

PROGRESS IN NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY

Progress in Nuclear Magnetic Resonance Spectroscopy 41 (2002) 187-231

www.elsevier.com/locate/pnmrs

NMR screening techniques in drug discovery and drug design

Brian J. Stockman^{a,*}, Claudio Dalvit^b

^aStructural, Analytical & Medicinal Chemistry, Pharmacia, 301 Henrietta St., Kalamazoo, MI 49001, USA ^bChemistry Department, Pharmacia, Viale Pasteur 10, Nerviano, Milano, Italy

Accepted 26 July 2002

Contents

1.	Intro	duction
2.	NMI	R screening techniques
	2.1.	Monitoring the macromolecule: SAR by NMR
		2.1.1. Drug discovery and design
		2.1.2. Practical considerations
	2.2.	Monitoring the small molecules
		2.2.1. Longitudinal relaxation
		2.2.2. Transverse relaxation
		2.2.3. Transverse relaxation with spin labels
		2.2.4. Double-quantum relaxation
		2.2.5. Diffusion experiments
		2.2.6. Intermolecular and intramolecular magnetization transfer
		2.2.7. Competition binding experiments
3.	Desi	gn of small molecule screening libraries
	3.1.	General considerations
	3.2.	SHAPES screening library: design and application
4.	Impa	cting chemistry and biology
	4.1.	Privileged molecules
	4.2.	NMR screening as a secondary assay 225
	4.3.	Selecting for minimal HSA binding
	4.4.	Genomics-derived proteins of unknown function

Abbreviations: CMC, comprehensive medicinal chemistry; CPMG, Carr–Purcell–Meiboom–Gill; DECODES, diffusion encoded spectroscopy; FABP, fatty acid binding protein; FKBP, FK506 binding protein; HSA, human serum albumin; HTS, high throughput screening; ICAM, intracellular adhesion molecule; IMPDH, inosine-5'-monophosphate dehydrogenase; ITC, isothermal titrating calorimetry; LED, longitudinal eddy-current delay; LFA, leukocyte function-associated antigen; NMR-DOC, nuclear magnetic resonance docking of compounds; NMR-SOLVE, nuclear magnetic resonance structurally oriented library valency engineering; PFG-SE, pulsed field gradient spin echo; PFG-STE, pulsed field gradient stimulated echo; RECAP, retrosynthetic combinatorial analysis procedure; SAR, structure–activity relationship; SAR by NMR, structure–activity relationship by nuclear magnetic resonance; SLAPSTIC, spin labels attached to protein side chains as a tool to identify interacting compounds; STD, saturation transfer difference; TEMPO, 2,2,6,6-tetramethyl-1-piperidinyloxy; WaterLOGSY, water–ligand observed via gradient spectroscopy.

Corresponding author. Tel.: +1-616-833-1882; fax: +1-616-833-1822.

E-mail address: brian.j.stockman@pharmacia.com (B.J. Stockman).

5. Prospects for automation	227
6. Future directions	227
References	228

Keywords: Protein-ligand interaction; Nuclear magnetic resonance screening; Macromolecules; Screening libraries

1. Introduction

Pharmaceutical and academic nuclear magnetic resonance (NMR) groups have implemented NMR screening techniques as a powerful approach to identify and to investigate protein/ligand interactions. Pharmaceutical groups in particular have incorporated NMR screening strategies into their drug discovery and drug design programs. This stems from the fact that NMR screening is naturally synergistic with combinatorial or medicinal chemistry, high throughput screening (HTS), structure-based drug design, and genomics [1]. This review will describe progress in the field since the SAR by NMR (structure-activity relationship by nuclear magnetic resonance) approach using 2D ${}^{1}H{-}^{15}N$ HSQC spectra of ¹⁵N-labeled proteins was described by Fesik's group in 1996 [2]. Others have also written at length on NMR screening techniques [3-11]. The focus of this comprehensive review will be to first provide a physical and mathematical basis of the various NMR screening techniques and then to describe examples from the literature where the techniques have been applied to biological systems. Emphasis will be placed on applications in drug discovery and drug design. A discussion on NMR screening library design is also included, with particular emphasis on the elegant SHAPES library and its applications [12]. The review will conclude with sections on NMR screening's impact on chemistry and biology, prospects for automation and future directions.

2. NMR screening techniques

NMR screening can be defined as the identification of small molecule ligands for macromolecular targets by observation of a change in an NMR parameter that occurs upon their interaction. NMR screening methods can be divided into those that detect interactions by observation of either macromolecule NMR parameters or small molecule NMR parameters. In the case of macromolecules, the parameter that can be monitored is limited typically to chemical shifts. For small molecules, the choice of NMR parameters is more diverse. These include longitudinal, transverse, and double-quantum (DQ) relaxation; diffusion coefficients; and intermolecular and intramolecular magnetization transfer. The latter includes transferred NOE, NOE pumping and reverse NOE pumping, saturation transfer, and WaterLOGSY (water-ligand observed via gradient spectroscopy) experiments. Screening methods that monitor the macromolecular target are exemplified by the SAR by NMR technique [2] and will be described first. Screening methods based on small molecule resonances will be described second.

2.1. Monitoring the macromolecule: SAR by NMR

The SAR by NMR technique [2] is based on the use of chemical shift changes to screen for low-affinity ligands, in combination with using structural information to direct a linked-fragment approach for achieving binding affinity enhancement. While the method has so far been demonstrated only for protein targets, it is in principle applicable to nucleic acid targets as well. A schematic of the procedure is outlined in Fig. 1 [2]. In the first step, a library of small organic molecules is screened to identify molecules that bind to the protein (Fig. 1, Step 1). Binding is detected by comparing 2D ¹H-¹⁵N HSQC spectra of the ¹⁵N-enriched target protein in the absence and presence of ligand to elucidate ligand-induced changes in chemical shift. Typically, perturbations are considered significant if $\Delta \delta > 0.1$ ppm for at



least two peaks in the spectrum [13], where

$$\Delta \delta = \sqrt{[(\delta({}^{1}\text{H}, \text{ppm})_{\text{free}} - \delta({}^{1}\text{H}, \text{ppm})_{\text{obs}})^{2} + 0.04(\delta({}^{15}\text{N}, \text{ppm})_{\text{free}} - \delta({}^{15}\text{N}, \text{ppm})_{\text{obs}})^{2}]}.$$
(1)

Binding constants for identified ligands can be determined by monitoring chemical shift changes as a function of ligand concentration [14]. Data are then fitted using a least-squares grid search varying the values of $K_{\rm D}$ and the chemical shift of ligand-saturated protein according to

$$K_{\rm D} = \{([{\rm P}]_0 - x)([{\rm L}]_0 - x)\}/x$$
⁽²⁾

where $[P]_0$ and $[L]_0$ are the total molar concentrations of target and ligand, respectively, and x is the molar concentration of the target/ligand



Fig. 1. Outline of the SAR by NMR method. Reprinted with permission from Shuker et al. [2]. © 1996 American Association for the Advancement of Science.

complex given by

$$x = (\delta_{\rm obs} - \delta_{\rm free})/\Delta \tag{3}$$

189

where δ_{obs} and δ_{free} are the chemical shifts for the target molecule at each ligand concentration and in the absence of ligand, respectively, and Δ is the chemical shift difference at saturating levels of ligand [15].

Once an initial lead compound is identified, binding constants or activity measurements for close analogues are determined in order to optimize affinity for this site (Fig. 1, Step 2). Next, a second low-affinity ligand is identified based on amide ¹H/¹⁵N chemical shift changes for a different set of residues in either the initial screen or in a second screen carried out in the presence of the optimized ligand from the first screen (Fig. 1, Step 3). Assaying related analogues as was done for the first ligand then optimizes the second ligand (Fig. 1, Step 4). Following identification of the two lead fragments, their location and orientation in the ternary complex is determined experimentally by either isotope-edited or transfer NOE methods. On the basis of this information, the two lead compounds are synthetically linked together in a manner that maintains the spatial orientation of the two ligands with respect to each other and to the target protein to produce the final high-affinity ligand (Fig. 1, Step 5) [2]. The use of SAR by NMR to identify highaffinity ligands for a wide range of protein targets is best described by examples. The advantages afforded by the method will be readily apparent: lack of background signals from test compounds because of ¹⁵N spectral editing, applicability to any class of compound (providing that the aqueous solubility is greater than about 1 mM), concomitant identification of ligand binding site location, rapid SAR development, and simple binding assays with no need to develop functional assays or to know the target's function. The requirements for ¹⁵N-labeled target protein, backbone ¹H-¹⁵N resonance assignments (which are

typically limited to proteins with MW < 30 kDa), and knowledge of the target protein's structure are the main disadvantages of this technique. Also, the 2D $^{1}H^{-15}N$ HSQC experiments used by this method typically require a longer acquisition time than do the experiments used in the various 1D NMR screening methods.



2.1.1. Drug discovery and design

FKBP. The SAR by NMR technique was demonstrated with the discovery of a compound that binds to FK506 binding protein (FKBP) with a $K_{\rm D}$ of 19 nM by linking two molecules with binding affinities of 2 and 100 µM [2]. FKBP blocks T-cell activation when complexed to the immunosuppressant FK506 [16]. To identify the first ligand, [¹⁵N]FKBP was screened against a \sim 1000-compound library. A number of ligands were identified in the high µM to low mM affinity range. A pipecolinic acid derivative (1) with a $K_{\rm D}$ of 2 µM was selected as the first ligand. Ligandinduced chemical shift changes indicated that (1) binds FKBP at the same location as the pipecolinic acid moiety of FK506. FKBP with saturating levels of (1) was then screened against the compound library. A benzanilide derivative (2) was identified with a K_D of 0.8 mM. Ligandinduced chemical shift changes indicated that the binding sites of (1) and (2) are located close to one another. Limited SAR was then developed for (2) by testing a number of related proprietary and commercial compounds for FKBP binding affinity. The compound (3) ultimately chosen as the second-site ligand has a $K_{\rm D}$ value of 100 μ M. In order to design a linkage between the two ligands, a model of the ternary complex of FKBP with (1) and (3) was derived on the basis of isotope-filtered NMR experiments. Five linked compounds were ultimately synthesized, and all were found to have nM affinity for FKBP. The best compound (4) has an affinity of 19 nM. Intermolecular NOEs suggested that the linked compounds bind to FKBP in the same location as the unlinked pieces, with only slight variations in position caused by





the biaryl screening hits (6) was determined using 3D NMR methods. NOE data indicated that the biaryl compounds do in fact bind in the desired S'_1 subsite and also defined the relative orientation of the compounds with respect to (5). On the basis of this structural data and the available biphenyl SAR, a series of linked compounds exemplified by (7) were synthesized. The linker length was varied from one to four carbon atoms, with two carbon atoms found to be optimal. This compound has an IC_{50} value of 15 nM. The linked compounds bind in the expected manner based on the solution structure determined for one of the binary complexes. The importance of linker length was evaluated by determining the enthalpic and entropic contributions to the binding energy as a function of the number of linking carbon atoms [18]. It was found that the 2-carbon linker results in significant gains in enthalpy compared to other linker lengths, but that there is very little entropic variation with linker length. Given the extended nature of the stromelysin active site, it is not unimaginable that another round of screening could be carried out to identify a third ligand fragment that would occupy the S_1 or S'_2 subsite. Provided that the final singlemolecule compound retains drug-like properties, repeated cycles of NMR screening could be used to maximize affinity and/or specificity for this enzyme.



E2. Human papillomaviruses (HPVs) cause anogenital warts, and certain high-risk strains are implicated in cervical cancers [19]. The E2 protein, which is required for replication, contains a DNAbinding domain (DBD) that is functionally competent as an isolated domain. Small molecules that would disrupt the ability of this DBD to specifically bind DNA might be useful therapeutic agents for HPV infection. NMR screening of [¹⁵N]E2-DBD was carried out using a library of ~2000 small organic molecules [20]. Since it was desired that the ligands disrupt the ability to bind DNA, ligand-induced chemical shift changes were used to locate the ligand binding sites on the E2-DBD surface. By these criteria, two classes of ligands exemplified by (8) and (9) were found to bind in a similar location on the DNA recognition helix, while a third class was found to bind the E2-DBD β -barrel. Compounds (8) and (9) were also shown to disrupt E2-DBD binding to DNA in filter-binding assays. Conclusive proof of the ligand binding site was obtained from isotope-edited NMR experiments on E2-DBD complexed to (8). Series of derivatives for both compounds were then synthesized and tested for affinity in order to develop a limited SAR. Since the two initial leads bind at the same site, the SAR was combined into a single molecule (10). This compound was found to have an IC₅₀ value of 10 µM in filter-binding assays. This level of activity was not found in a traditional high throughput assay of more than 100,000 compounds. For this target protein, the ability to identify weak affinity ligands for focused chemistry efforts was a distinct advantage afforded by the NMR methodology.

191



LCK SH2 domain. Src homology 2 (SH2) domains, which interact with peptides and proteins that contain phosphotyrosine (11) to mediate signal transduction pathways, are a potential target to disrupt a number of biological processes [21]. NMR screening was used in an attempt to identify novel phosphotyrosine mimetics as lead chemical templates for medicinal chemistry [22]. First, the binding site for phosphotyrosine was identified based on ligand-induced chemical shift changes in the Lck [¹⁵N]SH2 domain. This step insured that subsequently identified ligands could be validated as binding in the desired site by virtue of inducing chemical shift changes in these same residues. A library of about 3500 small, diverse compounds was then screened for binding. At least 15 compounds were identified as ligands that bind in the phosphotyrosine binding site. Although most ligands identified exist multiply charged at physiological pH, compounds (12), (13) and (14) contain only a single negative charge and bind with only slightly less affinity than phosphotyrosine. These phthalamate analogs represent novel phosphotyrosine mimetics that might have tolerable pharmacokinetic properties

and good affinity.



Erm methyltransferase. NMR screening was used to identify novel lead chemical templates for Erm methyltransferases [23]. This class of enzymes confers resistance to macrolide-lincosamide-streptogramin antibiotics by methylating 23S ribosomal RNA in the macrolide binding site [24]. Inhibitors of Erm methyltransferases can sensitize resistant bacteria to macrolide antibiotics and may thus provide part of a combination therapy for these resistant-strain infections. [15N]ErmAM methyltransferase was screened against a library of small organic molecules. Several classes of compounds, including a triazine compound (15), with affinities in the mM range were identified. Ligand-induced chemical shift changes indicated that the identified compounds bind ErmAM at the same location as the natural inhibitor Sadenosyl-L-homocysteine (SAH) (16). Limited SAR was developed around (15) by determining affinities for related proprietary compounds, ultimately leading to the synthesis of (17). This compound has a $K_{\rm D}$ value of less than 100 µM. The triazine series was further optimized using parallel syntheses to prepare more than 600 derivatives of (17) with various substitutions at the three exocyclic positions. A number of these compounds, including (18) and (19), have $K_{\rm I}$ values for ErmAM of less than 10 µM. To provide a starting point for structurebased drug design, the structures of ErmAM complexed to (18) and ErmC' complexed to (19) were determined by NMR spectroscopy and X-ray crystallography, respectively. The triazine (18) and aminopyrimidine (19) groups were found to overlay the adenine moiety of SAH. The specific protein-ligand interactions observed in both complexes are very similar. For these novel lead chemical templates, unoccupied portions of the active site compared to that containing SAH provide opportunities for ligand extensions in order to obtain improved affinity and/or better specificity. Importantly, the NMR-derived lead chemical templates are non-nucleosides and may thus

avoid some of the selectivity problems encountered with SAH-based derivatives.



Urokinase. Urokinase is implicated in a variety of malignancies by virtue if its capacity to activate plasminogen in the cascade mechanism that results in basement membrane degradation and tumor metastasis [25]. Small molecule inhibitors of urokinase are thus of interest for cancer therapy. NMR screening was applied to this enzyme in an attempt to find suitable lead chemical templates that possess improved pharmacokinetic properties compared to available inhibitors [13]. The known inhibitors all contain either an amidine or guanidine group that have pK_a values greater than 9.0. At physiological pH, the positive charge of these inhibitors leads to poor bioavailability. The [15N]urokinase used in this study was prepared from mammalian cells using selected ¹⁵N-labeled amino acids as precursors. This resulted in approximately 70 $^{1}\text{H}^{-15}\text{N}$ cross peaks that could be monitored during the NMR screening process. A library of \sim 3000 compounds was then screened for binding. The binding assays were carried out under saturating conditions of phenylguanidine (20), a known inhibitor, to prevent autolytic degradation during the experiments. Based on the observed pattern of chemical shift changes, 2-aminobenzimidazole (21) was identified as a ligand that binds in the same location as (20). Compound (21) has a pK_a of 7.5 and was found to have an IC_{50} value of 200 μ M. Medicinal chemistry elaborations of this new template yielded (22) which retains the improved pK_a value and also lowers the IC₅₀ value to 10 μ M. The pK_a values of \sim 7.5 for this compound class indicate that they will be uncharged at physiological pH, and thus represent a lead chemical template with a better



chance of downstream pharmacokinetic success. As an entry point for structure-based drug design, the Xray crystal structure of urokinase complexed to (**22**) was subsequently determined.

$$(20) (21) (22) (21) (22)$$

Adenosine kinase. Inhibitors of adenosine kinase, an enzyme that is involved in the metabolism of adenosine, may have therapeutic use as anticonvulsants and antinociceptives [26]. Although some inhibitors of adenosine kinase are available, they have properties that make them poor drugs. For instance, (23) is very potent with an in vitro IC₅₀ value of 1.7 nM, but has poor solubility and undesirable side effects. Rather than start from scratch, an NMR screening approach was taken where replacements for one ligand fragment of (23) were identified and then linked back to the rest of the inhibitor [27]. The goal was to identify a new lead chemical template that retains the potency but loses the liabilities of the original compound. Compound (23) can be fragmented into (24) and (25). NMR screening in the presence of one fragment should identify compounds that bind in the unoccupied portion of the active site. To find replacements for (25), adenosine kinase was screened against ~ 2000 compounds in the presence of saturating amounts of (26). Compound (26), which is a fragment from another high-affinity adenosine kinase inhibitor, was used instead of (24) in order to make sure that there would be no steric interference to prevent ligand binding at the unoccupied site. Prior to carrying out the screening experiments, residues of the protein in this unoccupied portion of the active site were identified based on differential chemical shifts observed for the binding of compound pairs such as (23) and (24). A number of compounds were identified as ligands in the NMR screen that bound in the unoccupied portion of the active site, including (27). Competition binding experiments and simple modeling studies were used to suggest ways to link (24) and (27) into a single-molecule inhibitor. Compound (28) was one of the linked compounds created. It has an in vitro IC₅₀ value of 10 nM, promising in vivo activity, and may have improved pharmacokinetic properties compared to (23). The approach described here provides a very rapid method to explore chemistry space in part of an active site without having to synthesize a large number of compounds. In fact, only two compounds were synthesized in this study. By first identifying ligand fragments from a library of desirable molecules, medicinal chemistry resources are focused on compounds with the greatest potential.



