Published in final edited form as:

Methods Enzymol. 2011 ; 493: 241–275. doi:10.1016/B978-0-12-381274-2.00010-8.

Binding site identification and structure determination of protein-ligand complexes by NMR

Joshua J. Ziarek, Francis C. Peterson, Betsy L. Lytle, and Brian F. Volkman*

Department of Biochemistry, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, Wisconsin, 53226 USA

Abstract

Over the last fifteen years, the role of NMR spectroscopy in the lead identification and optimization stages of pharmaceutical drug discovery has steadily increased. NMR occupies a unique niche in the biophysical analysis of drug-like compounds because of its ability to identify binding sites, affinities, and ligand poses at the level of individual amino acids without necessarily solving the structure of the protein-ligand complex. However, it can also provide structures of flexible proteins and low-affinity ($K_d > 10^{-6}$ M) complexes, which often fail to crystallize. This article emphasizes a throughput-focused protocol that aims to identify practical aspects of binding site characterization, automated and semi-automated NMR assignment methods, and structure determination of protein-ligand complexes by NMR.

1 Introduction

The incorporation of structure-based knowledge in drug design has led to a three-fold increase in identification of high potency compounds ($IC_{50} < 100$ nM) from fragment leads (Hajduk and Greer, 2007). Nuclear magnetic resonance spectroscopy (NMR) is an invaluable tool for drug discovery, particularly with the growth of fragment-based screening in the pharmaceutical industry. The SAR (structure-activity relationships) by NMR technique (Shuker et al., 1996) refocused the niche of NMR in drug design from structure determination to characterization and optimization of lead compounds (Peng et al., 2001). With steady progress in production of isotope-labeled proteins (Frederick et al., 2007; Jeon et al., 2005; Tyler et al., 2005a; Tyler et al., 2005b; Zhao et al., 2004), instrument sensitivity (Hajduk et al., 1999; Jensen et al., 2010), and methods for rapid data collection and analysis (Frydman et al., 2002; Rovnyak et al., 2004a; Rovnyak et al., 2004b; Schanda and Brutscher, 2005), NMR is being utilized earlier in the process of identifying lead molecules because of its ability to quickly recognize and characterize ligand binding regardless of target protein function (Hajduk and Greer, 2007).

This chapter presents a workflow for spectroscopists interested in examining protein-ligand interactions with an emphasis on identifying the binding site and solving the structure of the complex. We first present an efficient process for spectral acquisition, processing, and assignment of a target protein or complex. We then describe an approach to protein-based compound screening that includes experimental validation of *in silico* docking results and quantitative affinity measurements to prioritize further studies. Finally, we detail how to solve the structure of protein-ligand complexes. Various approaches toward completely automating NMR structure determination have been recently reviewed and discussed and are beyond the scope of this chapter (Baran et al., 2004; Guntert, 2009; Markley et al., 2009a; Williamson and Craven, 2009). Instead, we focus on the semi-automatic approach used in

*Corresponding Author, bvolkman@mcw.edu, 414-955-8400.

our laboratory to determine protein structures and on its application to solving protein-ligand complexes.

2 Automated and semi-automated chemical shift assignment

2.1 Acquisition and processing of NMR spectra

Before NMR data collection begins, ^{15}N -labeled proteins must first be screened by 1D and 2D NMR to assess folding, aggregation state, stability, ligand binding, and to detect unstructured regions (see the chapter in this volume for information on NMR sample preparation methods). If necessary, individual protein domains can be isolated from other regions of the protein to increase solubility and reduce aggregation (Peterson et al., 2007; Waltner et al., 2005). Proteins that are judged amenable to NMR structure determination are then produced in [U - ^{15}N , ^{13}C]-labeled form and put into a general workflow for NMR-based structure determination as outlined in Figure 1.

A standardized scheme for acquisition of 2D and 3D spectra sufficient for NMR structure determination is outlined in Table I. On a 600 MHz spectrometer equipped with a cryogenic probe, these experiments can be acquired in 7-10 days for proteins under 25 kDa at concentrations > 0.5 mM. This time is significantly shortened (to as little as 30 h) if the protein concentration is high enough that data can be recorded without signal averaging or using nonlinear sampling methods. Continuous data collection on a single sample reduces spectral variations, resulting in improved peak picking and automated chemical shift assignment later on. In the case of protein-ligand complexes, several additional experiments, outlined in Section 2.5, are necessary for ligand assignment. Additionally, intermolecular distance constraints for protein-ligand complexes are determined from one additional experiment, the 3D F1- ^{13}C / ^{15}N -filtered/F3- ^{13}C -edited NOESY-HSQC. This experiment selects for (edits) protons attached to ^{13}C in the F3 dimension, and selects against (filters) protons bound to ^{13}C and ^{15}N in the F1 dimension. Thus, if unlabeled ligand is bound to labeled protein, this experiment will yield only intermolecular NOEs between protein and ligand protons.

As outlined in Table I, all data for the assignment and structure analysis are acquired in three blocks, starting with the NOESY spectra, followed by backbone triple-resonance experiments and then TOCSY-based experiments that provide sidechain assignment information. 1D ^1H and 2D ^{15}N - ^1H HSQC spectra are collected before and after every 3D experiment to provide a record of sample and instrument stability over the data collection period. Standard data acquisition parameters, including up-to-date pulse calibrations and spectral parameters, are stored in a common repository on each NMR instrument, so that data acquisition for each new protein can be started rapidly after only a few routine steps (VT adjustment, ^2H lock, $^1\text{H}/^{15}\text{N}/^{13}\text{C}$ tuning, shimming, ^1H 90° pulse calibration). Accurate chemical shift referencing in the direct and indirect dimensions using 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) is important for consistency between spectra (Markley et al., 1998). Standard 2D and 3D pulse sequences are provided in pulse sequence libraries from the NMR spectrometer manufacturers and are also available from the BioMagResBank (BMRB) website (http://www.bmrb.wisc.edu/tools/choose_pulse_info.php).

As with data collection, we use a standardized set of scripts to convert and process the time-domain NMR data using NMRpipe (Delaglio et al., 1995). Conversion of the Bruker-formatted data to NMRpipe format is accomplished with the `bruk2pipe` script, supplied with NMRpipe. For processing, we typically zero-fill the direct ^1H dimension to 1024 complex points and apply both forward-backward linear prediction and zero-filling in each indirect dimension to 128 or 256 complex points. In depth analysis of spectral processing is beyond the scope of this text and interested readers are directed to Hoch and Stern's treatise (Hoch

and Stern, 1996). After transformation, the resulting frequency-domain spectra are converted to XEASY format (Bartels et al., 1995); however NMRpipe is compatible with any of the other commonly used spectral analysis programs such as Sparky (Goddard and Kneller, 2001), CARS (Keller, 2004/2005), or NMRView (Johnson and Blevins, 1994).

Reduced dimensionality methods for data collection encode multiple indirect frequencies simultaneously rather than independently, enabling the acquisition of four- or higher-dimensional spectra in a short time and reducing the number of necessary experiments for structure determination. Examples include the G-matrix Fourier transform (GFT) method (Kim and Szyperski, 2003; Shen et al., 2005) and HIFI-NMR (Eghbalnia et al., 2005). Non-uniform sampling (NUS), in which data is collected for only a subset of all incremented evolution periods at non-fixed intervals, is another approach to decreased data collection times that is especially promising for larger proteins (Barna et al., 1987; Rovnyak et al., 2004a; Rovnyak et al., 2004b). The recently released TopSpin 3.0 Bruker software package incorporates NUS as part of a routine workflow, including optimized sampling schemes and data processing via Multi-Dimensional Decomposition (Orekhov et al., 2003). Alternatively, processing via maximum entropy reconstruction (Mobli et al., 2007) has been made more accessible to the general protein NMR community by use of the Rowland NMR Toolkit Script Generator, available as a web server at: http://sbtools.uchc.edu/nmr/nmr_toolkit/.

2.2 Peak picking

Once all 2D and 3D data have been processed, the position and intensity of peaks in each experiment must be recorded for submission to an automated chemical shift assignment protocol. Peak picking is one step for which there are relatively few reliable automated programs. As has been discussed by others this is due to issues with noise, artifacts, and strong peak overlap (Altieri and Byrd, 2004; Williamson and Craven, 2009). Because accurate peak picking is crucial for attaining good results with automated assignment programs, most labs traditionally use manual or semi-automated methods assisted by software such as NMRView (Johnson and Blevins, 1994), XEASY (Bartels et al., 1995), or Sparky (Goddard and Kneller, 2001). AUTOPSY (Koradi et al., 1998) is one of the few commonly used automated peak picking programs; however, even with this sophisticated algorithm, peak lists may still require further editing especially in regions of overlapped peaks (Malmodin et al., 2003). It remains to be seen whether recent developments such as the PICKY automated peak picking program (Alipanahi et al., 2009b) or more error-tolerant backbone assignment programs such as MARS (Jung and Zweckstetter, 2004) and IPASS (Alipanahi et al., 2009a) will have an effect on currently established peak picking methods.

Our method for picking peaks of the 3D heteronuclear spectra is a semi-automated approach that uses the automated peak picker of XEASY. The basic steps of our peak picking protocol are outlined below:

1. Pick the ^{15}N -HSQC peaks using the automatic in-phase peak picker of XEASY.
2. After making any necessary manual adjustments to the ^{15}N -HSQC peak list, use this list as the basis for picking the HNCO peaks. The HNCO spectrum enables greater resolution of overlapped ^{15}N -HSQC peaks and thus the ^1H and ^{15}N dimensions of its peaks are used as the 2D root for picking peaks in all 3D heteronuclear experiments and the 3D ^{15}N -edited NOESY.
3. Manually inspect the HNCO peak list in XEASY, checking for any missing or extra peaks.
4. Automatically pick the peaks in the HNCA, HNCACB, HN(CO)CA, HN(CA)CO, C(CO)NH, H(CCO)NH, and HBHA(CO)NH spectra. Manually inspect the spectra for erroneous peaks.

5. The ^{13}C -edited aromatic NOESY is picked separately using the ^{13}C aromatic HSQC peak list as the 2D root. For the ^{15}N -edited NOESY and ^{13}C -edited aliphatic NOESY, we wait to pick the peaks until fully assigned HNCO and HCCH-TOCSY peak lists (see Sections 2.3 and 2.4) are available for use as a starting point for manual peakpicking.

