J Biomol NMR (2015) 63:53–58
DOI 10.1007/s10858-015-9962-3

ARTICLE

Statistical removal of background signals from high-throughput ^{1}H NMR line-broadening ligand-affinity screens

Bradley Worley[1] · Nicholas J. Sisco[1,2,3] · Robert Powers[1]

Received: 15 April 2015 / Accepted: 30 June 2015 / Published online: 9 July 2015

Abstract NMR ligand-affinity screens are vital to drug discovery, are routinely used to screen fragment-based libraries, and used to verify chemical leads from high-throughput assays and virtual screens. NMR ligand-affinity screens are also a highly informative first step towards identifying functional epitopes of unknown proteins, as well as elucidating the biochemical functions of protein–ligand interaction at their binding interfaces. While simple one-dimensional ^{1}H NMR experiments are capable of indicating binding through a change in ligand line shape, they are plagued by broad, ill-defined background signals from protein ^{1}H resonances. We present an uncomplicated method for subtraction of protein background in high-throughput ligand-based affinity screens, and show that its performance is maximized when phase-scatter correction is applied prior to subtraction.

Keywords NMR · High-throughput screening · Ligand-based affinity screening · Phase-scatter correction

Electronic supplementary material The online version of this article (doi:10.1007/s10858-015-9962-3) contains supplementary material, which is available to authorized users.

✉ Robert Powers
rpowers3@unl.edu

[1] Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE 68588-0304, USA

[2] Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287, USA

[3] Magnetic Resonance Research Center, Arizona State University, Tempe, AZ 85287, USA

Introduction

SAR by NMR (Shuker et al. 1996) spurred a revolution for the role of NMR in drug discovery. Like X-ray crystallography, NMR had been primarily used as a means to determine protein and protein–ligand structures as part of a structure-based drug discovery effort (Ferentz and Wagner 2000). NMR is now an important alternative to traditional high-throughput assays (HTS) for identifying drug-like chemical leads (Pellecchia et al. 2002; Powers 2009). By combining NMR ligand-affinity screens with fragment-based libraries, a dramatic increase in chemical diversity is achieved (from 10^{6} to 10^{63}), while also minimizing resources, increasing hit-rates and improving the drug-like qualities of the resulting chemical leads (Baker 2013; Hajduk and Greer 2007). Consequently, NMR fragment-based screens have significantly benefited the pharmaceutical industry by leading to a number of clinical-stage compounds (Baker 2013).

NMR ligand-affinity screening is also a powerful platform for protein functional annotation during the search for novel drug targets (Mercier et al. 2009; Powers et al. 2008). A significant percentage of the human proteome and the proteomes of other infectious organisms is comprised of functionally uncharacterized proteins (Muller et al. 2002). Undoubtedly hidden among this multitude of unannotated proteins are novel drug targets that may lead to new treatments or new means of overcoming mechanisms of drug resistance. Besides verifying that a functionally unannotated protein is druggable, NMR ligand affinity screening also identifies the functional epitope and the classes of ligands that bind the uncharacterized protein. This information may then be leveraged to infer a function through structural similarities with functionally annotated proteins (Powers et al. 2006, 2011).

NMR spectroscopy reports a multitude of time-averaged physical observables that carry information relating to the

nature of interactions between small molecule ligands and protein targets (Lepre et al. 2004). A number of one-dimensional (1D) ^{1}H NMR pulse sequences have been developed to probe these distinct features of binding, including differences in free and bound ligand diffusion and relaxation properties (Hajduk et al. 1997), and saturation transfers from water (Dalvit et al. 2000) and protein (Mayer and Meyer 2001) resonances. As part of an NMR high-throughput screen, these 1D ^{1}H NMR pulse sequences present a number of unique challenges that include high false positive rates, low throughput, and high demand for protein samples (Harner et al. 2013; Lepre 2011). However, at suitably chosen concentrations of ligand and protein, a standard, unedited 1D ^{1}H NMR experiment may be used to detect binding interactions through enhanced relaxation rates of ligand spins (Mercier et al. 2006, 2009; Powers et al. 2008).