LFA-1/ICAM-1. Leukocyte function-associated antigen-1 (LFA-1) is a cell surface adhesion receptor involved in inflammatory and specific T-cell immune responses [28]. These processes are mediated by interaction with intracellular adhesion molecules (ICAMs) located on endothelial and antigen-presenting cells [29]. In particular, it is known that the LFA-1 I domain, consisting of \sim 200 amino acids, is required for binding to ICAMs [28]. Preventing the LFA-1/ ICAM interaction may have therapeutic value for treating inflammatory diseases [30]. Using HTS, parylthio cinnamide compounds were identified as antagonists of the LFA-1/ICAM-1 interaction [31]. For example, (29) has an IC₅₀ value of 44 nM, but it has poor solubility and oral bioavailability. Fragmentbased NMR screening was applied to improve the pharmaceutical properties of these compounds [32]. That the *p*-arylthic cinnamide compounds do in fact bind to the I domain was indicated by chemical shift changes induced in [15N]I domain in the presence of these compounds. Protein-ligand NOEs were then used to ascertain that (29) binds in a hydrophobic

pocket of the I domain allosteric site in a similar manner as does Lovastatin [33] and structurally related molecules. The isopropyl phenyl group was found to bind in a relatively hydrophilic portion of the binding site, suggesting that modifications to this part of the antagonist might improve potency and other properties. To identify possible replacements for the isopropyl phenyl group, [¹⁵N]I domain was screened against ~ 2500 compounds with molecular weight less than 150 Da under saturating levels of a truncated inhibitor (30). Several classes of ligands with $K_{\rm D}$ values in the mM range were identified. Structural studies on ternary complexes containing I domain, (30) and one of the identified ligands (31) indicated that each binds in the desired portion of the binding site. A linking strategy based on prior chemical SAR was used to create a number of compounds. Several of these are quite potent, including (32) which has an IC₅₀ value of 40 nM. Importantly, (32) has increased solubility and oral bioavailability compared to (29). Other identified ligands provide further opportunities for chemical diversity exploration.



Other targets. Although not yet described in detail in the literature, ¹⁵N-based NMR screening has been carried out at least to the point of identifying ligands for several other protein targets [15,34,35]. These include the Ras-binding domain of the RAF protein [15], the receptor binding domain of the vascular endothelial growth factor [36], the DBD of NFATc [37], the peptidyl prolyl *cis*-*trans* isomerase Pin-1 [38], the antiapoptotic proteins Bcl-xL [39] and Bcl-2 [40], the phosphotyrosyl phosphatase PTP-1B [41], and cyclophilin A [35].

2.1.2. Practical considerations

Details of the SAR by NMR process have been

described in various published applications [2, 13-15,20,22,23,27,32]. Some key points are summarized here. Stock solutions of each compound are made in 1 M DMSO. If necessary, the pH of the 1 M stock solution is adjusted with acetic acid or ethanolamine so that no pH change is observed upon a 1/10 dilution into phosphate buffer. This is important since small changes in pH can lead to chemical shift changes that could complicate analysis of the spectra. Reference spectra of the target can also be collected at slightly different pH values in order to distinguish pH-induced chemical shift changes from ligand-induced chemical shift changes [9]. Stock solutions are then prepared in DMSO that contain 8-10 compounds at a concentration of 100 mM each. For the screening, NMR samples are prepared by adding $4 \mu L$ of this DMSO stock solution to 0.4 mL buffered protein solution. Typically, the final concentration of protein is 0.3 mM and the final concentration of each organic compound is 1 mM. Using a sample changer, 100–120 ¹H–¹⁵N HSOC spectra can be collected in a 24 h period, thus allowing about 1000 compounds to be screened per day. In cases where addition of the mixture of 8-10 compounds results in differences in the ¹H-¹⁵N HSQC spectrum, each compound is screened separately to identify which specific compound or compounds in the mixture binds to the target protein. If desired, the $K_{\rm D}$ for identified ligands, including the second ligand in the presence of the first, can be determined in the fast exchange regime according to Eq. (2) by titrating ligand into the protein solution and measuring the changes in chemical shift as a function of protein/ligand ratio.

In order to be a truly high-throughput screening technique, it should be possible to screen more than 100,000 compounds in about a week. This could be achieved by screening 50 μ M protein solutions against mixtures of 100 compounds at 50 μ M each in about 10 min using cryoprobe technology [42]. The concentrations of protein and small molecules must be reduced to these levels so that the total combined concentration of all 100 compounds will still be at a practical level of 5 mM. The sensitivity advantage of the cryoprobe compared to a normal probe makes these concentrations detectable without lengthening the total experiment time. The 10-fold gain in throughput increases the total number of compounds that can be screened to about 10,000 per day. The





Fig. 2. Comparison of the (a) ${}^{1}\text{H}/{}^{13}\text{C}$ -HSQC and (b) ${}^{1}\text{H}/{}^{15}\text{N}$ -TROSY spectra collected at 800 MHz on a 50 μ M sample of ${}^{13}\text{C}(\text{methyl})/\text{U}-{}^{15}\text{N},{}^{2}\text{H}$ -labeled maltose binding protein. The total experiment time was 10 min for each spectrum. Reprinted with permission from Hajduk et al. [43]. © 2000 American Chemical Society.

reduction in protein and compound concentrations used has two other important ramifications. First, the relatively high solubility requirements originally imposed on the small molecules can be relaxed. The compounds no longer possess aqueous solubilities greater than 1 mM. Lowering the solubility requirement down to around 100 µM increases the diversity of compounds suitable for screening. Second, the affinity cutoff for a compound to be detected as a binding ligand is reduced to $\sim 150 \,\mu$ M. This compares to $\sim 1 \text{ mM}$ at the higher concentrations of protein and ligand typically used. Greater stringency is important since otherwise the hit rate would be too high for screening such a large compound library. A low hit rate is important in order to minimize time spent deconvoluting the 100-compound mixtures to identify the ligands.

The potential for ¹³C-based NMR screening has also been demonstrated [43]. This approach extends the applicability of target-based NMR screening to larger systems because of the favorable relaxation properties of ¹³C in methyl groups. 2D ¹H-¹³C spectra collected on proteins are analyzed analogous to the 2D $^{1}\text{H}-^{15}\text{N}$ spectra of traditional SAR by NMR. Rather than having all possible C-H groups ¹³Clabeled, only a subset of methyl groups are labeled. An isotopic enrichment method that was developed previously to produce valine, leucine and isoleucine (δ_1) methyl-protonated ¹⁵N-, ¹³C-, ²H-labeled proteins [44] was adapted to result in ¹³C-labeling of only valine, leucine and isoleucine (δ_1) methyl groups. This labeling pattern reduces spectral complexity, and, since there are three protons on each ¹³C and

there is no one-bond ${}^{13}C-{}^{13}C$ coupling interactions, the sensitivity is potentially threefold higher than the ¹H-¹⁵N approach. Based on a study of 191 nondegenerate X-ray crystallographic structures of protein/ligand complexes, this labeling pattern also provides sufficient coverage of the protein surface so that ligand binding can be detected. In that retrospective study, 92% of the ligands had a heavy atom within 6 Å of a methyl group that would be ¹³Clabeled. This compares favorably with only 82% of the ligands having a heavy atom within 6 Å of a backbone ¹⁵N atom. The enrichment protocol developed for preparing the ¹³C-labeled proteins resulted in a cost that is comparable to that of ¹⁵N-labeled protein. The bacterial expression protocol uses $[3^{-13}C]$ - α -ketobutyrate and $[3,3'^{-13}C]$ - α -ketoisovalerate as ¹³C sources. Synthetic routes to produce these two precursors in a cost effective manner were developed. For proteins with molecular weight less than 30 kDa, the ¹³C-based technique was found to be three times more sensitive than the ¹⁵N-based method. 2D ¹H-¹³C HSQC spectra could be collected in 10 min on a 50 μ M protein sample. ¹H-¹⁵N HSQC spectra suitable for screening cannot be collected in this short amount of time unless a cryoprobe is used. For larger proteins with molecular weight greater than 40 kDa, the ¹³C-based technique is still more sensitive than the ¹⁵N-based method, but the overall sensitivity is not sufficient for screening. However, required sensitivity could be recovered by also perdeuterating the ¹³C-labeled protein. In this situation ${}^{1}H{}^{-13}C$ HSQC spectra suitable for screening were obtained on a 300 µM solution of a 100 kDa protein in 30 min

using a cryoprobe on a 500 MHz spectrometer. In this case, the sensitivity of the ¹³C-based method was sevenfold greater than the ¹⁵N-based method. Similar comparisons to ¹H–¹⁵N TROSY experiments indicated that the ¹³C-based method is substantially more sensitive. This is clearly demonstrated in Fig. 2 [43] which compares ¹H–¹³C HSQC and ¹H–¹⁵N TROSY spectra acquired in 10 min at 800 MHz on a 50 μ M sample of [¹³C-methyl, ¹⁵N, ²H]maltose binding protein.

Larger proteins can also be made amenable to ¹H-¹⁵N HSQC-based screening through the introduction of selective labels. Using the ${}^{13}C_i - {}^{15}N_{i+1}$ labeling strategy (where *i* corresponds to the residue number in the linear amino acid sequence) of Kainosho and Tsuji [45], the ¹H-¹⁵N correlation for V115 in the active site of fatty acid binding protein-4 (FABP-4) was identified [46]. The ¹H-¹⁵N HSQC spectrum for this 16.4 kDa protein contains only valine resonances since only [¹³C, ¹⁵N]valine was used in the growth media, drastically simplifying the spectrum. The occurrence of a unique valine-valine pair at V114-V115 led to the easy assignment of V115 and thus a single active site resonance to follow during screening. Binding of one component from a mixture was demonstrated by observed perturbations of V115 in the ¹H-¹⁵N HSQC spectrum. Perturbation of the V115 resonance is indicative of not only ligand binding, but ligand binding in the active site. With the labeling strategy used, the V115 resonance could be monitored with either ¹H-¹⁵N HSQC or ¹H-¹⁵N-¹³C HNCO spectra.

The SAR by NMR technique is a powerful combination of NMR techniques to first screen for weak binding ligands and then to structurally direct the linkage of two low-affinity ligands to create a high-affinity ligand. In contrast to combinatorial chemistry, which is also a building block approach in the design of molecules, the SAR by NMR technique requires much less synthetic chemistry resources [2,47]. In the SAR by NMR paradigm, medicinal chemistry efforts are concentrated on those molecules demonstrated to bind to the target receptor, and are directed by structural information regarding the location and orientation of the small molecule fragments [2,47]. Portions of the SAR by NMR process may also prove useful in situations where a lead molecule

has been discovered by traditional screening methods or some method other than NMR spectroscopy. Traditional NMR methods can then be used to locate its binding site on the target molecule and the SAR by NMR process could then be used to find low-affinity molecular fragments to link to the original lead compound to increase binding affinity. This may be particularly useful if optimization of the original lead compound proves difficult by traditional medicinal chemistry.

Anecdotal and conference reports suggest that ¹H-¹⁵N HSQC-based screening to identify initial hits has found some application in the pharmaceutical industry [48,49]. Academic groups have also employed ¹H-¹⁵N HSQC-based screening to investigate complex biological systems. Targeting of the Vam7 PX domain to yeast vacuoles, which is one component of protein sorting and membrane trafficking, has been studied by screening lipids for binding to the Vam7 PX domain [50]. To define the structural basis for specific membrane binding by Vam7, a set of soluble lipids was screened for interactions with the PX domain. Only dibutanoyl phosphatidylinositol 3-phosphate was found to induce chemical shift changes in the PX domain, suggesting that only this specific interaction has biological relevance. Chemical shift perturbations in the PX domain bound to dibutanoyl phosphatidylinositol 3-phosphate induced by addition of dodecylphosphocholine identified residues thought to interact with the interface and interior of the lipid bilayer. Signal transduction regulation by small molecules has also been studied by screening a library of 850 compounds for binding to PAS-A, a domain of PAS kinase [51]. Ligands with $K_{\rm D}$ values ranging from low μM to low mM were identified.

Ligand-induced chemical shift changes provide a quick way to locate the ligand's binding site provided that resonance assignments have been made for the protein. More detailed information regarding ligand orientation and specific protein– ligand interactions is then derived from protein– ligand NOEs. In some cases, such as for weak affinity or poorly soluble ligands, or for resonance degeneracies, protein–ligand NOEs cannot be obtained. Ligand orientation information, however, can be obtained through analysis of differential chemical shifts [52] or simulations of chemical shift perturbations [53].

The differential chemical shifts method relies on comparison of induced chemical shift changes for a series of closely related analogs of the ligand in question [52]. The largest differences in ligandinduced chemical shift changes will occur for the protein residues that interact with the portion of the ligand that is chemically varied. Conversely, similar ligand-induced chemical shift changes will occur for protein residues that interact with the portion of each ligand held constant. This method was demonstrated using a series of ascomycin analogs that bind to FKBP and a series of alanine mutant Bak peptides that bind to Bcl-xL [52]. This technique not only identifies which portions of a ligand interact with which protein residues, but also which portions of a ligand do not interact with protein and instead point away from the binding site. In the latter case, lack of differential chemical shifts indicates that the chemically varied portion of the ligand does not contact the protein. This type of information is useful since it defines portions of the ligand that can be modified to improve solubility or pharmacokinetic properties without losing affinity.

Weakly interacting ligands can also be aligned to protein surfaces using simulations of chemical shift perturbations [53]. In this technique, calculated chemical shift perturbations are compared to those measured experimentally. The position of the ligand is then iterated to achieve the best match between the calculated and experimental chemical shift perturbations. In contrast to differential chemical shifts, this method does not require analogs of the compound in question. However, in its simplest form only ring current contributions to chemical shift changes are used. This limits its applicability to ligands with aromatic groups. For simplification, common aromatic groups are used to mimic more complex ligand aromatic groups. For example, the binding site of the W-7 ligand was mapped onto Ca²⁺-bound calmodulin using a tryptophan ring system to mimic the W-7 naphthalene ring system [53]. As parameterization methods improve, calculations involving the real ligand and more encompassing contributions to chemical shift can be envisioned. The mapped probe position can be represented in coordinate space by an

electron current density termed the j-surface [49]. If analogs of a compound are available, their j-surfaces can be compared to deduce the binding orientation [49]. Chemical shift perturbations along with restraints from residual dipolar coupling have also been used to dock protein–protein complexes [54].

197

2.2. Monitoring the small molecules

When monitoring small molecules, the choice of physical mechanisms manifested in measurable NMR parameters is diverse. These include longitudinal, transverse, and DQ relaxation; diffusion coefficients; and intermolecular and intramolecular magnetization transfer. The latter includes transferred NOE, NOE pumping and reverse NOE pumping, saturation transfer, and WaterLOGSY experiments. The main advantage of monitoring the small molecule resonances is that there is no need to isotopically enrich the target macromolecule and no upper limit to the size of target that can be screened. Also, unlike the targetbased SAR by NMR method, the identity of a ligand contained in a mixture of compounds can often be obtained directly from the screening data without the need to deconvolute the mixture. The main disadvantages are that ligands with high affinity will be missed and that no information regarding ligand binding sites is directly available from the screening data. This stems from the fact that the binding interaction is manifested in the signals observed for free ligand. The ligand must be in medium to fast exchange between the free and bound states for the binding interaction to be detected, thus ligands in slow exchange will appear as non-binders. Recently, however, competition binding versions for several of these experiments have been developed that allow the detection of highaffinity ligands and their binding site locations. Competition binding experiments also make possible extremely HTS. Descriptions of each method are presented along with selected applications in drug discovery.

2.2.1. Longitudinal relaxation

The observed longitudinal relaxation rate $R_{1,obs}$ for a resonance of a small molecule interacting with a





Fig. 3. Plot of $R_{1,ns}$, $R_{1,s}$ and R_2 as a function of correlation time obtained from Eqs. (7), (8) and (10). In the diagram $k = \hbar^2 \gamma_{\rm H}^4 / 10 r_{ij}^6$, A proton Larmor frequency of 600 MHz was used for the simulation.

macromolecule is given by the equation [55]

$$R_{1,\text{obs}} = \frac{[\text{EL}]}{[\text{L}_{\text{TOT}}]} R_{1,\text{bound}} + \left(1 - \frac{[\text{EL}]}{[\text{L}_{\text{TOT}}]}\right) R_{1,\text{free}} \quad (4)$$

where $[L_{TOT}]$ and [EL] are the total ligand concentration and the bound ligand concentration, respectively. $R_{1,bound}$ and $R_{1,free}$ are the longitudinal relaxation rates for the ligand in the bound and free states, respectively. Longitudinal relaxation experiments can be performed either by inverting all resonances contained in a spectrum or by selectively inverting one single resonance. The former will be referred to as the longitudinal non-selective relaxation rate $R_{1,ns}$, while the latter will be referred to as the longitudinal selective relaxation rate $R_{1,s}$.

The $R_{1,ns}$ and $R_{1,s}$ for a proton *i* refer to a sum of uncorrelated pairwise proton dipole–dipole interactions, and the possible contribution by other relaxation mechanisms are grouped into an extra term, ρ_i^*

$$R_{1,\mathrm{ns}}^{i} = \sum_{j \neq i} \rho_{ij} + \sum_{j \neq i} \sigma_{ij} + \rho_{i}^{*}$$
(5)

$$R_{1,s}^i = \sum_{j \neq i} \rho_{ij} + \rho_i^* \tag{6}$$

where ρ_{ij} and σ_{ij} are the self- and cross-relaxation rates for any H_i-H_j dipole-dipole interaction, respectively, and the sum is extended to all protons that are dipolar coupled to proton *i*.

Eqs. (5) and (6) can be written in terms of spectral

densities [56] obtaining

$$R_{1,\rm ns}^{i} = \sum_{j \neq i} \frac{\gamma^{4} \hbar^{2}}{10 r_{ij}^{6}} \left\{ \frac{3\tau_{\rm c}}{1 + \omega^{2} \tau_{\rm c}^{2}} + \frac{12\tau_{\rm c}}{1 + 4\omega^{2} \tau_{\rm c}^{2}} \right\}$$
(7)

$$R_{1,s}^{i} = \sum_{j \neq i} \frac{\gamma^{4} \hbar^{2}}{10 r_{ij}^{6}} \left\{ \frac{3\tau_{c}}{1 + \omega^{2} \tau_{c}^{2}} + \frac{6\tau_{c}}{1 + 4\omega^{2} \tau_{c}^{2}} + \tau_{c} \right\}$$
(8)

where \hbar is the reduced Planck constant, γ is the proton gyromagnetic ratio, ω is the proton Larmor frequency, τ_c is the correlation time, and r_{ij} is the internuclear distance between protons *i* and *j*.

The plot of $R_{1,ns}$ and $R_{1,s}$ as a function of the correlation time obtained from Eqs. (7) and (8) is shown in Fig. 3. When a small molecule is bound to a large macromolecule with slow reorientation $(\omega \tau_c \gg 1)$, there is a substantial contribution to $R_{1,s}$. It is thus evident that $R_{1,s}$ experiments can be used to monitor binding of small molecules to a receptor. According to Eq. (7) and the diagram of Fig. 3, $R_{1,ns}$ lacks a direct τ_c dependence and therefore cannot be used for screening purposes.

 $R_{1,s}$ experiments are carried out with the selective inversion of one ligand resonance. This methodology has been applied to the study of single molecules as well as to pairs of molecules interacting with a receptor [55,57–63]. Measured $R_{1,s}$ changes upon complexation to the target biomolecule can provide useful information about the binding mode. In addition, measured $R_{1,s}$ values can be used to extract the binding constant. A Lineweaver-Burk type plot of $1/\Delta R_{1,s}$ vs. [L_{TOT}], where $\Delta R_{1,s}$ is the difference between the observed $R_{1,s}$ in the presence and absence of the protein, permits one to extrapolate the values of $K_{\rm D}$ since $1/\Delta R_{1,s} = 0$ at $[L_{\rm TOT}] = -K_{\rm D}$ [64]. This is clearly valid only in the limit of fast exchange and for $[L_{TOT}] \gg [E_{TOT}]$, where $[E_{TOT}]$ is the total protein concentration.

