dimension F_3, and signals of the unlabeled ligand are suppressed. In the F_1 dimension, however, all signals including unlabeled ligand signals are recorded. Therefore, intra-protein NOEs will be present as well as ligand-to-protein NOEs. In contrast, no intra-ligand NOEs will be recorded.

Theoretically, intra-protein and inter-molecular signal can be distinguished, although they are in the same spectral regions. Signals with a diagonal symmetric partner signal could be assigned as intra-protein NOEs. Conversely, ligand signals could be identified because the diagonal-symmetric signal, from the reverse magnetization transfer, is missing. However, in practice there are many reasons for missing signals in NOESY spectra and therefore a missing diagonal-symmetric signal is not a valid condition for identifying a ligand signal. This is why the filtered spectra are needed for unambiguously identifying the desired signals in simplified spectra.

5.2. NMR experimental elements for isotope editing and filtering

In the following, the basics of isotope filtering and editing techniques will be described, following the historical evolution starting with half-filter experiments, which are based on phase cycling – with potential subtraction artefacts – and work best for just one well-defined value of scalar coupling. To overcome these shortcomings, gradient-based purging schemes were developed that allowed filtering in a single scan; and adiabatic inversion pulses with matched sweep rates were employed that allowed dealing with largely different scalar coupling values of C-H moieties. Finally, our lab has developed complementary editing techniques, which allow inclusion of $^1H^{!X}$ resonances in isotope resolved dimensions, allowing to integrate intra-ligand NOEs and inter-molecular NOEs in sensitive 3D NOESY spectra.

5.2.1. Half-filter experiments

We will now turn to the details of the individual techniques. At the onset of NMR structural work on protein-ligand complexes, selection and rejection was achieved with so-called half-filter elements [35]. They exploit that ^{1}H nuclei that are covalently bound to ^{13}C or ^{15}N nuclei evolve according to the scalar one-bond coupling between them (e.g. $^1J_{HC}$ = 125–220 Hz and $^1J_{HN}$ = 93 Hz). Half-filter elements are simple spin echo sequences of defined length and appropriate pulsing, that refocus chemical shift but evolve J-couplings. After a period of $1/^1J_{HX}$, $^1H^X$ nuclei will have evolved into opposite phase than $^1H^{!X}$ nuclei bound to ^{12}C or ^{14}N (Fig. 7A). Therefore, the two species of ^{1}H nuclei can be distinguished. Experimentally, in one scan the *J*-coupling to the heteronucleus is allowed to evolve, in a second scan the *J*-coupling is refocused, with the effect that now signals of $^1H^{!X}$ and $^1H^X$ have the same phase. If scan one and two are subtracted, the $^1H^{!X}$ signals will be cancelled and $^1H^X$ signals will add up, leading to a spectral dimension containing exclusively $^1H^X$ signals. Conversely, by making the sum of the two scans, $^1H^X$ signals are cancelled and $^1H^{!X}$ signals add up (Fig. 7D).

5.2.2. Practical considerations when recording filtered experiments

The process for editing and filtering ^{1}H signals is based on subtraction of signals. For clean subtraction of signals a very stable

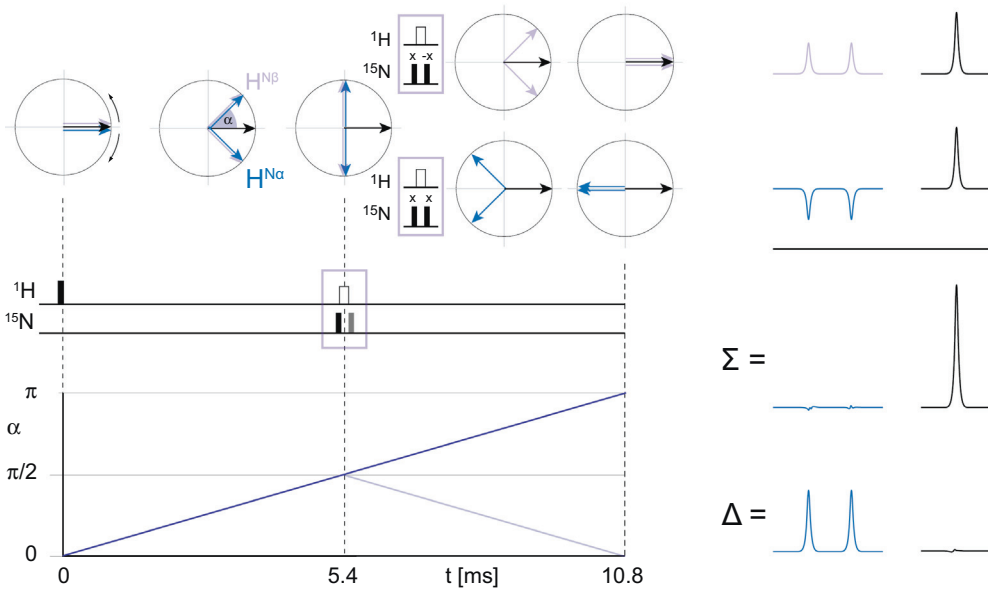

Fig. 7. Theoretical basis of half-filter experiments. (A) On top, magnetization vector diagrams of the x-y plane are shown at the time points indicated with dotted lines in the pulse scheme immediately below. Blue vectors represent H^N nuclei, black vectors $H^{!N}$ nuclei. (B) In the pulse scheme, 90° pulses and 180° pulses are represented by filled and empty squares, respectively. The phase of the grey 90° pulse on ^{15}N in the red box is alternated from one scan to the other. (C) Below the pulse sequence, a diagram showing the time evolution of the scalar coupling is drawn from $t = 0$ to $t = 1/^1J_{HN} = 10.8$ ms. The red box shows the two different transients that can be recorded with this experiment, depending on whether the second 90° pulse on ^{15}N has a phase of −x (light blue vectors and signals) or x (blue vectors and signals). For both transients, the further evolution of the magnetization vectors until the end of the refocusing period is shown (Only evolution due to scalar coupling is shown, since chemical shift evolution is refocused in this spin echo element). On the right-hand side, H^N (light blue, blue) and $H^{!N}$ (black) signals resulting from the two transients of the experiments are shown. Below the sum and the difference spectra are shown, illustrating how H^N or $H^{!N}$ signals can be suppressed selectively in the final spectrum. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

spectrometer setup is essential, otherwise subtraction artefacts lead to strong t_1 noise, which makes spectra difficult to interpret – especially NOESY spectra, where weak signals just above the noise level are rich in information. Filtered and edited NOESY experiments are therefore among the most demanding experiments for NMR hardware. Measures to avoid t_1 noise start with a vibration-free environment, stable temperature, stable shims (on modern NMR spectrometers we use automatic shimming features "autoshim", which we wouldn't have used in the past), powerful lock systems (make sure that enough lock substance is used) and end with proper setup of experiments. For the latter, it is important to include enough dummy scans. Small ligands often have long T_1 relaxation times and it takes several dozens of scans (>128) to reach a steady-state magnetization with stable subtraction results. A large number of dummy scans is also important in order to reach a stable temperature. Often, decoupling sequences heat the sample and it typically takes minutes for the temperature of the NMR sample to re-equilibrate. An additional aspect to subtraction artefacts is the resolution of the signals: in practice, broad signals (>10 Hz) tend to yield clean subtraction, while sharp signals (1 Hz) do lead to strong artefacts. It may often lead to visually cleaner spectra if strong (10 Hz) line broadening is also applied to rather sharp ligand signals in $^1H^{!X}$ dimensions.

5.2.3. Improvements on half-filter experiments: purging schemes and matched adiabatic pulses

The half-filter experiments described above are powerful methods, but they have two limitations: they may suffer from subtraction artefacts, which require rather long phase cycles to suppress these; and they only work well for a single defined value of the heteronuclear *J*-coupling. Both problems were addressed with technical advances. For isotope-filtering purge pulses or gradients were employed, which efficiently de-phase $^1H^X$ coherences. Here, a spin echo of a duration of $1/2J$ is used, which leads to anti-phase magnetization on $^1H^X$ that is orthogonal to $^1H^{!X}$ magnetization (Fig. 8a). At this stage, a spinlock pulse with phase y can be applied to purge the magnetization along the x-axis [36] (Fig. 8c). (Alternatively, the $^1H^X$ magnetization can be converted into unobservable multi-quantum terms by applying a 90° pulse on the X-nucleus. This approach is however not further discussed in this paper). The spinlock pulse has some disadvantages in terms of transverse relaxation and undesired effects on the water. Therefore, a refined variant of this element uses a −90° pulse along the x-axis to bring the $^1H^{!X}$ magnetization along the + z axis (Fig. 8d), while $^1H^{!X}$ magnetization stays in the plane, where it is de-phased by the gradient (Fig. 8e).

This gradient-based filtering element is only half as long as a half-filter element and allows filtering in a single scan. It is however typically applied twice in order to further suppress residual $^1H^X$ signals, which leads to superior results than half-filter elements. That is, if a certain moiety is only suppressed to 10% in a single purge filter element, it will be reduced to 1% by repeating the filter element.