2.3 Backbone chemical shift assignments

With the advent of numerous semi-automated and automated assignment protocols over the last decade, the backbone resonance assignment process, utilizing standard triple-resonance NMR data, has become a relatively routine process in many laboratories. What once took weeks can now be completed in a day or less for proteins under 25 kDa. Automated resonance assignment programs include GARANT (Bartels et al., 1996; Bartels et al., 1997), AutoAssign (Zimmerman et al., 1994; Zimmerman et al., 1997), MARS (Jung and Zweckstetter, 2004), ABACUS (Grishaev et al., 2005b), MATCH (Volk et al., 2008), and PINE (Bahrami et al., 2009), as well as many others which have been previously reviewed (Altieri and Byrd, 2004; Baran et al., 2004; Grishaev et al., 2005a). The programs PINE and AutoAssign have the additional advantage of being readily available as web-based servers: <http://pine.nmrfam.wisc.edu/> and <http://nmr.cabm.rutgers.edu/autoassign/>, respectively. Furthermore, the AutoAssign web server has the option to submit jobs simultaneously to PINE, the results of which can be shown in a comparison chart where the user can easily identify resonance assignment differences between the two programs.

Much of our past work has utilized GARANT for backbone assignments, but PINE is a convenient new alternative that generates equally good results. PINE accepts peak lists from up to 13 3D heteronuclear experiments in either XEASY or Sparky format. The output of PINE is a probabilistic assignment for every residue. This means that it is possible for some residues to have more than one candidate spin system for their chemical shifts. In these cases, an associated probability is assigned to each candidate spin system assigned to a residue. Assigned XEASY peak lists are obtained using PINE as follows:

1. The input files are the sequence file (*.seq), empty chemical shift list (*.prot) and peak lists from HNCA, HNCACB, HN(CO)CA, HN(CA)CO, C(CO)NH, H(CCO)NH, and HBHA(CO)NH experiments, all in XEASY format. The empty *.prot file is required in order to obtain assigned, XEASY-formatted peak lists from the PINE server.
2. In cases where prior partial assignments are available, such as backbone assignments that are unchanged between the apo and bound forms of the protein, a pre-assignment file in BMRB format can be uploaded.
3. The files are uploaded to the server: <http://miranda.nmrfam.wisc.edu/PINE/>. Depending on the size of the protein, the quality of input data, and number of current jobs, run times vary from one hour to one day.
4. The assignment results, which include secondary structure identification and assigned peak lists, are returned via email.
5. The peak lists are then loaded into XEASY for inspection and verification of backbone assignments, and any missing assignments are completed manually. Alternatively, the PINE-SPARKY extension ((Lee et al., 2009); available at <http://pine.nmrfam.wisc.edu/PINE-SPARKY/>) can be used to complete the assignments interactively. This interface allows the user to simultaneously visualize the spectra and the identity of probable peak assignments.

2.4 Sidechain chemical shift assignments

In contrast to the backbone assignments, side chain assignments typically require greater manual intervention due to incomplete data and degeneracy of aliphatic chemical shifts. With ideal C(CO)NH, H(CCO)NH and HCCH-TOCSY peak lists, both PINE and GARANT are capable of generating automatic sidechain assignments. In reality, however, at least 90% completeness of ordered regions is required for automated structure calculations and, therefore, nearly complete assignments require manual analysis (Jee and Guntert, 2003). First, the partial aliphatic sidechain assignments obtained from PINE are verified. From these assignments, the side chain ^1H - ^{13}C crosspeaks in the patterns expected for an HCCH-TOCSY spectrum can be generated using GARANT. This peak list forms the starting point for manual completion of all aliphatic sidechain assignments; depending on the completeness of HCCH-TOCSY assignments the manual processing can be completed within one day. Finally, aromatic side chain assignments are completed manually using the ^{13}C aromatic NOESY and the previously determined aliphatic side chain assignments.

1. Verify the assignments for the HSQC, HNCO, HNCACB, HN(CA)CO, C(CO)NH, HBHA(CO)NH and H(CCO)NH peak lists. Add any missing assignments and save the resulting *.prot file.
2. Generate the HCCH-TOCSY peak list with the genPeaks command of GARANT or an equivalent utility:

```
genPeaks HCCH24 protein.seq protein.prot hcch.peaks
```
3. Examine the HCCH-TOCSY peak list in XEASY or another spectral assignment program, picking peaks in any unassigned spin systems. Correct any incorrect or missing assignments; only the diagonal peaks need to be assigned. The methionine H^ϵ peaks can typically be recognized by their characteristic ^1H and ^{13}C shifts and the lack of any other correlations. These and other sidechain chemical shifts that cannot be assigned due to overlapped or missing resonances are completed later during analysis of the ^{13}C -edited NOESY.
4. Assign NH_2 chemical shifts in the ^{15}N -HSQC on the basis of $\text{H}^\beta\text{-NH}_2$ and $\text{H}^\gamma\text{-NH}_2/\text{H}^\beta\text{-NH}_2$ NOE correlations in the ^{15}N -edited NOESY for asparagine and glutamine, respectively.
5. Assign the $\text{C}^\delta\text{H}^\delta$ groups of the ^{13}C aromatic NOESY peak list (diagonal peaks only) according to NOE correlations with intraresidual H^N , H^α , and H^β resonances (Lin et al., 2006). Once the chemical shifts of the $\text{C}^\delta\text{H}^\delta$ groups are known, their neighboring $\text{C}^\text{x}\text{H}^\text{x}$ or $\text{N}^\text{x}\text{H}^\text{x}$ groups can be assigned.
6. Create a final *.prot file by incorporating assigned chemical shifts from the following experiments in this order:
 1. HCCH-TOCSY
 2. ^{13}C aromatic NOESY
 3. HNCO
 4. HSQC
7. Proline ^{13}C chemical shifts can identify cis Xaa-Pro peptide bonds (Stanczyk et al., 1989). Checking for the cis versus trans proline conformation can be done automatically by using the CYANA macro 'cisprocheck' or a program such as POP (Schubert et al., 2002). Prolines in the cis conformation should be renamed 'cPRO' in the XEASY/CYANA format sequence file. Likewise, any cysteine sidechains known to participate in a disulfide bridge should be renamed 'CYSS'.

2.5 Ligand assignments

The assignment strategy for small molecules is complicated by the lack of isotopic enrichment of ^{15}N and ^{13}C . In the free state, assignment of all protons in the small molecule can be completed using a set of 2D homonuclear spectra that include 2D ^1H -TOCSY, 2D ^1H -NOESY and/or 2D ^1H -ROESY. However, ligand binding to its target protein induces chemical shift changes in both the protein and ligand with varying magnitudes that are dependent on the local chemical environment. To assign an unlabeled ligand in the bound state, one can use a 2D F1/F2- $^{13}\text{C}/^{15}\text{N}$ -doubly-filtered NOESY spectrum (Ikura and Bax, 1992) as previously described (Alam et al., 1998; Peterson et al., 2010). In the doubly-filtered NOESY spectrum, only NOE cross peaks involving ^{12}C -attached protons as the source and destination are present in the final spectrum. Since ^{12}C -attached protons only occur within the small molecule, the doubly-filtered NOESY spectrum will only show intramolecular NOEs for the ligand. These NOEs can then be used to sequentially assign the proton shifts for the small molecule.

2.6 Validation

Because the completeness and accuracy of the chemical shift assignments is crucial for the robustness of the subsequent steps of structure determination, the assignments should be checked for errors before proceeding with structure determination. TALOS and CYANA, programs described in Sections 4.2.1 and 4.2.2, automatically check for chemical shifts that are significantly outside the typical range for the particular atom/residue and also report statistics on the completeness of the assignments. A variety of other programs available for validating chemical shift referencing and/or assignments have recently been evaluated (Wang et al., 2010).

3 Identification of ligand binding sites by chemical shift mapping

3.1 HSQC titrations with fragments, compounds and mixtures

Fragment-based drug design is most effective for identifying relatively weak ($K_d > 10^{-6}$ M) compounds with the potential for development into high-affinity molecules. NMR is an optimal technique to test the quality of these fragment leads because it can provide atom-specific information that is most sensitive to low-affinity interactions. A basic outline of how NMR can contribute to the rational drug development process is outlined in Figure 2. Once the chemical shift assignments of the target protein are known (see Section 2), [$U\text{-}^{15}\text{N},^{13}\text{C}$]-enriched protein can be titrated with a putative ligand or binding partner and the resulting changes in chemical shift monitored by 2D correlation spectroscopy. This procedure is called *chemical shift mapping* and can be used to determine binding constants, stoichiometry, and ligand position. 2D ^1H - ^{15}N heteronuclear single-quantum coherence (HSQC) spectroscopy is the most common method for mapping backbone chemical shifts because every residue type (except proline) contains a highly sensitive probe to changes in local magnetic environment. HMQC (Mandal and Majumdar, 2004) and TROSY (Pervushin et al., 1997) are alternative pulse schemes that provide advantages in certain cases and supply equivalent spectral information. As most protein-ligand interactions are sidechain mediated, 2D ^1H - ^{13}C HSQCs can provide additional information, but the sidechain signals suffer from overlap in large target proteins. Thus selective ^{13}C -labeling of Ile, Leu, and Val residues can significantly improve resolution without compromising sensitivity (Meyer and Peters, 2003).

Titration are typically performed two ways: 1) add known ligand volumes to a target-protein sample, or 2) prepare multiple samples with a constant protein concentration and increasing ligand concentrations. The former method consumes substantially less protein, but suffers from less accurate binding constants due to changing protein concentrations.

Using cryoprobe technology without a sample-handling robot, three unique compounds can typically be analyzed per day using this scheme:

1. Collect initial HSQC spectra on 50 μ M protein sample.
2. Collect additional HSQC spectra with ligand additions of 25, 50, and 800 μ M; comparison of four data points facilitates exchange regime determination and binding affinity quantitation (Fielding, 2003).
3. Overlay spectra and qualitatively examine magnitude of chemical shift perturbation, specificity, and exchange regime (Fig. 3A).
4. Measure chemical shift values of each assigned resonance throughout titration, determine adjusted chemical shift (Eq. 5), and plot as a function of residue number (Fig. 3B). Use chemical shift perturbations to quantify binding affinity (Eq. 6; Fig. 3C). If the target protein structure is known (see Section 4 for structure determination), residues with the most significant perturbations can be highlighted (Fig. 3D); this is particularly valuable for identifying ligand specificity when compound docking models have been calculated.

NMR validation can be the most time-consuming aspect of screening large numbers of compounds. To improve throughput, compounds can be grouped together and simultaneously screened by NMR. The most promising mixtures are then analyzed individually. Alternatively, “chromatographic” NMR experiments have been described to use 2D diffusion information to identify high-affinity small molecules from mixtures (Gonnella et al., 1998; Gounarides et al., 1999; Hajduk et al., 1997a; Hajduk et al., 1997b). The gain in efficiency of mixture analysis strategies is difficult to assess; Ross and colleagues argue the high binding success rate ($\sim$ 1:3) of compounds suggests the extra effort spent to deconvolute spectra should be redirected to initially preparing more reagents for individual analysis (Ross et al., 2000). Automated sample handling and rapid 2D NMR acquisition schemes, such as SOFAST-HMQC (Schanda and Brutscher, 2006; Schanda et al., 2005) and single-scan spatial-encoding (Frydman et al., 2002; Gal et al., 2006), can be expected to increase the throughput of NMR-based ligand screening in the near future.