Although these experiments are very powerful, they have not been used in NMR screening because of the problems of achieving selective excitation for a large library of chemically diverse compounds. In Section 2.2.7, it will be shown that the $R_{1,s}$ experiment can be used in an efficient way to perform HTS.

2.2.2. Transverse relaxation

In the moderately fast exchange limit the observed transverse relaxation rate, $R_{2,obs}$, for a resonance of a





Fig. 4. Identification by 500 MHz ¹H NMR spectra of a ligand for FKBP from a mixture containing eight other non-binding compounds using transverse relaxation-editing. (a) Transverse relaxation-edited ¹H NMR spectrum of the nine compounds in the absence of FKBP. (b) Transverse relaxation-edited ¹H NMR spectrum of the nine compounds in the presence of FKBP from a similar spectrum recorded on just FKBP. (c) Difference spectrum of (a) minus (b). (d) Reference spectrum of the ligand. (e) Difference spectrum similar to spectrum (c) recorded on solutions containing only the eight non-binding compounds. Dashed lines in (c) and (d) correspond to the ligand resonances. Reprinted with permission from Hajduk et al. [70]. © 1997 American Chemical Society.

small molecule interacting with a macromolecule is provided by the equation [65]

$$R_{2,\text{obs}} = \frac{[\text{EL}]}{[\text{L}_{\text{TOT}}]} R_{2,\text{bound}} + \left(1 - \frac{[\text{EL}]}{[\text{L}_{\text{TOT}}]}\right) R_{2,\text{free}} + \frac{[\text{EL}]}{[\text{L}_{\text{TOT}}]} \times \left(1 - \frac{[\text{EL}]}{[\text{L}_{\text{TOT}}]}\right)^2 \frac{4\pi^2 (\delta_{\text{free}} - \delta_{\text{bound}})^2}{K_{-1}} \quad (9)$$

where $R_{2,\text{bound}}$ and $R_{2,\text{free}}$ are the transverse relaxation rate constants for the ligand in the bound and free states, respectively, δ_{bound} and δ_{free} are the chemical shifts for the resonance of the ligand in the bound and free states, respectively, and $1/K_{-1}$ is the residence time of the ligand bound to the protein. The last term disappears in the true fast exchange limit.

In terms of spectral densities, the transverse relaxation rate, R_2 , for in-phase magnetization is given by the equation [56,66]:

$$R_{2}^{i} = \sum_{j \neq i} \frac{\gamma^{4} \hbar^{2}}{10r_{ij}^{6}} \\ \times \left\{ \frac{15\tau_{c}}{2(1+\omega^{2}\tau_{c}^{2})} + \frac{3\tau_{c}}{1+4\omega^{2}\tau_{c}^{2}} + \frac{9}{2}\tau_{c} \right\}.$$
 (10)

The third term contains the direct τ_c dependence (i.e. spectral density calculated at zero frequency) that can be used for the screening. The plot of R_2 as a function of the correlation time is shown in Fig. 3. It should be pointed out that Eq. (10) is an approximation. For multiplet resonances, dipole-dipole and CSA-dipole cross correlation terms must also be considered. These terms contain spectral densities calculated at zero frequency and therefore for long τ_c they will contribute to the observed differential linewidth of the lines comprising a multiplet [67]. The last term in Eq. (9) accounts for exchange broadening in the intermediate exchange regime. This contribution is very small and is therefore safely neglected for molecules with weak binding affinity (high µM) and when the ligand concentration is in high excess compared to the protein concentration. However, this term can be efficiently used in the competition binding experiments discussed later.

Compounds interacting with a receptor display an enhanced R_2 value. Since the linewidth is equal to R_2/π , the resonances of a binding molecule display an increased broadening. Often this is visible from a simple inspection of the 1D ¹H spectrum. The broadening effects are more pronounced with large receptors or with receptors bound to a solid matrix (i.e. macromolecules that have a long overall correlation time τ_c). This results from the significant contribution of the third term of Eq. (10) that is proportional to τ_c . For lower molecular weight receptors the broadening may not be as large. In these cases, binders can be identified using the so-called T_2 - (or $T_{1\rho}$ -) filtered experiments that have





Fig. 5. Principle of second-site screening using a spin-labeled first ligand. Spin-labeled ligand, denoted by compound (34), occupies part of the Bcl-xL binding site. Ligands that bind in an adjacent location, denoted by compound (35), experience a paramagnetic relaxation enhancement. Reprinted with permission from Jahnke et al. [74]. © 2000 American Chemical Society.

macromolecule resonances in mixtures containing macromolecules and small molecules [68,69].

The potential of these methods for NMR screening has been documented by Fesik and coworkers [70]. Fig. 4 [70] shows how a ligand for FKBP with an affinity of 200 µM was identified from a mixture containing eight other compounds that were known not to bind FKBP. This method has also been utilized as part of a flow-injection screening process to identify ligands [8]. A Carr-Purcell-Meiboom-Gill (CPMG) spin echo train sequence [71] is introduced before the acquisition period in order to reduce or eliminate the signals of the protein and bound ligands, without significantly affecting the signals of unbound molecules. If the CPMG sequence is used, the delay between successive 180° pulses, $\tau_{\rm CPMG}$, must satisfy $|4\pi J^{\text{HH}} \tau_{\text{CPMG}}| \ll 1$ in order to minimize scalar coupling (J^{HH}) evolution [72]. Short spin echo train durations will preferentially detect high-affinity ligands, while longer spin echo train durations are used for the detection of weaker binding ligands. High-affinity ligands can be detected with this method only when the concentration of the ligand and protein are comparable. The experiments require the acquisition of the spectra of the chemical mixture in the absence and presence of the protein. The two spectra must be acquired with the same experimental conditions in order to allow for a direct comparison.

2.2.3. Transverse relaxation with spin labels

A very sensitive R_2 experiment for NMR screening utilizes spin labels covalently attached to the side chain of certain amino acids of the protein. In the fast exchange limit the observed transverse relaxation rate, $R_{2,obs}$, for a resonance of a nucleus in a small molecule interacting with a spin-labeled macromolecule is provided by the equation [73]

$$R_{2,\text{obs}} = \frac{[\text{EL}]}{[\text{L}_{\text{TOT}}]} R_{2,\text{bound}} + \left(1 - \frac{[\text{EL}]}{[\text{L}_{\text{TOT}}]}\right) R_{2,\text{free}} + \frac{[\text{EL}]}{[\text{L}_{\text{TOT}}]} \times \left(1 - \frac{[\text{EL}]}{[\text{L}_{\text{TOT}}]}\right)^2 \frac{4\pi^2 (\delta_{\text{free}} - \delta_{\text{bound}})^2}{K_{-1}} + \frac{[\text{EL}]}{[\text{L}_{\text{TOT}}]} R_{2,\text{para}}$$
(11)

where $R_{2,para}$ describes additional transverse relaxation resulting from paramagnetic relaxation enhancement when the ligand is bound to the spin-labeled target protein. Spin-spin relaxation rates are proportional to the product of the squares of the gyromagnetic ratios of the two spins. Since the gyromagnetic ratio of an unpaired electron is 658 times greater than that of a proton, an unpaired electron will have a drastically greater relaxation effect on nearby protons than will another proton. A spin label incorporated strategically onto the surface of a protein can thus be used to identify ligands that bind in proximity to this location [73]. Likewise, a spin label appropriately placed on a known ligand can be used to identify ligands that bind simultaneously with and in close proximity to the spin-labeled ligand [74].

The principle of second-site screening by using a spin-labeled first-site ligand is illustrated for the antiapoptotic protein Bcl-xL in Fig. 5 [74]. In this example, HTS had already identified (33) as an inhibitor with a $K_{\rm I}$ value of 140 μ M for the disruption of Bcl-xL/Bax association. Traditional medicinal chemistry was not successful in improving the affinity of this lead compound. However, since NMR methods had determined that (33) occupies only a portion of the natural antagonist's binding location, NMR screening experiments were designed to identify

B.J. Stockman, C. Dalvit / Progress in Nuclear Magnetic Resonance Spectroscopy 41 (2002) 187-231



Fig. 6. Identification of *p*-hydroxybenzanilid as a ligand for FKBP from a mixture containing four other non-binding compounds using transverse relaxation-editing enhanced by spin labels. Transverse relaxation-edited ¹H NMR spectra were recorded at 600 MHz with 10 ms (top row) and 200 ms (bottom row) spin-lock periods. Solutions contained either no FKBP (left), unmodified FKBP (middle), or spin-labeled FKBP (right). Solid arrows identify resonances arising from *p*-hydroxybenzanilid, while dashed arrows identify resonances arising from a second, weakly binding compound. Reprinted with permission from Jahnke et al. [73]. © 2001 American Chemical Society.

ligand fragments that would bind in the remainder of the active site. The low solubility of (33) prevented screening Bcl-xL under saturating conditions of (33) that would block this portion of the binding site. As an alternative, (34) was synthesized. This compound is an adduct of (33) with a TEMPO spin label that retains similar affinity to Bcl-xL. The binding mode of (34) was determined by NMR and indicated that the spin label is properly positioned toward the empty part of the binding site. A library of eight compounds was then screened for simultaneous binding with (34) to Bcl-xL using 1D ¹H relaxation-editing techniques. In these experiments, library compounds, (34) and BclxL were present at 500, 50 and 100 µM, respectively. Second-site ligands were identified by paramagnetic relaxation enhancement. In control experiments without Bcl-xL, no paramagnetic relaxation enhancement was observed. This indicates that the second-site ligand and (34) must be binding simultaneously to Bcl-xL. One of the second-site ligands identified was (35) which has an affinity of about 1 mM. Chemical shift analyses and protein-ligand NOEs confirmed that (35) does in fact bind in the desired part of the active site. The orientation of (35) with respect to (34)was determined by analyzing the paramagnetic relaxation enhancements for the individual protons of (35). This orientation is that shown in Fig. 5 and suggests that (33) and (35) could be linked by an appropriate spacer to form (36) in an attempt to create a single-compound inhibitor of Bcl-xL/Bax association. To test the limits of this technique, the authors carried out a series of experiments and model calculations. These indicated that (35) would have been detected at Bcl-xL concentrations as low as 10 µM. Moreover, ligands with higher affinity or that bind closer to the spin label would result in a further drop in required protein concentrations. With cryoprobes, the necessary protein concentration could drop into the nM range. The main advantages of this method are the requirement for simultaneous ligand binding and the first-site/second-site ligand orientation information obtained. Simultaneous binding rules out false-positive compounds that might be binding in the unsaturated first-site instead of the desired site. This is most critical when saturating concentrations of the first-site ligand cannot be obtained because of poor solubility. This technique is also extremely sensitive, amenable to automation, and insensitive to slight ligand-induced variations in sample condition such as pH. The big disadvantage of this method is the need to synthesize a spin-labeled derivative of the first ligand that retains good binding

affinity.



Spin labels incorporated into proteins and used as probes for primary NMR screening have the potential to drastically reduce the quantity of a protein needed for the primary screen [73]. The SLAPSTIC method (spin labels attached to protein side chains as a tool to identify interacting compounds) has recently been described using FKBP as a model system [73]. For this technique to work, the spin-labeled side chain must be reasonably close to the desired ligand binding site, thus some structural information must already be known about the target protein. For FKBP, the authors chose to spin label lysine side chains since there are several within 12-15 Å of the binding site of the pipecolinic acid moiety. Mixtures containing phydroxybenzanilid, a known ligand with an affinity of ~ 1.1 mM, and four non-binding compounds were then screened against spin-labeled FKBP using 1D¹H relaxation-editing techniques. In these experiments, the compounds and spin-labeled FKBP were present at 50 and 20 µM, respectively. Binding of phydroxybenzanilid was easily detected by complete reduction of its resonances as shown in Fig. 6 [73]. In the presence of 60 µM FKBP that was not spinlabeled, p-hydroxybenzanilid resonances were also noticeably reduced. A similar level of reduction was observed using a 1 µM solution of spin-labeled FKBP, suggesting that the spin-labeling technique can reduce protein quantity needs by a factor of 50. Compound concentrations of $\sim 50 \ \mu M$ are sufficient to detect binding in as little as 2 min per mixture. Disadvantages of the SLAPSTIC method include the necessity for appropriately spin-labeled protein targets that retain their biochemical integrity and function, and, as with other ligand-based techniques, its inability to detect tight-binding ligands and to obtain information about ligand binding sites.

2.2.4. Double-quantum relaxation

In the fast exchange limit the observed DQ relaxation rate $R_{DQ,obs}$ for two scalar coupled protons *i* and *j* of a small molecule interacting with a macromolecule is provided by the equation

$$R_{\text{DQ,obs}} = \frac{[\text{EL}]}{[\text{L}_{\text{TOT}}]} R_{\text{DQ,bound}} + \left(1 - \frac{[\text{EL}]}{[\text{L}_{\text{TOT}}]}\right) R_{\text{DQ,free}} + 4\pi^2 \frac{[\text{EL}]}{[\text{L}_{\text{TOT}}]} \left(1 - \frac{[\text{EL}]}{[\text{L}_{\text{TOT}}]}\right)^2 \times \left\{ \frac{(\delta_{\text{free}}^i - \delta_{\text{bound}}^i)^2}{K_{-1}} + \frac{(\delta_{\text{free}}^j - \delta_{\text{bound}}^j)^2}{K_{-1}} + \frac{2(\delta_{\text{free}}^i - \delta_{\text{bound}}^i)(\delta_{\text{free}}^j - \delta_{\text{bound}}^j)}{K_{-1}} \right\}$$
(12)

where $R_{\text{DQ,bound}}$ and $R_{\text{DQ,free}}$ are the DQ relaxation rate constants for the ligand in the bound and free states, respectively. $\delta^{i}_{\text{bound}}$, $\delta^{j}_{\text{bound}}$ and δ^{i}_{free} , δ^{j}_{free} are the chemical shifts for the resonances of the protons *i* and *j* of the ligand in the bound and free states, respectively.

Neglecting dipole-dipole and CSA-dipole cross correlation, the DQ relaxation rate is provided by

$$R_{\rm DQ} = \frac{\gamma^4 \hbar^2}{10r_{ij}^6} \left\{ \frac{3\tau_{\rm c}}{(1+\omega^2\tau_{\rm c}^2)} + \frac{6\tau_{\rm c}}{1+4\omega^2\tau_{\rm c}^2} \right\} \\ + \sum_{k \neq i,j} \frac{\gamma^4 \hbar^2}{10r_{ik}^6} \left\{ \frac{9\tau_{\rm c}}{2(1+\omega^2\tau_{\rm c}^2)} + \frac{3\tau_{\rm c}}{1+4\omega^2\tau_{\rm c}^2} + \frac{5}{2}\tau_{\rm c} \right\} \\ + \sum_{l \neq i,j} \frac{\gamma^4 \hbar^2}{10r_{jl}^6} \left\{ \frac{9\tau_{\rm c}}{2(1+\omega^2\tau_{\rm c}^2)} + \frac{3\tau_{\rm c}}{1+4\omega^2\tau_{\rm c}^2} + \frac{5}{2}\tau_{\rm c} \right\}$$
(13)

where *k* are all of the protons that are dipolar coupled with the ligand proton *i*, and *l* are all of the protons with a dipolar interaction with the ligand proton *j* [75–78]. Like R_2 , DQ relaxation has a direct dependence upon the correlation time of the molecule. This contribution does not come from the dipolar





Fig. 7. 2D ¹H (a) *z*- or magic angle pulsed field gradient DQ and (b) DQC relaxation weighted DQ pulse sequences. Gradients G_1 to G_3 are coherence selection gradients and should be applied at the magic angle if triple-axis gradient hardware is available. The first two weak PFGs are used for suppression of radiation damping. The delay τ is set between 30 and 50 ms whereas the delay τ_1 is set between 10 and 100 ms. A very weak (<20 Hz) H₂O presaturation field can be applied before the excitation DQ period in the z-PFG version of the DQ experiment if the H₂O suppression is not optimal. Technical details of the experiments have been reported in a recent review article [83].

interaction between the two protons *i* and *j* directly involved in the DQ coherence. Instead, it originates from all of the dipolar interactions that these two protons experience with other protons. A significant difference between the two relaxation rates is represented by the exchange contribution [79-82]. For DQ coherence three terms contribute to the relaxation. In addition to the exchange terms deriving from each proton, there is a cross term. When the chemical shift difference of the resonances of the two protons i and j have the same sign (i.e. both resonances are shifted upfield or downfield for the ligand bound to the protein), all three terms will add and the exchange term can become significant. When the two chemical shift differences have opposite sign (i.e. one resonance is shifted upfield and the other downfield for the ligand bound to the protein), the exchange term is reduced.

The DQ experiments require first the creation of DQ coherence via antiphase magnetization and then its conversion to single quantum (SQ) coherence for detection. Therefore these experiments are more appropriately defined as transverse SQ and DQ relaxation filter experiments. The spectra can be recorded in 1D or 2D versions. The sensitivity improvement of the 1D vs. the 2D version of the DQ experiment is only $\sqrt{2}$. In addition, problems of overlap and the absence of information about the spin networks in the 1D spectrum limits its utilization to only simple mixtures comprised of a few compounds. The scalar connectivities observed in the 2D DQ spectrum allow the direct identification of the molecules interacting with the target protein without the need to deconvolute the complex mixture. Suppression of the protein signals is very efficient not only because of the presence of the excitation DQ period and the relaxation of DQ coherence during t_1 , but also because of the antiphase character of the multiplets along ω_2 . The large linewidth of the protein resonances results in extensive cancellation of the positive and negative lobes along ω_2 .

The 2D DQ experiment is a cosine modulated experiment and therefore maximum signal is observed in the first t_1 increments. The experiment can be recorded rapidly by acquiring only a limited number of t_1 increments. Linear prediction can then be applied in the t_1 dimension for improving the resolution. Extensive folding can also be performed in ω_1 thereby reducing the spectral width and the number of t_1 increments. The pulse sequences for these experiments are shown in Fig. 7 [83]. Coherence selection field gradients are used for good solvent suppression and for reducing artefacts, thus allowing a direct comparison of the different DQ spectra. Remarkable





Fig. 8. Panel (a) depicts the fingerprint region of the 2D DQ 600 MHz ¹H spectrum of a three compound mixture: Ac-Tyr-Val-Asn-Val-OH (resonances A), Ac-Tyr(PO₃H₂)-Val-NH₂ (resonances B) and Ac-(*N*-Me)Phe(*p*-CH₂PO₃H₂)-Val-Asn-NH-CH₂-CH(CH₃)₂ (resonances C) in the absence of the target protein. Panel (b) depicts the same 2D DQ fingerprint region of the mixture in the presence of the Grb2 SH2 domain. Each of the molecules is present at a molar ratio to the target protein of 2:1 (100 μ M:50 μ M). The rectangles in (b) indicate the positions of the missing cross peaks. The spectra were recorded with the pulse sequence of Fig. 7a.

multiple solvent suppression is then achieved with coherence selection gradients tilted at the magic angle [83]. This becomes particularly relevant in the observation of very dilute samples used in NMR screening, where the dynamic range of the receiver needs to be maximized in an optimal way. An advantage of using samples dissolved in H_2O is the observation of the NH doublet resonances. The NH

proton signals are in a spectral region with little or no overlap with other signals and therefore can be used for assignment purposes in complicated mixtures. Furthermore, the chemical shift difference of a ligand NH proton free in solution and when hydrogen bonded to a protein is typically large. Therefore the exchange terms in Eq. (12) can make a large contribution to the DQ relaxation.