The other limitation of half-filter experiments is that they only work efficiently for the one value of the scalar coupling constant to which the spin echo delay was matched. This works well for the rather homogenous *J*-couplings in H-N moieties, but doesn't so for C-H moieties, where *J*-couplings vary form 125 Hz for methyl groups, over 150 Hz for the α–position to values above 200 Hz for aromatic groups (Fig. 9A and B). To solve this problem, adiabatic pulses which sweep through the spectral window and invert ^{13}C nuclei in a time dependent manner are used [37,38]. They are designed such that aromatic carbon nuclei are inverted first and aliphatic nuclei last. This leads to different effective periods during which the scalar coupling is evolving and refocused. Interestingly, there is an approximate correlation between chemical shift of ^{13}C and the $^1J_{CH}$ coupling. This was exploited by matching the rate of the adiabatic frequency sweep to the optimal refocusing period

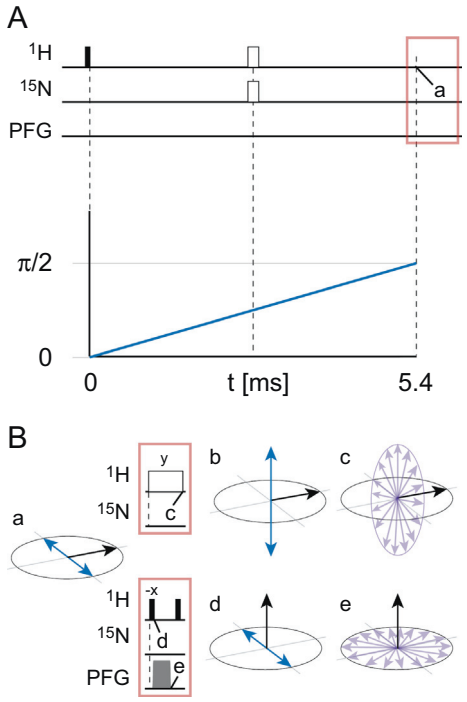

Fig. 8. Purging elements as shorter alternatives to half-filter elements. Similar representation as in Fig. 7 for the half-filter experiment and drawn to scale. In purging elements, magnetization is only evolved into anti-phase magnetization (time point a), which requires only half the time as the full refocusing to in-phase magnetization as in the half-filter experiments. At the end of the spin echo, either a spin lock pulse with phase y is applied leading to purging of the magnetization along the x-axis (time point c, upper vector model). Alternatively, a 90° pulse with phase −x can be applied on ^{1}H, flipping the $H^{!N}$ magnetization along the +z-axis (d). Now a gradient pulse is applied, efficiently dephasing the magnetization of H^{N} nuclei (e). Purging by gradients is preferred, because the desired $H^{!N}$ magnetization is stored along the z-axis for the gradient duration, leading to lower relaxation losses than if the magnetization is along the y-axis during purging, where transverse relaxation is active.

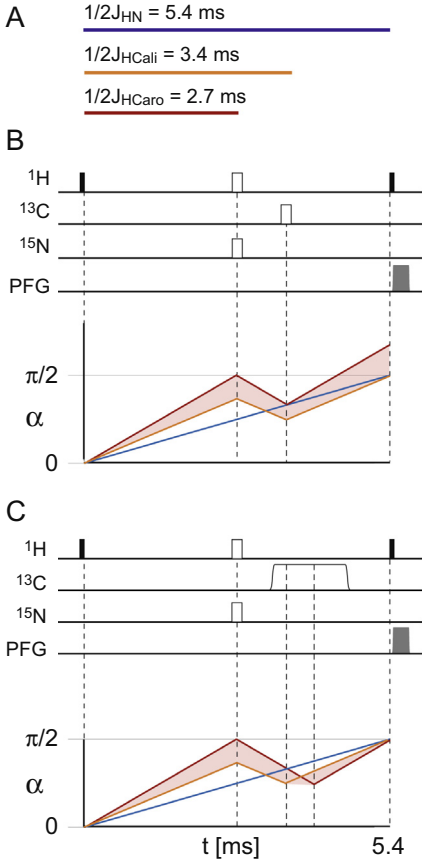

Fig. 9. Matched sweep adiabatic pulses for refocusing different J-coupling values. (A) Values of scalar couplings for H-N, H-C(aliphatic) and H-C(aromatic) moieties, and the respective duration of the scalar coupling evolution delay of 1/2*J*. (B) Similar representation as in Fig. 8. The half-filter (Fig. 7) and purging elements (Fig. 8) only work for one value of *J*-coupling (or integer multiples of it). In the upper scheme, a gradient purging element is shown, which simultaneously supresses ^{1}H nuclei bound to ^{15}N and ^{13}C. This element is optimized for aliphatic ^{13}C with scalar coupling values (bright red line) of $^{1}J_{CH} \approx 125$ Hz. With these parameters, ^{1}H nuclei bound to aromatic carbons (dark red line) are only partially suppressed. To also fully suppress aromatic signals, the same purging element needs to be repeated with optimal parameters for aromatic $^{1}J_{CH} \approx 190$ Hz. (C) In the lower scheme, an adiabatic pulse is used on ^{13}C, which inverts aliphatic and aromatic carbons sequentially, with a timing defined by its frequency-sweeping rate. If the frequency sweeping rate is matched to the empirical value of ~220 ppm/ms, scalar couplings to aromatic and aliphatic carbons are both optimally refocused. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of specific groups [37]. This allows efficient suppression of $^{1}H^{C}$ (Fig. 9C).

In the case of large proteins (MW > 50 kDa), the length of the filtering element may be reduced to account for serious loss of signal due to transverse relaxation. The perfectly matched adiabatic pulse is rather long and often applied twice for optimal suppression keeping the desired signal transverse for a long period of time (several milliseconds). It may then be required to reduce the length of the adiabatic pulse (to ~500 μs for example) and the refocusing delay, but it is at the cost of the protein resonances filter efficiency.

The discussed isotope filter and editing elements are highly efficient, but it must be emphasized that there will always be leakage of residual ^{12}C and ^{14}N-bound protons, from slowly exchanging hydroxyl protons and bound water, and from non-matched *J*-coupling values.

5.3. Variants of isotope editing experiments

Double purge elements using adiabatic pulses with matched frequency sweeps are the state of the art in isotope filtering and editing methods. Nowadays, simultaneous ^{15}N- and ^{13}C-filtering and -editing are performed in order to save measurement time. These isotope -editing, -filtering and -resolving elements can be combined in different ways to selectively yield intra-protein, intra-ligand or inter-molecular NOEs. Table 2 gives an overview of variants of ^{1}H,^{1}H-NOESY spectra with different combinations of heteronuclear -editing, -filtering and -resolving elements, and lists the NOEs, which can be identified in the respective spectra.

In the following, a number of aspects specific to individual experiments are discussed. This includes the directionality of inter-molecular NOEs – whether protein-to-ligand NOEs or ligand-to-protein NOEs are recorded – and alternative ways of isotope editing and filtering based on time-proportional phase incrementation of ligand signals and by using transferred NOEs.

5.3.1. Filtered-edited or edited-filtered NOESY experiments for recording inter-molecular NOEs?

For recording inter-molecular NOEs, a 2D F_1e,F_2f NOESY spectrum (4) (or the 3D version F_1r,F_3f NOESY with chemical shift evolution in F_2 for resolution on the heteroatom) can be recorded yielding $^{1}H^{X}$–$^{1}H^{!X}$ NOEs. Theoretically, the reversed version of this experiment, a 2D F_1f,F_2e spectrum (3) (or its 3D version), would yield equivalent $^{1}H^{!X}$–$^{1}H^{X}$ NOEs. At a closer look however, the two pathways above are not equivalent: In the former F_1e,F_2f experiment NOEs originating on the protein are recorded on the ligand; in the latter F_1f,F_2e experiment NOEs originating on the

Table 2
Variants of isotope-filtered and -edited $^{1}H,^{1}H$-NOESY experiments.

	$^{1}H,^{1}H$ NOESY[a]	Indirect ^{1}H dimension[b]	Direct ^{1}H dimension[b]	NOEs[c]	High resolution ^{1}H dimension[d]	Scheme[e]
1	2D tr-NOE			$^{1}H \rightarrow {}^{1}H$	Ligand	
2	2D F_1f,F_2f	filtered	filtered	$^{1}H^{!X} \rightarrow {}^{1}H^{!X}$	Ligand	
3	2D F_1f,F_2e 3D F_1f,F_3r	filtered	edited (2D) resolved (3D)	$^{1}H^{!X} \rightarrow {}^{1}H^{X}$	Protein	
4	2D F_1e,F_2f 3D F_1r,F_3f	edited (2D) resolved (3D)	filtered	$^{1}H^{X} \rightarrow {}^{1}H^{!X}$	Ligand	
5	3D F_3r		resolved	$^{1}H \rightarrow {}^{1}H^{X}$	Protein	
6	3D F_1r	resolved		$^{1}H^{X} \rightarrow {}^{1}H$	Ligand and Protein	
7	4D F_1f/r, F_4f/r	edited/filtered	edited/filtered	$^{1}H^{X} \rightarrow {}^{1}H^{X}$ $^{1}H^{!X} \rightarrow {}^{1}H^{X}$ $^{1}H^{X} \rightarrow {}^{1}H^{!X}$ $^{1}H^{!X} \rightarrow {}^{1}H^{!X}$	Ligand and Protein	
8	3D F_1r^{AI}	resolved (X and !X)		$^{1}H^{X} \rightarrow {}^{1}H$ $^{1}H^{!X} \rightarrow {}^{1}H^{X}$ $^{1}H^{!X} \rightarrow {}^{1}H^{!X}$	Ligand and Protein	
9	3D F_3r^{AI}		resolved (X and !X)	$^{1}H \rightarrow {}^{1}H^{X}$ $^{1}H^{!X} \rightarrow {}^{1}H^{X}$ $^{1}H^{!X} \rightarrow {}^{1}H^{!X}$	Ligand and Protein	

[a] The short name of the experiment describes the dimension (F_1, F_2, F_3 ...) and the isotope editing that is applied to this dimension (f = filtered, e = edited, r = resolved).
[b] Editing-scheme of the two ^{1}H dimension of a $^{1}H,^{1}H$ NOESY experiment.
[c] Directionality and the involved ^{1}H nuclei giving rise to an NOE cross peak visible in the respective spectrum are described: ^{1}H = all ^{1}H, $^{1}H^{!X}$ = ^{1}H not bound to ^{13}C and ^{15}N, $^{1}H^{X}$ = ^{1}H bound to ^{13}C or ^{15}N).
[d] The direct dimension is the ^{1}H dimension with the highest resolution. In the case described in d, signals of protein or ligand will be in the direct dimension.
[e] Scheme depicting observed NOEs and resulting spectrum (see Fig. 10 for more details). For a complex consisting of a labelled protein (blue) and unlabeled ligand (yellow), the types of NOEs that are observed are shown as arrows. If the originating or destination nuclei can be identified as protein ($^{1}H^{X}$) or ligand ($^{1}H^{!X}$) protons, the color of the arrow is set accordingly. A schematic 2D $^{1}H,^{1}H$ NOESY spectrum is shown, showing the four signals of the protein-ligand complex on the left. Signals clearly identifiable as originating from the ligand and the protein are shown in yellow and blue, respectively. Signals where this identification is not possible purely from the type of spectrum are shown in grey.