While it is possible to detect ligand binding using standard 1D ^{1}H NMR, the resulting spectra are a combination of free and bound ligand and protein signals, a fact which makes them difficult to interpret. Broad, rolling baselines arising from slowly tumbling protein spins are particularly problematic during interpretation, as they often mask changes in ligand signal broadness and intensity. This masking effect due to protein baselines is exacerbated at protein–ligand concentration ratios nearing or exceeding unity, forcing the use of excess ligand during analysis and increasing the false negative rate during screening. To mitigate these issues, we present a statistical method, called Uncomplicated Statistical Spectral Remodeling (USSR), which removes protein baselines from high-throughput ligand-based screening datasets by leveraging inter-sample reproducibility of protein signals. In addition, we show that the use of phase-scatter correction (PSC) (Worley and Powers 2014b) greatly improves inter-sample protein baseline reproducibility and reduces the false-positive rate incurred by subsequent USSR-based analyses. Our reported combination of PSC and USSR enables a rapid analysis of standard 1D ^{1}H NMR screening data, especially in difficult cases having a high protein–ligand concentration ratio.

Materials and methods

Sample preparation and NMR acquisition

A set of 117 samples containing free ligand mixtures and a set of 117 samples containing bovine serum albumin (BSA) and ligand mixtures were prepared based on previously published procedures (Mercier et al. 2009; Powers et al. 2008). In summary, each mixture contained no more than four ligands, each ligand had a concentration of 100 μM, and BSA had a concentration of 200 μM. All NMR samples were prepared to 600 μL total volume in a buffer containing 10 mM bis–tris-d_{19}, 1 mM NaCl, 1 mM KCl, 1 mM $MgCl_2$, and 10 μM TMSP in D_2O at pH 7.0 (uncorrected). Samples were loaded into standard 5 mm NMR tubes for spectral acquisition.

All NMR experiments were collected on a Bruker Avance DRX 500 MHz spectrometer equipped with a 5 mm inverse triple-resonance (^{1}H, ^{13}C, ^{15}N) cryoprobe with a z-axis gradient. A Bruker BACS-120 sample changer and ICON-NMR software were used to automate NMR data collection. Standard 1D ^{1}H NMR spectra were collected for each sample using a SOGGY water suppression pulse sequence (Hwang and Shaka 1995; Nguyen et al. 2007). All experiments were performed at 20 °C with 256 scans, 8 dummy scans, a carrier frequency offset of 2352.1 Hz, a 5482.5 Hz spectral width, and a 1.0 s inter-scan delay. Free induction decays were collected with 4 k complex data points resulting in a total acquisition time of 8 min per experiment.

NMR processing

Acquired NMR spectra were loaded and processed in batch inside the GNU Octave 3.6 programming environment (Eaton et al. 2008) using functions available in the MVAPACK software package (Worley and Powers 2014a). Free induction decays were loaded in from Bruker DMX binary format and corrected for group delay errors by a circular shift. All decays were then zero filled twice, Fourier transformed and automatically phase corrected using a simplex optimization routine. PSC (Worley and Powers 2014b) was applied to a copy of the screen spectral data, and spectral remodeling was performed in parallel on the uncorrected and corrected datasets for the purposes of comparison.

Phase-scatter correction

The PSC algorithm is a spectral normalization method that includes zero-order and first-order phase correction terms in its objective function (Worley and Powers 2014b). As a normalization method, PSC scales each 'target' spectrum in a dataset to match a reference spectrum, which is usually the mean. Differences in relative phase between spectra, which confound the identification of ideal scaling values, are also corrected by PSC within a Levenberg–Marquardt nonlinear optimization of the following objective:

$$Q(s|\beta,\varphi_0,\varphi_1) = \sum_{k=1}^{K} \left|\beta s_k e^{i(\varphi_0+\varphi_1 k)} - r_k\right|^2 \quad (1)$$

where s_k and r_k are the k-th data points of the K-element target and reference spectra, respectively. The PSC algorithm identifies an optimal scale factor (β) and phases (φ_0, φ_1) for each spectrum in the dataset, and returns the corrected data matrix after applying the identified scaling and phase correction to each spectrum.