Fig. 9. Panel (a) depicts the fingerprint region of the 2D DQ 600 MHz ¹H spectrum of the four compound mixture containing Ac-Tyr-Val-Asn-Val-OH (resonances A), Ac-Tyr(PO₃H₂)-Val-NH₂ (resonances B), indole (resonances C) and Ac-Tyr(PO₃H₂)-Val-NH-CH₂-CONH₂ (resonances D). The resonances of indole are not visible in this spectral region. Panel (b) depicts the same 2D DQ fingerprint region of the four compound mixture in the presence of the Grb2-SH2 domain. Panel (c) depicts the 2D difference spectrum obtained by subtracting the two 2D DQ spectra in (a) and (b). Each of the molecules is present at a molar ratio to the target protein of 6:1 (300 μ M:50 μ M). The resonances of the indole are not affected by the presence of the protein (data not shown).

Fig. 8 shows the DQ fingerprint region of a compound mixture recorded with the pulse sequence of Fig. 7a in the absence and presence of the Grb2-SH2 domain. This represents a challenging test case because of the small size of the protein employed (MW = 11.8 kDa). Close comparison of the two spectra allows rapid identification of Ac-(*N*-Me)Phe(*p*-CH₂PO₃H₂)-Val-Asn-NH-CH₂-CH(CH₃)₂ (resonances labeled C) as a ligand for Grb2-SH2. The experiment, in particular the versions recorded with $\beta = 45^{\circ}/135^{\circ}$ or $\beta = 60^{\circ}/120^{\circ}$, represents the most sensitive 2D NMR experiment for detection of binding of small and medium size molecules.

Furthermore, the absence of strong diagonal peaks and singlet resonances with the concomitant absence of t_1 noise and excellent solvent suppression in DQ spectra, allows a more precise and reliable use of the 2D subtraction technique. This can be appreciated in Fig. 9, which shows an expanded region of the DQ spectra recorded on a four compound mixture. Despite the strong overlap of cross peaks A and D (Fig. 9a), it is possible to identify peak D in the DQ difference spectrum (Fig. 9c) as the cross peak that is affected by the presence of the protein. A drawback of this methodology is the requirement for the presence of two doublets in the spectrum of the molecule to be screened.

2.2.5. Diffusion experiments

The translational diffusion rate can also be used to detect complex formation. For a sphere of radius *r* in a continuous medium of viscosity η , the translational diffusion rate *D* is provided by the Stokes–Einstein equation [84]

$$D = \frac{KT}{6\pi\eta r} \tag{14}$$

where K is the Boltzmann constant and T is the absolute temperature.

Owing to the 1/r dependence, small molecules that do not aggregate have D values which are about an order of magnitude larger compared to the D value of a large macromolecule. When a small molecule binds to a large macromolecule it will transiently have the Dvalue of the receptor. In the fast exchange limit the observed translation diffusion coefficient D_{obs} is then given by the equation

$$D_{\rm obs} = \frac{[\rm EL]}{[\rm L_{TOT}]} D_{\rm bound} + \left(1 - \frac{[\rm EL]}{[\rm L_{TOT}]}\right) D_{\rm free} \qquad (15)$$

where D_{bound} and D_{free} are the diffusion coefficients of the ligand in the bound and free states, respectively. For high molecular weight targets, the D_{bound} contribution to the equation becomes negligible. D values can be measured using either the pulsed field gradient spin echo (PFG-SE) [85] or the pulsed field gradient stimulated echo (PFG-STE) [86] methods. The former uses a spin echo scheme with the defocusing gradient applied immediately after the first hard 90° pulse and the refocusing gradient applied just before acquisition. The latter uses the 90°-90°-90° scheme with the defocusing gradient applied immediately after the first rf pulse and the refocusing gradient applied after the third rf pulse. An advantage of the stimulated echo experiment is that the magnetization is placed for most of the time along the z-axis, thus avoiding extensive signal losses resulting from rapid transverse relaxation in those experiments that require long diffusion delays. However, the stimulated echo suffers from the drawback that only half of the signal is detected. Technical improvements such as the use of bipolar gradients and the use of the longitudinal eddy-current delay (LED) sequence in the stimulated echo version reduce artifacts originating from eddy currents [87].

The *D* value is calculated according to the equation [88]

$$\ln \frac{I(G)}{I(0)} = -\gamma^2 G^2 D \delta^2 \left(\Delta - \frac{\delta}{3} \right) \tag{16}$$

for gradients G of rectangular shape and

$$\ln \frac{I(G)}{I(0)} = -\gamma^2 G^2 D \delta^2 \left(\frac{4\Delta - \delta}{\pi^2}\right) \tag{17}$$

for gradients G of sine shape.

I(G) and I(0) are the signal intensities measured in the presence and absence of gradients, respectively, γ is the proton gyromagnetic ratio, G and δ are the amplitude and duration of the two PFGs, respectively, and Δ is the time between the rising edges of the two gradients. NMR spectra are acquired with different gradient amplitudes. A plot of the natural logarithm of the signal intensity versus G^2 yields a straight line





Fig. 10. Identification by 500 MHz ¹H NMR spectra of 4-cyano-4'hydroxybiphenyl as a ligand for stromelysin from a mixture containing eight other non-binding compounds using diffusionediting. (a) Diffusion-edited ¹H NMR spectrum of the nine compounds in the absence of stromelysin recorded with low gradient strength. (b) Diffusion-edited ¹H NMR spectrum of the nine compounds in the presence of stromelysin recorded with low gradient strength after subtracting a similar spectrum recorded with high gradient strength (to remove the protein signals). (c) Difference spectrum of (a) minus (b). (d) Reference spectrum of the ligand. (e) Difference spectrum similar to spectrum (c) recorded on solutions containing only the eight non-binding compounds. Dashed lines in (c) and (d) correspond to the ligand resonances. Buffer impurities are denoted with asterisks. Reprinted with permission from Hajduk et al. [70]. © 1997 American Chemical Society.

whose slope can be used to calculate the self-diffusion coefficient according to Eqs. (16) or (17).

Instead of a quantitative analysis, a qualitative one serves most purposes in the early stages of NMR screening to identify compounds that bind to a receptor from a mixture of non-binding molecules [70]. First, a PFG-STE or PFG-SE spectrum of the chemical mixture in the absence of the protein is recorded at a low gradient strength. Next, the same spectra for the chemical mixture in the presence of the protein are recorded at low and high gradient strengths and subtracted to produce a spectrum that contains only the signals of the compounds not interacting with the protein. The resulting subtracted spectrum is then subtracted from the spectrum of the chemical mixture recorded in the absence of the protein to obtain a spectrum that contains only the signals of the molecule that binds to the receptor. The utility of diffusion editing has been demonstrated using this procedure with stromelysin [70] as shown in Fig. 10. 4-Cyano-4'-hydroxybiphenyl, which binds to stromelysin with a dissociation constant of 20 µM in the presence of (5), was easily identified from a mixture containing eight other non-binding compounds. Simpler approaches directly compare the residual ligand signals in diffusion edited spectra recorded with strong gradients in the absence and presence of receptor, or spectra recorded with weak and strong gradients in the presence of receptor [89]. Clearly this is possible only if there is no strong overlap of the resonances of the different ligands comprising the mixture. The effectiveness of these methods has been demonstrated by detecting known tetrapeptide ligands for vancomycin from a mixture containing eight other non-binding tetrapeptides [90], and by detecting the binding of a known ligand to a DNA dodecamer from a mixture containing three other non-binding small molecules [91]. A 2D experiment known as diffusion encoded spectroscopy (DECODES) that combines diffusion editing with TOCSY permits determination of the ligand structures [92]. This could be quite useful in situations where the compounds in the mixtures are not completely characterized in advance, such as might be the case for plant and fungi extracts or for combinatorial chemistry libraries. Diffusionediting has also been used to identify small molecules that interact with polymers [93].

A variant of the experiment is the coupling of the diffusion filter with an isotope filter experiment. The proton signals of ${}^{13}C/{}^{15}N$ -labeled protein are suppressed with an isotope filter allowing the measurement, without interference from the protein signals, of the diffusion coefficient of the unlabeled ligand [94, 95]. The method was demonstrated using ${}^{13}C/{}^{15}N$ -labeled stromelysin by detection of only known ligand resonances from a mixture containing one other non-binding small molecule [95]. Resonances arising from the protein and non-binding small molecule were suppressed.

NMR screening with the diffusion-based approach



Fig. 11. 2D 500 MHz ¹H NOESY spectra of a 15-member oligosaccharide library in the (a) absence and (b) presence of agglutinin. Observed NOEs are positive in (a) and negative in (b). In (b), a spin-lock filter was used to remove protein resonances, and transfer NOE correlations are observed only for the oligosaccharide α -L-Fuc-(1 \rightarrow 6)- β -D-GlcNAc-OMe. Reprinted with permission from Meyer et al. [101]. © 1997 Federation of European Biochemical Societies.

suffers from some important limitations. In the approach described above a signal reduction of more than fourfold is observed compared to the acquisition of a normal 1D spectrum. A factor of two is lost in the PFG-STE experiment and another factor of two is lost in the two spectral subtraction procedures. In addition there are signal losses originating from longitudinal relaxation during the diffusion period. Differences in diffusion coefficients are also not very large. For example, hemoglobin, a protein of MW ~ 65 kDa, has a *D* value of $0.69 \times 10^{-6} \text{ cm}^2/\text{s}$ [84], which is only seven times smaller compared to sucrose (D = 5.2×10^{-6} cm²/s) and 12 times smaller compared to alanine $(D = 8.6 \times 10^{-6} \text{ cm}^2/\text{s})$ [96]. Therefore it is necessary to acquire diffusion-filtered spectra with a high signal-to-noise ratio in order to detect small changes in the diffusion coefficient of a small molecule interacting with a protein. Molecules binding weakly to the receptor are difficult to detect since the fraction of bound ligand in the NMR screening experiments will be very small, unless a high protein concentration and a small $[L_{TOT}]/[E_{TOT}]$ ratio are used.

A useful application of the diffusion coefficient D is in the determination of the dissociation binding constant of weak affinity ligands. This is possible from the knowledge of the concentration of bound ligand calculated with Eq. (15). $K_{\rm D}$ is then given by the

expression:

$$K_{\rm D} = \frac{[{\rm E}_{\rm TOT}][{\rm L}_{\rm TOT}] - [{\rm E}_{\rm TOT}][{\rm EL}] + [{\rm EL}]^2 - [{\rm L}_{\rm TOT}][{\rm EL}]}{[{\rm EL}]}.$$
(18)

2.2.6. Intermolecular and intramolecular magnetization transfer

2.2.6.1. Transferred NOE. The 1D and 2D transferred NOE experiments have been used extensively to determine the conformation of weak to medium affinity ligands when bound to a large protein [97–100]. Recently the method has been proposed as a tool for performing NMR screening [101]. The observed ligand intramolecular cross-relaxation rate $\sigma_{\text{NOE,obs}}$ is given by the equation

$$\sigma_{\text{NOE,obs}} = \frac{[\text{EL}]}{[\text{L}_{\text{TOT}}]} \sigma_{\text{NOE,bound}} + \left(1 - \frac{[\text{EL}]}{[\text{L}_{\text{TOT}}]}\right) \sigma_{\text{NOE,free}}$$
(19)

where $\sigma_{\text{NOE,free}}$ and $\sigma_{\text{NOE,bound}}$ are the cross relaxation rates for the ligand in the free and bound states, respectively, and σ_{NOE} is provided





Fig. 12. Pulse sequences used in the reverse NOE pumping experiment. Reprinted with permission from Chen and Shapiro [106]. © 2000 American Chemical Society.

by the equation [56]

$$\sigma_{\rm NOE} = \frac{\gamma^4 \hbar^2}{10 r_{ij}^6} \left\{ \frac{6\tau_{\rm c}}{1 + 4\omega^2 \tau_{\rm c}^2} - \tau_{\rm c} \right\}$$
(20)

where r_{ij} is the distance between the two ligand protons.

From this equation two important considerations can be made. First, for small molecules the cross relaxation rate for the bound ligand is of opposite sign to the cross relaxation rate for the free ligand. Second, the absolute magnitude of the rate is significantly larger for the bound state. Thus in an equilibrium of bound and free ligand molecules, the total intramolecular ligand NOE is predominantly determined by the bound form. Fig. 11 shows the first use of the transferred NOE method to identify a ligand for a macromolecule from a mixture of small molecules [101]. The oligosaccharide α -L-Fuc-(1 \rightarrow 6)- β -D-GlcNAc-OMe was identified as a ligand for agglutinin from two separate mixtures of oligosaccharides. The method has also been used to screen mixtures of polysaccharides against the protein E-selectin [102]. In the latter, 2D TrNOE spectra were recorded for the mixture in the absence and presence of the protein. Close comparison of the two spectra allowed the identification of the NOE cross peaks that change sign or intensity. Analysis of all cross peak patterns identified the active component (37) in the mixture, which is a mimetic of the known ligand sialyl Lewis^x (38). The bioactive conformation of (37) had been determined previously from similar TrNOE experiments [103]. This suggests that in the absence of severe overlap with signals from other compounds, it is also possible to determine the bound conformation

of the ligand from the same 2D spectrum collected to identify the ligand. To reduce problems of overlap, a 3D-TOCSY-TrNOESY experiment has also been proposed [104]. This 3D variant of the experiment is currently not of routine application, but may be useful for the deconvolution of large chemical mixtures.



A drawback of this method is the requirement of large ligand concentrations. Therefore it is not suitable for the identification of ligands with poor solubilities. Another disadvantage is the presence in the spectrum of strong diagonal peaks that hamper the observation of cross peaks between ligand resonances having similar chemical shifts. In addition, the strong diagonal peaks may introduce t_1 noise or baseline problems that interfere with the observation of weak cross peaks.

2.2.6.2. NOE pumping and reverse NOE pumping. The 1D NOE pumping experiment and its related reverse NOE pumping experiment have been proposed by Chen and Shapiro [105,106] as a method for primary NMR screening. In the former, magnetization transfer from the receptor to the ligand is observed, and in the latter, the reverse process of magnetization transfer from the ligand to the receptor is monitored. A diffusion filter is introduced in the NOE pumping experiment before the NOE mixing time to destroy completely the ligand magnetization while preserving most of the receptor magnetization. During the mixing time the Zeeman magnetization of the receptor will relax and part of the magnetization will be transferred via intermolecular cross relaxation to the ligand. The choice of the optimal mixing time depends upon the protein longitudinal relaxation value. The method was demonstrated on a solution containing human serum albumin (HSA) and a mixture of three compounds: salicylic acid, L-ascorbic acid, and glucose [105]. Signals for salicylic acid, which is a known binding ligand, and water were observed. Signals for Lascorbic acid and glucose were not observed. Standard diffusion experiments failed to distinguish





Fig. 13. Reverse NOE pumping 500 MHz ¹H spectra of a mixture containing octanoic acid, glucose and HSA recorded with the pulse sequences shown in (a) Fig. 12b and (b) Fig. 12a. The difference spectrum of (a) minus (b) is shown in (c). Reprinted with permission from Chen and Shapiro [106]. © 2000 American Chemical Society.

salicylic acid as the ligand in this mixture because the apparent diffusion coefficients for the three small molecules were nearly identical. For the identification of weak affinity ligands, the NOE pumping experiment is more robust than standard diffusion experiments [105].

The reverse NOE pumping experiment requires the acquisition of two experiments using the pulse sequences shown in Fig. 12 [106]. The first experiment uses a transverse relaxation filter followed by the NOE mixing time and the detection pulse. The relaxation filter suppresses all of the protein signals while inverting the ligand proton signals. During the mixing time the ligand magnetization will be partially transferred via intermolecular cross relaxation to the receptor. A second experiment is then recorded where the order of the mixing time and transverse relaxation filter is reversed. In this experiment, no net magnetization transfer from the ligand to the receptor takes place. Therefore the magnetization of a binding molecule decays at a slower rate when compared to the first experiment. A subtraction of the two spectra, recorded with interleaved acquisition, results in a spectrum containing only the signals of the molecules interacting with the protein. The method was demonstrated on HSA solutions containing the non-binding small molecule glucose and one of a series of



Fig. 14. Identification of methyl β -D-galactopyranoside as a ligand for the RCA₁₂₀ lectin using STD ¹H NMR spectroscopy at 500 MHz. (a) Reference 1D ¹H NMR spectrum of RCA₁₂₀ lectin. (b) STD NMR spectrum of RCA₁₂₀ lectin. (c) Reference 1D ¹H NMR spectrum of RCA₁₂₀ lectin recorded with a spin-lock filter. (d) Reference 1D ¹H NMR spectrum of RCA₁₂₀ lectin in the presence of a 30-fold excess of methyl β -D-galactopyranoside. (e) STD NMR spectrum of RCA₁₂₀ lectin plus methyl β -D-galactopyranoside. (f) STD NMR spectrum as in (e) but with the addition of a spin-lock filter. Reprinted with permission from Mayer and Meyer [113]. © 2001 American Chemical Society.

unbranched fatty acid molecules that are known ligands as shown in Fig. 13 [106]. Only signals for the fatty acid molecules were observed. The amount of signal pumped increased with fatty acid chain length, indicative of increasing affinity, suggesting that a series of ligands can be ranked-ordered by affinity using this technique [106].

2.2.6.3. Saturation transfer experiments. The saturation transfer experiments developed in the early 1960s [107] have recently been applied to NMR screening with good results [108,109]. In these experiments an rf field is applied selectively to a resonance for a long time in order to equilibrate the populations of the two energy levels. Selective saturation of some protein proton resonances results in saturation of all protein protons via flip-flop energy transitions. The rate of these transitions is directly proportional to the protein correlation time (second





Fig. 15. Titration curve showing the STD amplification factor of a given ligand proton versus the ligand concentration. Reprinted with permission from Mayer and Meyer [113]. © 2001 American Chemical Society.

term of Eq. (20)), and therefore for large proteins or proteins coupled to a solid matrix the spreading of magnetization throughout the protein is rapid [110, 111]. The pulse sequence of the 1D saturation transfer difference (STD) method corresponds to the 1D truncated driven NOE difference experiment [112]. Spectra are acquired with selective saturation of some protein proton resonances (on-resonance), typically the upfield shifted methyl group resonances, and with selective saturation in an empty spectral region (offresonance). The subtraction of the spectra is performed internally via phase cycling after every scan to minimize the subtraction artifacts. The resulting subtracted spectrum contains only the signals of the molecules interacting with the protein. The signals of the protein are suppressed with the use of a transverse relaxation filter applied before the acquisition. An example is shown in Fig. 14 for the interaction of methyl β -D-galactopyranoside with the RCA₁₂₀ lectin [113]. The efficiency of STD for NMR screening has been demonstrated by detection of N-acetylglucosamine as a ligand for wheat germ agglutinin from a mixture containing six other non-binding saccharides [108]. The STD scheme can also be implemented as a 2D homonuclear ${}^{1}H-{}^{1}H$ [108] or heteronuclear ¹H-¹³C [114] experiment. Recording of 2D STD spectra allows unambiguous identification of ligand molecules directly from mixtures.