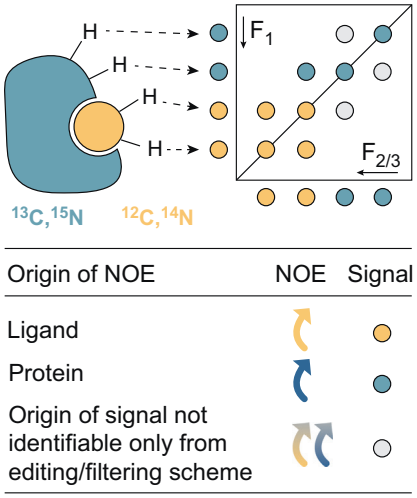

Fig. 10. Scheme used to describe variants of edited filtered experiments in Table 2. On the left-hand side, a complex of a protein (blue) and a ligand (yellow) is depicted. Selected ^{1}H nuclei of the protein and the ligand are shown (H). On the right-hand side, a 2D ^{1}H,^{1}H NOESY spectrum of this complex is shown. The position of the signals in the spectrum of the four individual hydrogen nuclei are assigned with the dashed arrows. NOEs and their directionality are represented by arrows. Signals are colored according to the origin of the NOE giving rise to the signal. If the signal can be identified based on the editing/filtering scheme of the spectrum its color is set accordingly (yellow for ligand, blue for protein). If the origin of the NOE cannot be identified by the editing/filtering scheme, the signal is shown in grey. e.g. in typical 3D F_3r NOESYs, ligand signals are recorded in F_1, but are a priori not distinguishable from protein signals. (Strictly speaking, ligand signals could be identified in this case because the diagonal-symmetric signal is missing. However, in practice there are many reasons for missing signals in NOESY spectra and therefore a missing diagonal-symmetric signal is not a valid condition for identifying a ligand signal.) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

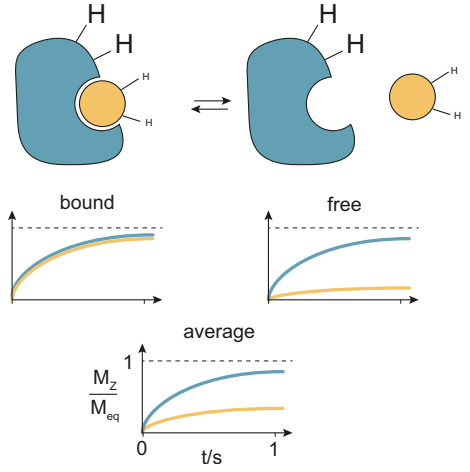

Fig. 11. Illustration of magnetization recovery in the case of fast exchanging small ligands. In the upper panel a scheme of a protein (blue) and a ligand (yellow) in fast binding exchange are shown. Selected ^{1}H nuclei of the protein and the ligand are shown (H). The size of the H-character indicates the magnetization available at the beginning of a scan. The diagrams below show the time course of magnetization of protein (blue) and ligand (yellow) during the recovery delay of e.g. 1 s in the bound and free states. While the protein has essentially the same R_1 value in the free and bound states, the ligand has a much smaller R_1 in the free state ($R_{1,F}$) and less efficient longitudinal relaxation, resulting in little magnetization after a recovery delay of only 1 s. For the complex in fast exchange the population weighted average ($R_{1,avg} = p_B R_{1,B} + p_F R_{1,F}$) is relevant. In low-affinity complexes a larger proportion of ligand is in the free state, therefore the $R_{1,avg}$ is dominated by the inefficient R_1 of the free state, leading to low magnetization on the ligand. Therefore, in most cases of fast exchanging complexes, inter-molecular NOE experiments should be started on the protein. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

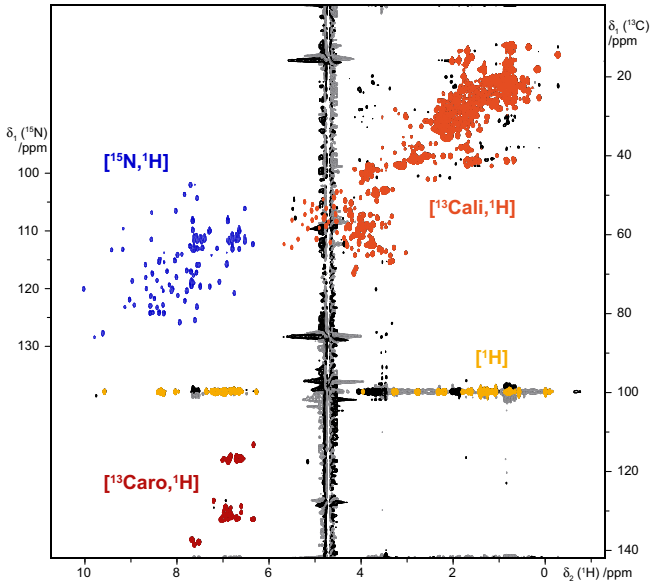

Fig. 12. Inclusion of ligand signals in a 3D All-Inclusive NOESY. The first 2D [X,^{1}H] plane of an All-Inclusive 3D $F_1 r^{Al}$ ^{1}H,^{1}H-NOESY spectrum is shown. [X,^{1}H]-correlations of aliphatic (bright red), aromatic (dark red) and amide (blue) moieties are colored. Unlabeled ligand signals are all recorded at an artificial chemical shift of 100 ppm in the ^{13}C dimension. In a single combined 3D NOESY therefore, all NOE cross peaks of a protein-ligand complex are recorded, including both equivalent pathways and the diagonal signals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ligand are recorded on the protein. These pathways are equivalent in tightly bound complexes, where protein and ligand have same T_1 and T_2 relaxation times. For weakly bound complexes, however, the small ligand spends a large fraction of time in the free state and the T_1 of the ligand is therefore much longer than the T_1 of the protein. In a typical NOESY experiment with relatively short relaxation delays (1 s), ligand signals recover only to a fraction of their equilibrium intensity, and therefore NOEs originating on the ligand will be much weaker than those originating on the protein. Therefore, for small weakly binding ligands in fast exchange, F_1e, F_2f experiments should be run. This yields strong inter-molecular NOEs. The low starting magnetization of the ligand can be balanced by increasing the ligand concentration. Theoretically, at concentrations exceeding the ratio of the averaged T_1 relaxation time of the ligand and the T_1 of the protein, it should be favorable to start the experiment on the ligand. In our experience, however, the F_1e,F_2f pathway is nearly always superior, which may be due to factors we haven't considered.

5.3.2. All-inclusive NOESY

In general, filtered-edited experiments have the advantage of simplification of spectra and clear identification of resonances. Each filtering and editing step however, lowers the sensitivity of an experiment. Filtering elements have a fixed length (e.g. 10.8 ms for suppressing ^{15}N-bound ^{1}H signals) during which transverse relaxation is active. For medium sized proteins, 10 ms can lead to a reduction of the signal to half its size. Additionally, several pulses are applied during this element. Since pulses are not perfect, each pulse also reduces the signal by a few percent. That is why in total, filtered-edited experiments with two long filtering or editing elements are rather insensitive compared to 3D X-resolved NOESY spectra, where only one long editing (resolving) element is present.

Our group recently developed a 3D NOESY, the All-inclusive NOESY, which has the same sensitivity as a normal 3D X-resolved NOESY but includes the ^{1}H$^{!X}$ ligand signals in the

Table 3
Suggested ^{1}H,^{1}H NOESY experiments for characterizing different parts of a protein-ligand complex depending on different exchange kinetics. The experiments are identified by the short notations following the same nomenclature as in Table 2.The bold numbers indicate the NMR experiment described in Table 2. The reasons for using individual experiments are given in the text. For reasons of simplicity, combined ^{13}C, ^{15}N editing and filtering is assumed [41].

		Part of complex to be studied			
		Bound ligand	Inter-molecular	Bound protein	Full complex
Exchange rate	k_{ex} fast	2D tr-NOE (**1**) (or 2D F_1f,F_2f (**2**))	2D F_1e,F_2f (**4**) or 3D F_1r,F_3f (**4**)	3D F_3r (**5**) or 3D F_1r (**6**)	3D F_1r^{Al} (**8**) (and **4**) or 4D F_1r/f, F_4r/f (**7**) or (**1**), (**4**) and (**5**/**6**)
	k_{ex} slow	2D F_1f,F_2f (**2**)	2D F_1f,F_2e (**3**) or 3D F_1f,F_3r (**3**)	3D F_3r (**5**) or 3D F_1r (**6**)	3D F_1r^{Al} (**8**) or 4D F_1r/f, F_4r/f (**7**) or (**2**), (**3**) and (**5**/**6**)

X-resolved dimension at a singular heteronuclear chemical shift (e.g. 100 ppm in the ^{13}C dimension, Fig. 12) (Gossert et al., to be published; pulse sequence and parameter set in Bruker format, setup description and a processing script are available online from the Bruker user library [39]). This combines the advantages of filtered-edited spectra – identification of inter- and intra-molecular NOEs – with the advantage of higher sensitivity of spectra with only one edited dimension.

The pulse-sequence of the All-Inclusive NOESY is essentially identical to a conventional 3D NOESY. However, in a conventional 3D NOESY several measures are taken to suppress unlabeled signals, which typically are unwanted intense buffer and solvent signals. These measures are gradients, phase-cycles on the heteronuclei and a TPPI scheme (time proportional phase incrementation). TPPI artificially moves residual unlabeled signals to the edges of the spectrum (producing so-called axial peaks). In order to observe signals of the unlabeled ligand, gradients and phase cycles are omitted. The now observable signals will however show up as axial signals at the extremes of the spectrum due to the TPPI procedure. Yet, by modifying the TPPI procedure these signals can artificially be moved to any part of the spectrum. In the case shown in Fig. 12, ligand signals were made to appear at 100 ppm in the carbon dimension. This spectral region is typically empty and therefore ligand signals will not overlap with signals of the labelled protein. Additionally, potential ridges from residual protonated buffer substances will not distort the baseline in regions of interest.

Since, in the All-inclusive NOESY, unlabeled signals of the ligand are not suppressed but appear at 100 ppm in the ^{13}C dimension, intra-ligand NOEs and inter-molecular ligand-to-protein NOEs are included and can as well be clearly identified. Therefore, the All-inclusive NOESY has a similar information content as an entire set of filtered-edited NOESY spectra, and is usually more sensitive than the experiments with two filtering-editing elements.