3.2 Discrimination between specific and non-specific binding

As many small molecules are hydrophobic, the solubility of ligands is often improved by reconstitution in an organic solvent; thus, the solvent itself may interact with the target protein. These interactions may be non-saturable but can produce perturbations of substantial magnitude, and may even reveal the locations of druggable sites in the protein (Byerly et al., 2002; Dalvit et al., 1999; Liepinsh and Otting, 1997). Solvent perturbations are relatively simple to detect by performing a solvent-only titration for comparison with spectra reporting on candidate ligands (Fig. 3B).

Another facet of non-specific binding is the orientation or mode that a compound adopts when associated with a target protein. Lead compounds that exhibit multiple-mode binding should be modified to maintain a single-mode prior to optimization. Analysis of titration lineshapes can quickly identify binding modes (Reibarkh et al., 2006); single-mode ligands with micromolar affinity should produce line-broadening at substoichiometric concentrations that sharpen upon saturation. Protein resonances that do not sharpen generally reflect multiple-mode ligand binding. However, this does not preclude the possibility that ligands can modulate the oligomeric state (Veldkamp et al., 2005) or conformation of the protein (Vajpai et al., 2008).

As chemical shift perturbations can result from both ligand-binding and ligand-induced conformational changes, discrimination between direct and indirect effects is crucial to

effective chemical shift mapping. Residues with the largest changes in chemical shift are presumed to be specifically involved in the binding interaction as ligand-induced conformation changes are rare in low-affinity interactions (Cavanagh, 2007). However, the binding site of ligands interacting with target proteins with large conformation changes can be identified by the use of differential chemical shifts. Fesik and colleagues outline a protocol to compare chemical shift perturbations of highly similar compounds (Medek et al., 2000). The chemical shift changes resulting from conformation change should be similar while ligand-induced perturbations will differ by chemical uniqueness. We performed a similar analysis of differential chemical shifts between sulfated and unsulfated peptides to identify the sulfotyrosine recognition site of CXCL12 in the absence of explicit structural data (Veldkamp et al., 2006).

3.3 Quantitation of binding constants by NMR

The determination of binding constants during compound screening is crucial to ranking compounds for optimization and lead selection. The ability to quantitate affinity by NMR is highly dependent on the binding kinetics. The following is a general outline of the theoretical principals for determining binding reactions on the NMR time scale [for detailed texts on NMR kinetics, see (Cavanagh, 2007; Levitt, 2008)]. Assuming a simple two-state equilibrium, for a protein (*P*) binding to a ligand (*L*) to form a complex (*PL*), with second-order binding kinetics,

where k_{on} and k_{off} are the rate constants of association and dissociation, respectively. The rate constant for exchange (k_{ex}), or the chemical exchange lifetime ($\tau_{ex} = 1/k_{ex}$), is calculated as

$$k_{ex} = k_{on}[L] + k_{off} \quad (2)$$

Comparison of the exchange rate constant (k_{ex}) with the difference in angular frequency ($\Delta\omega$) between an unbound (ω_p) and bound (ω_{PL}) resonance,

$$\Delta\omega = \omega_p - \omega_{PL}, \quad (3)$$

categorizes the binding interaction into three basic exchange regimes:

$$\begin{aligned} \text{Fast exchange: } k_{ex} &\gg |\Delta\omega| \\ \text{Intermediate exchange: } k_{ex} &\sim |\Delta\omega| \\ \text{Slow exchange: } k_{ex} &= |\Delta\omega| \end{aligned} \quad (4)$$

The particular exchange regime a binding reaction belongs to can be readily identified by observation of each resonance throughout a titration; it is important to note that each atom can exhibit its own rate of exchange for a particular binding reaction.

Protein resonances in fast exchange do not change significantly in intensity and continuously exhibit chemical shift changes through the titration. Fast-intermediate exchange is similar to fast-exchange with the exception of diminished intensity and linewidth as the chemical shift changes from unbound and bound. Resonances that significantly diminish, possibly below detection, and then reappear at a different chemical

shift with increasing ligand concentrations are in intermediate exchange. In practice, resonances in intermediate exchange may not reappear because it is not possible to get high enough ligand concentrations depending on solubility and protein concentration. Slow exchange is characterized by a resonance that decreases in intensity but does not change chemical shift with greater ligand concentration; at the same time, a resonance with a unique chemical shift will begin to increase in intensity. Visual inspection of binding kinetics provide a qualitative interpretation of binding affinity as fast exchange is weak ($K_d > 10^{-6}$ M), intermediate exchange is generally between 10 and 100 μ M, and slow exchange is strong ($K_d < 10^{-6}$ M).

To facilitate high-throughput quantitative analysis we will only consider binding reactions in fast exchange. We presume this is valid because most compounds screened will have affinities weaker than 10^{-6} M; however, to minimize error the need to ensure fast-exchange binding kinetics cannot be understated. Interested readers are directed toward excellent reviews on quantitating slow and intermediate exchange kinetics (Feeney et al., 1979; Fielding, 2003). This is especially important during lead optimization as compounds are expected to possess high affinities. HSQC chemical shift changes are quantified for each titration point using an adjusted chemical shift value (Δ) of the form,

$$\Delta = \sqrt{(\Delta\delta_H)^2 + (c\Delta\delta_N)^2} \quad (5)$$

in which $\Delta\delta_H$ and $\Delta\delta_N$ are the chemical shift change of amide proton and nitrogen, respectively. Many values for the scaling factor, c , have been proposed ranging from 0.1 to 0.25 (Geyer et al., 1997; Mulder et al., 1999); we typically employ a value of 0.2 to balance the contributions of ^1H and ^{15}N to the magnitude of the perturbation. More complex formulations that use scaling factors specific to each amino acid have also been proposed (Geyer et al., 1997; Schumann et al., 2007). A similar range of scaling factors has been proposed for the measurement of combined ^1H - ^{13}C shift perturbations (Schumann et al., 2007). The increase in adjusted chemical shift perturbation as a function of ligand concentration is then fitted by non-linear regression using an equation that accounts for ligand depletion, single-site binding, and non-specific binding:

$$\Delta = (\Delta_{\max} / 2[P_T]) \times \left\{ K_d + [P_T] + [L_T] - \sqrt{(K_d + [P_T] + [L_T])^2 - 4[P_T][L_T]} \right\} + NS[L_T] \quad (6)$$

in which Δ is the adjusted chemical shift change, $\Delta_{\max}$ is the maximum adjusted chemical shift change at saturation, $[P_T]$ is total protein concentration, K_d is the equilibrium dissociation constant, $[L_T]$ is the total ligand concentration, and NS is non-specific binding.

3.4 Pharmacophore modeling from comparative shift perturbation analysis

The chemical features of a protein-ligand interface that are responsible for a given biological function define a pharmacophore. They include the hydrogen bond vectors, hydrophobic volumes, and steric properties of active ligands that identify potential inhibitors. It is important to note that pharmacophores do not correspond to common functional groups, and therefore occupy a unique niche in structure-based drug design. A pharmacophore model is often used as a pre-filter for *in silico* screening because it can be evaluated more rapidly than explicit docking and may be less prone to susceptible to false positive hits. In a ligand-based computational process, the pharmacophore definition can be enhanced by NMR titration and structure data. Specifically, using the method for comparative chemical shift perturbation analysis (Medek et al., 2000) outlined in Section 3.2, one can identify the binding site of a

series of related lead compounds. Once the binding modes of several compounds have been determined by comparative analysis, the pharmacophore elucidation can be performed by any number of programs [as reviewed by (Leach et al., 2010)]. Structural knowledge of a protein provides invaluable information of exclusion regions (Greenidge et al., 1998; Rella et al., 2006; Tintori et al., 2008); knowledge of binding site volumes accelerates compound docking considerably as only molecules that will fit are examined.

3.5 Epitope definition for docking and *in silico* screening

Both traditionally druggable proteins (enzymes), with distinct binding pockets, and protein-protein interfaces, with relatively flat surfaces, are being successfully targeted for small-molecule inhibitors (Veldkamp et al., 2010; Wells and McClendon, 2007). An integral part of rational drug design is identification of biologically relevant binding surfaces or epitopes. When target protein structures are known, knowledge of epitopes are invaluable for directed compound docking and reducing the computational load of screening chemical libraries. Traditionally, epitope definition utilized structures solved by crystallography; however, NMR is an established alternative that is especially suited for solving weak complexes that are frequently not crystallizable. When full structure determination of the complex is not feasible, NMR can still yield epitope information from HSQC titrations. In the absence of high-resolution structural data, ligand binding epitopes can be discerned through site-directed mutagenesis or antibody inhibition.

A variety of docking programs have been described for computational screening of chemical libraries, including AutoDock (Morris et al., 1998), DOCK (Ewing et al., 2001), eHiTs (Zsoldos et al., 2006), FlexX (Rarey et al., 1996), and Fred (McGann et al., 2003). In collaboration with other experts, we have successfully implemented epitope-directed compound screening against cytokine-receptor signaling systems using both the DOCK and AutoDock packages. These programs can be used for screening of both private and public chemical libraries. For example, we recently used the NMR structure of the chemokine CXCL12 in complex with a 38-residue peptide of the CXCR4 receptor (Veldkamp et al., 2008) to identify a small molecule ligand for CXCL12 that inhibits CXCR4-mediated calcium flux (Veldkamp et al., 2010). Specifically, we used the CXCL12-binding site of two CXCR4 residues to define an epitope. Using DOCK, we targeted 1.4 million compounds from the ZINC database (Irwin and Shoichet, 2005), a free virtual library of commercially available small molecules, at this site. In an NMR-based assay of the top-ranked molecules, four of the five tested compounds bound specifically to the targeted epitope, including one with a K_d value of 64 μ M. The success of this study and others (Arkin and Wells, 2004; Berg, 2008; Fuller et al., 2009) demonstrates the value of NMR binding data to identify epitopes for virtual screening within the rational drug development process.

4 3D structure determination by NMR

4.1 NMR-derived structural constraints

Solution NMR structure calculations rely mainly on two types of geometric constraints: (1) distance constraints based on analysis of 3D ^{15}N - and ^{13}C -edited NOESY spectra and (2) dihedral angle constraints predicted from chemical shifts. When disulfide bonds, bound metal ions or other cofactors of known geometry are present, additional constraints are typically included in the form of generic upper and lower distance limits to enforce the covalent crosslinks or coordinating bonds. The use of other restraints, such as residual dipolar couplings (RDCs) and pseudocontact shifts, can be directly incorporated into CYANA 3.0, an updated version of the CYANA structure determination program (Herrmann et al., 2002). However, in our experience these constraint types are not essential

to obtain high quality structural models for most diamagnetic globular domains under ~25 kDa.