The fractional STD effect $(I_0 - I_{sat})/I_0$ (where I_0 is the intensity of one signal in the off-resonance spectrum and I_{sat} is the intensity of the same signal in the on-resonance spectrum) expresses the signal intensity in the STD spectrum as a fraction of the intensity in an unsaturated reference spectrum. A better description of the observed phenomenon is provided by the STD amplification factor (STD_{af}) defined as the fractional STD effect multiplied by the ligand excess, [L_{TOT}]/[E_{TOT}]). The size of STD_{af} is not directly correlated to the affinity of the ligand. Tight binders may produce a small STD_{af} because of their low K_{-1} (i.e. long residence time of the ligand bound to the protein), whereas a weak affinity ligand could produce a larger STD_{af}. Analysis of the different STD_{af} for all of the proton resonances of a ligand will provide important structural information to discern the ligand surface that is directly in contact with the protein [113]. This epitope mapping is possible only for weak affinity ligands. If the residence time of the ligand is long compared to the longitudinal relaxation of the ligand in the bound state, then spreading of the magnetization over many protons is observed, resulting in very small differential STD effects.

Titration and competition binding experiments can also be performed with STD [113]. Both experiments can be used to derive a value for the binding constant of a ligand. In the titration binding experiment the STD_{af} is plotted against [L_{TOT}] as shown in Fig. 15. A dose response curve is obtained and the K_D can be extracted by fitting the data with the equation

$$STD_{af} = \frac{-STD_{af}^{max}}{1 + \left(\frac{L}{K_{D}}\right)} + STD_{af}^{max}$$
(21)

where $\text{STD}_{af}^{\text{max}}$ is the maximum STD_{af} effect, and *L* is the ligand concentration.

It is also possible to derive the K_D value of a ligand from the IC₅₀ value of a competitive molecule with a known dissociation constant K_I according to the equation [115]

$$K_{\rm D} = \frac{[{\rm L}]K_{\rm I}}{{\rm IC}_{50} - K_{\rm I}}$$
(22)

where [L] is the ligand concentration for which the $K_{\rm D}$ should be calculated, and IC₅₀ is the concentration of the competitive inhibitor at which the STD_{af} observed for *L* is reduced by half. The IC₅₀ is calculated from STD experiments performed at a fixed ligand concentration [L] and increasing competitive inhibitor concentrations. Eq. (22) was derived with the approximation that [L] is in high excess compared to the protein and $K_{\rm D}$. In Section 2.2.7, a more general equation will be described that allows determination





Fig. 16. Plot of the intermolecular cross-relaxation rate σ versus water residence time for macromolecules with correlation times of 5, 10, 20 and 40 ns. A proton Larmor frequency of 600 MHz was used for the simulation. Reprinted with permission from Dalvit et al. [117]. © 2001 Kluwer Academic Publishers.

of the binding constant from a single-point measurement.

2.2.6.4. WaterLOGSY. The WaterLOGSY (Water– ligand observed via gradient spectroscopy) experiment [116,117] utilizes the large bulk water magnetization to transfer magnetization via the protein– ligand complex to the free ligand in a selective manner. In this experiment, the resonances of nonbinding compounds appear with opposite sign and tend to be weaker than those of the interacting ligands.

In all of the high resolution X-ray structures of protein–ligand complexes, water molecules are found linking the ligand to the protein, with most of the water molecules (~80%) making three or more hydrogen bonds [118]. In addition to the bridging water molecules, other water molecules are identified at the binding site. These water molecules occupy small hydrophobic voids or holes left unfilled by the imperfect matching of the ligand and protein surfaces. The residence times of water in protein cavities invariably seem to range between a few ns to a few hundred μ s [119–126]. As shown in Fig. 16, this time span is long compared to the effective correlation time where intermolecular water–protein NOEs change

sign (ca. 300 ps) at 600 MHz, and short compared to the chemical shift time scale, where a separate resonance for the bound water would be observed (ms). Thus, selective excitation of the water signal followed by NOE mixing effectively transfers magnetization from the bulk water to the protein with the same sign as the starting magnetization. The transfer of magnetization from bound water to nearby protons of the protein is more efficient for longer water residence times and for larger proteins (i.e. longer correlation times) as shown in Fig. 16.

A second important mechanism for magnetization transfer from water to the protein–ligand complex is by chemical exchange with labile carboxyl, amino, hydroxyl, imidazole, guanidinium and amide protons [112,127,128]. In physiological solutions typically used in the NMR screening experiments, this exchange is very rapid for the solvent exposed protons of the protein. Like the intermolecular NOE with buried water molecules, this magnetization transfer pathway conserves the sign of the magnetization. Both processes act constructively to transfer magnetization from the bulk water to the protein. The large number of exchangeable protons and buried, yet exchangeable, water molecules in a protein–ligand



Fig. 17. Pulse sequences used in the WaterLOGSY experiment. The sequences are based on the ePHOGSY experiment without (upper) and with (lower) a water flip-back pulse [157]. Reprinted with permission from Dalvit et al. [117]. © 2001 Kluwer Academic Publishers.

complex could explain the high sensitivity of the WaterLOGSY experiment for the selective detection of ligands. The method is a powerful tool for primary screening of compound mixtures by NMR. signal. Selective inversion of the water signal can be achieved in different modes [124]. ePHOGSY [129, 130], which is based on a water-selective 180° refocusing pulse between two pulsed field gradients as shown in Fig. 17, is one of the technically most robust schemes to achieve efficient selective water

WaterLOGSY experiments are performed by either selective decoupling or inversion of the water



Fig. 18. Identification of a ligand using the WaterLOGSY technique. Top is an expanded region of the 1D 1 H NMR reference spectrum at 600 MHz of a mixture of 10 compounds in the absence of protein. Bottom is the corresponding WaterLOGSY spectrum in the presence of the protein cdk2. The positive intensity signal indicated by an arrow in the WaterLOGSY spectrum identifies this compound as a ligand for cdk2.





Fig. 19. Diagram of the approach used for HTS with the competition binding experiments.

excitation. The scheme effectively suppresses radiation damping that interferes with the selective excitation and contributes to introduction of artifacts in the spectrum. In addition, the scheme defocuses the magnetization of all resonances that are not near the water chemical shift. WaterLOGSY spectra obtained with this method are devoid of artifacts, and even small effects can be analyzed with confidence. As shown in Fig. 18, non-binders are characterized by negative intensity in WaterLOGSY spectra, while ligands are characterized by positive intensity.

Titration binding experiments can be performed in order to extract an approximate value of the binding constant of weak affinity ligands. However, particular care must be taken in the analysis of the titration experiments, since the signal intensity in the Water-LOGSY spectra contains an off-setting effect deriving from the hydration of the free ligand. This contribution can be calculated in a complementary experiment recorded for the ligand in the absence of the protein. The subtraction of this experimentally derived contribution from the measured Water-LOGSY signals in the presence of the protein permits the calculation of the binding constant [117]. A drawback of the method is that very low-affinity ligands cannot be detected when the ratio $[L_{TOT}]/[E_{TOT}]$ is high because of the competing effect arising from hydration of the free ligand. This effect is more pronounced for protons adjacent to ligand exchangeable protons. However, weak affinity ligands are usually not interesting since they are not specific and they may simply interact transiently with the surface of the protein. When screening large libraries, one is more interested in the detection of medium or high-affinity ligands suitable for combinatorial chemistry or medicinal chemistry follow-up.

2.2.7. Competition binding experiments

The Achilles' heel of all the ligand-observed NMR screening techniques is their inability to detect strongly binding ligands with slow dissociation rates. This stems from the fact that the experiments are carried out with a greater than 10-fold excess of ligand over protein. In the crude approximation of a diffusion-limited on-rate (K_{on}) of 10⁸ M⁻¹ s⁻¹ [131], the upper limit of detection, in the most favorable case, is represented by molecules with dissociation binding constants in the 100 nM range. Compounds binding tighter to the protein, compounds that have a slow on rate, and compounds that bind covalently to

B.J. Stockman, C. Dalvit / Progress in Nuclear Magnetic Resonance Spectroscopy 41 (2002) 187-231



Fig. 20. Series of 1D WaterLOGSY ¹H NMR spectra recorded at 600 MHz for (a) 50 μ M 6-CH₃ D,L-Trp, (b) 50 μ M 6-CH₃ D,L-Trp with 5 μ M HSA, (c) 50 μ M 6-CH₃ D,L-Trp with 5 μ M HSA and the three compound mixture of 10 μ M sucrose, 10 μ M 7-CH₃ D,L-Trp and 10 μ M diazepam, (d) 50 μ M 6-CH₃ D,L-Trp with 5 μ M HSA, (e) 50 μ M 6-CH₃ D,L-Trp with 5 μ M HSA and 10 μ M sucrose, (f) 50 μ M 6-CH₃ D,L-Trp with 5 μ M HSA, (e) 50 μ M 6-CH₃ D,L-Trp with 5 μ M HSA and 10 μ M sucrose, (f) 50 μ M 6-CH₃ D,L-Trp with 5 μ M HSA and 10 μ M diazepam. The displayed spectral regions contain the 6-CH₃ signal of the tryptophan derivative. The length of the double spin echo was 25.2 ms in order to destroy most of the protein signals and obtain a flat baseline. Reprinted with permission from Dalvit et al. [132]. © 2002 American Chemical Society.

the protein will not be detected because the residence time of these compounds within the protein is long and therefore the turnover between free and bound state of the ligand is slow. Additionally, compounds with poor solubilities that are potential ligands are difficult to detect since the methods require the observation of the ligand signals.

Competition binding experiments have been used to test the specificity of an identified ligand and for extracting with titration experiments the dissociation binding constant according to Eq. (22). Recently, it has been shown that these experiments, when properly designed, can be utilized efficiently to perform HTS without the drawbacks associated with direct ligand detection [132,133]. In addition, these experiments provide a rapid estimate of the binding constant of the detected ligand.

The HTS with NMR strategy is diagrammed in Fig. 19. The strategy requires first the identification of a

weak affinity ligand and, when possible, its full characterization. In our laboratory, a small library containing a few hundred soluble, well-characterized molecules is first screened using the WaterLOGSY method shown in Fig. 18. The identified binders are subsequently studied with isothermal titrating calorimetry (ITC) in order to determine their binding constants. One of these compounds, based on its binding constant, NMR spectrum, and, when available, on its X-ray structure complexed with the protein, is then selected as the reference compound for the competition binding experiments. Additional preliminary NMR experiments are performed on this molecule in the presence of the protein to optimize the experimental conditions for screening and for quantifying, according to binding strength, the hits originating from the next round of screening. Chemical mixtures are then screened against the target macromolecule in the presence of the reference compound.





Fig. 21. Plot of WaterLOGSY signal attenuation of the reference compound as a function of the dissociation binding constant $K_{\rm I}$ of the competing compound. The simulation was performed using Eq. (23) with a competitor concentration of 50 μ M. The protein and reference compound concentrations were 2 and 50 μ M, respectively. The value of 1 on the *y*-axis corresponds to the signal of the reference compound observed in the absence of the competing compound plus the offset arising from the hydration of the free ligand. The value of 0 on the *y*-axis corresponds, in the approximation of only one protein binding site for the reference compound, to the WaterLOGSY signal of the reference compound in the absence of the protein. Simulations were performed for four reference compound $K_{\rm D}$ values as indicated on the graph.

2.2.7.1. C-WaterLOGSY. The C (Competition)-WaterLOGSY experiments are particularly suited for the detection of high-affinity ligands. In the presence of a competitive inhibitor the proteinbound concentration of the reference compound diminishes [131]. The WaterLOGSY signal intensity ratio for a reference compound in the presence and absence of a competitor is given by the equation [132] reference compound and the competitive ligand, respectively. The equation is valid only in the assumption of a simple competitive mechanism.

A WaterLOGSY spectrum is first acquired for the selected reference compound in the absence of the protein. This is necessary for extracting the hydration correction term discussed above. Then, an identical spectrum is acquired for the selected reference

$$\frac{I_{\text{WLOGSY}}(+)}{I_{\text{WLOGSY}}(-)} = \frac{[\text{E}_{\text{TOT}}] + [\text{L}_{\text{TOT}}] + K_{\text{D}}\left(1 + \frac{[\text{I}]}{K_{\text{I}}}\right) - \sqrt{\left\{[\text{E}_{\text{TOT}}] + [\text{L}_{\text{TOT}}] + K_{\text{D}}\left(1 + \frac{[\text{I}]}{K_{\text{I}}}\right)\right\}^{2} - 4[\text{E}_{\text{TOT}}][\text{L}_{\text{TOT}}]}{[\text{E}_{\text{TOT}}] + [\text{L}_{\text{TOT}}] + K_{\text{D}} - \sqrt{\left\{[\text{E}_{\text{TOT}}] + [\text{L}_{\text{TOT}}] + K_{\text{D}}\right\}^{2} - 4[\text{E}_{\text{TOT}}][\text{L}_{\text{TOT}}]}}$$
(23)

where $I_{WLOGSY}(+)$ and $I_{WLOGSY}(-)$ are the experimentally corrected (i.e. after subtracting the contribution of the hydration of the free ligand) intensity of the reference compound in the presence and absence of the competitive small molecule, respectively. The quantities $[E_{TOT}]$, $[L_{TOT}]$, and [I] are the protein, reference compound, and competitive compound concentrations, respectively. The quantities K_D and K_I are the dissociation binding constants for the compound in the presence of the protein. These two spectra need to be acquired only once. NMR screening experiments with c-WaterLOGSY are then performed on different chemical mixtures against a macromolecular target. Fig. 20 shows this concept with a series of spectra that demonstrate the binding of diazepam to HSA. Fig. 20b indicates that 6-CH₃ D,L-Trp binds to HSA and can be used as the reference compound. WaterLOGSY signal attenuation of the reference

B.J. Stockman, C. Dalvit / Progress in Nuclear Magnetic Resonance Spectroscopy 41 (2002) 187-231



Fig. 22. Expanded region of the 1D ¹H NMR spectra at 600 MHz recorded at different ligand concentrations in the presence of 1.5 μ M p21-activated kinase (PAK). The ligand concentrations used were (a) 40 μ M, (b) 80 μ M, and (c) 140 μ M. Reprinted with permission from Dalvit et al. [133]. © 2002 American Chemical Society.

compound in the presence of a three compound mixture (Fig. 20c) is an indication that one or more compounds comprising the mixture is a potent ligand and displaces the reference compound from the protein. A simple deconvolution (Fig. 20e-g) of the chemical mixture performed with c-WaterLOGSY permits the identification of the active molecule in the mixture.

The calculation of the signal reduction (with the proper correction) and the knowledge of the dissociation binding constant of the reference compound provide an estimation or a lower limit, according to Eq. (23) and graph shown in Fig. 21, of the dissociation binding constant of the identified ligand.

Protein and ligand concentrations as low as 2 and 5 μ M, respectively, can be used. This permits the identification of strong inhibitors that are only marginally soluble. Eq. (23) is not limited to the c-WaterLOGSY experiments. It can also be used for STD-based competition binding experiments by simply replacing $I_{WLOGSY}(+)$ with STD_{af}(+) and $I_{WLOGSY}(-)$ with STD_{af}(-).

A simulation according to Eq. (23) of the Water-LOGSY signal of the reference compound as a function of K_D and K_I permits the proper choice of the reference compound. Therefore it is possible to detect indirectly the presence of a strong inhibitor in a chemical mixture simply by monitoring the WaterLOGSY signal of a reference compound. The lower limit in affinity strength for the detection can be tuned by properly selecting the reference compound (i.e. different K_D) and/or different $[I_{TOT}]/[L_{TOT}]$ ratios according to Eq. (23).

2.2.7.2. Transverse relaxation. A different approach is used with the transverse and selective longitudinal relaxation competition binding experiments [133]. The strategy requires first the acquisition of R_2 or $R_{1,s}$ experiments for the reference molecule in the presence of the protein at different ligand concentrations as shown in Fig. 22, or at different protein concentrations. The latter is preferred when the reference compound is not soluble at high concentrations or when the molecule has a second low-affinity binding site that will start to be partially populated at high concentrations. Broadening is most clearly observed for the ligand resonance at 7.13 ppm. The observed linewidth, the R_2 value measured with CPMG, the signal intensity ratio of a resonance in two R_2 -filtered experiments

recorded with different CPMG spin echo train durations, or, as in the example of Fig. 23, the intensity ratio of one resonance in the fast exchange regime and one in the intermediate exchange regime, is plotted against the ratio $[EL]/[L_{TOT}]$. This ratio can be calculated with the knowledge of the ligand binding constant derived from an ITC measurement and the total protein $[E_{TOT}]$ and ligand $[L_{TOT}]$ concentrations according to the equation:

constant K_D^{app} of the reference compound in the presence of the competing molecule according to the equation:

$$K_{\rm D}^{\rm app} = \frac{[{\rm E}_{\rm TOT}][{\rm L}_{\rm TOT}] - [{\rm E}_{\rm TOT}][{\rm EL}] + [{\rm EL}]^2 - [{\rm L}_{\rm TOT}][{\rm EL}]}{[{\rm EL}]}.$$
(25)

In the assumption of a simple competitive mechanism, the K_D^{app} is then used to extract the binding constant K_I of the competitive ligand

$$\frac{[\text{EL}]}{[\text{L}_{\text{TOT}}]} = \frac{[\text{E}_{\text{TOT}}] + [\text{L}_{\text{TOT}}] + K_{\text{D}} - \sqrt{([\text{E}_{\text{TOT}}] + [\text{L}_{\text{TOT}}] + K_{\text{D}})^2 - 4[\text{E}_{\text{TOT}}][\text{L}_{\text{TOT}}]}{2[\text{L}_{\text{TOT}}]}$$
(24)

NMR screening competition experiments are performed at a fixed protein and reference molecule concentration. The presence in a chemical mixture of a competing molecule, I, will result in a decrease of [EL]. This can be appreciated in Fig. 24b, which shows the NMR spectrum of the reference molecule in the presence of a seven compound chemical mixture. The significant sharpening of the resonance at 7.13 ppm, compared to the corresponding spectrum recorded in the absence of the seven compound mixture shown in Fig. 24a, reveals the presence of a strong binder in the mixture. Deconvolution in the presence of the reference compound allows the identification of the high-affinity ligand as shown in Fig. 24c and d.

At the concentration of ligand employed in the screening, complete displacement of the reference compound is achieved with high-affinity ligands (Fig. 24), and partial displacement is achieved with medium to weak affinity ligands. For deriving the binding constant of the strong inhibitor and not simply a lower limit of its K_{I} , it is sometimes necessary to record another spectrum with lower competing molecule concentration. The measured parameter is then analyzed according to the graph obtained from the titration experiments for the reference compound. Since [L_{TOT}] in the NMR screening experiments is known and fixed, the concentration of [EL] for the reference compound in the presence of the competing molecule can be calculated from the fitting function of Fig. 23. The knowledge of [L_{TOT}], [EL], and [E_{TOT}] permits determination of the apparent dissociation binding

according to the equation [131]

$$K_{\rm I} = \frac{[I]K_{\rm D}}{K_{\rm D}^{\rm app} - K_{\rm D}} \tag{26}$$

where [I] is the concentration of the competing molecule. In our studies, binding constants for two compounds with binding affinities in the μ M to nM range were determined by both NMR and ITC measurements. Although the NMR-derived values were extracted using a single point measurement, there is good agreement with the values obtained by a complete ITC measurement (2.9 μ M compared to 2.5 μ M and 0.15 μ M compared to 0.25 μ M).