Statistical spectral remodeling

Our method of spectral remodeling (USSR) capitalizes on the reproducibility of the protein baseline and the low likelihood that ligand signals will dominate any given spectral data point across multiple samples. For each pair of free mixture ($\boldsymbol{f_i}$) and screen (mixture plus protein, $\boldsymbol{p_i}$) ^{1}H NMR spectra, a difference spectrum ($\boldsymbol{d_i}$) is computed using a simple point-wise subtraction. The central tendency ($\boldsymbol{\mu}$) and dispersion ($\boldsymbol{\sigma}$) of the difference spectra are then robustly estimated using the median and median absolute deviation, respectively. Figure 1 shows the statistical baseline computed by USSR from an analysis of ligand binding to BSA. Once a statistical baseline is established for a given dataset, each spectrum $\boldsymbol{p_i}$ in the screen is remodeled to maximally remove interference from protein signals. Each spectral data point in $\boldsymbol{p_i}$ is compared to $\boldsymbol{\mu} \pm \boldsymbol{\sigma}$ using a Bonferroni-corrected Student's t test (Dunn 1961). The resulting p value provides a measure of how distinguishable the corresponding data point is from the statistical baseline. Based on a preselected level of significance (α), data points having low p values are retained (less the statistical baseline) in the remodeled spectrum ($\boldsymbol{r_i}$) and data points having high p values are modeled as Gaussian white noise. Figure 2 shows an example remodeled spectrum from the ligand binding analysis of BSA.

Statistical hit determination

For each peak in each remodeled spectrum from USSR, a K_D was computed based on the intensity ratio between free and remodeled ligand signals. First, in the limit of fast exchange between free and bound ligand states relative to the NMR timescale, the fraction of bound ligand (f_B) was computed:

$$f_B = \left(\frac{I_F}{I_B} - 1\right)\left(\frac{\nu_B}{\nu_F} - 1\right)^{-1} \quad (2)$$

where I_F and I_B are the intensities of free and remodeled (bound) ligand signals, and ν_F and ν_B are the estimated NMR line widths of the free and remodeled ligand signals, respectively (Shortridge et al. 2008). This fast-exchange assumption may be safely regarded as valid in most high-throughput 1D ^{1}H NMR protein–ligand affinity screening experiments (Lepre et al. 2004), where the width and intensity of each ligand signal is a population-weighted sum of its values in the free and bound states. Without any assumptions about relative concentrations of ligand and protein, the fraction of bound ligand is related to total protein concentration $[\mathrm{P}]_\mathrm{T}$, total ligand concentration $[\mathrm{L}]_\mathrm{T}$ and K_D via the following equation (Shortridge et al. 2008):[1]

$$f_B = \left[1 + \frac{2K_D}{\left([P]_T - [L]_T - K_D\right) + \sqrt{\left([P]_T - [L]_T - K_D\right)^2 + 4K_D[P]_T}}\right]^{-1} \quad (3)$$

The symbolic solution of the above equation for K_D (Wolfram Research 2014) yields the following result:

$$K_D = \frac{(f_B - 1)\left(f_B[L]_T - [P]_T\right)}{f_B} \quad (4)$$

which was then used to compute per-peak K_D values for each remodeled spectrum $\boldsymbol{r_i}$. Finally, the per-peak K_D values were used to compute sample mean and standard deviation K_D values for each ligand. Hit detection was accomplished by comparing per-ligand mean and standard deviation K_D values against a threshold via a Student's t test, where a resulting p value less than a predefined significant p value was reported as binding.