4.1.1 Dihedral angles inferred from chemical shifts—After the chemical shift assignments have been completed, constraints for the backbone ϕ and ψ dihedral angles are generated using the program TALOS+ (Shen et al., 2009). Alternatively, 3J -scalar coupling measurements can be used to estimate dihedral angles from empirically-defined relationships (Karplus, 1959); however, our streamlined approach employs chemical shift-derived dihedral angle constraints as they can be obtained without the need for additional spectral acquisition. TALOS+, like the original program TALOS (Cornilescu et al., 1999), compares the H^N , N , C' , H^α , C^α , and C^β chemical shifts from the protein against a database of chemical shifts and backbone geometry in tripeptide segments from known protein structures. The TALOS+ database currently contains 200 proteins. As with the original TALOS program, the ten best database matches for each residue (based on up to 18 individual chemical shifts: six atoms each from three amino acids) are compared to see if they cluster in a consistent region of the Ramachandran (ϕ vs. ψ angle) plot. TALOS+ utilizes two additional parameters for making predictions. The first is a neural network component whose output is used as an additional empirical term in the conventional TALOS database search. The second is an estimated backbone order parameter S^2 derived from the chemical shifts using the random coil index method (RCI; (Berjanskii and Wishart, 2005)), which biases against the prediction of backbone dihedral angles in flexible regions. An interface for inspecting and modifying the results is provided, after which the acceptable predictions are converted into a dihedral angle constraint file for use in structure calculations (Cornilescu et al., 1999). TALOS+ also provides an additional feature to pre-check chemical shift referencing and possible chemical shift errors. Tabular chemical shift data in the TALOS input format is easily generated either by using the CYANA macro '*taloslist*' for chemical shifts in the CYANA/XEASY(.prot) format or with the *bmr2talos* script supplied with the TALOS+ software. We have further simplified the data conversion and TALOS+ calculation process with an in-house script (available upon request).

4.1.2 NOE distance constraints—Distance restraints derived from proton-proton nuclear Overhauser effects (NOEs) are the primary NMR data used to define the secondary and tertiary structure of a protein. NOEs develop due to through-space dipole-dipole interactions rather than through-bond interactions. Thus, they contain information on the distances between hydrogen atoms separated by 5 Å or less in space even though the residues may be distal in primary sequence. The intensity of a NOE, that is the volume of the corresponding crosspeak in a NOESY spectrum, is proportional to the inverse of the sixth power of the distance between the two nuclei due to averaging caused by rotational motion. However, because the NOE is not always a precise reflection of the internuclear distance, structure calculation programs generally translate NOESY crosspeak intensities into upper bounds on interatomic distances rather than fixed distance restraints.

With the availability of a variety of automated assignment programs, the generation of NOE distance constraints from NOESY spectra is a relatively robust and automatic process, at least for the calculation of initial structures. These structures can then be used, if necessary, to expand and correct the preliminary list of NOE assignments for input into another round of structure calculations. The process continues in an iterative fashion, making adjustments to the input parameters and experimental constraints until an acceptable final ensemble of structures is obtained. Programs used for automated assignment include NOAH (Mumenthaler and Braun, 1995), ARIA (Linge et al., 2003a; Nilges, 1995; Rieping et al., 2007), CYANA (Guntert, 2003, 2004; Herrmann et al., 2002), PASD (Kuszewski et al., 2004; Kuszewski et al., 2008), and AutoStructure (Huang et al., 2005; Huang et al., 2006), all of which have been reviewed previously (Guntert, 2003; Williamson and Craven, 2009).

Although the automated NOE assignment methods require NOESY peak lists that are accurately referenced relative to the chemical shift assignments, some programs can deal robustly with noisy peak lists. It has been demonstrated that the NOEASSIGN module of CYANA, version 2.1, can yield reliable NOE distance constraints even when one-half to two-thirds of the crosspeaks in the initial NOESY peak lists are noisy or artifactual (Lopez-Mendez and Guntert, 2006), while PASD can tolerate up to 80% incorrectly picked peaks (Kuszewski et al., 2004).

The NOEASSIGN module of CYANA is our preferred tool for automated NOESY peak assignments. Interproton NOE distance constraints generated by NOEASSIGN are derived from three different 3D NOESY spectra: ^{15}N -edited NOESY-HSQC, ^{13}C -aliphatic-edited NOESY-HSQC, and ^{13}C -aromatic-edited NOESY-HSQC. An additional experiment, the 3D F1- $^{13}\text{C}/^{15}\text{N}$ -filtered/F3- ^{13}C -edited NOESY, is added for protein-ligand structures. Because the NOEASSIGN module can cope with a high number of artifacts in the peak lists, only minor manual editing should be necessary at this stage as long as the NOESY spectral referencing matches that of the ^{15}N -HSQC/HNCO and HCCH-TOSCY spectra used to obtain the chemical shift values. Because automated NOE assignment is inseparable from the structure calculation process, the details are described in Section 4.2.

4.2 Automated NMR structure refinement with CYANA

4.2.1 Solving the structure of the free protein—Over the past six years, we have successfully used NOEASSIGN-generated distance constraints, in combination with dihedral angle constraints, to solve more than 20 unique structures with CYANA (de la Cruz et al., 2007; Lytle et al., 2004; Lytle et al., 2006; Peterson et al., 2007; Peterson et al., 2006; Peterson et al., 2005; Tuinstra et al., 2007; Vinarov et al., 2004; Waltner et al., 2005). This list includes not only single domain proteins but also symmetrical homodimers (Peterson et al., 2010; Peterson et al., 2006) and protein-ligand complexes (Veldkamp et al., 2008). The NOEASSIGN algorithm introduced in CYANA 2.0 is an improved version of the original CANDID algorithm (Herrmann et al., 2002) that is based on a probabilistic treatment of the NOE assignment process. An overall probability of the correctness of possible NOE assignments is calculated as the product of the probabilities for each of the criteria for assignment: agreement between the chemical shift values and the peak positions, compatibility with a preliminary 3D structure, and network-anchoring, i.e. the extent of embedding in the network of all other NOE assignments. Automated NOE assignment and structure calculation by the CYANA torsion angle dynamics (TAD) algorithm (Guntert et al., 1997) are combined in an iterative process that comprises seven cycles of automated NOE assignment, followed by a final structure calculation using only unambiguously assigned distance constraints. The structural result of a given cycle is used to guide the NOE assignments in the following cycle, with the precision of the structure normally improving with each subsequent cycle. In the first two cycles, “constraint combination” is applied to medium- and long-range distance restraints in order to reduce the impact of erroneous distance restraints on the resulting structure. In all seven cycles, ambiguous distance constraints are used to generate conformational restraints from NOESY cross peaks with multiple possible assignments.

In our experience the NOEASSIGN routine, given the proper input, consistently results in a converged set of initial structures for monomeric proteins. In favorable cases, the result is a well-defined structural ensemble, with backbone RMSD $< 1 \text{ \AA}$ and target function $< 5 \text{ \AA}^2$. Structures can typically be obtained in less than an hour and often under thirty minutes, depending on the size of the protein and available computing power. For example, for a 160-residue protein computation time is 50 minutes using an 8-core Apple Mac Pro or about 15 minutes with a 16 node/32 processor Linux cluster.

An example script for running a CYANA calculation with the automated NOE assignment module can be found on the CYANA homepage (http://www.cyana.org/wiki/index.php/Main_Page). The order of the tolerance parameters used by NOEASSIGN is (1) direct ^1H , (2) indirect ^1H , and (3) indirect $^{15}\text{N}/^{13}\text{C}$. We typically use values of 0.03 ppm for both ^1H dimensions and 0.3 ppm for the ^{15}N and ^{13}C dimensions. By leaving the calibration parameter line blank, these parameters are determined automatically such that the median of the upper distance limits for each peak list equals the value of the reference distance variable *dref*, which has a default value of 4.0 Å. In each cycle, and in the final structure calculation, 100 conformers are calculated using the standard simulated annealing schedule with 10000 torsion angle dynamics steps per conformer. The 20 conformers with the lowest final target function values are analyzed using a variety of statistics, including overall target function, restraint totals, and ensemble coordinate precision (atomic r.m.s.d). In our protocol, NOEASSIGN is initially run in triplicate with identical input files (sequence file, prot file, aco file, and peak lists), changing only the value of '*randomseed*' for each run. The results of the three NOEASSIGN calculations are then evaluated. While each final structure should converge to fairly similar RMSD and target function values, as reported in the *final.ovw* file, it is more important to assess whether all three structural ensembles display the same overall tertiary fold by using a structure visualization program, such as PyMol or Molmol (Koradi et al., 1996). The presence of any significant deviations is indicative of misassigned NOEs, which may reflect inaccuracies in the chemical shift assignment list. These misassignments can have a significant impact on the structure that may go undetected as refinement proceeds; therefore, it is necessary that they be corrected at the outset. Constraint combination applied during the first two NOEASSIGN cycles helps to minimize these potential structural distortions, but it is not a panacea. It is also important to inspect the statistics of the initial structural ensemble using the '*cyanatable*' script, because a poorly defined structure ensemble from cycle 1 may ultimately converge in later cycles on a precise but inaccurate final ensemble. NOEASSIGN output can be further evaluated based on the following two criteria for a successful calculation: 1) Less than 25% of the long-range NOEs have been discarded by the automated NOESY assignment routine for the final structure calculation and 2) the backbone RMSD to the mean for the structure bundle of cycle 1 is < 3 Å (Guntert, 2004).

When a satisfactory, consistent ensemble has been obtained using NOEASSIGN, it is typically necessary inspect and edit the NOESY peak assignments in XEASY (starting from the peaklists generated in cycle 7), focusing on the largest consistent distance and angle violations listed in the *.ovw file as well as any disordered regions or inconsistent conformations between replicate ensembles. Erroneous or missing chemical shift assignments should also be corrected by looking for spin systems (NOESY strips) that remain largely unassigned. NOEASSIGN calculations may then be repeated to satisfy the two criteria defined above. After correcting and verifying all peak lists and chemical shift assignments, a lack of convergence in cycle 1 could indicate the presence of a homodimeric assembly that was not apparent from earlier analysis of the spectra or hydrodynamic measurements of oligomeric state. After obtaining a satisfactory ensemble of structures by NOEASSIGN, manual structure refinement is performed in order to prepare the ensemble from torsion angle dynamics for a final Cartesian-space refinement calculation in explicit water and subsequent validation steps.

Manual structure refinement proceeds iteratively, using CYANA to calculate an ensemble of structures based on the assigned peak lists (an example script can be found at the CYANA homepage). At the outset of manual refinement, the calibration of NOE intensity/distance conversions should be optimized to decrease both the RMSD and target function. Initial runs are performed with automated calibration in which the parameters for the backbone, sidechain, and methyl calibration functions are calculated automatically by the CYANA

macro ‘*calibration*’. The calculated parameters found in the output file are then used as a starting point for fine-tuning the calibration. This is accomplished by testing a grid of calibration values (at least three for each peaklist; nine calculations altogether) to determine the combination that results in the lowest target function and RMSD. The calibration should also be checked by comparing known interproton distances within secondary structures from well-resolved X-ray structures (Hur and Karplus, 2005) to the corresponding distances in the upper limit distance constraint (.upl) file.