5.3.3. Filtering based on fast exchange: Transferred NOEs yield intra-ligand and inter-molecular protein-to-ligand NOEs

An additional, but completely different way of obtaining only a defined subset of NOEs of a protein-ligand complex is based on transferred NOEs [33]. Transferred NOEs can be recorded on small ligands in fast exchange (k_{ex} > 1000 s^{-1}, K_D typically >1 μM). Samples are prepared with a massive excess of ligand ([*L*]:[*P*] = 5–50). In samples with such high ligand to protein ratios, only ligand signals are observed, this is also due to the much sharper signals of the ligand which spends a large proportion of time in the free state. Therefore, no filtering in any dimension is needed and a simple 2D NOESY experiment can be recorded. The NOEs observed on the ligand signals are dominated by the positive cross peaks that build up in the bound state. For a small ligand of 300 Da bound to a protein of 30 kDa the positive NOE cross peak from the bound state is about 20-fold more intense than the negative NOE cross peak of the free state. Therefore, if the bound fraction is larger than 5%, positive cross peaks result, reporting on the structure of the bound state. Additionally, protein-to-ligand NOEs can be recorded, which also arise from the bound population. However, these inter-molecular signals are rather weak, because they arise from a small concentration of protein. A further advantage of this experiment with extreme ligand to protein ratio is that the ligand resonances are hardly different from the free state and assignment of ligand signals is trivial. The transferred NOEs observed in this simple experiment, are therefore often sufficient to determine the structure of the bound form of the ligand, and in some special cases, even the binding mode can be determined.

5.4. Selecting the appropriate experiments for recording intra-ligand, inter-molecular and intra-protein NOEs

From the above account of versions of ^{1}H,^{1}H NOESY spectra, with different isotope-filtering, -editing and -resolving elements, the suitable experiment for a given experimental situation needs to be chosen. An overview of appropriate experiments for a given task is given in Table 3. As for the optimal ratio of protein and ligand in a sample, the choice of the experiment type depends on two major considerations: the type of NOEs that should be recorded (only intra-ligand, intra-protein, inter-molecular or all together) and whether the complex exhibits fast or slow exchange kinetics on the NMR time scale.

5.4.1. Complexes with fast exchanging ligands

In the case of fast exchange kinetics (k_{ex} > 1000 s^{-1}), the ligand and the protein experience averaged properties of the bound and the free state ($R_{avg} = p_B R_B + p_F R_F$). This applies for NOEs, but also for R_2 and R_1. Therefore, always an average of the two states is observed. Sample preparation determines the fraction of free and bound states of ligand and protein, and therefore their relaxation properties and the intensity of NOEs that are observed. Since the relaxation properties change strongly for the ligand, and hardly for the protein, the focus of this part of the discussion lies on the ligand.

- Recording intra-ligand NOEs of the bound ligand
  Here, a 2D tr-NOE spectrum (1) is best suited. The ligand is observed, which has a favorable averaged R_2, which yields sharp lines. Additionally, this experiment is very efficient as it allows working with low concentrations of protein, exploiting an amplification factor of several visits of ligands to a single protein. The contribution of the free state to the NOE is very small (see text above) and can easily be identified because the NOE cross-peaks are negative.
- Inter-molecular NOEs
  Here a 3D F_1r,F_3f spectrum is usually suited best (4). Protein signals are resolved with their heteronucleus, which allows highly unambiguous assignment. In some cases, the 2D version of the experiment may be enough. For ligands in fast exchange the directionality of the experiment is important: the originating nucleus should be on the protein, the destination one on the ligand. This is due to the unfavorable long T_1 relaxation time of the ligand, which leads to inefficient relaxation during the

rather short recovery delay of these experiments. Therefore, the starting magnetization on the ligand will typically only be a fraction of that on the protein, and thus it is important to run the inter-molecular NOE experiment for fast exchanging small ligands as F_1r,F_3f. The disadvantage might be that the direct dimension with highest resolution is where the ligand NOE signals are recorded. If the ligand has a low-complexity spectrum compared to the protein, this can be seen as a waste of resolution. In our experience, however, this directionality of the inter-molecular NOE experiment allows detection of signals that are not visible if the ^{1}H on the ligand are the originating nuclei, and high resolution is often needed on the ligand, e.g. in peptides and drug molecules. See the text above for a more detailed discussion on the directionality of inter-molecular NOESY experiments.

- Recording intra-protein NOEs of the bound protein
  The standard 3D X-resolved NOESY spectra, as used for normal protein structure determination, are best suited. Either combined ^{13}C,^{15}N evolution is used or separate spectra are obtained, which are individually more sensitive, but may hamper analysis due to slight chemical shift mismatches. For the latter reason, we generally use combined evolution periods, and compensate the slightly lower sensitivity by recording for longer measurement times.
  For obtaining intra-protein NOEs, strictly speaking, both dimensions should be edited, but this comes at such a large loss in sensitivity that it is normally not applied. The extra inter-molecular NOE signals recorded by not editing the second ^{1}H dimension represent no major issue, as they are typically less than 100, which represents less than 5% of the total NOE signals, and are easily filtered out by the automated assignment routines. If later the entire complex structure should be solved, the inter-molecular signals obviously represent a valuable source of information.For these experiments, the positioning of the isotope-resolving element
  (i.e. the HSQC-element) should be considered, because it has consequences for the optimal resolution of the signals. The resolving-element can either be placed in the indirect dimension (6) or in the direct dimension (5). For the former case, two indirect low-resolution dimensions are describing the originating nucleus, and one high-resolution dimension – the direct dimension – is used for describing the destination nucleus. From first principles, this yields more precise peak positions since one signal is defined by a single dimension, but this has high resolution. In the reversed case (5), the originating nucleus is defined by a single dimension, but now this dimension is low resolution, increasing the ambiguity for potential NOE assignments. However, in practice, water suppression for experiment (5) is much more easily implemented than for experiment (6). We nevertheless encourage use of strategy (6) as it simplifies analysis by overall higher precision of signal positions, which offsets the additional work required for carefully setting up the water suppression scheme.
- Recording the entire set of NOEs required for full complex structure determinationIn essence, for obtaining all necessary NOEs, all the above-mentioned spectra need to be recorded. However, there are alternative ways of obtaining the entire set of NOEs in less measurement time. One alternative is to record a 4D spectrum (7) [40], which can be processed in different ways in order to yield different sub-spectra. This approach allows identifying all types of inter- and intra-molecular NOE signals, but has lower sensitivity for the intra-protein NOEs compared to the standard 3D NOESYs. Our preferred alternative is the All-Inclusive NOESY (8), where all inter- and intra-molecular NOE signals are also recorded in a single spectrum. It is sensitive as it is based on a 3D NOESY experiment with only one resolving element, in contrast to standard experiments for recording inter-molecular NOEs which always have two filtering and editing elements. In special cases of fast exchanging ligands yielding low inter-molecular NOE intensities for ligand-to-protein signals, an additional edited-filtered NOESY may be recorded in order to identify more inter-molecular NOE signals. This is due to the fact that if the All-Inclusive NOESY is recorded with the resolving step in F_1 in order to obtain optimal resolution, only ligand-to-protein NOEs are easily identified, which however may be very weak in this special case. The more intense protein-to-ligand NOEs are recorded, but are not distinguishable from protein-to-protein NOEs, as in a conventional 3D spectrum.
  In such cases one can also record the reversed version of this experiment with F_3r^{Al}. This will yield strong inter-molecular signals in the ligand plane at 100 ppm ^{13}C chemical shift, but comes at the expense of sub-optimal resolution.

5.4.2. Complexes with slow exchanging ligands

- Recording intra-ligand NOEs of the bound ligand
  For slowly exchanging ligands a ^{1}H,^{1}H NOESY with filtering elements in both proton dimension (2) is the method of choice.
- Inter-molecular NOEs
  The considerations of unfavorable ligand T_1-relaxation of the fast exchange case don't apply in the slow exchange case and one is free to choose the directionality of the inter-molecular NOE experiment. Therefore, here, a ligand-to-protein 3D (F_1f, F_3r) spectrum is probably ideal (3). In the low-resolution F_1 dimension, the few ligand signals will be recorded and the spectral window can be reduced accordingly; and the high-resolution dimension is used for the protein. In most cases this is appropriate, because ligand signals can be resolved even in the indirect dimension. In cases where there is signal overlap on the ligand is critical, one may reverse the dimensions (4). Here, the high-resolution dimension is then used to resolve a few ligand signals. The resolution of the protein signals is still acceptable in a 3D NOESY, as they are defined by two chemical shifts.
- Recording intra-protein NOEs of the bound protein
  Here, the same considerations as for the fast exchange case apply.
- Recording the entire set of NOEs required for full complex structure determination
  In principle, the same considerations as for the fast exchange case apply also here. In the case of the All-Inclusive NOESY, no additional inter-molecular NOE spectrum will be needed, which was needed for some cases of unfavorable ligand T_1-relaxation in the fast exchange case.

6. Experimental optimizations

6.1. Optimizing NOE mixing time

Choosing the mixing time for the NOESY experiment is a critical step. The mixing time depends on the experimental aim: whether contacts between ligand and protein just are to be proven or whether the intensity of the NOE cross peaks will be used to derive distance restraints for the structure calculation. In the first case, often very long mixing times are chosen ($\tau_m \geq 200$ ms). In practice, often long mixing times are used to compensate for the low sensitivity of filtered-filtered and filtered-edited experiments for detecting intra-molecular and inter-molecular NOEs. However, while it's true that inter-molecular NOEs are typically longer range than intra-molecular ones and take longer to build up, the major factor leading to low intensity signals in these spectra, is the heavy