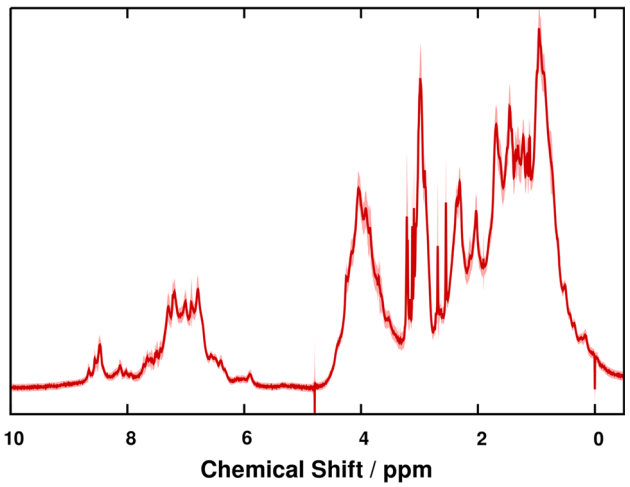

Fig. 1 Statistical baseline (mean plus or minus four standard deviations) computed from the NMR ligand-based screen against BSA. The mean baseline is traced in *deep red*, while the baseline is filled in *light red* underneath

Analysis of dataset size

A small simulation study was performed to assess the quality of USSR statistical baseline estimates over a range of sample sizes (number of spectral pairs). For sizes from 2 to 116, the BSA dataset was randomly subsampled, without replacement, to produce a smaller dataset. For each resultant dataset, the statistical baseline was estimated, and its Pearson correlation to the true statistical baseline was computed and stored. Over all numbers of spectral pairs in the simulation, the median baseline correlations were computed, and are reported in Supplementary Figure S-1.

[1] The equation reported in the main text of Shortridge et al. (2008) (Eq. 3) contains a typographical error. The correct equation (equation A8) may be found in the Appendix of the aforementioned work.

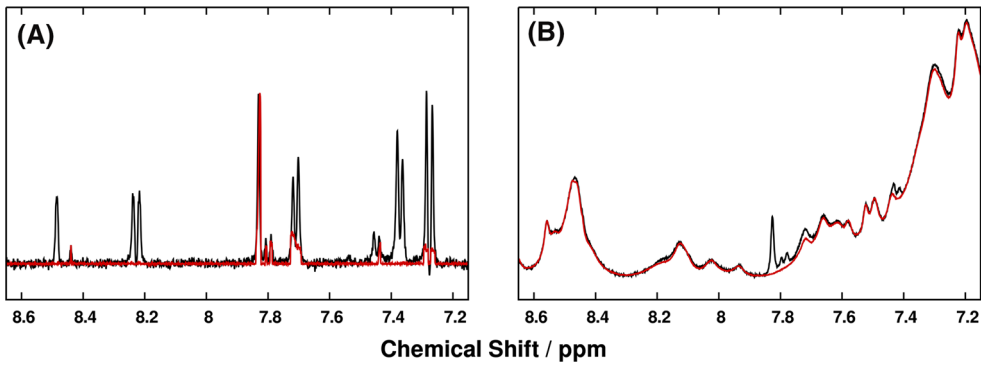

Fig. 2 An example spectral remodeling result of tolazamide, dimethyl 4-methoxyisophthalate, 1,7-dimethylxanthine and oxolinic acid in the presence of BSA, showing **a** the free ligand spectrum (*black*) and the remodeled spectrum (*red*) resulting from removing the statistical baseline (*red*) from the screen spectrum (*black*) in **b**. The remodeled pseudo-spectrum readily indicates that several peaks from dimethyl 4-methoxyisophthalate have broadened into the baseline due to interaction with BSA

Results

From the USSR analysis of ligand binding to BSA, 43 compounds were classified as hits from the library of 456 compounds. All classified hits were determined to bind BSA with at least 1.0 mM affinity at a statistical confidence level of 99 %. A summary of the hits, along with their estimated K_D and p values, is provided in Supplementary Table S-1. Comparison of results from both PSC-corrected and uncorrected USSR datasets reveals that the use of PSC normalization prior to USSR modeling greatly reduces the effective positive rate of statistical hit determination: 195 hits were identified from the PSC-uncorrected spectra. Closer examination of hits identified without PSC correction indicates that USSR failed to fully subtract the statistical baseline from the screen spectra (e.g. Figure S-2), resulting in residual baseline intensity passing into Eq. (2) during K_D calculation and hit determination. In short, the use of PSC normalization prior to USSR enables more effective baseline subtraction by decreasing both dilution- and phase-related protein baseline intensity variation in collected ^{1}H NMR spectra (Fig. 3). Baseline estimates obtained by collecting a spectrum of pure protein will suffer from the same phase-induced variation, which would also increase the false positive rate during hit determination. Our introduced combination of PSC and USSR provides a more reliable means of baseline identification, without the collection of a free protein spectrum.