Iterative rounds of CYANA calculations are then performed until no further improvement can be made in the structure. Between each CYANA calculation, the NOESY peak lists are inspected and peak assignments are added, removed, or corrected based on the results of the structure calculations. The consistently violated constraints should be corrected by reassigning crosspeaks, referring to the structure as necessary when there are multiple possibilities. PyMol or Molmol can be used to obtain a list of all atoms within 5.0 Å of an atom of interest. Special attention should be given to verifying the aromatic side chain assignments, since these residues often occupy central locations in the hydrophobic core. In addition to adding and correcting NOE assignments, any consistently violated dihedral angle constraints should be reevaluated or eliminated. Since TALOS constraints rely on chemical shift-based predictions with a finite error rate, a conservative approach preserves only those dihedral angle constraints that correspond to regions of recognizable secondary structure.

4.2.2 Defining CYANA library entries for ligands—Automated NOE structure refinement of protein-ligand complexes using CYANA requires the creation of a molecular topology description for each component of the complex that is not found in the standard CYANA library (e.g. amino acids and nucleotides). The library definition or topology file for a small molecule ligand contains a description of atom types and nomenclature, covalent connectivities, dihedral angle definitions, and standard geometry of the small molecule. The requirement for such topology files in solving the structures of protein-ligand complexes has increased in parallel with structure-based drug design and was initially addressed by the PRODRG server (Schüttelkopf and van Aalten, 2004) (<http://davapc1.bioch.dundee.ac.uk/prodrg/>). The PRODRG server takes a PDB file or simple text drawings and generates topology files for GROMOS, GROMACS, WHAT_IF, REFMAC5, CNS, O, SHELX, HEX, and MOL2. However, these formats are not suitable for a CYANA topology file. The pdb2reslib module of Olivia (<http://fermi.pharm.hokudai.ac.jp/olivia/>) addresses this need and allows the user to generate a CYANA topology file from a PDB file. In this section, we describe the process of generating the CYANA topology description for an arbitrary ligand molecule.

Ligand library entries can be generated using a PDB coordinate file as a convenient, widely available description of molecular structure. If a PDB file is not readily available, one can be created by importing a simplified molecular input line entry specification (SMILES) string (Weininger, 1988; Weininger et al., 1989) for the molecule into a graphical chemical editor such as Marvin Sketch (<http://www.chemaxon.com/products/marvin/marvinsketch/>). The ligand model can then be exported in standard Protein Data Bank format and converted to the CYANA library format using the pdb2reslib module of Olivia. After conversion, the new ligand definition should be inspected and edited to add missing dihedral angles (or remove unnecessary angles), adjust atom nomenclature and numbering, and adjust dihedral angle naming prior to use in CYANA (Figure 4). The edited entry can then be appended to the standard CYANA library file with a unique residue identifier (e.g. DRG). Finally, because molecular dynamics calculations in torsion angle space requires that all components of the system be linked by covalent or virtual bonds in a single chain, the ligand must be incorporated into the protein complex by appending linker residues (e.g. PL, LL, LL2, etc.)

and the ligand residue to the sequence (.seq) file, which CYANA will automatically interpret and use to link the ligand to the polypeptide chain.

4.2.3 Structure determination of a protein-ligand complex—The structure of a protein-ligand complex is highly valuable in the early stages of structure-based drug design, where lead compounds that interact weakly with their targets may not readily crystallize. In general, the NMR structure of a protein-ligand complex is solved in the same way as the protein itself, as described in the previous sections. The availability of chemical shift and NOE assignments for the ligand-free protein, along with ligand-induced chemical shift perturbation data, simplifies the assignment process for the protein complex. However, structure determination of a protein-ligand complex presents a challenge similar to what we have previously described for NMR structure determination of symmetric protein homodimers (Markley et al., 2009b). As with a protein homodimer, the identification and assignment of intermolecular NOEs between protein and ligand is the principal hurdle, as they are indistinguishable from intramolecular protein NOEs in standard isotope-edited NOESY spectra. Chemical shift degeneracy and their low relative abundance add to the difficulty of detecting NOE contacts that define protein-ligand interface. To overcome these problems, we acquire a 3D F1- $^{13}\text{C}/^{15}\text{N}$ -filtered, F3- ^{13}C -edited NOESY spectrum (Stuart et al., 1999) on a sample of $^{15}\text{N}/^{13}\text{C}$ labeled protein saturated with unlabeled ligand (Figure 5). This experiment detects only NOEs arising between protons directly bound to ^{13}C (protein) and protons bound to any non- ^{15}N or ^{13}C atom (ligand) while suppressing all other cross-peaks by isotope filtering and editing. Constraints generated from the filtered NOESY spectrum can therefore be unambiguously assigned to protons on the small molecule ligand. Additionally, CYANA 3.0 now allows the user to specify NOEs as intermolecular by setting the ‘color’ value of the peak list (column 5) to 9. If $^{13}\text{C}/^{15}\text{N}$ labeled ligand is available, a 3D F1- $^{13}\text{C}/^{15}\text{N}$ -filtered, F3- ^{13}C -edited NOESY spectrum can be collected on a sample containing unlabeled protein and labeled ligand, thereby providing a complementary dataset that further disambiguates atom assignments for intermolecular NOEs. We employed this approach to solve the structures of two protein-peptide complexes, for which [$U\text{-}^{15}\text{N}/^{13}\text{C}$] peptide ligands were produced using standard methods for expression and labeling in *E. coli* (Tyler et al., 2010; Veldkamp et al., 2008).

4.3 Structure refinement in explicit water and validation

The final stages of refinement are reached when the desired goals for precision (e.g. backbone RMSD below ~ 0.6 Å), agreement with experimental constraints (none violated by more than 0.5 Å or 5°), and good torsion angle geometry (no residues in disallowed regions of ϕ/ψ space) have been achieved. Typically, this requires manual inspection and editing of the NOE assignments that result in consistently violated constraints as we have previously described (Markley et al., 2009b). NMR structures defined by $> 15 - 20$ non-trivial constraints/residue typically exhibit high coordinate precision, with RMSD values of 0.5 – 0.7 Å for backbone atoms and 0.8 – 1.2 Å for all non-hydrogen atoms. However, it is important to assess the local mobility of the polypeptide backbone using ^{15}N relaxation measurements such as ^{15}N - ^1H heteronuclear NOE, since flexible regions will often be devoid of long-range NOEs. Consequently, the local NOE constraint density in flexible regions is expected to be low and no extra effort should be made to identify long-range NOE constraints for those residues. Once a consistent set of experimental constraints has been obtained, and the resulting structural ensemble meets the basic acceptance criteria, a final refinement calculation is performed in explicit water with physical force field parameters.

To simplify the water refinement routine, we have developed a script that converts the output of a CYANA calculation into input for water refinement and initiates the water refinement routine as previously described (Markley et al., 2009b). Our protocol utilizes a

method for Cartesian space molecular dynamics in explicit water using XPLOR-NIH that was shown previously to improve the stereochemical quality of NMR structural ensembles generated with simple force fields (Linge et al., 2003b). Improvements in a variety of validation criteria were observed upon water refinement of NMR ensembles, particularly for the percentage of residues in the most favored region of the Ramachandran (ϕ vs. ψ angle) plot. While this routine is not fully integrated into CYANA calculations, the user can initiate a water refinement run with a single script once a consistent set of experimental constraints has been achieved. The initial structural model and the model refined using restrained molecular dynamics in explicit water are validated against the distance and dihedral angle constraints using WHAT IF and PROCHECK-NMR. Comparison of the input and refinement statistics shows that the Ramachandran statistics of the TAD ensemble from CYANA and the WHAT IF RMS Z scores are systematically improved by refinement in explicit water. Additional information, protocols and scripts for final structural refinement in explicit water can be obtained from the Northeast Structural Genomics wiki (http://www.nmr2.buffalo.edu/nescg.wiki/Structure_Refinement_Using_CNS_Energy_Minimization_With_Explicit_Water).

5 Conclusion

Solution NMR is a powerful tool for the characterization and structural determination of protein-ligand complexes. As a result of structural genomics efforts aimed toward automation of the principal bottlenecks in protein structure determination by NMR, structures of proteins and domains under ~ 30 kDa can now be solved in a few weeks. Consequently, NMR spectroscopy is a feasible alternative to X-ray structure determination for drug discovery applications, particularly for flexible proteins or membrane proteins that are not amenable to crystallization. The efficiency of our assignment and structure determination protocol is greatly enhanced by maintaining the essential tools: NMR pulse programs and instrument calibrations, spectral conversion and processing scripts, peak picking scripts, and standardized scripts for running programs such as TALOS+, CYANA, and NIH-XPLOR, all of which are available from the authors upon request. In addition to its capacity to define the structures of protein complexes, NMR remains a robust method to screen and characterize protein-ligand complexes at the atomic level. Each of these applications will be useful as new targets are validated and enter the pipelines for structure-based drug discovery.