Another method that can be used with competition binding experiments is the single or multiple-selective R_2 experiment described in Fig. 25 (top). In this experiment, the same resonance or resonances of the reference molecule are selectively inverted using a shaped 180° pulse. Typically two spectra are recorded, one with a short spin echo period and another with a longer period that is close to the R_2 of the reference molecule in the presence of the protein. The intensity ratios of the signals extracted from the two spectra are used for the titration measurements and for the screening process as described above. For the system used in Fig. 22, it was sufficient to record selective R_2 experiments with a single spin echo period (data not shown). This was possible because of the differential R_2 of the two selectively inverted ligand resonances at 6.65 and 7.13 ppm. The signal intensity ratios of these two resonances extracted from





Fig. 23. Plot of the signal intensity ratio of the two ligand resonances, $I_{7.13}/I_{6.65}$, shown in Fig. 22, as a function of the ratio [EL]/[L_{TOT}]. The resonance at 7.13 ppm is a singlet (one proton) that undergoes significant exchange broadening in the presence of the protein, while the resonance at 6.65 ppm is a doublet (one proton) with a small *J*. The ratio [EL]/[L_{TOT}] was calculated using an ITC-derived K_D value of 7.1 μ M for the ligand. The first point on the left corresponds to the value in the absence of protein. The curve represents the best fit of the experimental points. Similar results were obtained using the signal intensity ratio of the two ligand resonances extracted from multiple-selective R_2 spectra (with selective inversion of the resonances at 7.13 and 6.65 ppm) recorded with the pulse sequence of Fig. 25 (top). Reprinted with permission from Dalvit et al. [133]. © 2002 American Chemical Society.

the selective R_2 experiments was used as described above.

Finally, the 1D selective or multiselective TOCSY or COSY experiments can also be used in association with the competition binding experiments. These experiments are useful in cases of severe overlap. However, it is applicable only if the reference compound contains scalar coupled spin systems. The selective excitation is achieved with the same scheme used in the selective R_2 experiment. For the TOCSY, two experiments with fixed spin-lock periods are recorded, one with $\tau = 0$ and the other with a long τ duration.

2.2.7.3. Selective longitudinal relaxation. Although $R_{1,s}$ experiments have not been used in ligand-based screening because of the problems mentioned earlier, they are particularly well-suited for performing NMR



Fig. 24. NMR screening and deconvolution performed with 50 μ M reference ligand in the presence of 2 μ M PAK. Expanded region of the 1D ¹H NMR spectra at 600 MHz recorded in the (a) absence of any additional compounds, (b) presence of a seven compound mixture with each molecule at 20 μ M, (c) presence of the same chemical mixture but without SU-13901, and (d) presence of only SU-13901. The spectra were acquired with 128 scans and a 2.82 s repetition time. An arrow at 7.13 ppm indicates the resonance of the reference compound that experiences significant broadening in the presence of the protein. Reprinted with permission from Dalvit et al. [133]. © 2002 American Chemical Society.

screening with competition binding experiments [133]. The pulse sequence for these experiments recorded in aqueous solutions is shown in Fig. 25 (bottom). In the experiments it is sufficient to selectively invert the same resonance of the reference molecule for every mixture screened. Because the reference molecule is in excess compared to the protein and the molecule is a weak to medium affinity ligand, the observed chemical shifts of the ligand resonances correspond to the chemical shifts of the free ligand. This permits acquisition of the experiments in automation using the same excitation frequency for the selective inversion. For greater sensitivity of the experiment, the resonance to be selectively inverted should be chosen to be one of the ligand resonances that displays the largest difference





Fig. 25. Pulse sequences used for the selective R_2 (top) and R_1 (bottom) experiments performed in aqueous solution. Selective inversion of single or multiple resonances of the ligand is performed with shaped pulses. A short gradient can be applied in the R_1 filter experiment immediately after the selective 180° pulse in order to destroy magnetization not aligned along the z-axis. Particular care should be taken when selective inversion with shifted laminar pulses is applied simultaneously to multiple ligand resonances. In the R_1 experiments, the ligand resonances simultaneously inverted should correspond to protons that are not close in space. In the R_2 experiments, the resonances to be simultaneously inverted should not be mutually scalar coupled in order to avoid evolution under scalar coupling during the spin echo period (i.e. for a two proton spin system with scalar coupling it will be possible to invert only one of the two protons). Water suppression is performed with the double spin echo scheme [158].

of $R_{1,s}$ in the free and bound states. When possible, a singlet resonance is preferred because of the greater intensity and reduced problems of overlap. It is possible, however, to determine the presence of a ligand in both the R_2 and $R_{1,s}$ experiments even in the presence of overlap. In these particular cases it is also necessary to acquire the same experiments for the mixture in the absence of the reference compound. Subtraction of the spectra recorded in the presence and absence of a ligand.

The calculated selective $R_{1,s}$, the signal intensity ratio of an inverted resonance compared to a noninverted resonance in an $R_{1,s}$ experiment recorded with a single τ value, or the signal intensity ratio of an inverted resonance in two $R_{1,s}$ experiments recorded with different τ values, for the reference molecule is plotted as a function of the ratio [EL]/[L_{TOT}] in the same way as for R_2 . The titration experiments are performed either by keeping the protein concentration



Fig. 26. Plot of the signal intensity ratio of the inverted resonance (Trp C2-H) in two $R_{1,s}$ filtered experiments recorded with $\tau = 1.91$ s and $\tau = 0.48$ s (top) and plot of the signal intensity ratio for the Trp C2-H and C4-H resonances in an $R_{1,s}$ filtered experiment (inversion of the C2-H resonance) recorded with $\tau = 1.91$ s versus the ratio [EL]/[L_{TOT}]. The L-Trp concentration was kept constant at 100 μ M and the protein (HSA) concentration was varied from 0 to 12 μ M. For the top diagram the magnetization is positive with $\tau = 1.91$ s and negative with $\tau = 0.48$ s. The concentration ratio [EL]/[L_{TOT}] was calculated using the equilibrium dialysis-derived $K_{\rm D}$ value of 23.0 μ M for Trp measured at 20 °C [159]. The first point on the left corresponds to the value in the absence of protein. The curves represent the best fits of the experimental points.

fixed and varying the reference compound or by keeping the ligand concentration fixed and varying the protein concentration. This is shown in the example of L-Trp in the presence of HSA in Fig. 26 where the L-Trp concentration was kept constant and the protein concentration was varied. Screening is performed using an $R_{1,s}$ experiment recorded with either a single or two τ values. We refer to these experiments as the





ողոպոպոպոպոպոպոտ 7.365 ppm

Fig. 27. NMR screening and deconvolution performed with a 1D 600 MHz ¹H $R_{1,s}$ filtered experiment for 100 μ M Trp in the presence of 8 μ M HSA. Expanded spectral regions containing only the selectively inverted Trp C2-H resonance recorded in the (a) absence of any additional compounds, (b) presence of a four compound mixture with each molecule at 30 μ M, (c) presence of the same chemical mixture but without diazepam, (d) presence of only diazepam, and (e) absence of HSA and any additional compounds. The spectra were acquired with 128 scans and a 5.82 s repetition time. The τ value was 0.955 s, which corresponds closely to the null point for the spectrum in (a). Reprinted with permission from Dalvit et al. [133]. © 2002 American Chemical Society.

 $R_{1,s}$ filtered experiments. For maximum sensitivity, this delay should correspond to the τ value at which the largest intensity difference is observed between the $R_{1,s}$ measurements for the reference compound in the absence and presence of the macromolecular target. For rapid visual inspection, the τ value to the corresponding null point (i.e. $1 - 2 \exp(-\tau R_{1,s}) = 0)$ can also be selected as shown in Fig. 27a. The presence in the chemical mixture of a competing molecule will result in an $R_{1,s}$ filtered spectrum with a residual negative signal (Fig. 27b). The $R_{1,s}$ of the reference compound becomes smaller because of partial displacement of the reference compound from the protein. Deconvolution of the mixture carried out with the same selective $R_{1,s}$ filtered experiment identifies the competitive inhibitor (Fig. 27d). A measurement of the $R_{1,s}$, the signal intensity ratio of an inverted resonance compared to a non-inverted resonance in an $R_{1,s}$ filtered experiment recorded with a single τ value, or the signal intensity ratio of an inverted resonance in two $R_{1,s}$ filtered experiments recorded with different τ values, for the

reference molecule in the presence of the competing molecule permits, according to the diagram of Fig. 26 and following the same procedure described for the transverse relaxation experiment, an estimate of the binding constant of the competing molecule [133].

When problems of overlap are encountered it is possible to use the $R_{1,s}$ filtered selective TOCSY experiment. However, this method can be applied only if the inverted resonance is a multiplet.

A drawback of the $R_{1,s}$ competition binding experiments is the detection of only molecules competing with the reference molecule for the same protein binding site (in the absence of allosteric effects). The other two competition binding techniques, R_2 and c-WaterLOGSY, permit the simultaneous detection of weak to medium affinity ligands binding at different sites. This is clearly possible if the concentration of the compounds comprising the mixture is comparable to the concentration of the reference molecule. It is also conceivable to combine the $R_{1,s}$ and R_2 schemes into a single experiment. This could be achieved by inserting a CPMG period into the pulse sequence of Fig. 25 before the double spin echo portion.

The competition binding experiments for performing HTS and for deriving the binding constants of the NMR hits [133] are clearly not limited to ¹H detection. When the reference molecule contains a fluorine atom, it is possible to perform exactly the same competition binding experiments using ¹⁹F NMR [134]. The longitudinal relaxation of ¹⁹F is not a good parameter for the competition binding experiments since it lacks the direct τ_c dependence needed for identifying small molecules interacting with a macromolecule. However, ¹⁹F transverse relaxation represents an excellent parameter for the competition binding experiments since it contains spectral densities calculated at zero frequency for the heteronuclear ¹⁹F-¹H dipolar interactions and for the ¹⁹F CSA interaction. Since the CSA of ¹⁹F is large, the latter spectral density will contribute significantly to the transverse relaxation of the fraction of bound ligand [135]. CSA contribution to relaxation is directly proportional to the square of the magnetic field. Therefore, the effect is more pronounced at higher magnetic fields. The titration experiments for the reference molecule as a

function of the fraction of bound ligand are performed by either acquiring 1D 19 F R_2 filtered experiments or simply 1D 19 F experiments with ¹H decoupling. Screening is then carried out by monitoring changes in the transverse relaxation (either via the R_2 filtered experiments or via analysis of the linewidth) of the 19 F signal of the reference molecule as described above for the ¹H experiments [134].

The competition binding experiments can also be carried out with the use of two compounds, one representing the reference molecule and the other representing a small molecule that does not interact with the receptor. The latter serves as an internal standard that is used for quantitating, with a single experiment, changes in the signal of the reference compound [134].

Approaches that are based on properly designed competition binding experiments permit rapid screening of thousands of compounds against protein or DNA and RNA fragments in a short period of time. The methodology can also be applied to the screening of plant and fungi extracts. As can be appreciated from Figs. 23 and 26, the quantity of protein required for screening is small and protein concentrations as low as 500 nM to 1 µM can be used. The use of cryoprobe technology will further speed up the screening process. It is expected that in this case the limiting step will be represented simply by the time required to change the sample, to equilibrate the temperature and to shim each sample in the magnet. Indeed, for the simple 1D ¹H NMR experiments used in the transverse relaxation with competition binding, mixtures of 10 compounds can be screened in a total time of less than two minutes per mixture. This translates into ~ 7000 compounds per day. If mixtures of up to 100 compounds were used, the throughput would exceed \sim 70,000 compounds per day. Since the concentration of each compound in the mixture to be screened is typically less than 10 µM, mixtures containing 100 compounds will only have a total small molecule concentration of 1 mM. The low compound concentrations used by these techniques permits screening of compounds with poor solubility. This increases the chemical diversity of compounds amenable to NMR screening.

3. Design of small molecule screening libraries

221

3.1. General considerations

A common feature of all NMR screening paradigms is the need for a collection of small molecules to comprise a screening library. The greater amounts of both target molecules and screening compounds needed compared to traditional HTS limits the number of compounds that can be efficiently screened. As a result, published NMR screens have used compound libraries containing typically less than $\sim 10,000$ compounds. Although the recently introduced competition binding NMR screening methods [133] may provide for efficient screening of a much greater number of compounds than this, if one could find suitable hits from a smaller collection of wellchosen compounds it may not be necessary to expend the time and chemical resources to screen an entire corporate compound library against every single target. Hits so identified could then be used to focus further screening efforts or to direct combinatorial syntheses, saving both time and chemical resources [8].

Library design for NMR-based screening has recently been reviewed, including both primary screening and follow-up libraries [136]. Certain key aspects of primary screening libraries will be discussed here. In general terms, several types of libraries can be envisioned: broad screening libraries applicable to many types of target proteins; directed libraries that are designed with the common features of an active site in mind that might be useful for screening a series of targets from the same protein class, such as protease enzymes; and 'functional genomics' libraries composed of known substrates, cofactors and inhibitors for a diverse array of enzymes that might be useful for defining the function of genomics-identified targets. Rather than just randomly choosing compounds for a broad screening library, several rationale approaches have been implemented. These include diversity libraries designed using various chemical informatics tools and the SHAPES library developed by Fejzo and coworkers that is composed largely of molecules that represent frameworks commonly found in known drug molecules [12].

Having a choice for which molecules to include in

a screening library is a luxury that should not be wasted. In addition to structural diversity, there are other important characteristics to consider when selecting the subset molecules. These include purity, identity, reactivity, toxicological properties, molecular weight, water solubility, and suitability for chemical elaboration by traditional or combinatorial chemistry methods. It makes sense to populate the screening library with compounds of high integrity that are not destined for failure down the road. Time spent upfront to insure purity, stability, and identity with LC-MS or LC-NMR analyses will save resources downstream. Filtering tools should be used to avoid compounds that are known to be highly reactive, toxic or to have poor metabolic properties. Lack of reactivity is critical since compounds can be screened more efficiently as mixtures. Most labs typically pool their selected small molecules into mixtures of up to 10 compounds for screening.

Compounds for diversity-based NMR screening libraries can be selected based on their drug-like character. Using this protocol, which selects compounds that have similar characteristics as known drugs, hits are predisposed to have desirable pharmacokinetic characteristics [136]. Compounds with drug-like characteristics can also be selected based on scaffolds that occur frequently in known drugs. This approach, known as the SHAPES method, is described in detail below. Since primary NMR screens typically identify low-affinity binders, diversity libraries that consist of lead-like as opposed to drug-like molecules [137] may actually be best. It is often the case that chemical elaborations to improve affinity also increase molecular weight and decrease solubility [137]. The molecular weight of the compounds therefore should not exceed 350. Since most hits obtained will have affinities for their target in the $\sim 100 \,\mu\text{M}$ range, low molecular weight will leave room for chemical elaboration to build in more affinity and selectivity. Using larger molecular weight drug-like compounds would not substantially improve affinity of the hits and could easily preclude obtaining lead chemical templates of reasonable size. By contrast, lead-like hits that are reasonably water soluble allow for chemical elaboration that results in modest increased lipophilicity of the final therapeutic entity [137]. Water solubility is also important since it enhances the potential success of downstream studies such as calorimetry, enzymology, co-crystallization and NMR structural studies. Compound solubility is especially important for flow-injection NMR methods in order to prevent clogging of the capillary lines [8].

Compounds should also be chosen with their suitability for chemical elaboration by traditional or combinatorial chemistry methods in mind. Hits with facile handles for synthetic chemistry will be of more interest and will allow more efficient use of medicinal chemistry resources. An example of this is the 'SHAPES linking library' which consists of compounds containing scaffolds along with linkers and side chains that are accessible using reactions implemented by the combinatorial chemistry group at Vertex [136]. Libraries constructed using the retrosynthetic combinatorial analysis procedure (RECAP) offer a similar approach [138]. Databases of biologically active molecules are fragmented around bonds formed by common combinatorial chemistry reactions into their biologically relevant or privileged motifs. The compounds thus include a combinatorial handle that allows rapid expansion around a given motif once it is identified as a weak affinity ligand. Finally, input from medicinal chemists regarding which compounds to include is critical. From an extreme perspective, favorite compounds suggested by chemists should be included in the screening library.

3.2. SHAPES screening library: design and application

The SHAPES strategy is an NMR-based approach to lead generation that includes the rational design of a compound library, collection and analysis of NMR screening data sets, and follow-up screening and/or medicinal chemistry [12]. Unlabeled target molecules and library compounds are used. The SHAPES library is a small collection of diverse, low molecular weight, water soluble compounds whose molecular shapes



Fig. 28. Molecular frameworks used for selecting compounds in the original SHAPES library. Sidechain attachment points are denoted by single electrons or lone pairs. X represents a C, N, O or S atom. Reprinted with permission from Fejzo et al. [12]. © 1999 Elsevier Science.

represent those most commonly found in known drug molecules [139]. From an analysis of the comprehensive medicinal chemistry (CMC) database, it was determined that 32 different frameworks describe $\sim 50\%$ of all known drugs [139]. When atom type was included, 41 frameworks describe 24% of all known drugs. Framework classification was combined with similar data on the most common drug side chains [140] to create the SHAPES library. The 132 compounds in the original SHAPES library are based on the molecular frameworks shown in Fig. 28 [12]. All compounds were commercially available, soluble, non-aggregated at 1 mM, chemically and isomerically pure, and non-reactive. It should also be pointed out that many frameworks from the CMC analysis were NOT incorporated into the SHAPES library because of either synthetic complexity or poor solubility. The molecular weight range of the selected compounds is 68-341 Da. Compounds were also required to have at least two protons within 5 Å apart in order to be amenable to

transfer NOE-based screening. For screening, the compounds were pooled into mixtures of 1-4 compounds. A subsequent SHAPES library contained ~500 compounds, each with a combichemaccessible linker or side chain [136].

Two screening methods are used in the SHAPES approach: 1D¹H line broadening and 2D transfer NOE. Both methods rely on fast exchange between the free and bound species of any ligand present. Line broadening methods, like relaxation-editing, have the disadvantage that chemical shift differences between the free and bound forms can complicate the analysis. Nevertheless, Fig. 29 [12] shows the line broadening observed for a ligand in the presence of the 224 kDa protein inosine-5'-monophosphate dehydrogenase (IMPDH). SHAPES screens are typically collected at concentrations of 50 µM protein and 1 mM small molecule. The total time to screen the SHAPES library using 2D transfer NOE spectra is only a few days. Since neither method of binding detection can distinguish between specific and non-specific binding, competition experiments or enzymatic assays must be carried out to





Fig. 29. 1D 800 MHz ¹H NMR spectra of a small molecule free in solution (bottom spectrum) and in the presence of inosine-5'-monophosphate dehydrogenase (top spectrum). Reprinted with permission from Fejzo et al. [12]. © 1999 Elsevier Science.

confirm identified hits. If a known high-affinity ligand exists for the target of interest, repeating the binding experiment in its presence should eliminate the binding observed for the weaker affinity SHAPES ligand. In cases where more than one SHAPES hit is identified, it can be quite useful to rank order their affinities. A straightforward manner to accomplish this without the need for time-intensive titration experiments is with NMR diffusion measurements [12,141].