Our cursory analysis of the robustness of the USSR statistical baseline during random subsampling of the BSA dataset indicated that the PSC/USSR methodology can reliably operate at very low dataset sizes (i.e. 10–20 spectral pairs). Pearson correlations between true and subsampled baselines did not appreciably decrease even after harsh subsampling (Supplementary Figure S-1), and correlations computed from PSC-normalized data

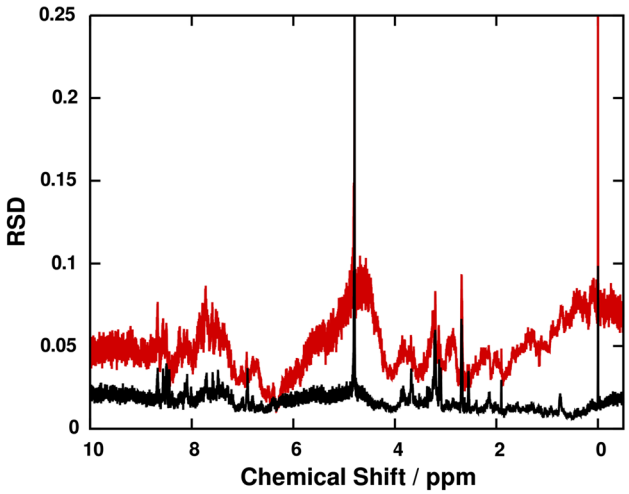

Fig. 3 Relative standard deviations (RSDs) of the statistical baselines computed before (*red*) and after (*black*) PSC, which substantially decreases inter-sample variability of the protein baseline signals

maintained significantly higher values than those from non-normalized data. While it would be possible to obtain a statistical baseline from fewer than ten spectral pairs, this is not recommended, as it will decrease the effectiveness of the Bonferroni-corrected t test that USSR performs during remodeling. Therefore, as a general rule of thumb, PSC/USSR analyses may be performed on high-throughput screening datasets having as few as ten spectral pairs, and higher sample sizes only serve to further increase the reliability of remodeled results.

Discussion and conclusions

While the saturation transfer difference (STD) NMR experiment (Mayer and Meyer 2001) is a popular choice for ligand-based NMR affinity screens, a 1D ^{1}H NMR line-

broadening experiment is often a more efficient alternative. A standard 1D ^{1}H NMR spectrum requires only a few seconds to acquire, making it an ideal choice for high-throughput screening. STD experiments require significantly longer acquisition times (upwards of hours) in order to acquire difference spectra with sufficient signal-to-noise to reduce false negatives. A particular strength of STD is the minimal amount of protein required per experiment, making it practical to screen a reasonably large chemical library (upwards of thousands of compounds) with only a few milligrams of protein. Through a judicious choice of protein and ligand concentrations coupled with the use of cryoprobes and high magnetic fields, it is also possible to minimize protein requirements in 1D ^{1}H line-broadening screens. While STD experiments still tend to require less protein than line-broadening experiments, the higher false positive rate of STD screening easily negates any advantages of minimal protein usage. This high false positive rate arises due to the tendency of STD experiments to emphasize weak binding affinities commonly encountered during aggregation and nonspecific binding (Harner et al. 2013; Lepre 2011).

NMR line-broadening experiments take advantage of the molecular-weight dependence of T_2 relaxation and the resultant measurable difference in line-widths between proteins and the compounds in a screening library (Hajduk et al. 1997). Upon binding a protein target, the ^{1}H NMR resonances of a compound will broaden significantly or even disappear. In principal, this spectral broadening is easily observable and binding is readily identified. In practice, the background signals from the protein can confound the data analysis. This background interference increases with the size and concentration of the protein and leads to an increase in false negative rates. Apparent line-width differences between free and bound ligands *also* increase with protein size and concentration, making the optimal experimental conditions for NMR line-broadening screens exactly the same conditions which confound manual interpretation. Clearly, the ability to accurately remove the protein background from an NMR line-broadening experiment will improve both the utility and reliability of the technique, especially at relatively high protein–ligand concentration ratios where binding is more apparent.