References

- Alam SL, Volkman BF, Markley JL, Satterlee JD. Detailed NMR analysis of the heme-protein interactions in component IV Glycera dibranchiata monomeric hemoglobin-CO. *J Biomol NMR*. 1998; 11:119–133. [PubMed: 9679291]
- Alipanahi, B.; Gao, X.; Karakoc, E.; Balbach, F.; Donaldson, L.; Arrowsmith, C.; Li, M. Technical Report, No CS-2009-16. University of Waterloo; 2009a. IPASS: Error tolerant NMR backbone resonance assignment by linear programming. <http://www.cs.uwaterloo.ca/research/tr/2009>
- Alipanahi B, Gao X, Karakoc E, Donaldson L, Li M. PICKY: a novel SVD-based NMR spectra peak picking method. *Bioinformatics*. 2009b; 25:i268–275. [PubMed: 19477998]
- Altieri AS, Byrd RA. Automation of NMR structure determination of proteins. *Curr Opin Struct Biol*. 2004; 14:547–553. [PubMed: 15465314]
- Arkin MR, Wells JA. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat Rev Drug Discov*. 2004; 3:301–317. [PubMed: 15060526]
- Bahrami A, Assadi AH, Markley JL, Eghbalian HR. Probabilistic interaction network of evidence algorithm and its application to complete labeling of peak lists from protein NMR spectroscopy. *PLoS Comput Biol*. 2009; 5:e1000307. [PubMed: 19282963]

- Baran MC, Huang YJ, Moseley HN, Montelione GT. Automated analysis of protein NMR assignments and structures. *Chem Rev.* 2004; 104:3541–3556. [PubMed: 15303826]
- Barna JCJ, Laue ED, Mayger MR, Skilling J, Worrall SJP. Exponential sampling: an alternative method for sampling in two dimensional NMR experiments. *J Magn Reson.* 1987; 73:69.
- Bartels C, Billeter M, Güntert P, Wüthrich K. Automated Sequence-specific Assignment of Homologous Proteins using the program GARANT. *J Biomol NMR.* 1996:207–213. [PubMed: 22911044]
- Bartels C, Güntert P, Billeter M, Wüthrich K. GARANT - a General Algorithm for Resonance Assignment of Multidimensional Nuclear Magnetic Resonance Spectra. *J Comp Chem.* 1997; 18:139–149.
- Bartels C, Xia T, Billeter M, Güntert P, Wüthrich K. The program XEASY for computer-supported NMR spectral-analysis of biological macromolecules. *J Biomol NMR.* 1995; 6:1–10. [PubMed: 22911575]
- Berg T. Small-molecule inhibitors of protein-protein interactions. *Curr Opin Drug Discov Devel.* 2008; 11:666–674.
- Berjanskii MV, Wishart DS. A simple method to predict protein flexibility using secondary chemical shifts. *J Am Chem Soc.* 2005; 127:14970–14971. [PubMed: 16248604]
- Byerly DW, McElroy CA, Foster MP. Mapping the surface of Escherichia coli peptide deformylase by NMR with organic solvents. *Protein Sci.* 2002; 11:1850–1853. [PubMed: 12070337]
- Cavanagh, J. Protein NMR spectroscopy: principles and practice. 2nd. Academic Press; Amsterdam; Boston: 2007.
- Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR.* 1999; 13:289–302. [PubMed: 10212987]
- Dalvit C, Floersheim P, Zurini M, Widmer A. Use of organic solvents and small molecules for locating binding sites on proteins in solutions. *J Biomol NMR.* 1999; 14:23–32. [PubMed: 10419292]
- de la Cruz NB, Peterson FC, Lytle BL, Volkman BF. Solution structure of a membrane-anchored ubiquitin-fold (MUB) protein from Homo sapiens. *Protein Sci.* 2007; 16:1479–1484. [PubMed: 17567738]
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR.* 1995; 6:277–293. [PubMed: 8520220]
- Eghbalnia HR, Bahrami A, Tonelli M, Hallenga K, Markley JL. High-resolution iterative frequency identification for NMR as a general strategy for multidimensional data collection. *J Am Chem Soc.* 2005; 127:12528–12536. [PubMed: 16144400]
- Ewing TJA, Makino S, Skillman AG, Kuntz ID. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des.* 2001; 15:411–428. [PubMed: 11394736]
- Feeney J, Batchelor JG, Albrand JP, Roberts GCK. The effects of intermediate exchange processes on the estimation of equilibrium constants by NMR. *Journal of Magnetic Resonance* (1969). 1979; 33:519–529.
- Fielding L. NMR methods for the determination of protein-ligand dissociation constants. *Curr Top Med Chem.* 2003; 3:39–53. [PubMed: 12577990]
- Frederick RO, Bergeman L, Blommel PG, Bailey LJ, McCoy JG, Song J, Meske L, Bingman CA, Ritters M, Dillon NA, Kunert J, Yoon JW, Lim A, Cassidy M, Bunge J, Aceti DJ, Primm JG, Markley JL, Phillips GN Jr, Fox BG. Small-scale, semi-automated purification of eukaryotic proteins for structure determination. *J Struct Funct Genomics.* 2007; 8:153–166. [PubMed: 17985212]
- Frydman L, Scherf T, Lupulescu A. The acquisition of multidimensional NMR spectra within a single scan. *Proc Natl Acad Sci U S A.* 2002; 99:15858–15862. [PubMed: 12461169]
- Fuller JC, Burgoyne NJ, Jackson RM. Predicting druggable binding sites at the protein-protein interface. *Drug Discov Today.* 2009; 14:155–161. [PubMed: 19041415]
- Gal M, Mishkovsky M, Frydman L. Real-time monitoring of chemical transformations by ultrafast 2D NMR spectroscopy. *J Am Chem Soc.* 2006; 128:951–956. [PubMed: 16417386]

- Geyer M, Herrmann C, Wohlgemuth S, Wittinghofer A, Kalbitzer HR. Structure of the Ras-binding domain of RalGEF and implications for Ras binding and signalling. *Nat Struct Biol.* 1997; 4:694–699. [PubMed: 9302994]
- Goddard, TD.; Kneller, DG. Sparky 3. University of California; San Francisco: 2001.
- Gonnella N, Lin M, Shapiro MJ, Wareing JR, Zhang X. Isotope-Filtered Affinity NMR. *J Magn Reson.* 1998; 131:336–338. [PubMed: 9571109]
- Gounarides JS, Chen A, Shapiro MJ. Nuclear magnetic resonance chromatography: applications of pulse field gradient diffusion NMR to mixture analysis and ligand-receptor interactions. *J Chromatogr B Biomed Sci Appl.* 1999; 725:79–90. [PubMed: 10226879]
- Greenidge PA, Carlsson B, Bladh LG, Gillner M. Pharmacophores incorporating numerous excluded volumes defined by X-ray crystallographic structure in three-dimensional database searching: application to the thyroid hormone receptor. *J Med Chem.* 1998; 41:2503–2512. [PubMed: 9651155]
- Grishaev, A.; Llin-s, M.; Thomas, LJ. Protein Structure Elucidation from Minimal NMR Data: The CLOUDS Approach, *Methods Enzymol.* Academic Press; 2005a. p. 261–295.
- Grishaev A, Steren CA, Wu B, Pineda-Lucena A, Arrowsmith C, Llinas M. ABACUS, a direct method for protein NMR structure computation via assembly of fragments. *Proteins.* 2005b; 61:36–43. [PubMed: 16080153]
- Guntert P. Automated NMR protein structure calculation. *Prog Nucl Magn Reson Spectrosc.* 2003; 43:105–125.
- Guntert P. Automated NMR structure calculation with CYANA. *Methods Mol Biol.* 2004; 278:353–378. [PubMed: 15318003]
- Guntert P. Automated structure determination from NMR spectra. *Eur Biophys J.* 2009; 38:129–143. [PubMed: 18807026]
- Guntert P, Mumenthaler C, Wuthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol.* 1997; 273:283–298. [PubMed: 9367762]
- Hajduk PJ, Dinges J, Miknis GF, Merlock M, Middleton T, Kempf DJ, Egan DA, Walter KA, Robins TS, Shuker SB, Holzman TF, Fesik SW. NMR-based discovery of lead inhibitors that block DNA binding of the human papillomavirus E2 protein. *J Med Chem.* 1997a; 40:3144–3150. [PubMed: 9379433]
- Hajduk PJ, Gerfin T, Boehlen JM, Haberli M, Marek D, Fesik SW. High-throughput nuclear magnetic resonance-based screening. *J Med Chem.* 1999; 42:2315–2317. [PubMed: 10395471]
- Hajduk PJ, Greer J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat Rev Drug Discov.* 2007; 6:211–219. [PubMed: 17290284]
- Hajduk PJ, Meadows RP, Fesik SW. Discovering high-affinity ligands for proteins. *Science.* 1997b; 278:497–499. [PubMed: 9381145]
- Herrmann T, Guntert P, Wuthrich K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol.* 2002; 319:209–227. [PubMed: 12051947]
- Hoch, JC.; Stern, AS. NMR data processing. Wiley-Liss; New York: 1996.
- Huang YJ, Moseley HN, Baran MC, Arrowsmith C, Powers R, Tejero R, Szyperski T, Montelione GT. An integrated platform for automated analysis of protein NMR structures. *Methods Enzymol.* 2005; 394:111–141. [PubMed: 15808219]
- Huang YJ, Tejero R, Powers R, Montelione GT. A topology-constrained distance network algorithm for protein structure determination from NOESY data. *Proteins Struct Funct Bioinform.* 2006; 62:587–603.
- Hur O, Karplus K. Methods of translating NMR proton distances into their corresponding heavy atom distances for protein structure prediction with limited experimental data. *Protein Eng Des Sel.* 2005; 18:597–605. [PubMed: 16246822]
- Ikura M, Bax A. Isotope-filtered 2D NMR of a protein-peptide complex: Study of a skeletal muscle myosin light chain kinase fragment bound to calmodulin. *J Am Chem Soc.* 1992; 114:2433–2440.
- Irwin JJ, Shoichet BK. ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model.* 2005; 45:177–182. [PubMed: 15667143]

- Jee J, Guntert P. Influence of the completeness of chemical shift assignments on NMR structures obtained with automated NOE assignment. *J Struct Funct Genomics*. 2003; 4:179–189. [PubMed: 14649302]
- Jensen DR, Woytovich C, Li M, Duvnjak P, Cassidy MS, Frederick RO, Bergeman LF, Peterson FC, Volkman BF. Rapid, robotic, small-scale protein production for NMR screening and structure determination. *Protein Sci*. 2010; 19:570–578. [PubMed: 20073081]
- Jeon WB, Aceti DJ, Bingman CA, Vojtik FC, Olson AC, Ellefson JM, McCombs JE, Sreenath HK, Blommel PG, Seder KD, Burns BT, Geetha HV, Harms AC, Sabat G, Sussman MR, Fox BG, Phillips GN Jr. High-throughput purification and quality assurance of Arabidopsis thaliana proteins for eukaryotic structural genomics. *J Struct Funct Genomics*. 2005; 6:143–147. [PubMed: 16211511]
- Johnson B, Blevins R. NMRView: a computer program for the visualization and analysis of NMR data. *J Biomol NMR*. 1994; 4:603–614. [PubMed: 22911360]
- Jung YS, Zweckstetter M. Mars -- robust automatic backbone assignment of proteins. *J Biomol NMR*. 2004; 30:11–23. [PubMed: 15452431]
- Karplus M. Contact Electron-Spin Coupling of Nuclear Magnetic Moments. *The Journal of Chemical Physics*. 1959; 30:11–15.
- Keller R. Optimizing the Process of Nuclear Magnetic Resonance Spectrum Analysis and Computer Aided Resonance Assignment, A dissertation submitted to the Swiss Federal Institute of Technology Zurich (ETH Zürich) for the degree of Doctor of Natural Sciences. Diss ETH Nr 15947. 2004/2005
- Kim S, Szyperski T. GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information. *J Am Chem Soc*. 2003; 125:1385–1393. [PubMed: 12553842]
- Koradi R, Billeter M, Engeli M, Guntert P, Wuthrich K. Automated peak picking and peak integration in macromolecular NMR spectra using AUTOPSY. *J Magn Reson*. 1998; 135:288–297. [PubMed: 9878459]
- Koradi R, Billeter M, Wuthrich K. MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph*. 1996; 14:51–55. 29–32. [PubMed: 8744573]
- Kuszewski J, Schwieters CD, Garrett DS, Byrd RA, Tjandra N, Clore GM. Completely automated, highly error-tolerant macromolecular structure determination from multidimensional nuclear Overhauser enhancement spectra and chemical shift assignments. *J Am Chem Soc*. 2004; 126:6258–6273. [PubMed: 15149223]
- Kuszewski JJ, Thottungal RA, Clore GM, Schwieters CD. Automated error-tolerant macromolecular structure determination from multidimensional nuclear Overhauser enhancement spectra and chemical shift assignments: improved robustness and performance of the PASD algorithm. *J Biomol NMR*. 2008; 41:221–239. [PubMed: 18668206]
- Leach AR, Gillet VJ, Lewis RA, Taylor R. Three-dimensional pharmacophore methods in drug discovery. *J Med Chem*. 2010; 53:539–558. [PubMed: 19831387]
- Lee W, Westler WM, Bahrami A, Eghbalian HR, Markley JL. PINE-SPARKY: graphical interface for evaluating automated probabilistic peak assignments in protein NMR spectroscopy. *Bioinformatics*. 2009; 25:2085–2087. [PubMed: 19497931]
- Levitt, MH. Spin dynamics: basics of nuclear magnetic resonance. 2nd. John Wiley & Sons; Chichester, England; Hoboken, NJ: 2008.
- Liepinsh E, Otting G. Organic solvents identify specific ligand binding sites on protein surfaces. *Nat Biotechnol*. 1997; 15:264–268. [PubMed: 9062927]
- Lin Z, Xu YQ, Yang S, Yang DW. Sequence-specific assignment of aromatic resonances of uniformly ¹³C,¹⁵N-labeled proteins by using ¹³C- and ¹⁵N-edited NOESY spectra. *Angew Chem Int Ed*. 2006; 45:1960–1963.
- Linge JP, Habeck M, Rieping W, Nilges M. ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics*. 2003a; 19:315–316. [PubMed: 12538267]
- Linge JP, Williams MA, Spronk CA, Bonvin AM, Nilges M. Refinement of protein structures in explicit solvent. *Proteins*. 2003b; 50:496–506. [PubMed: 12557191]
- Lopez-Mendez B, Guntert P. Automated protein structure determination from NMR spectra. *J Am Chem Soc*. 2006; 128:13112–13122. [PubMed: 17017791]