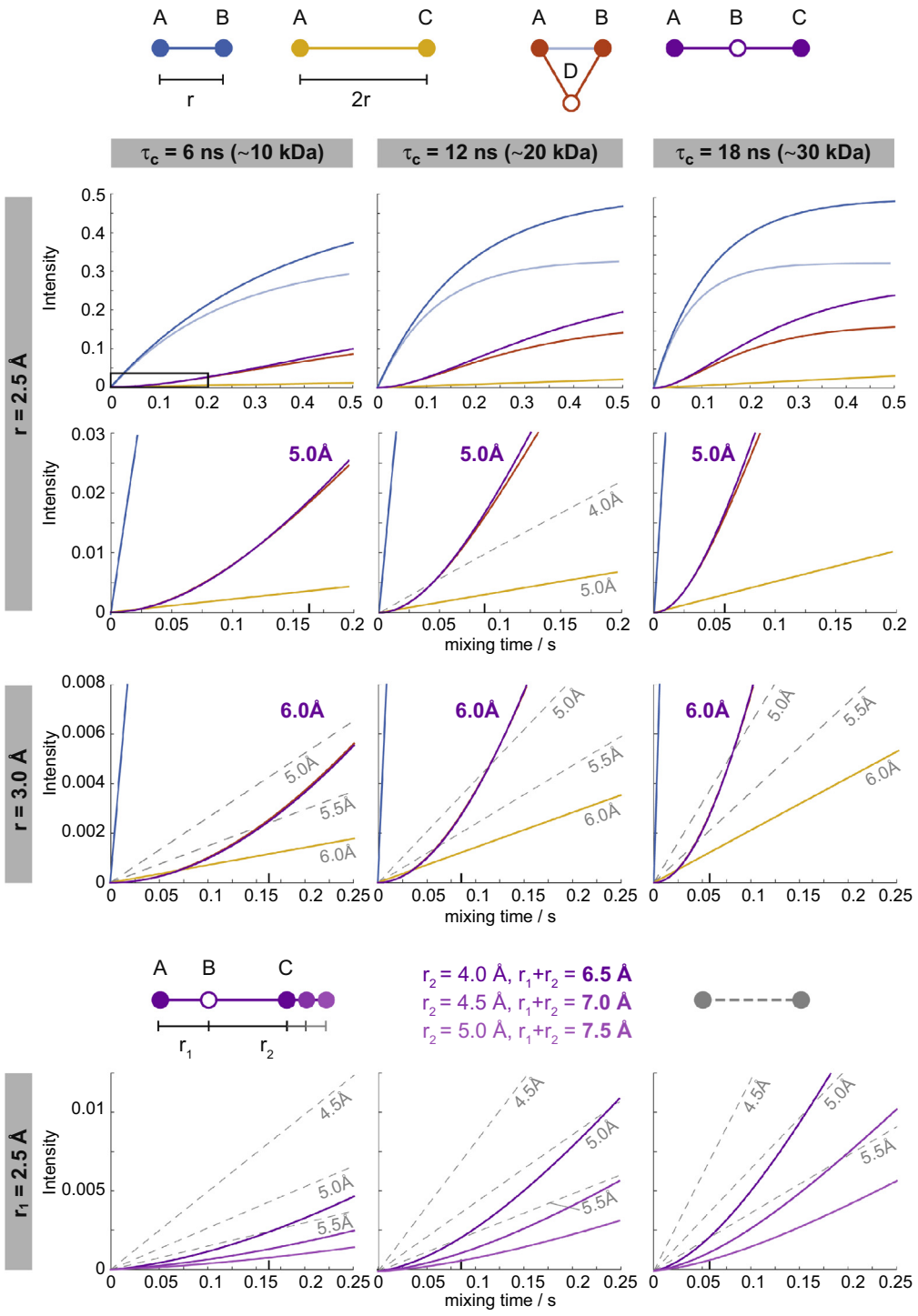

Fig. 13. Direct versus indirect magnetization transfer for different nuclear arrangements and distances. Direct magnetization transfer for different protein sizes during the NOE mixing time between two nuclei at distance r (blue curve) and $2r$ (yellow curve) are calculated in the upper diagrams for r = 2.5 Å and 3.0 Å. NOE build ups due to spin diffusion are shown in red and magenta for triangular and linear arrangements of the intermediate relay nucleus (see top of figure for topology). The two pathways are nearly equivalent. For the r = 2.5 Å case, also the decrease of the direct NOE due to spin diffusion is shown (pale blue line). The lower r = 2.5 Å diagram is an enlargement of the initial build-up time, indicated by the black square in the first diagram. Below an asymmetric linear arrangement is shown, which may be more typical for protein-ligand interfaces (see text). In dashed grey lines, build-up curves for direct NOEs for the indicated distances are shown, where 5.5 Å represents the detection limit. The different magenta curves show spin diffusion for the different distances indicated above in the same color. Significance: Since NOE signal intensities are related to distances by r^{-6}, doubling and halving of the intensity will only lead to +12% and −11% errors on the distance. We therefore regard intensity changes smaller than a factor of 2 as tolerable. For the equilateral triangular geometry, there is therefore no issue with spin diffusion, as the spin diffusion contribution never is as large as the direct NOE. The linear geometry in contrast represents the worst case. In the r = 2.5 Å scenario, spin diffusion and direct NOE are equal already at mixing times <20 ms and therefore unavoidable. E.g. at 50 ms mixing time for a 20 kDa protein the 5 Å spin diffusion signal (=2 × 2.5 Å) will have the same intensity as a direct 4 Å signal (grey curve), at 200 ms the intensity corresponds to a 3.5 Å distance. Additionally, NOE signals corresponding to (typically unobservable) 6 Å will be equivalent to 5.5 Å NOEs at ~50 ms and to 5.0 Å NOEs at ~100 ms (grey curves). Therefore, the mixing time should be kept as short as possible. The thick tick marks indicate the rule of thumb for maximal mixing time (1 s/(τ_c/ns)). The mixing time can only be increased for measuring purely inter-molecular or intramolecular NOEs for fast exchanging ligands according to the occupancy. If intra-protein NOEs appear in the same spectrum rigorously short mixing times should be used. (For the calculations, a field of 700 MHz was assumed).

Please cite this article in press as: J. Orts, A.D. Gossert, Methods (2018), https://doi.org/10.1016/j.ymeth.2018.01.019

transverse relaxation during the two filtering/editing elements of the pulse sequence. Compensating low sensitivity from relaxation and imperfect pulses with longer mixing times is not correct: by choosing long mixing times, the chances are increased to observe these weak signals, but by doing so, one loses the accurate distance information, encoded in the intensity of these signals.

For determining structures, it must be ensured, that the $1/r^6$ relation of the signal intensity to the inter-nuclear distance is conserved. The major factor hampering a clean translation of NOE cross peak intensities and distances is spin diffusion, i.e. relayed NOEs over shorter distances, which lead to overly intense signals for nuclei that are separated by a longer distance. To avoid spin diffusion, short mixing times should be used (40–80 ms) depending on the protein size (Fig. 13). Longer mixing times should only be chosen for protein-ligand complexes with low affinity, where the bound state may only be visited during a fraction of the mixing time. If the bound population is known, it is legitimate to correct the mixing time by this factor.

6.2. Simplified description of relaxation processes during mixing time to assess useful duration of mixing time

The time evolution of the intensity of NOE cross peaks during the mixing time can be theoretically described rather well. However, precise calculation of the relevant relaxation processes requires the treatment of all cross- and auto-relaxation rates between all nuclei in a protein-ligand complex and all the kinetic exchange processes associated with it. This can be calculated using the full relaxation matrix approach, which we will outline here. However, in day-to-day work, simplified expressions are needed for assessing parameters like suitable mixing times for structure determination or for simply detecting highest NOE intensity, which we will derive in the following.

NOESY peak intensities can be calculated using the full relaxation matrix analysis that includes all relaxation and chemical exchange rates for all spins written in a single matrix **R** (without losing generality and for simplicity the kinetics matrix, **K**, which includes protein ligand-exchange kinetics, is incorporated into the relaxation matrix, **R**) [42,43]. The longitudinal magnetization follows the modified Solomon equations,

$$\mathbf{M}(\tau_m) = (\mathbf{M}(0) - \mathbf{M}^{eq})e^{-\mathbf{R}\tau_m} + \mathbf{M}^{eq} \tag{10}$$

where **R** is the relaxation matrix including terms describing auto- and cross-relaxation as well as exchange kinetics, τ_m the mixing time, $\mathbf{M^{eq}}$ contains the equilibrium magnetization for each spin and **M**(0) the magnetization of each spin at the beginning of the mixing time. This explicit treatment has allowed calculation of very precise distances from protein NOEs and has recently also been applied to protein-ligand complexes [44–48].

To simplify the discussion, we will treat kinetic exchange rates separately and focus first on auto- and cross relaxation rates. Individual elements of the relaxation matrix are expressed with the usual equations:

$$R_{ii} = \rho_i = \sum_{j \neq i} b^2 (J(0) + 3J(\omega) + 6J(2\omega)) \tag{11a}$$

$$R_{ij} = \sigma_{ij} = b^2 (6J(2\omega) - J(0)) \tag{11b}$$

with $b = \frac{\mu_0 \hbar \gamma_H^2}{8\pi} \frac{1}{r_{ij}^3}$ describing the dipolar interaction and $J(\omega) = \frac{2}{5}(\frac{\tau_c}{1+(\omega\tau_c)^2})$ the spectral density of motional frequencies assuming isotropic tumbling of a rigid protein – an approximation, which in most cases is not fully valid though, as most proteins are neither rigid nor spherical.

ρ_i is the auto-relaxation rate of nucleus i, σ_{ij} the cross-relaxation rate between nuclei i and j, r_{ij} the proton(i)-proton(j) distance, μ_0 the permeability of free space, ħ the reduced Planck constant, γ_H the gyromagnetic ratio of the proton and τ_c the rotational correlation time of a spherical particle [42,43].

The integrated rate equation (Eq. (10)) can be simplified by describing the exponential term with a Taylor expansion. As an example, we consider the case were three spins i, j and k are present. The Taylor series of Eq. (10) up to the second order gives the following magnetization transfer, following the same notations as defined above:

$$\frac{\Delta M_{ii}(\tau_m)}{\Delta M_{ii}(0)} = 1 - \rho_i \tau_m + \frac{1}{2}(\rho_i^2 + \sigma_{ij}^2 + \sigma_{ik}^2)\tau_m^2 \tag{12a}$$

$$\frac{\Delta M_{ij}(\tau_m)}{\Delta M_{ii}(0)} = -\sigma_{ij} \tau_m + \frac{1}{2}((\rho_i + \rho_j)\sigma_{ij} + \sigma_{ik}\sigma_{kj})\tau_m^2 \tag{12b}$$

where the terms containing the subscript k account for the loss of magnetization of spin i to spin k and spin diffusion from spin i to spin j via spin k. For the sake of the explanations, to describe the NOE cross peak intensities, we use the first term of the Taylor expansion to describe the cross-relaxation (Eq. (12b)) and keep the exponential form for the auto-relaxation of the originating and destination nucleus, yielding Eq. (12c,d)

$$\Delta M_{ij}(\tau_m) \approx -\sigma_{ij}\tau_m \Delta M(0) e^{-\rho\tau_m} \propto \frac{\tau_c \tau_m}{r^6} \Delta M(0) e^{-\rho\tau_m} \tag{12c, d}$$

The cross-relaxation rate σ_{ij} in Eq. (12c,d) is proportional to τ_c/r^6 for large molecules, since $J(2\omega) \ll J(0)$, σ_{ij} is dominated by $J(0)$ (Eq. (11b)), which is proportional to the correlation time τ_c. From this simplified description, we see that the NOE build-up curve intensity exhibits a linear increase depending on the mixing time, τ_m, and the correlation time of the complex, τ_c, and is damped by an exponential decrease due to relaxation with the rate constant ρ (Fig. 13).