By removing interfering protein baseline signals, our USSR method for analyzing 1D ^{1}H NMR line-broadening spectra for high-throughput screening provides a straightforward means to visually or computationally analyze screening results. In fact, the outcome of applying our USSR method to an extremely challenging and atypical test case is rather dramatic: our NMR line-broadening screen using BSA and a chemical library of 456 compounds identified 43 binders, despite the BSA background signals completely obscuring the ligand spectra. An example screening result of tolazamide, dimethyl 4-methoxyisophthalate, 1,7-dimethylxanthine and oxolinic acid against BSA is illustrated in Fig. 2. Removal of the interfering protein statistical baseline from the screen spectrum (Fig. 2b) yielded a high-quality pseudo-spectrum of the ligand mixture in the presence of BSA. Overlaying the remodeled NMR spectrum with the free ligand mixture spectrum indicated that the two spectra were essentially identical for the non-binding ligands (Fig. 2a). Only dimethyl 4-methoxyisophthalate, which binds BSA, exhibited any difference after remodeling. The USSR method of baseline estimation and subtraction is expected to perform equally well under any conditions where a common, highly reproducible spectral feature exists within a dataset. Our application of PSC/USSR to high-throughput protein–ligand affinity screening is but one example of its potential uses.

However, reliable identification of the protein baseline from screening data requires highly reproducible sample preparation, data collection and processing. The last of these requirements is met by the use of PSC (Worley and Powers 2014b) prior to remodeling, which brings protein baselines from all spectra into closer agreement with each other and minimizes the number of false hits identified during analysis. It is important to note that PSC is only an effective pre-treatment for USSR when protein baseline signals are of comparable intensity to ligand signals. PSC normalization is designed to maximize statistical agreements *between* spectra by phasing and normalization correction, and its use of the L_2 norm as a criterion for 'agreement' implies that higher-intensity features will be preferentially corrected. Thus, PSC achieves the best results prior to USSR when protein signals are a major spectral component, as is the case when protein–ligand concentration ratios are near or greater than unity. In effect, the combined use of PSC and USSR expands the range of protein–ligand concentration ratios which may be probed by ^{1}H line-broadening experiments for the purposes of high-throughput screening.

Finally, it cannot be under-emphasized that the single-point K_D computations employed by USSR during statistical hit determination are only order-of-magnitude estimates of the true dissociation rates, and can carry significant systematic and random errors. In particular, the fraction of bound ligand—and by extension, the dissociation constant—depends exquisitely on the estimated free and bound ligand linewidths, ν_F and ν_B. Thus, any imprecision in the linewidth estimates will propagate into a systematic bias in the final dissociation constants. If required, verification of initial hits may be achieved to higher accuracy via multiple-point estimation of the K_D through nonlinear least squares (Shortridge et al. 2008).

An implementation of the USSR algorithm is available in open-source GNU Octave code as a part of the MVAPACK toolbox, downloadable at http://bionmr.unl.edu/mvapack.php.

Acknowledgments This work was supported, in part, by funds from the National Institutes of Health (Grant number R21AI081154). The research was performed in facilities renovated with support from the National Institutes of Health (Grant number RR015468-01).

References

Baker M (2013) Fragment-based lead discovery grows up. Nat Rev Drug Discov 12:5–7. doi:10.1038/nrd3926

Dalvit C, Pevarello P, Tato M, Veronesi M, Vulpetti A, Sundstrom M (2000) Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. J Biomol NMR 18:65–68. doi:10.1023/A:1008354229396

Dunn OJ (1961) Multiple comparisons among means. J Am Stat Assoc 56:52–64. doi:10.2307/2282330

Eaton JW, Bateman D, Hauberg S (2008) GNU octave manual version 3. Network Theory Limited, Bristol

Ferentz AE, Wagner G (2000) NMR spectroscopy: a multifaceted approach to macromolecular structure. Q Rev Biophys 33:29–65. doi:10.1017/s0033583500003589