- Lytle BL, Peterson FC, Qiu SH, Luo M, Zhao Q, Markley JL, Volkman BF. Solution structure of a ubiquitin-like domain from tubulin-binding cofactor B. *J Biol Chem*. 2004; 279:46787–46793. [PubMed: 15364906]
- Lytle BL, Peterson FC, Tyler EM, Newman CL, Vinarov DA, Markley JL, Volkman BF. Solution structure of *Arabidopsis thaliana* protein At5g39720.1, a member of the AIG2-like protein family. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 2006; 62:490–493.
- Malmodin D, Papavoine CH, Billeter M. Fully automated sequence-specific resonance assignments of hetero- nuclear protein spectra. *J Biomol NMR*. 2003; 27:69–79. [PubMed: 12878842]
- Mandal PK, Majumdar A. A comprehensive discussion of HSQC and HMQC pulse sequences. *Concepts in Magnetic Resonance Part A*. 2004; 20A:1–23.
- Markley JL, Aceti DJ, Bingman CA, Fox BG, Frederick RO, Makino S, Nichols KW, Phillips GN Jr, Primm JG, Sahu SC, Vojtik FC, Volkman BF, Wrobel RL, Zolnai Z. The Center for Eukaryotic Structural Genomics. *J Struct Funct Genomics*. 2009a; 10:165–179. [PubMed: 19130299]
- Markley, JL.; Bahrami, A.; Eghbalian, HR.; Peterson, FC.; Ulrich, EL.; Westler, WM.; Volkman, BF. Macromolecular structure determination by NMR spectroscopy. In: Gu, J.; Bourne, PE., editors. *Structural bioinformatics*. 2nd. Wiley-Blackwell; Hoboken, N.J.: 2009b. p. 93-142.
- Markley JL, Bax A, Arata Y, Hilbers CW, Kaptein R, Sykes BD, Wright PE, Wüthrich K. Recommendations for the Presentation of NMR Structures of Proteins and Nucleic Acids. *Pure Appl Chem*. 1998; 70:117–142.
- McGann MR, Almond HR, Nicholls A, Grant JA, Brown FK. Gaussian docking functions. *Biopolymers*. 2003; 68:76–90. [PubMed: 12579581]
- Medek A, Hajduk PJ, Mack J, Fesik SW. The Use of Differential Chemical Shifts for Determining the Binding Site Location and Orientation of Protein-Bound Ligands. *J Am Chem Soc*. 2000; 122:1241–1242.
- Meyer B, Peters T. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew Chem Int Ed Engl*. 2003; 42:864–890. [PubMed: 12596167]
- Mobli M, Maciejewski MW, Gryk MR, Hoch JC. An automated tool for maximum entropy reconstruction of biomolecular NMR spectra. *Nat Methods*. 2007; 4:467–468. [PubMed: 17538627]
- Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem*. 1998; 19:1639–1662.
- Mulder FA, Schipper D, Bott R, Boelens R. Altered flexibility in the substrate-binding site of related native and engineered high-alkaline *Bacillus subtilis*ins. *J Mol Biol*. 1999; 292:111–123. [PubMed: 10493861]
- Mumenthaler C, Braun W. Automated assignment of simulated and experimental NOESY spectra of proteins by feedback filtering and self-correcting distance geometry. *J Mol Biol*. 1995; 254:465–480. [PubMed: 7490763]
- Nilges M. Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *J Mol Biol*. 1995; 245:645–660. [PubMed: 7844833]
- Orekhov VY, Ibraghimov I, Billeter M. Optimizing resolution in multidimensional NMR by three-way decomposition. *J Biomol NMR*. 2003; 27:165–173. [PubMed: 12913413]
- Peng JW, Lepre CA, Fejzo J, Abdul-Manan N, Moore JM. Nuclear magnetic resonance-based approaches for lead generation in drug discovery. *Methods Enzymol*. 2001; 338:202–230. [PubMed: 11460549]
- Pervushin K, Riek R, Wider G, Wüthrich K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A*. 1997; 94:12366–12371. [PubMed: 9356455]
- Peterson FC, Baden EM, Owen BA, Volkman BF, Ramirez-Alvarado M. A single mutation promotes amyloidogenicity through a highly promiscuous dimer interface. *Structure*. 2010; 18:563–570. [PubMed: 20462490]

- Peterson FC, Deng Q, Zettl M, Prehoda KE, Lim WA, Way M, Volkman BF. Multiple WASP-interacting protein recognition motifs are required for a functional interaction with N-WASP. *J Biol Chem.* 2007; 282:8446–8453. [PubMed: 17229736]
- Peterson FC, Hayes PL, Waltner JK, Heisner AK, Jensen DR, Sander TL, Volkman BF. Structure of the SCAN domain from the tumor suppressor protein MZF1. *J Mol Biol.* 2006; 363:137–147. [PubMed: 16950398]
- Peterson FC, Lytle BL, Sampath S, Vinarov D, Tyler E, Shahan M, Markley JL, Volkman BF. Solution structure of thioredoxin h1 from *Arabidopsis thaliana*. *Protein Sci.* 2005; 14:2195–2200. [PubMed: 15987893]
- Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. *J Mol Biol.* 1996; 261:470–489. [PubMed: 8780787]
- Reibarkh M, Malia TJ, Wagner G. NMR distinction of single- and multiple-mode binding of small-molecule protein ligands. *J Am Chem Soc.* 2006; 128:2160–2161. [PubMed: 16478139]
- Rella M, Rushworth CA, Guy JL, Turner AJ, Langer T, Jackson RM. Structure-based pharmacophore design and virtual screening for novel angiotensin converting enzyme 2 inhibitors. *J Chem Inf Model.* 2006; 46:708–716. [PubMed: 16563001]
- Rieping W, Habeck M, Bardiaux B, Bernard A, Malliavin TE, Nilges M. ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics.* 2007; 23:381–382. [PubMed: 17121777]
- Ross A, Schlotterbeck G, Klaus W, Senn H. Automation of NMR measurements and data evaluation for systematically screening interactions of small molecules with target proteins. *J Biomol NMR.* 2000; 16:139–146. [PubMed: 10723993]
- Rovnyak D, Frueh DP, Sastry M, Sun ZY, Stern AS, Hoch JC, Wagner G. Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction. *J Magn Reson.* 2004a; 170:15–21. [PubMed: 15324754]
- Rovnyak D, Hoch JC, Stern AS, Wagner G. Resolution and sensitivity of high field nuclear magnetic resonance spectroscopy. *J Biomol NMR.* 2004b; 30:1–10. [PubMed: 15452430]
- Schanda P, Brutscher B. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *J Am Chem Soc.* 2005; 127:8014–8015. [PubMed: 15926816]
- Schanda P, Brutscher B. Hadamard frequency-encoded SOFAST-HMQC for ultrafast two-dimensional protein NMR. *J Magn Reson.* 2006; 178:334–339. [PubMed: 16269260]
- Schanda P, Kupce E, Brutscher B. SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J Biomol NMR.* 2005; 33:199–211. [PubMed: 16341750]
- Schubert M, Labudde D, Oschkinat H, Schmieder P. A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on ¹³C chemical shift statistics. *J Biomol NMR.* 2002; 24:149–154. [PubMed: 12495031]
- Schumann FH, Riepl H, Maurer T, Gronwald W, Neidig KP, Kalbitzer HR. Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions. *J Biomol NMR.* 2007; 39:275–289. [PubMed: 17955183]
- Schuttelkopf AW, van Aalten DM. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr.* 2004; 60:1355–1363. [PubMed: 15272157]
- Shen Y, Atreya HS, Liu G, Szyperski T. G-matrix Fourier transform NOESY-based protocol for high-quality protein structure determination. *J Am Chem Soc.* 2005; 127:9085–9099. [PubMed: 15969587]
- Shen Y, Delaglio F, Cornilescu G, Bax A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR.* 2009; 44:213–223. [PubMed: 19548092]
- Shuker SB, Hajduk PJ, Meadows RP, Fesik SW. Discovering high-affinity ligands for proteins: SAR by NMR. *Science.* 1996; 274:1531–1534. [PubMed: 8929414]