6.3. Maximizing NOE cross peaks

For obtaining maximal intensities for inter-molecular NOE cross peaks, three measures can be taken: (i) choosing the most sensitive experiment, (ii) reducing T_1 relaxation by measuring in D_2O and (iii) choosing a long mixing time. These three measures all come with some disadvantages, as explained in the following in more detail.

(i) In our experience, the most sensitive experiment for obtaining inter-molecular NOEs is a 2D F_2f [^{1}H,^{1}H] NOESY. By omitting editing in the first dimension the experiment is shortened considerably, leading to less T_2 relaxation and therefore higher signal to noise. However, mixed inter-molecular and intra-ligand NOEs will be recorded. Inter-molecular NOEs can in principle be identified by the missing diagonal-symmetric signal. But for the sake of testing whether inter-molecular NOEs can be obtained at all for a system, this limitation is not so relevant.

(ii) Eq. (12c,d) describes the NOE peak intensity as a build-up with the cross-relaxation rate σ, which is damped by the exponential decay of the diagonal signal due to auto-relaxation ρ. The value of ρ can be lowered by measuring in D_2O, and concomitantly the NOE peak intensity will be larger[6]. Additionally, a higher receiver gain can be chosen and some experiments can be shortened because no water suppression element is required. The downside of this procedure is that exchangeable protons are not detectable anymore,

[6] This is traditionally exploited in saturation transfer difference (STD) experiments [49], where much higher signal intensity is obtained when measuring in D_2O, as leakage to the solvent is reduced during the long mixing time of 1–2 s.

most notably amide moieties. Furthermore, tiny chemical shift changes might occur. (We thank reviewer 1 for pointing out the advantages and disadvantages of working in D_2O).

(iii) The maximal signal that can be achieved, can approximately be calculated by taking the derivative of Eq. (12c,d) for a two-spin system and setting it equal to zero. In this simplified description, all terms will disappear and lead to the expression:

$$\tau_{m,\max} = 1/\rho \tag{13}$$

The resulting mixing time for maximal cross-peak intensity is in the order of seconds, for long distance NOEs it is even longer, as the cross peak intensity essentially continues growing asymptotically, for as long as the diagonal signal is stronger than the cross peak. The calculation of the cross peak intensities depends on accurate values for σ and ρ. While the cross-relaxation rates σ can be rather well predicted, the auto-relaxation rates ρ often contain a so-called leakage factor of 1–2 s^{-1} stemming from other relaxation sources, such as paramagnetic relaxation due to oxygen in the solution. Therefore, the decay rate of a cross peak which determines the position of the maximal NOE value, is inaccurate in our calculations for long build-up curves and should be taken with a grain of salt. We therefore suggest mixing times around 500 ms.

These long mixing times will yield intense NOE signals, but due to massive spin diffusion and auto-relaxation it will not be possible to translate those signals into meaningful distance restraints.

6.4. NOE mixing time for quantification of inter-nuclear distances

For deriving distances, NOE signal intensities should be recorded at the beginning of the build-up curve during the quasi-linear regime, which allows translation into distances. Additionally, the cross peak intensity should not be influenced strongly by spin diffusion. However, it is not possible to avoid spin diffusion, no matter how short the mixing time is (see Fig. 13 for r = 2.5 Å). This can be rationalized using our simplified treatment: during the initial build up, the intensity of the NOE cross peak is proportional to $-\sigma\tau_m$ (Eq. (12c,d)). The relayed spin diffusion, i.e. two consecutive NOE transfer steps, is then described by $½(\sigma\tau_m)^2$ (this term appears in the Taylor expansion of $e^{-R\tau_m}$, see Eq. (12b)). Spin diffusion can therefore be identified in principle by measuring build-up curves, as there is an initial exponential lag-phase while magnetization builds up on the intermediate nucleus. For cases with short distances (2.5 Å) on larger proteins (>50 kDa), this lag time however, can be below 5 ms and experimental identification is nearly impossible [50]. However, for long-distance relayed NOEs, there is a clear lag-phase and the spin diffusion NOE only reaches higher intensity than the direct NOE after more than 50 ms (Fig. 13). Since in most structure determination efforts just one mixing time will be recorded, spin diffusion won't be identified based on the curve shape of an NOE build-up curve, therefore it needs to be limited by experimental measures.

Fortunately, since the NOE cross peak intensity relates to the inverse sixth power of the inter-nuclear distance, $I \propto r^{-6}$, even a rather large change of the cross peak intensity has a small influence on the derived distance. For example, if the intensity is doubled by spin diffusion, the distance is decreased by mere 11%. Ideally therefore, to obtain distances with errors below 10% the contribution from spin diffusion should not be larger than the direct NOE ($\sigma_{AC}\tau_m \geq \sigma_{AB}\tau_m\sigma_{BC}\tau_m$). Nuclei arranged in an equilateral triangle do fulfil this condition for short mixing times and therefore don't represent an important problem in terms of spin diffusion, even considering that multiple pathways can lead to combined spin diffusion (Fig. 13). However, essentially all arrangements with the intermediate nucleus within the equilateral triangle will yield a larger spin diffusion contribution at longer mixing times than the direct NOE. In particular, the linear sequence of three equally spaced nuclei (A, B, C) represents the worst-case scenario, yielding NOE cross peak intensities corresponding to nearly half the actual distance. Or, put in other words, apparent NOEs to nuclei 6–7 Å apart can result from spin diffusion in a linear arrangement (see Fig. 13 for more details and examples). More sophisticated approaches to simulate spin diffusion or to interpret build-up curves containing spin diffusion are available, but are outside the scope of this review [33,44,51,52].

However, due to the lag-phase of long-range spin diffusion NOEs, short mixing times help reducing the amount of artificial long-range peaks. From Eq. (10) it can be derived, that if the correlation time of the protein is doubled, the mixing time can be roughly halved. Therefore, as a rule of thumb, the mixing time should not exceed 1 ns/τ_c in seconds, i.e. for a 20 kDa protein with a τ_c of 12 ns the mixing time should be smaller than 1/12 s ≈ 80 ms. Using this rule of thumb for the length of the mixing time helps keeping errors from spin diffusion to below 10% for most geometries. However, for worst case linear topologies including geminal protons at 1.8 Å distance, large spin diffusion errors can simply not be avoided, but can often be resolved by structure determination programs. Therefore, the aim of this rule of thumb is to limit errors of non-covalent NOEs.

Additionally, short range spin diffusion, will reduce the intensity of the direct NOE, but typically by less than a factor of 2. (Reducing cross peak intensity by factor 2 will lead to distance errors of +12%). One therefore needs to bear in mind that long range cross peaks are biased towards too short distances and short-range NOEs often towards too long distances, the latter case being less severe. Structure calculation programs take this fact into account to a certain degree, by allowing elasticity of NOE-derived upper distance limits [11].

6.5. Special considerations for inter-molecular NOEs

Inter-molecular proton-proton distances tend to be larger on average than intramolecular ones, because very short distances (shorter than the actual van der Waals radius) as for example those from covalently bound geminal hydrogens (r = 1.8 Å) are missing at molecular interfaces. Therefore, it is tempting to increase the mixing time for optimal detection of long range NOEs. For inter-molecular NOEs, the more typical spin diffusion situation is that two protons are closely spaced on the protein and a third proton is on the ligand at larger distance. It is true that for this case, the spin diffusion error is a bit less pronounced. Nevertheless, staying with our example of a 20 kDa protein, a 6.5 Å actual distance will give rise to an apparent NOE of 5.5 Å at 80 ms mixing time, and therefore actually show up in the spectrum. This is severe enough spin diffusion not to increase the mixing time, although it seems tempting. Considering the high sensitivity of contemporary NMR equipment with high magnetic fields and cryogenically cooled probes, the suggested mixing times are sufficient for obtaining a full set of data.

The mixing time can only be increased for measuring purely inter-molecular or intramolecular NOEs for fast exchanging ligands according to the occupancy. For all other experiments, lower occupancy is corrected afterwards, when calibrating NOEs.

6.6. Changing temperature and viscosity to increase the NOE

If ligands are in an intermediate exchange regime, the build-up curves often cannot be interpreted with a simple model, such as the isolated two spin model, but require the full relaxation matrix formalism with the exchange matrix [53,54]. In that case, we recommend to change the sample condition e.g. the temperature and viscosity to reach either the fast or slow exchange regime as described in the section on sample preparation.

Viscosity and temperature will also affect the rotational correlation time τ_c of the molecules [55,56].

$$\tau_c = \frac{4\pi\eta r_H^3}{3k_B T} \tag{14}$$

which itself depends on temperature (T) and the viscosity of the solvent (η), all else equal. Empirically, the temperature dependence of the rotational correlation time can be calculated as follows:

$$\tau_c(T) = \frac{\eta_T}{\eta_{293K}} \frac{293K}{T} \tau_c(293K) \tag{15a}$$

where η_T, the viscosity of water at a given temperature can be either looked up in tables or calculated approximately using the following formula [56]:

$$\eta_T = 1.7753 - 0.0565T + 1.075110^{-3}T^2 - 9.2222 \cdot 10^{-6}T^3 \tag{15b}$$

For complexes with small proteins (<10 kDa), it can help to increase the viscosity in order to increase the NOE. Viscosity of the solvent can be roughly doubled by adding 30% glycerol and increased further by lowering temperature.

Note on the inter-scan delay: The optimal recovery delay between two transients in order to maximize the signal per unit of time is $1.3 \times T_1$ [57]. While the T_1 relaxation time for protein is typically below 1 s, the T_1 relaxation time for ligand is in the order of a few seconds. It is quite common to use a recovery delay of 1 s but this is not advisable for every experiment.

If the magnetization is arising from the ligand and transferred to the protein or to the ligand itself, then the recovery delay should be should be matched to the ligand T_1, and a recovery delay of 2–2.5 s is reasonable.