Screening hits from the SHAPES library are used to direct further screening by choosing existing proprietary or acquirable compounds to be assayed enzymatically, choosing compounds from virtual screens to be acquired and assayed, or suggesting combinatorial or medicinal chemistry elaborations. If there is more than one SHAPES hit that can be confirmed to bind at the same location, limited SAR may already be apparent to select or synthesize the next set of compounds to screen. The SHAPES hits can also serve as input for substructure searches. The core of the SHAPES hit is held reasonably constant but is extended by addition of common rings and linkers. The greater number of potential interactions with the target will lead to some of the resulting larger molecules having increased affinity compared to the original NMR hit. In six SHAPES screens carried out at Vertex [12], hit rates of 10-20% have been observed. Of these hits, 30-40% were subsequently determined to be true inhibitors with IC₅₀ values ranging from 150 µM to 5 mM. When follow-up libraries of 100-300 compounds were assayed enzymatically at 30 μ M, the hit rate was 5–6%. This represented a 3-5 fold enhancement compared to non-SHAPES compounds in the same assay. Structures of several of the most potent hits complexed to their target were determined by X-ray crystallography. In a retrospective analysis of HTS data, it was determined that compounds containing scaffolds from SHAPES hits had a 2.5-4 fold higher hit rate at the 30% inhibition level and an 8-10 fold higher hit rate at 50% inhibition. The hit-enrichment observed suggests that SHAPES-directed screening can save substantial time and chemical resources in high throughput assays.

SHAPES hits can also lead directly into ligand design. For example [12], SHAPES hits (39) and (40) have low mM affinities for p38 MAP kinase. Synthesized compounds (41) and (42) have affinities in the $\sim 200 \,\mu\text{M}$ range. Consolidating (41) and (42) into (43) resulted in an inhibitor with a $K_{\rm I}$ value of 200 nM. SHAPES hits can also be linked together in a similar manner as in the SAR by NMR method if two SHAPES hits give rise to interligand NOEs in the transfer NOE spectrum. Provided that the interligand NOEs observed arise from direct magnetization transfer, the two ligands are in close proximity and can be chemically linked together. Much like the SAR by NMR method, one SHAPES hit could be used to saturate the target and re-screen for a second site SHAPES hit. The best SHAPES success story to date has been with JNK3 MAP kinase [142]. No leads were found in the HTS assay, but 17 SHAPES hits were identified. Based on the SHAPES hits, about 100 compounds were identified for follow-up assays. Of these, eight had IC_{50} values less than 20 μ M. Subsequent medicinal chemistry efforts resulted in lead compounds with 1 μ M and 3 nM potency, with an X-ray crystallographic structure of at least one



complex now determined.



Documented applications of the SHAPES approach have been limited to soluble proteins. Since only ligand signals are used to monitor binding, the same methodology should be just as applicable to membrane-bound proteins in micelles or detergents, nucleic acids, subcellular structures, and targets cross-linked to a solid support [12]. SHAPES screening of a structured RNA target using the WaterLOGSY technique has been reported [142]. The SHAPES strategy offers several advantages compared to the SAR by NMR method: isotopically enriched protein is not required, proteins of any molecular weight are suitable (in fact, larger is better), and the structure or identity of the target is not required. The main disadvantage of the SHAPES strategy compared to SAR by NMR is that no information about the ligand binding site location or protein-ligand interactions is directly obtained. As the SHAPES inventors note, the most important advantage of SHAPES is its applicability to very large proteins [12]. If the majority of drug targets turn out to have molecular weights greater than 30 kDa, this will be an important distinction for NMR labs involved in drug discovery research. Experimentally derived data from SHAPES screening can leverage computational methods such as virtual screening, making the latter more information-based [12]. The synergistic use of SHAPES, computational chemistry and medicinal chemistry provides for the most comprehensive use of each individual technique.

4. Impacting chemistry and biology

4.1. Privileged molecules

The Abbott group has screened a sufficiently large

number of targets that a retrospective analysis identified certain motifs, termed privileged molecules, that are preferred for protein binding [34]. This statistical analysis involved screening results for 11 protein targets against a diversity-based screening library of 10,080 compounds having an average molecular weight of 200 Da. The screening library compounds were first broken down into 104 fragments using a modification of the RECAP algorithm [138]. Of these fragments, a total of 12 were found to bind with significantly higher frequency to at least one of the target proteins. In particular, the carboxylic acid (44), biphenyl (45) and diphenylmethane (46) substructures were found to preferentially bind to six, five and three of the target proteins, respectively. Most of the protein targets that preferentially bind to the carboxylic acid motif have DNA, RNA or phosphorylated peptides as their natural substrates. Chemical shift analyses indicated that the carboxylic acid ligands do in fact bind in the substrate binding site, suggesting that the negative charge of the carboxylic acid mimics that of the natural substrate. By contrast, preferential biphenyl binding was observed for a diverse set of protein targets. In each case, chemical shift analyses indicated binding of the biphenyl ligands at or near the natural substrate site. Interestingly, the same set of biphenyl compounds did not preferentially bind to each of the five proteins. Instead, specific and distinct biphenyl SAR was observed. This indicates that the biphenyl motif provides an excellent starting point for both high affinity and specificity for a wide range of protein targets [34].

225



4.2. NMR screening as a secondary assay

In addition to their role as a primary screening assay, NMR-based binding experiments can be very useful as a secondary assay. Hits resulting from completion of a traditional high throughput screen are often triaged in some manner before consideration as lead chemical templates. For instance, if the HTS assay is multicomponent, one would like to know

whether the hits actually bind to the target of interest. NMR-binding assays using the isolated target protein and the HTS hits can provide this distinction. Moreover, if the target has more than one binding site, such as for substrate and cofactor, NMR-binding assays can be used to distinguish between them. Simple yes/no answers from secondary NMR-binding assays can be an important aspect of the triage funnel. NMR-derived $K_{\rm D}$ values can also be used to rank order HTS hits. For enzyme targets, this data can corroborate $K_{\rm I}$ values. For non-enzyme targets or for enzyme targets difficult to assay, this data may not be obtainable using other methods. Identification of novel DNA gyrase inhibitors is an excellent example of where NMR was used in a secondary assay sense [143]. In this study, HTS hits were screened using ¹H-¹⁵N HSQC spectra to test not only for binding, but also for binding at the targeted ATP binding site. Several compound classes were ruled out because of lack of binding to the ATP site. Selected hits from remaining compound classes were then rank ordered by affinity using NMR-derived K_D values. NMR screening can also be used to validate or complement virtual screening results. For example, compounds selected in a virtual screen can be tested for binding to the target. Compounds that pass the NMR-binding test can then be used to direct similarity or substructure searches to identify a larger set of compounds for testing in HTS assays. It has been demonstrated using FKBP as the test protein that NMR-derived K_D values can be used to validate virtual screening scoring functions [144]. A good correlation between scoring rank and K_D value validates the virtual screen. For a novel target, it may be beneficial to test 5-10% of the virtual screening hits (over a wide scoring range) in an NMR-based binding assay in order to validate the virtual screen prior to acquiring any of the identified compounds for HTS assay.

4.3. Selecting for minimal HSA binding

NMR-binding assays have recently been introduced for rational design of compounds that retain their desired potency but have minimal HSA binding propensity [145]. HSA binds to a wide range of compounds, and, since it is the most abundant protein in serum plasma, it can drastically affect the distribution and efficacy of pharmacological molecules [146]. In this pioneering work, the binding sites of diflunisal on HSA were determined using tripleresonance NMR methods in conjunction with the HSA X-ray crystal structure [147]. Based on this structure, derivatives of diflunisal with anticipated reduced HSA affinities were synthesized. NMRderived dissociation constants for these compounds were measured using chemical shift titrations. Compounds that retained good activity but had reduced HSA binding were identified. Often times, one has an assortment of compounds with nearly equal potency. Competition binding experiments [132–134] can also be used as an efficient and rapid method to rank order the compounds of a given class for HSA binding affinity. This information would be highly valuable in drug development.

4.4. Genomics-derived proteins of unknown function

In today's era of high throughput genome sequencing, complete genomes of tens of organisms have already been sequenced and work on hundreds more is in progress.¹ This has led to identification of thousands of new proteins. However, the potential of these proteins as drug targets cannot be fully assessed without knowledge regarding their function and importance in biological processes. It may be possible to infer function by determining the structures of these proteins. It may also be possible to infer the function of these proteins based upon what types of small molecules that they bind. Specific interactions between macromolecules and smaller molecular weight ligands are important in all biochemical processes. Enzymes require specific binding of cofactors and/or substrates to carry out the reactions that they catalyze. Inhibitors are designed to specifically bind enzymes and receptors in or around the active site, and they often are analogous to substrates or cofactors. A functional genomics library of about 200 small molecule ligands that consists of known cofactors, substrates, substrate analogs and inhibitors of proteins of known function could prove quite useful for NMR screening against proteins of unknown function [148]. Binding of one or more compounds to the unknown function protein may provide clues about its functional characteristics. A similar

¹ http://wit.integratedgenomics.com/GOLD/.

approach has been developed for use with the ThermoFluor® screening technology [149]. The potential for this type of screening to impact biology has been demonstrated for the unknown function protein HI0719 from *Haemophilus influenzae* [150]. The structure of this protein provided no clues to its function. However, biological evidence suggested that it might be involved in regulation of isoleucine biosynthesis and related pathways. NMR screening of small molecules in these pathways, as well as commonly occurring small molecules in cells, provided functional insights. From the ¹H–¹⁵N HSQC screens, α -ketobutyrate and several analogs were identified as ligands [150].

5. Prospects for automation

As NMR screening methods become more prevalent in the pharmaceutical industry, it is highly desirable to automate as much of the process as possible. Ross and Senn have recently parsed NMR screening into seven steps for automation: just-in-time sample preparation, transfer of the sample to the magnet, preparation for data collection, data collection, sample recovery, data processing, and data analysis [9]. Just-in-time sample preparation is important to prevent any time-dependent changes in small molecules (aggregation, chemical reactions) or target molecule (precipitation, loss of activity) that might occur after mixing the small molecule and target solutions but prior to data collection. Methods to accomplish this include using a Gilson liquids handler coupled via capillary tubing to a flowinjection probe [8] or a Genesis sample-handling robot to mix the samples in an NMR tube [9]. Transfer of the just-mixed sample to the magnet is accomplished by either a capillary line leading to a flow-injection probe or by a robotic sample arm in the case of individual NMR tubes. Once the sample is in place and sufficient time has elapsed for temperature equilibration, automated locking and shimming are carried out in advance of data collection. Then, a simple 1D ¹H spectrum can be acquired to provide a reference spectrum to monitor shimming, sample integrity and even pH. This is followed by automated collection of the

desired 1D or 2D screening data sets. After data acquisition, sample is recovered via the capillary lines or robotic transfer. This cycle can be repeated until the entire screening library has been completed. Automated data processing can be accomplished with a variety of software packages. These tend to work well on 2D spectra, but 1D spectra typically require manual intervention because of the large solvent resonance [9]. Automated data analysis of 1D spectra works best for methods where binding is detected by appearance of a ligand signal, such as in the WaterLOGSY experiment or the competition binding experiments described above. For 2D spectra, several methods of automated analysis have been proven effective [9,151].

A series of 2D ¹H-¹⁵N HSQC spectra can be automatically compared for chemical shift perturbations indicative of ligand binding using several approaches. These include non-statistical methods based on scalar products of data vectors, statistical methods based on principal component analysis, and methods based on peak matching [9]. The first two approaches were applied to a series of ~ 300 2D ¹H⁻¹⁵N HSQC spectra, each collected on an ¹⁵Nlabeled protein in the presence of a single, different test compound [151]. Bucketing calculations, which divide spectra into small regions termed buckets and extract the information content, were used for data reduction. A table ranked by the correlation coefficients was generated, but no clear correlations were observed using this non-statistical method. Subsequently, integration patterns for all 300 small molecule spectra were analyzed to generate a data matrix of N integration regions times 300. A statistical software package was then used to analyze this data matrix using principal component analysis. Of the two classes of spectral changes that emerged, one was found to correspond to pH changes caused by certain small molecules while the other corresponded to small molecules binding to the target protein [151].

6. Future directions

Continued advances in instrumentation and methodologies will lead to faster throughput and to the application of NMR screening to ever larger systems.

Screening proteins in living cells will also provide novel information for drug discovery.

The potential for cryoprobe technology to increase throughput [42] has already been discussed. Throughput can also be increased by simultaneous data collection on multiple samples. This has been demonstrated using 19 capillary lines formed into a bundle that is 5 mm in diameter [9]. Each capillary contains target and a different screening library mixture. NMR screening data sets are collected for each capillary independently using chemical shift imaging-based methods [152]. Increased throughput could also be achieved by using a recently described four-coil, flow-through multiplex sample NMR probe [153].

The SEA-TROSY (solvent exposed amides with TROSY) method may extend the applicability of ¹⁵Nbased SAR by NMR to much higher molecular weight proteins [154]. In this experiment, only backbone amide groups that are in fast exchange with solvent are observed. Even for a very large protein, this selected subset of ${}^{1}H{-}^{15}N$ resonances results in significant spectral simplification. Since most protein–ligand interactions involve surface residues, this subset of amide groups is sufficient to detect binding interactions. In favorable cases, ligand binding sites may be located by combining the SEA element with triple resonance TROSY-type experiments to sequentially assign stretches of exchangeable amide groups [154].

The recently described nuclear magnetic resonance docking of compounds (NMR-DOC) method may also extend NMR screening to larger systems [155]. In this strategy, protein is selectively labeled with ${}^{13}C^{\epsilon}/{}^{1}H$ methionine, ${}^{13}C^{\delta}/{}^{1}H$ isoleucine and ${}^{13}C/{}^{1}H$ threonine in an otherwise deuterated background. Perturbations of the observed ¹H-¹³C correlations with known ligands or inhibitors are used to assign active site resonances. New ligands are then identified using saturation transfer methods. Since the majority of the protein is deuterated, saturation transfer is quite specific for ligand binding near the protonated residues. Identified ligands can then be docked to the protein via transfer NOEs to the protonated residues. The related method, nuclear magnetic resonance structurally oriented library valency engineering, (NMR-SOLVE) has the potential to design inhibitors, even in the absence of target structural information, for targets that possess two adjacent ligand binding pockets [155]. Identification of NOEs between a target interface residue and two ligands, one on each side of the interface, indicates that the two ligands could potentially be linked to form a single, high-affinity ligand. Ligand/ligand orientation information could also be used to explore the two binding pockets with combinatorial chemistry.

The SAR by NMR methodology may also be applicable to proteins within living cells [156]. This was demonstrated by selective ¹⁵N-labeling and overexpression of the N-terminal metal-binding domain of MerA in *Escherichia coli* cells. A 2D $^{1}\text{H}-^{15}\text{N}$ HSQC spectrum with good resolution could be acquired in 10 min on a 15% bacterial slurry. An attractive feature to this method is that proteins can be studied in the presence of other proteins and endogenous small molecules, which represents more physiologically correct conditions. The authors propose using this method to study membrane permeability and target protein affinity of potential drugs [156].

Acknowledgements

We thank Ad Bax, Yiqing Feng, and Robert Konrat for reviewing the manuscript; Daniel Sem, Johan Weigelt, and Daniel Wyss for providing preprints of their manuscripts; and Philip Hajduk, Wolfgang Jahnke, Bernd Meyer, Jonathan Moore, and Michael Shapiro for contributing figures from their published works.

References

- [1] B.J. Stockman, Prog. NMR Spectrosc. 33 (1998) 109.
- [2] S.B. Shuker, P.J. Hajduk, R.P. Meadows, S.W. Fesik, Science (Washington, DC) 274 (1996) 1531.
- [3] M.J. Shapiro, J.R. Wareing, Curr. Opin. Drug Discov. Dev. 2 (1999) 396.
- [4] P.J. Hajduk, R.P. Meadows, S.W. Fesik, Q. Rev. Biophys. 32 (1999) 211.
- [5] J.M. Moore, Curr. Opin. Biotechnol. 10 (1999) 54.
- [6] J.M. Moore, Biopolymers (Pept. Sci.) 51 (1999) 221.
- [7] J.W. Peng, C.A. Lepre, J. Fejzo, N. Abdul-Manan, J.M. Moore, Methods Enzymol. 338 (2001) 202.
- [8] B.J. Stockman, K.A. Farley, D.T. Angwin, Methods Enzymol. 338 (2001) 230.

- [9] A. Ross, H. Senn, Drug Discov. Today 6 (2001) 583.
- [10] T. Diercks, M. Coles, H. Kessler, Curr. Opin. Chem. Biol. 5 (2001) 285.
- [11] D.S. Sem, M. Pellecchia, Curr. Opin. Drug Discov. Dev. 4 (2001) 479.
- [12] J. Fejzo, C.A. Lepre, J.W. Peng, G.W. Bemis, Ajay, M.A. Murcko, J.M. Moore, Chem. Biol. 6 (1999) 755.
- [13] P.J. Hajduk, S. Boyd, D. Nettesheim, V. Nienaber, J. Severin, R. Smith, D. Davidson, T. Rockway, S.W. Fesik, J. Med. Chem. 43 (2000) 3862.
- [14] P.J. Hajduk, G. Sheppard, D.G. Nettesheim, E.T. Olejniczak, S.B. Shuker, R.P. Meadows, D.H. Steinman, G.M. Carrera Jr., P.A. Marcotte, J. Severin, K. Walter, H. Smith, E. Gubbins, R. Simmer, T.F. Holzman, D.W. Morgan, S.K. Davidsen, J.B. Summers, S.W. Fesik, J. Am. Chem. Soc. 119 (1997) 5818.
- [15] S.W. Fesik, P.J. Hajduk, US Patent 5698401.
- [16] J. Liu, J.D. Farmer Jr., W.S. Lane, J. Friedman, I. Weissman, S.L. Schreiber, Cell 66 (1991) 807.
- [17] J.W. Skiles, N.C. Gonnella, A.Y. Jeng, Curr. Med. Chem. 8 (2001) 425.
- [18] E.T. Olejniczak, P.J. Hajduk, P.A. Marcotte, D.G. Nettesheim, R.P. Meadows, R. Edalji, T.F. Holzman, S.W. Fesik, J. Am. Chem. Soc. 119 (1997) 5828.
- [19] W.C. Phelps, J.A. Barnes, D.C. Lobe, Antivir. Chem. Chemother. 9 (1998) 359.
- [20] P.J. Hajduk, J. Dinges, G.F. Miknis, M. Merlock, T. Middleton, D.J. Kempf, D.A. Egan, K.A. Walter, T.S. Robins, S.B. Shuker, T.F. Holzman, S.W. Fesik, J. Med. Chem. 40 (1997) 3144.
- [21] J.S. Brugge, Science (Washington, DC) 260 (1993) 918.
- [22] P.J. Hajduk, M.-M. Zhou, S.W. Fesik, Bioorg. Med. Chem. Lett. 9 (1999) 2403.
- [23] P.J. Hajduk, J. Dinges, J.M. Schkeryantz, D. Janowick, M. Kaminski, M. Tufano, D.J. Augeri, A. Petros, V. Nienaber, P. Zhong, R. Hammond, M. Coen, B. Beutel, L. Katz, S.W. Fesik, J. Med. Chem. 42 (1999) 3852.
- [24] E.A. Eady, J.I. Ross, J.H. Cove, Antimicrob. Chemother. 26 (1990) 461.
- [25] M.J. Duffy, Fibrinolysis 7 (1993) 295.
- [26] M. Williams, E. Kowaluk, S. Arneric, J. Med. Chem. 42 (1999) 1481.
- [27] P.J. Hajduk, A. Gomtsyan, S. Didomenico, M. Cowart, E.K. Bayburt, L. Solomon, J. Severin, R. Smith, K. Walter, T.F. Holzman, A. Stewart, S. McGaraughty, M.F. Jarvis, E.A. Kowaluk, S.W. Fesik, J. Med. Chem. 43 (2000) 4781.
- [28] C.G. Gahmberg, Curr. Opin. Cell. Biol. 9 (1997) 643.
- [29] T.M. Carlos, J.M. Harlan, Blood 84 (1994) 2068.
- [30] J.E. Sligh, C.M. Ballantyne, S.S. Rich, H.K. Hawkins, C.W. Smith, A. Bradley, A.L. Beaudet, Proc. Natl Acad. Sci. USA 90 (1993) 8529.
- [31] G. Liu, J.T. Link, Z. Pei, E.B. Reilly, S. Leitza, B. Nguyen, K.C. Marsh, G.F. Okasinski, T.W. von Geldern, M. Ormes, K. Fowler, M. Gallatin, J. Med. Chem. 43 (2000) 4025.
- [32] G. Liu, J.R. Huth, E.T. Olejniczak, R. Mendoza, P. DeVries, S. Leitza, E.B. Reilly, G.F. Okasinski, S.W. Fesik, T.W. von Geldern, J. Med. Chem. 44 (2001) 1202.