Hajduk PJ, Greer J (2007) A decade of fragment-based drug design: strategic advances and lessons learned. Nat Rev Drug Discov 6:211–219. doi:10.1038/nrd2220

Hajduk PJ, Olejniczak ET, Fesik SW (1997) One-dimensional relaxation- and diffusion-edited NMR methods for screening compounds that bind to macromolecules. J Am Chem Soc 119:12257–12261. doi:10.1021/Ja9715962

Harner MJ, Frank AO, Fesik SW (2013) Fragment-based drug discovery using NMR spectroscopy. J Biomol NMR 56:65–75. doi:10.1007/s10858-013-9740-z

Hwang TL, Shaka AJ (1995) Water suppression that works—excitation sculpting using arbitrary wave-forms and pulsed-field gradients. J Magn Reson A 112:275–279. doi:10.1006/jmra.1995.1047

Lepre CA (2011) Practical aspects of NMR-based fragment screening. Methods Enzymol 493:219–239. doi:10.1016/b978-0-12-381274-2.00009-1

Lepre CA, Moore JM, Peng JW (2004) Theory and applications of NMR-based screening in pharmaceutical research. Chem Rev 104:3641–3675. doi:10.1021/Cr030409h

Mayer M, Meyer B (2001) Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. J Am Chem Soc 123:6108–6117. doi:10.1021/Ja0100120

Mercier KA, Baran M, Ramanathan V, Revesz P, Xiao R, Montelione GT, Powers R (2006) FAST-NMR: functional annotation screening technology using NMR spectroscopy. J Am Chem Soc 128:15292–15299. doi:10.1021/Ja0651759

Mercier KA, Shortridge MD, Powers R (2009) A multi-step NMR screen for the identification and evaluation of chemical leads for drug discovery. Comb Chem High Throughput Screen 12:285–295

Muller A, MacCallum RM, Sternberg MJE (2002) Structural characterization of the human proteome. Genome Res 12:1625–1641. doi:10.1101/gr.221202

Nguyen BD, Meng X, Donovan KJ, Shaka AJ (2007) SOGGY: solvent-optimized double gradient spectroscopy for water suppression. A comparison with some existing techniques. J Magn Reson 184:263–274. doi:10.1016/j.jmr.2006.10.014

Pellecchia M, Sem DS, Wuthrich K (2002) NMR in drug discovery. Nat Rev Drug Discov 1:211–219. doi:10.1038/nrd748

Powers R (2009) Advances in nuclear magnetic resonance for drug discovery. Expert Opin Drug Discov 4:1077–1098

Powers R, Copeland JC, Germer K, Mercier KA, Ramanathan V, Revesz P (2006) Comparison of protein active site structures for functional annotation of proteins and drug design. Proteins Struct Funct Bioinform 65:124–135

Powers R, Mercier KA, Copeland JC (2008) The application of FAST-NMR for the identification of novel drug discovery targets. Drug Discov Today 13:172–179. doi:10.1016/j.drudis.2007.11.001

Powers R, Copeland JC, Stark JL, Caprez A, Guru A, Swanson D (2011) Searching the protein structure database for ligand-binding site similarities using CPASS v. 2. BMC Res Notes. doi:10.1186/1756-0500-4-17

Shortridge MD, Hage DS, Harbison GS, Powers R (2008) Estimating protein–ligand binding affinity using high-throughput screening by NMR. J Comb Chem 10:948–958. doi:10.1021/Cc800122m

Shuker SB, Hajduk PJ, Meadows RP, Fesik SW (1996) Discovering high-affinity ligands for proteins: SAR by NMR. Science (Washington, DC) 274:1531–1534

Wolfram Research I (2014) Mathematica, 10.0th edn. Wolfram Research Inc., Champaign

Worley B, Powers R (2014a) MVAPACK: a complete data handling package for NMR metabolomics. ACS Chem Biol 9:1138–1144. doi:10.1021/Cb4008937

Worley B, Powers R (2014b) Simultaneous phase and scatter correction for NMR datasets. Chemom Intell Lab Syst 131:1–6. doi:10.1016/j.chemolab.2013.11.005