If the magnetization is coming from the protein, the recovery delay can be much shorter, e.g. 0.8–1 s.

Recording of NOESY spectra

- Choose suited experiment from Table 3
- Use rule of thumb for mixing time $\tau_m \leq 1s/(\tau_c/ns) \approx 5s/(3\times MW/kDa)$*
- For higher sensitivity, record spectra in D_2O. For maximal sensitivity, use 2D F_2f spectrum and longer mixing time (do not use for structure calculation)

Deriving distance restraints

- Calibrate restraints in regular way and correct for occupancy if needed
- Mind that median inter-molecular distances are 0.25Å larger than intra-molecular ones (4.45 vs. 4.2Å, respectively)

*For non-geminal protons this limits the error on NOE peak intensity to below a factor of 2, which translates to ~10% error on distance

7. Deriving distance restraints

Distance restraints for protein-ligand complexes are derived from NOE data in a very similar way as for single protein structures. There are two major differences: First, inter-molecular distances at molecular interfaces are in general on average a bit longer than intra-molecular ones. That is because covalently bound nuclei like geminal protons or other such short distances are missing at interfaces. Therefore, inter-molecular NOEs should be calibrated to a slightly longer median distance. Second, in the case of fast exchange, one should correct for incomplete occupancy of the complex and shorten the apparent distances accordingly.

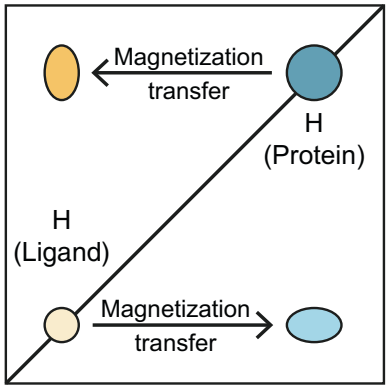

Fig. 14. Differential signal intensities depending on T_1- and T_2-relaxation during the NOESY experiment, and additionally in the fast exchange case differential linewidths. Schematic representation of a 2D NOESY spectrum. Diagonal and NOE cross peaks for two protons are shown. The arrows indicate the direction of the magnetization transfer and the shape of the peaks represent the line width of the resonances. Signal intensities (indicated by the saturation of the color) depend on T_1- and T_2-relaxation during the NOESY experiment, which can be differ for different nuclei. In particular, in the fast exchange case, the ligand and the protein have very different relaxation properties. This issue can be alleviated by normalizing signal intensities to the diagonal signals.

7.1. Correcting for incomplete occupancy

If the ligand is a weak binder we have to distinguish by purposes, either deriving the structure of the bound ligand or deriving the structure of the complex.

For intra-ligand NOEs and inter-molecular NOEs detected on the ligand, the effective cross relaxation rate that is the population average between the free and bound state of the ligand needs to be defined: [58]

$$\sigma^L_{avg} = p^L_{free}\sigma_{free} + p^L_{bound}\sigma_{bound} \tag{16}$$

Since the correlation time of the ligand is in the order of picoseconds and the correlation time of the complex is to a good approximation the same as the one of the protein, the first term can be neglected; e.g. σ is 20-fold smaller for a 300 Da free ligand than for a ligand bound in a 30 kDa complex. Consequently, the effective cross-relaxation rate is scaled by the bound population of the ligand and the cross-relaxation rate of the complex is defined using the correlation time of the protein.

$$\sigma^L_{avg} \approx p^L_{bound}\sigma_{bound} \tag{17a}$$

For inter-molecular NOEs detected on the protein, the effective cross relaxation rate is defined as,

$$\sigma^P_{eff} = p^P_{bound}\sigma_{bound} \tag{17b}$$

where the bound population of the ligand and the protein can be calculated from equation (Eqs. (3) and (4)) knowing the affinity of the complex.

Just to put these corrections into perspective: Due to the relation of the NOE cross peak intensity to the inter-nuclear distance, $I \propto r^{-6}$, even a rather large change on the cross peak intensity has a small influence on the derived distance. E.g. if the intensity is halved, the distance is increased by mere 12%.

If the ligand is a strong binder, namely it is in slow exchange with the protein, the correlation times used to interpret the NOE intensities of the ligand, the protein and the complex are the same and equal to the one of the protein. Therefore, no correction needs to be applied.

7.2. Accounting for differential relaxation of different nuclei

It is quite common that T_1 relaxation times vary from proton to proton within the same molecule, and even more so on two different molecules, like a protein and a ligand. This is leading to different recovery of the magnetization and consequently different initial magnetizations for various protons in the complex. We have previously seen that the NOE cross-peak intensity depends on the initial magnetization at the beginning of the mixing time $M(0)$. If the spins are not fully recovered and their initial magnetization is different, due to different T_1, their NOE cross peak intensities will be scaled differently and inconsistently. Therefore, a distribution of initial magnetizations will introduce errors in the distances derived from NOESY cross peak intensities. This effect is more pronounced when the recovery delay is too short or comparable to T_1.

A simple way to mitigate this problem is to normalize the cross-peak intensities from the diagonal peak where the magnetization originated from. Using a normalized intensity cancels the incomplete recovery because both the cross-peak and the diagonal peak have the same incomplete recovery [44–47]. The same idea applies to different relaxation pathways during the pulse sequence element, such as the INEPT or simply different T_2 during acquisition, where different protons may experience different signal relaxation (Fig. 14).

Ideally, all four intensities, the two cross-peak and diagonal peak intensities, should be combined in order to derive the highest possible accurate distance. This is however often not possible due to overlap. More details can be found in the review by B. Vögeli [48]. For protein-ligand complexes, this analysis is however hampered by lack of diagonal signals in several versions of edited and filtered experiments, and missing diagonal symmetric cross peaks (Fig. 10). Here, either separate 1D experiments are recorded with the same relaxation delay as in the NOESY spectrum in order to approximate diagonal peak intensities, or an All-inclusive spectrum is used, where all diagonal signals and equivalent pathways are preserved Fig. 11.

7.3. Calibration of NOEs

The NOE is a relaxation *rate*, and therefore best quantified by analyzing the build-up rate of the NOE cross peaks. To this end, NOESY spectra at several mixing times are measured, and accurate inter-proton distances can be calculated using the build-up curves, the corresponding diagonal decays and the bound population of the protein or the ligand (Eq. (17)). This formalism was recently used to derived proton distances with an accuracy of 0.1 Å[48].

But most of the time only one NOESY spectrum is measured and therefore the (normalized) cross peak intensities have to be directly converted to distances.

In that case, a way to derive distances from intensities consists of using known fixed intra-molecular distances within the ligand or the protein (e.g. distances from geminal protons or protons in an aromatic ring) to calibrate the NOE intensities:

$$r_{ij} = \left(\frac{I_{\text{known}}}{I_{ij}}\right)^{\frac{1}{6}} r_{\text{known}} \qquad (18)$$

A second way, currently used by the standard protocol in Cyana, is to use an empirical median distance and to calibrate the median intensity of all cross peaks collected in the NOESY spectrum, i.e. replace the known distance (r_{known}) in Eq. (18) by 3.9–4.2 Å and the known intensity (I_{known}) by the median of all intensities [11,12].

If the same procedure is applied to the protein and the ligand NOEs, the ligand NOE intensities should be scaled by the factors described in Eq. (17).

For inter-molecular NOEs, the median inter-proton distances are larger than 3.9–4.2 Å. From an analysis of complex structures in the protein structure database (pdb, www.rcsb.org), we observed that the median intra-molecular distances (4.2 Å) are slightly shorter than the inter-molecular distances and the median value of the inter-molecular distances is 4.4–4.6 Å, depending on the complex (overall median value 4.45 Å). This is due to lacking short distances at interfaces, as they occur e.g. in geminal protons. Therefore, the inter-molecular distances should be corrected accordingly.

The obtained distances can readily be used as upper distance constraints in a Cyana structure calculation protocol [11,12].

8. Structure calculation

Structure calculations of complexes are well-established and described in detail in the respective manuals of the software providers, therefore this will not be discussed in detail here [11,13]. As a potential template for the reader, we supply commented files for a structure determination protocol with the software Cyana in the supplementary material. This includes a library file describing the drug nutlin as a residue for the program Cyana, a residue sequence file and a structure calculation macro.

9. Conclusion

In this review, we laid out the topics that are most important to us when dealing with structure determination of protein-ligand complexes by NMR. Our considerations on sample preparation and choice of experiments are summarized in Tables 1 and 3, respectively. The background information and the theoretical considerations on which our choices are based are described in the text. In between, there are numerous practical tips on e.g. working with ligands with limited solubility, optimizing experimental conditions, choosing the NOE mixing time and calibration of inter-molecular NOEs.

Acknowledgements

We gratefully acknowledge inspiring discussions with Dr. Wolfgang Jahnke, Dr. Fred Damberger and Dr. Simon Rüdisser. We would like to thank the reviewers for their valuable suggestions on the manuscript, in particular the comments on measuring in D_2O, and the limitations of isotope filter experiments.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ymeth.2018.01.019.

References

[1] K. Wüthrich, NMR studies of structure and function of biological macromolecules (Nobel Lecture), Angew. Chem. Int. Ed. 42 (2003) 3340–3363.
[2] K. Wüthrich, NMR of Proteins and Nucleic Acids, Wiley Interscience, 1986.
[3] A.S. Arseniev, I.L. Barsukov, V.F. Bystrov, A.L. Lomize, Y.A. Ovchinnikov, ^{1}H-NMR study of gramicidin a transmembrane ion channel, FEBS Lett. 186 (1985) 168–174.