- [33] J. Kallen, K. Welzenbach, P. Ramage, D. Geyl, R. Kriwacki, G. Legge, S. Cottens, G. Weitz-Schmidt, U. Hommel, J. Mol. Biol. 292 (1999) 1.
- [34] P.J. Hajduk, M. Bures, J. Praestgaard, S.W. Fesik, J. Med. Chem. 43 (2000) 3443.
- [35] M. Maurer, D. Reichert, B. Kutscher, G. Quinkert, C. Griesinger, XIX International Conference on Magnetic Resonance in Biological Systems, 2000.
- [36] W.J. Fairbrother, M.A. Champe, H.W. Christinger, B.A. Keyt, M.A. Starovasnik, Protein Sci. 6 (1997) 2250.
- [37] S.A. Wolfe, P. Zhou, V. Dotsch, L. Chen, A. You, S.N. Ho, G.R. Crabtree, G. Wagner, G.L. Verdine, Nature 385 (1997) 172.
- [38] R. Ranganathan, K.P. Lu, T. Hunter, J.P. Noel, Cell 89 (1997) 875.
- [39] M. Sattler, H. Liang, D. Nettesheim, R.P. Meadows, J.E. Harlan, M. Eberstadt, H.S. Yoon, S.B. Shuker, B.S. Chang, A.J. Minn, C.B. Thompson, S.W. Fesik, Science (Washington, DC) 275 (1997) 983.
- [40] A.M. Petros, A. Medek, D.G. Nettesheim, D.H. Kim, H.S. Yoon, K. Swift, E.D. Matayoshi, T. Oltersdorf, S.W. Fesik, Proc. Natl Acad. Sci. USA 98 (2001) 3012.
- [41] D. Barford, A.J. Flint, N.K. Tonks, Science (Washington, DC) 263 (1994) 1397.
- [42] P.J. Hajduk, T. Gerfin, J.-M. Boehlen, M. Häberli, D. Marek, S.W. Fesik, J. Med. Chem. 42 (1999) 2315.
- [43] P.J. Hajduk, D.J. Augeri, J. Mack, R. Mendoza, J. Yang, S.F. Betz, S.W. Fesik, J. Am. Chem. Soc. 122 (2000) 7898.
- [44] N.K. Goto, K.H. Gardner, G.A. Mueller, R.C. Willis, L.E. Kay, J. Biomol. NMR 13 (1999) 369.
- [45] M. Kainosho, T. Tsuji, Biochemistry 21 (1982) 6273.
- [46] J. Weigelt, M. van Dongen, J. Uppenberg, J. Schultz, M. Wikström, J. Am. Chem. Soc. 124 (2002) 2446.
- [47] P.J. Hajduk, R.P. Meadows, S.W. Fesik, Science (Washington, DC) 278 (1997) 497.
- [48] B.T. Farmer II, W. Metzler, V. Goldfarb, P. McDonnell, S. Krystek, K. Constantine, L. Mueller, XIX International Conference on Magnetic Resonance in Biological Systems, 2000.
- [49] D.F. Wyss, Keystone Symposium on Structural Genomics: From Gene Sequence to Function, 2002.
- [50] M.L. Cheever, T.K. Sato, T. deBeer, T.G. Kutateladze, S.D. Emr, M. Overduin, Nat. Cell Biol. 3 (2001) 613.
- [51] C.A. Amezcua, S.M. Harper, K.H. Gardner, Keystone Symposium on Frontiers of Structural Biology, 2002.
- [52] A. Medek, P.J. Hajduk, J. Mack, S.W. Fesik, J. Am. Chem. Soc. 122 (2000) 1241.
- [53] M.A. McCoy, D.F. Wyss, J. Biomol. NMR 18 (2000) 189.
- [54] M.A. McCoy, D.F. Wyss, J. Am. Chem. Soc. 124 (2002) 2104.
- [55] G. Valensin, T. Kushnir, G. Navon, J. Magn. Reson. 46 (1982) 23.
- [56] R.R. Ernst, G. Bodenhausen, A. Wokaun, Principles of Nuclear Magnetic Resonance in One and Two Dimensions, Clarendon Press, Oxford, 1987.
- [57] C. Rossi, G. Valensin, A. Prugnola, N. Niccolai, in: G.

Naray-Szabo, K. Simon (Eds.), Steric Aspects of Biomolecular Interactions, CRC Press, Boca Raton, 1987.

- [58] E. Gaggelli, G. Valensin, K. Tamir, G. Navon, Magn. Reson. Chem. 30 (1992) 461.
- [59] C. Rossi, A. Donati, C. Bonechi, G. Corbini, R. Rappuoli, E. Dreassi, P. Corti, Chem. Phys. Lett. 264 (1997) 205.
- [60] G. Veglia, M. Delfini, M.R. Del Giudice, E. Gaggelli, G. Valensin, J. Magn. Reson. 130 (1998) 281.
- [61] C. Rossi, S. Bastianoni, C. Bonechi, G. Corbini, P. Corti, A. Donati, Chem. Phys. Lett. 310 (1999) 495.
- [62] M. Delfini, R. Gianferri, V. Dubbini, C. Manetti, E. Gaggelli, G. Valensin, J. Magn. Reson. 144 (2000) 129.
- [63] C. Rossi, C. Bonechi, S. Martini, M. Ricci, G. Corbini, P. Corti, A. Donati, Magn. Reson. Chem. 39 (2001) 457.
- [64] G. Valensin, P.E. Valensin, E. Gaggelli, in: J.W. Jaroszewski, K. Schaumburg, H. Kofod (Eds.), NMR Spectroscopy in Drug Research, Munksgaard, Copenhagen, 1988.
- [65] L.Y. Lian, G.C.K. Roberts, in: G.C.K. Roberts (Ed.), NMR of Macromolecules, Oxford University Press, Oxford, 1993.
- [66] T.E. Bull, Prog. NMR Spectrosc. 24 (1992) 377.
- [67] A. Kumar, P.K. Madhu, Conc. Magn. Reson. 8 (1996) 139.
- [68] D.L. Rabenstein, T. Nakashima, G. Bigam, J. Magn. Reson. 34 (1979) 669.
- [69] T. Scherf, J. Anglister, Biophys. J. 64 (1993) 754.
- [70] P.J. Hajduk, E.T. Olejniczak, S.W. Fesik, J. Am. Chem. Soc. 119 (1997) 12257.
- [71] S. Meiboom, D. Gill, Rev. Sci. Instrum. 29 (1958) 688.
- [72] R. Freeman, H.D.W. Hill, J. Chem. Phys. 55 (1971) 1985.
- [73] W. Jahnke, S. Rüdissor, M. Zurini, J. Am. Chem. Soc. 123 (2001) 3149.
- [74] W. Jahnke, L.B. Perez, C.G. Paris, A. Strauss, G. Fendrich, C.M. Nalin, J. Am. Chem. Soc. 122 (2000) 7394.
- [75] A. Bax, M. Ikura, L.E. Kay, D.A. Torchia, R. Tschudin, J. Magn. Reson. 86 (1990) 304.
- [76] T.J. Norwood, J. Boyd, J.E. Heritage, N. Soffe, I.D. Campbell, J. Magn. Reson. 87 (1990) 488.
- [77] R. Konrat, H. Sterk, Chem. Phys. Lett. 203 (1993) 75.
- [78] B. Boulat, G. Bodenhausen, J. Biomol. NMR 3 (1993) 335.
- [79] K. Kloiber, R. Konrat, J. Biomol. NMR 18 (2000) 33.
- [80] M. Tessari, G.W. Vuister, J. Biomol. NMR 16 (2000) 171.
- [81] D. Früh, J.R. Tolman, G. Bodenhausen, C. Zwahlen, J. Am. Chem. Soc. 123 (2001) 4810.
- [82] E. Chiarparin, S. Rüdisser, G. Bodenhausen, Chem. Phys. Chem. 1 (2001) 41.
- [83] C. Dalvit, J.M. Böhlen, Annu. Rep. NMR Spectrosc. 37 (1999) 203.
- [84] P. Stilbs, Prog. NMR Spectrosc. 19 (1987) 1.
- [85] E.O. Stejskal, J.E. Tanner, J. Chem. Phys. 42 (1965) 288.
- [86] J.E. Tanner, J. Chem. Phys. 52 (1970) 2523.
- [87] C.S. Johnson Jr., Prog. NMR Spectrosc. 34 (1999) 203.
- [88] W.S. Price, P.W. Kuchel, J. Magn. Reson. 94 (1991) 133.
- [89] M. Lin, M.J. Shapiro, J.R. Wareing, J. Am. Chem. Soc. 119 (1997) 5249.
- [90] K. Bleicher, M. Lin, M.J. Shapiro, J.R. Wareing, J. Org. Chem. 63 (1998) 8486.
- [91] R.C. Anderson, M. Lin, M.J. Shapiro, J. Comb. Chem. 1 (1999) 69.

- [92] M. Lin, M.J. Shapiro, J. Org. Chem. 61 (1996) 7617.
- [93] P. Hodge, P. Monvisade, G.A. Morris, I. Preece, J. Chem. Soc. Chem. Commun. 3 (2001) 239.
- [94] C. Dalvit, P. Ramage, U. Hommel, J. Magn. Reson. 131 (1998) 148.
- [95] N. Gonnella, M. Lin, M.J. Shapiro, J.R. Wareing, X. Zhang, J. Magn. Reson. 131 (1998) 336.
- [96] C. Dalvit, J.M. Böhlen, NMR Biomed. 10 (1997) 285.
- [97] P. Balaram, A.A. Bothner-By, J. Dadok, J. Am. Chem. Soc. 94 (1972) 4015.
- [98] G.M. Clore, A.M. Gronenborn, J. Magn. Reson. 48 (1982) 402.
- [99] A.P. Campbell, B.D. Sykes, Annu. Rev. Biophys. Biomol. Struct. 22 (1993) 99.
- [100] F. Ni, Prog. NMR Spectrosc. 26 (1994) 517.
- [101] B. Meyer, T. Weimar, T. Peters, Eur. J. Biochem. 246 (1997) 705.
- [102] D. Henrichsen, B. Ernst, J.L. Magnani, W.-T. Wang, B. Meyer, T. Peters, Angew. Chem. Int. Ed. 38 (1999) 98.
- [103] W. Jahnke, H.C. Kolb, M.J.J. Blommers, J.L. Magnani, B. Ernst, Angew. Chem. Int. Ed. 36 (1997) 2603.
- [104] L. Herfurth, T. Weimar, T. Peters, Angew. Chem. Int. Ed. 39 (2000) 2097.
- [105] A. Chen, M.J. Shapiro, J. Am. Chem. Soc. 120 (1998) 10258.
- [106] A. Chen, M.J. Shapiro, J. Am. Chem. Soc. 122 (2000) 414.
- [107] S. Forsén, R.A. Hoffman, J. Chem. Phys. 39 (1963) 2892.
- [108] M. Mayer, B. Meyer, Angew. Chem. Int. Ed. 38 (1999) 1784.
- [109] J. Klein, R. Meinecke, M. Mayer, B. Meyer, J. Am. Chem. Soc. 121 (1999) 5336.
- [110] A. Kalk, H.J.C. Berendsen, J. Magn. Reson. 24 (1976) 343.
- [111] J.D. Stoesz, A.G. Redfield, FEBS Lett. 91 (1978) 320.
- [112] K. Wüthrich, NMR of Proteins and Nucleic Acids, Wiley, New York, 1986.
- [113] M. Mayer, B. Meyer, J. Am. Chem. Soc. 123 (2001) 6108.
- [114] M. Vogtherr, T. Peters, J. Am. Chem. Soc. 122 (2000) 6093.
- [115] Y.-C. Cheng, W.H. Prusoff, Biochem. Pharmacol. 22 (1973) 3099.
- [116] C. Dalvit, P. Pevarello, M. Tatò, M. Veronesi, A. Vulpetti, M. Sundström, J. Biomol. NMR 18 (2000) 65.
- [117] C. Dalvit, G.P. Fogliatto, A. Stewart, M. Veronesi, B. Stockman, J. Biomol. NMR 21 (2001) 349.
- [118] C.S. Poornima, P.M. Dean, J. Comput. Aided Mol. Des. 9 (1995) 500.
- [119] G. Otting, K. Wüthrich, J. Am. Chem. Soc. 111 (1989) 1871.
- [120] G. Otting, E. Liepinsh, K. Wüthrich, Science (Washington, DC) 254 (1991) 974.
- [121] V.P. Denisov, B. Halle, J. Mol. Biol. 245 (1995) 682.
- [122] V.P. Denisov, B. Halle, J. Peters, H.D. Hörlein, Biochemistry 34 (1995) 9046.
- [123] J.A. Ernst, R.T. Clubb, H.-X. Zhou, A.M. Gronenborn, G.M. Clore, Science (Washington, DC) 267 (1995) 1813.
- [124] G. Otting, Prog. NMR Spectrosc. 31 (1997) 259.
- [125] B. Halle, V.P. Denisov, K. Venu, Biol. Magn. Reson. 17 (1999) 419.
- [126] S. Wiesner, E. Kurian, F.G. Prendergast, B. Halle, J. Mol. Biol. 286 (1999) 233.
- [127] G. Otting, E. Liepinsh, Acc. Chem. Res. 28 (1995) 171.

B.J. Stockman, C. Dalvit / Progress in Nuclear Magnetic Resonance Spectroscopy 41 (2002) 187-231

- [128] E. Liepinsh, G. Otting, Magn. Reson. Med. 35 (1996) 30.
- [129] C. Dalvit, J. Magn. Reson. B112 (1996) 282.
- [130] C. Dalvit, J. Biomol. NMR 11 (1998) 437.
- [131] A. Fersht, Enzyme Structure and Mechanism, Freeman, New York, 1985.
- [132] C. Dalvit, M. Fasolini, M. Flocco, S. Knapp, P. Pevarello, M. Veronesi, J. Med. Chem. 45 (2002) 2610.
- [133] C. Dalvit, M. Flocco, S. Knapp, M. Mostardini, R. Perego, B.J. Stockman, M. Veronesi, M. Varasi, J. Am. Chem. Soc. 124 (2002) 7702.
- [134] C. Dalvit, M. Flocco, M. Veronesi, B.J. Stockman, Comb. Chem. HTS (2002) in press.
- [135] J.T. Gerig, Prog. NMR Spectrosc. 26 (1994) 293.
- [136] C.A. Lepre, Drug Discov. Today 6 (2001) 133.
- [137] S.J. Teague, A.M. Davis, P.D. Leeson, T. Oprea, Angew. Chem. Int. Ed. 38 (1999) 3743.
- [138] X.Q. Lewell, D.B. Judd, S.P. Watson, M.M. Hann, J. Chem. Inf. Comput. Sci. 38 (1998) 511.
- [139] G.W. Bemis, M.A. Murcko, J. Med. Chem. 39 (1996) 2887.
- [140] G.W. Bemis, M.A. Murcko, J. Med. Chem. 42 (1999) 5095.
- [141] A.J. Lennon, N.R. Scott, B.E. Chapman, P.W. Kuchel, Biophys. J. 67 (1994) 2096.
- [142] J. Moore, N. Abdul-Manan, J. Fejzo, A. Heiser, C. Lepre, J. Peng, J. Pocas, J. Spencer, Y.-W. Wang, Keystone Symposium on Structural Genomics: From Gene Sequence to Function, 2002.
- [143] H.-J. Boehm, M. Boehringer, D. Bur, H. Gmuender, W. Huber, W. Klaus, D. Kostrewa, H. Kuehne, T. Luebbers, N. Meunier-Keller, F. Mueller, J. Med. Chem. 43 (2000) 2664.

- [144] I. Muegge, Y.C. Martin, P.J. Hajduk, S.W. Fesik, J. Med. Chem. 42 (1999) 2498.
- [145] H. Mao, P.J. Hajduk, R. Craig, R. Bell, T. Borre, S.W. Fesik, J. Am. Chem. Soc. 43 (2001) 10429.
- [146] M.M. Reidenberg, S. Erill, Drug–Protein Binding, Praeger, New York, 1986.
- [147] S. Curry, H. Mandelkow, P. Brick, N. Franks, Nat. Struct. Biol. 5 (1998) 827.
- [148] P.V. Sahasrabudhe, D.T.A. Hadden, J.D. Robarge, B.J. Stockman, Keystone Symposium on Structural Genomics: From Gene Sequence to Function, 2002.
- [149] F.R. Salemme, Keystone Symposium on Structural Genomics: From Gene Sequence to Function, 2002.
- [150] L. Parsons, D.C. Yeh, N. Bonander, E. Eisenstein, P. Reddy, J. Orban, Keystone Symposium on Structural Genomics: From Gene Sequence to Function, 2002.
- [151] A. Ross, G. Schlotterbeck, W. Klaus, H. Senn, J. Biomol. NMR 16 (2000) 139.
- [152] P. Mansfield, Magn. Reson. Med. 1 (1984) 370.
- [153] E. MacNamara, T. Hou, G. Fisher, S. Williams, D. Raftery, Anal. Chim. Acta 397 (1999) 9.
- [154] M. Pellecchia, D. Meininger, A.L. Shen, R. Jack, C.B. Kasper, D.S. Sem, J. Am. Chem. Soc. 123 (2001) 4633.
- [155] M. Pellecchia, D. Meininger, Q. Dong, E. Chang, R. Jack, D.S. Sem, J. Biomol. NMR 22 (2002) 165.
- [156] Z. Serber, A.T. Keatinge-Clay, R. Ledwidge, A.E. Kelly, S.M. Miller, V. Dötsch, J. Am. Chem. Soc. 123 (2001) 2446.
- [157] S. Grzesiek, A. Bax, J. Am. Chem. Soc. 115 (1993) 12593.
- [158] T.-L. Hwang, A.J. Shaka, J. Magn. Reson. A112 (1995) 275.
- [159] U. Kragh-Hansen, Biochem. J. 273 (1991) 641.