- Stanczyk SM, Bolton PH, Dell'Acqua M, Gerlt JA. Observation and sequence assignment of a cis prolyl peptide bond in unliganded staphylococcal nuclease. *J Am Chem Soc.* 1989; 111:8317–8318.
- Stuart AC, Borzilleri KA, Withka JM, Palmer AG. Compensating for variations in H-1-C-13 scalar coupling constants in isotope-filtered NMR experiments. *J Am Chem Soc.* 1999; 121:5346–5347.
- Tintori C, Corradi V, Magnani M, Manetti F, Botta M. Targets looking for drugs: a multistep computational protocol for the development of structure-based pharmacophores and their applications for hit discovery. *J Chem Inf Model.* 2008; 48:2166–2179. [PubMed: 18942779]
- Tuinstra RL, Peterson FC, Elgin ES, Pelzek AJ, Volkman BF. An engineered second disulfide bond restricts lymphotactin/XCL1 to a chemokine-like conformation with XCR1 agonist activity. *Biochemistry (Mosc).* 2007; 46:2564–2573.
- Tyler RC, Aceti DJ, Bingman CA, Cornilescu CC, Fox BG, Frederick RO, Jeon WB, Lee MS, Newman CS, Peterson FC, Phillips GN Jr, Shahan MN, Singh S, Song J, Sreenath HK, Tyler EM, Ulrich EL, Vinarov DA, Vojtik FC, Volkman BF, Wrobel RL, Zhao Q, Markley JL. Comparison of cell-based and cell-free protocols for producing target proteins from the *Arabidopsis thaliana* genome for structural studies. *Proteins.* 2005a; 59:633–643. [PubMed: 15789406]
- Tyler RC, Peterson FC, Volkman BF. Distal interactions within the par3-VE-cadherin complex. *Biochemistry (Mosc).* 2010; 49:951–957.
- Tyler RC, Sreenath HK, Singh S, Aceti DJ, Bingman CA, Markley JL, Fox BG. Auto-induction medium for the production of [U-15N]- and [U-13C, U-15N]-labeled proteins for NMR screening and structure determination. *Protein Expr Purif.* 2005b; 40:268–278. [PubMed: 15766868]
- Vajpai N, Strauss A, Fendrich G, Cowan-Jacob SW, Manley PW, Grzesiek S, Jahnke W. Solution conformations and dynamics of ABL kinase-inhibitor complexes determined by NMR substantiate the different binding modes of imatinib/nilotinib and dasatinib. *J Biol Chem.* 2008; 283:18292–18302. [PubMed: 18434310]
- Veldkamp CT, Peterson FC, Pelzek AJ, Volkman BF. The monomer-dimer equilibrium of stromal cell-derived factor-1 (CXCL 12) is altered by pH, phosphate, sulfate, and heparin. *Protein Sci.* 2005; 14:1071–1081. [PubMed: 15741341]
- Veldkamp CT, Seibert C, Peterson FC, De la Cruz NB, Haugner JC 3rd, Basnet H, Sakmar TP, Volkman BF. Structural basis of CXCR4 sulfotyrosine recognition by the chemokine SDF-1/CXCL12. *Sci Signal.* 2008; 1:ra4. [PubMed: 18799424]
- Veldkamp CT, Seibert C, Peterson FC, Sakmar TP, Volkman BF. Recognition of a CXCR4 sulfotyrosine by the chemokine stromal cell-derived factor-1alpha (SDF-1alpha/CXCL12). *J Mol Biol.* 2006; 359:1400–1409. [PubMed: 16725153]
- Veldkamp CT, Ziarek JJ, Peterson FC, Chen Y, Volkman BF. Targeting SDF-1/CXCL12 with a ligand that prevents activation of CXCR4 through structure-based drug design. *J Am Chem Soc.* 2010; 132:7242–7243. [PubMed: 20459090]
- Vinarov DA, Lytle BL, Peterson FC, Tyler EM, Volkman BF, Markley JL. Cell-free protein production and labeling protocol for NMR-based structural proteomics. *Nat Methods.* 2004; 1:149–153. [PubMed: 15782178]
- Volk J, Herrmann T, Wuthrich K. Automated sequence-specific protein NMR assignment using the memetic algorithm MATCH. *J Biomol NMR.* 2008; 41:127–138. [PubMed: 18512031]
- Waltner JK, Peterson FC, Lytle BL, Volkman BF. Structure of the B3 domain from *Arabidopsis thaliana* protein At1g16640. *Protein Sci.* 2005; 14:2478–2483. [PubMed: 16081658]
- Wang B, Wang Y, Wishart DS. A probabilistic approach for validating protein NMR chemical shift assignments. *J Biomol NMR.* 2010; 47:85–99. [PubMed: 20446018]
- Weininger D. SMILES, a chemical language and information system 1. Introduction to methodology and encoding rules. *J Chem Inf Comput Sci.* 1988; 28:31–36.
- Weininger D, Weininger A, Weininger JL. SMILES. 2. Algorithm for generation of unique SMILES notation. *J Chem Inf Comput Sci.* 1989; 29:97–101.
- Wells JA, McClendon CL. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature.* 2007; 450:1001–1009. [PubMed: 18075579]

- Williamson MP, Craven CJ. Automated protein structure calculation from NMR data. *J Biomol NMR*. 2009; 43:131–143. [PubMed: 19137264]
- Zhao Q, Frederick R, Seder K, Thao S, Sreenath H, Peterson F, Volkman BF, Markley JL, Fox BG. Production in two-liter beverage bottles of proteins for NMR structure determination labeled with either ^{15}N - or ^{13}C - ^{15}N . *J Struct Funct Genomics*. 2004; 5:87–93. [PubMed: 15263847]
- Zimmerman D, Kulikowski C, Wang L, Lyons B, Montelione GT. Automated sequencing of amino acid spin systems in proteins using multidimensional HCC(CO)NH-TOCSY spectroscopy and constraint propagation methods from artificial intelligence. *J Biomol NMR*. 1994; 4:241–256. [PubMed: 8019136]
- Zimmerman DE, Kulikowski CA, Huang Y, Feng W, Tashiro M, Shimotakahara S, Chien C, Powers R, Montelione GT. Automated analysis of protein NMR assignments using methods from artificial intelligence. *J Mol Biol*. 1997; 269:592–610. [PubMed: 9217263]
- Zsoldos Z, Reid D, Simon A, Sadjad BS, Johnson AP. eHITS: An innovative approach to the docking and scoring function problems. *Curr Protein Pept Sci*. 2006; 7:421–435. [PubMed: 17073694]

Figure 1. Flow chart describing the collection, assignment, and structure determination of a protein-ligand complex.

Figure 2.
Workflow for binding site characterization by chemical shift mapping.

Figure 3. An example of binding site identification by chemical shift mapping

A) Overlays of HSQC spectra monitoring the chemokine CXCL12 being titrated with 0 μM , 25 μM , 50 μM , and 800 μM (colored from black to gray) of 3-(naphthylene-2-carbonylthiocarbamoylamino)-benzoic acid (Ligand). B) CXCL12 adjusted chemical shift change (Eq. 5) for each residue as a function of 800 μM Ligand (black) or d-DMSO control (gray). The dotted line represents residues with the most significant chemical shift perturbations. C) Quantification of V49 affinity ($K_d = 64 \mu\text{M}$) by fitting of chemical shift perturbations to Eq. 6. D) Comparison of significant chemical shift perturbations (black) with predicted docking of the Ligand to CXCL12.

B

[illegible]

A) 2D representation of ethyl 3-[[naphthalen-2-ylformamido)methanethioyl]amino}benzoate with atom numbers indicated. Multiple protons bound to a single heavy atom have the potential for chemical shifts degeneracy (indicated by – gray box) and are given a pseudoatom identifier (colored gray). Rotatable bonds are indicated by arrows. B) CYANA library entry for the molecule described in panel A after conversion of the PDB file by the pdb2reslib module of Oliva and user editing. Additions to the original PDB file and user changes are highlighted by gray shading. Conversion of a PDB file by pdb2reslib adds the necessary psuedoatom definitions (indicated by 1 and gray shading), the dummy atoms required to connect the small molecule to the protein backbone (indicated by 2 and gray shading) and definitions for the dihedral angels (indicated by 3 and gray shading). However, pdb2reslib may not define all potential dihedral angles and the user may choose to add additional dihedral angle definitions (indicated by 4 and gray shading). In this example, the user has chosen to rename the atoms and renumber the molecule sequentially from left to right as shown in panel (A) and rename the angles identified by pdb2reslib (indicated by 3 and gray shading) to reflect the tree structure used in CYANA structure calculations. A detailed explanation for the file format can be found on the CYANAwiki (http://www.cyana.org/wiki/index.php/Residue_library_file).

Figure 5. Detection of intermolecular NOEs in a 3D F1- ^{13}C -filtered, F3- ^{13}C -edited NOESY spectrum

A) Strips extracted from a 3D ^{13}C -edited NOESY spectrum (left strip) contain both intra- and intermolecular NOEs while strips from the 3D F1- ^{13}C -filtered, F3- ^{13}C -edited NOESY spectrum (right strip) show only intermolecular NOEs. Strip pairs are shown for MZF1 residues L68, L62 and A116. Leucine 68 is outside the dimer interface and therefore does not have NOEs in the F1- ^{13}C -filtered, F3- ^{13}C -edited NOESY spectrum. Lower case letters correspond to NOEs shown in panel (B). B) Intermolecular NOEs involving Q59, L62, P112, E113 and A116. Lower case letters correspond to labeled peaks in (A). One monomer is colored in white and the other in gray. C) Strips extracted from 3D ^{13}C -edited NOESY (left strip) and 3D F1- ^{13}C -filtered, F3- ^{13}C -edited NOESY (right strip) spectra as in panel (A). Comparison of the strip pairs reveals intermolecular NOEs between a polysaccharide and domain-5 of cation independent-mannose-6 phosphate receptor (CI-MPR). Tyrosine 679 and N680 are in contact with the sugar while P704 lies just outside the binding pocket. D) Schematic showing the labeling scheme for the polysaccharide–domain-5 complex.

Table I
Modular scheme for NMR data collection for protein structure analysis

After a target protein or complex has been validated for NMR-based structure determination, the NOESY experiments are collected in a continuous block. The NOESY experiments are used to define distance constraints and therefore are the most sensitive to changes in sample stability. When solving a protein-ligand complex, the unlabeled ligand resonances and intermolecular contacts are detected using isotope-filtered from NOESY experiments collected in the next block. Finally, a series of heteronuclear correlation spectra for use in the backbone and sidechain chemical shift assignment stages are acquired.

NOESY	Ligand-complexes	Backbone	Sidechain
3D ¹⁵ N-NOESY-HSQC	2D F1/F2- ¹³ C/ ¹⁵ N-doubly-filtered NOESY	3D HNCO	3D C(CO)NH
3D ¹³ C-NOESY-HSQC	3D F1- ¹³ C/ ¹⁵ N-filtered, F3- ¹³ C-edited NOESY	3D HNCA	3D HCCH-TOCSY
2D ¹³ C-HSQC-aromatic		3D HN(CO)CA	2D ¹³ C-HSQC
3D ¹³ C-NOESY-HSQC-aromatic		3D HN(CA)CO	3D H(CCO)NH
		3D HNCACB	3D HBHA(CO)NH