[4] M.P. Williamson, T.F. Havel, K. Wüthrich, Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry, J. Mol. Biol. 182 (1985) 295–315.
[5] D.C. Muchmore, L.P. McIntosh, C.B. Russel, E. Anderson, F.W. Dahlquist, Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance, Methods Enzymol. 177 (1989) 44–73.
[6] R.A. Venters, T.L. Calderone, L.D. Spicer, C.A. Fierke, Uniform carbon-13 isotope labeling of proteins with sodium acetate for NMR studies: application to human carbonic anhydrase II, Biochemistry 30 (1991) 4491–4494.
[7] C. Spitzfaden, W. Braun, G. Wider, H. Widmer, K. Wüthrich, Determination of the NMR solution structure of the cyclophilin A-cyclosporin a complex, J. Biomol. NMR 4 (1994) 463–482.
[8] C. Weber, G. Wider, B. Von Freyberg, R. Traber, W. Braun, H. Widmer, K. Wüthrich, NMR structure of cyclosporin a bound to cyclophilin in aqueous solution, Biochemistry 30 (1991) 6563–6574.
[9] S.W. Fesik, R.T. Gampe, H.L. Eaton, G. Gemmecker, E.T. Olejniczak, P. Neri, T.F. Holzman, D.A. Egan, R. Edalji, NMR studies of [U-^{13}C] cyclosporin a bound to cyclophilin: bound conformation and portions of cyclosporin involved in binding, Biochem. ACS 30 (1991) 6574–6583.
[10] P. Güntert, Structure calculation of biological macromolecules from NMR data, Q. Rev. Biophys. 31 (1998) 145–237.
[11] P. Güntert, L. Buchner, Combined automated NOE assignment and structure calculation with CYANA, J. Biomol. NMR (2015).
[12] P. Güntert, C. Mumenthaler, K. Wüthrich, Torsion angle dynamics for NMR structure calculation with the new program DYANA, J. Mol. Biol. 273 (1997) 283–298.
[13] C.D. Schwieters, J.J. Kuszewski, N. Tjandra, G.M. Clore, The Xplor-NIH NMR molecular structure determination package, J. Magn. Reson. 160 (2003) 65–73.
[14] T. Herrmann, P. Güntert, K. Wüthrich, Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS, J. Biomol. NMR 24 (2002) 171–189.
[15] E.M. Yilmaz, P. Güntert, NMR structure calculation for all small molecule ligands and non-standard residues from the PDB Chemical Component Dictionary, J. Biomol. NMR 63 (2015) 21–37.
[16] A. Widmer, Wit!P: Molecular Modeling/Graphics Tool, Novartis, Basel, 2014.
[17] D. Nietlispach, H.R. Mott, K.M. Stott, P.R. Nielsen, A. Thiru, E.D. Laue, Structure determination of protein complexes by NMR, Protein NMR Tech. (2004) 255–288.
[18] H. Wu, L.D. Finger, J. Feigon, Structure Determination of Protein/RNA Complexes by NMR, Methods Enzymol. 394 (2005) 525–545.
[19] A. Marintchev, D. Frueh, G. Wagner, NMR Methods for studying protein–protein interactions involved in translation initiation, in: Methods in Enzymology, Elsevier, 2007, pp. 283–331.
[20] J.J. Ziarek, F.C. Peterson, B.L. Lytle, B.F. Volkman, Binding site identification and structure determination of protein–ligand complexes by NMR, in: Methods in Enzymology, Academic Press, Burlington, 2011, pp. 241–275.
[21] U. Schieborr, S. Sreeramulu, H. Schwalbe, NMR structure determination of protein–ligand complexes, in: I. Bertini, K.S. McGreevy, G. Parigi (Eds.), NMR of Biomolecules: Towards Mechanistic Systems Biology, John Wiley & Sons, 2012.
[22] R.A. Alberty, G.G. Hammes, Application of the theory of diffusion-controlled reactions to enzyme kinetics, J. Phys. Chem. 62 (1958) 154–159.
[23] G. Dahl, T. Akerud, Pharmacokinetics and the drug–target residence time concept, Drug Discovery Today 18 (2013) 697–707.
[24] G. Schreiber, G. Haran, H.-X. Zhou, Fundamental aspects of protein−protein association kinetics, Chem. Rev. 109 (2009) 839–860.
[25] G. Wider, L. Dreier, Measuring protein concentrations by NMR spectroscopy, J. Am. Chem. Soc. 128 (2006) 2571–2576.
[26] L. Fielding, NMR methods for the determination of protein–ligand dissociation constants, Prog. Nucl. Magn. Reson. Spectrosc. 51 (2007) 219–242.
[27] A.D. Gossert, W. Jahnke, NMR in drug discovery: A practical guide to identification and validation of ligands interacting with biological macromolecules, Prog. Nucl. Magn. Reson. Spectrosc. 97 (2016) 82–125.
[28] H. Kovacs, D. Moskau, M. Spraul, Cryogenically cooled probes—a leap in NMR technology, Prog. Nucl. Magn. Reson. Spectrosc. 46 (2005) 131–155.
[29] J.-P. Renaud, C. Chung, U.H. Danielson, U. Egner, M. Hennig, R.E. Hubbard, H. Nar, Biophysics in drug discovery: impact, challenges and opportunities, Nat. Rev. Drug Discovery 15 (2016) 679–698.
[30] T.P. Kenakin, A Pharmacology Primer, 4th ed., Academic Press, San Diego, 2014.
[31] M.P. Williamson, Using chemical shift perturbation to characterise ligand binding, Prog. Nucl. Magn. Reson. Spectrosc. 73 (2013) 1–16.
[32] A.P. Campbell, B.D. Sykes, Theoretical evaluation of the two-dimensional transferred nuclear Overhauser effect, J. Magn. Reson. 1969 (93) (1991) 77–92.
[33] F. Ni, Recent developments in transferred NOE Methods, Prog. Nucl. Magn. Reson. Spectrosc. 26 (1994) 517–606.
[34] A.L. Breeze, Isotope-filtered NMR methods for the study of biomolecular structure and interactions, Prog. Nucl. Magn. Reson. Spectrosc. 36 (2000) 323–372.
[35] G. Otting, K. Wüthrich, Heteronuclear filters in two-dimensional [^{1}H, ^{1}H]-NMR spectroscopy: combined use with isotope labelling for studies of macromolecular conformation and inter-molecular interactions, Q. Rev. Biophys. 23 (1990) 39.
[36] H. Kogler, O. Sørensen, G. Bodenhausen, R. Ernst, Low-pass J filters. suppression of neighbor peaks in heteronuclear relayed correlation spectra, J. Magn. Reson. 1969 (55) (1983) 157–163.
[37] C. Zwahlen, P. Legault, S.J. Vincent, J. Greenblatt, R. Konrat, L.E. Kay, Methods for measurement of inter-molecular NOEs by multinuclear NMR spectroscopy: application to a bacteriophage λ N-peptide/boxB RNA complex, J. Am. Chem. Soc. 119 (1997) 6711–6721.
[38] E.R. Valentine, F. Ferrage, F. Massi, D. Cowburn, A.G. Palmer, Joint composite-rotation adiabatic-sweep isotope filtration, J. Biomol. NMR. 38 (2007) 11–22.
[39] E. Kupce, Bruker User Library, <https://www.bruker.com/service/information-communication/nmr-pulse-program-lib/bruker-user-library.html>.
[40] G. Melacini, Separation of intra- and inter-molecular NOEs through simultaneous editing and *J* -compensated filtering: a 4D quadrature-free constant-time *J* -resolved approach, J. Am. Chem. Soc. 122 (2000) 9735–9738.
[41] R. Boelens, M. Burgering, R.H. Fogh, R. Kaptein, Time-saving methods for heteronuclear multidimensional NMR of (^{13}C, ^{15}N) doubly labeled proteins, J. Biomol. NMR. 4 (1994) 201–213.
[42] S. Macura, R.R. Ernst, Elucidation of cross relaxation in liquids by two-dimensional N.M.R. spectroscopy, Mol. Phys. 41 (1980) 95–117.
[43] I. Solomon, Relaxation processes in a system of two spins, Phys. Rev. 99 (1955) 559–565.
[44] J. Orts, B. Vögeli, R. Riek, Relaxation matrix analysis of spin diffusion for the NMR structure calculation with eNOEs, J. Chem. Theory Comput. 8 (2012) 3483–3492.
[45] J. Orts, M.A. Wälti, M. Marsh, L. Vera, A.D. Gossert, P. Güntert, R. Riek, NMR-based determination of the 3D structure of the ligand-protein interaction site without protein resonance assignment, J. Am. Chem. Soc. 138 (2016) 4393–4400.
[46] M.A. Wälti, R. Riek, J. Orts, Fast NMR-based determination of the 3D structure of the binding site of protein-ligand complexes with weak affinity binders, Angew. Chem. Int. Ed. 56 (2017) 5208–5211.
[47] D. Strotz, J. Orts, M. Minges, B. Vögeli, The experimental accuracy of the uni-directional exact NOE, J. Magn. Reson. 259 (2015) 32–46.
[48] B. Vögeli, The nuclear Overhauser effect from a quantitative perspective, Prog. Nucl. Magn. Reson. Spectrosc. 78 (2014) 1–46.
[49] M. Mayer, B. Meyer, Characterization of ligand binding by saturation transfer difference NMR spectroscopy, Angew. Chem. Int. Ed. 38 (1999) 1784–1788.
[50] D. Neuhaus, M.P. Williamson, The Nuclear Overhauser Effect in Structural and Conformational Analysis, 2nd ed., WILEY-VCH, n.d.
[51] J. Orts, C. Griesinger, T. Carlomagno, The INPHARMA technique for pharmacophore mapping: A theoretical guide to the method, J. Magn. Reson. 200 (2009) 64–73.
[52] D. Strotz, J. Orts, C.N. Chi, R. Riek, B. Vögeli, eNORA2 exact NOE analysis program, J. Chem. Theory Comput. 13 (2017) 4336–4346.
[53] F. Ni, Complete relaxation matrix analysis of transferred nuclear Overhauser effects, J. Magn. Reson. 1969 (96) (1992) 651–656.
[54] F. Ni, Y. Zhu, Accounting for ligand-protein interactions in the relaxation-matrix analysis of transferred nuclear Overhauser effects, J. Magn. Reson. B. 102 (1994) 180–184.
[55] D. Lavalette, C. Tétreau, M. Tourbez, Y. Blouquit, Microscopic viscosity and rotational diffusion of proteins in a macromolecular environment, Biophys. J. 76 (1999) 2744–2751.
[56] J. García de la Torre, M. Huertas, B. Carrasco, HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations, J. Magn. Reson. 147 (2000) 138–146.
[57] R.R. Ernst, W.A. Anderson, Application of fourier transform spectroscopy to magnetic resonance, Rev. Sci. Instrum. 37 (1966) 93–102.
[58] R.M. Lippens, C. Cerf, K. Hallenga, Theory and experimental results of transfer-NOE experiments. 1. the influence of the off rate versus cross-relaxation rates, J. Magn. Reson. 1969 (99) (1992) 268–281.