Journal of Medicinal Chemistry

Article

pubs.acs.org/jmc

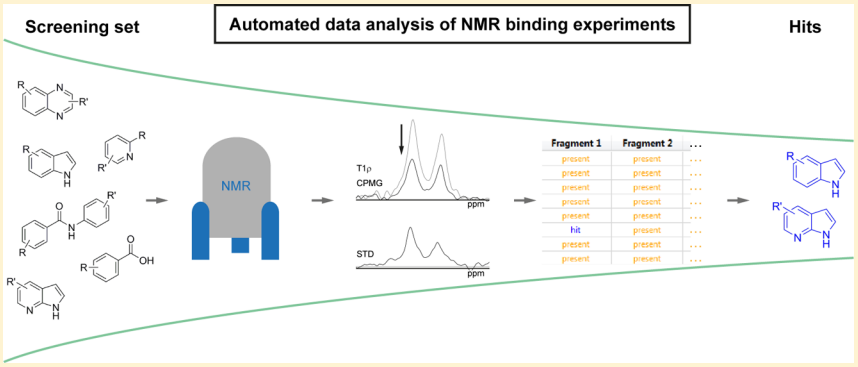
Fast and Efficient Fragment-Based Lead Generation by Fully Automated Processing and Analysis of Ligand-Observed NMR Binding Data

Chen Peng,[*,†,∥] Alexandra Frommlet,[‡,∥] Manuel Perez,[†] Carlos Cobas,[†] Anke Blechschmidt,[‡,§] Santiago Dominguez,[†] and Andreas Lingel[*,‡]

[†]Mestrelab Research S.L., Feliciano Barrera 9B − Baixo, 15706 Santiago de Compostela, Spain

[‡]Novartis Institutes for BioMedical Research, 5300 Chiron Way, Emeryville, California 94608, United States

Supporting Information



ABSTRACT: NMR binding assays are routinely applied in hit finding and validation during early stages of drug discovery, particularly for fragment-based lead generation. To this end, compound libraries are screened by ligand-observed NMR experiments such as STD, T1ρ, and CPMG to identify molecules interacting with a target. The analysis of a high number of complex spectra is performed largely manually and therefore represents a limiting step in hit generation campaigns. Here we report a novel integrated computational procedure that processes and analyzes ligand-observed proton and fluorine NMR binding data in a fully automated fashion. A performance evaluation comparing automated and manual analysis results on ^{19}F- and ^{1}H-detected data sets shows that the program delivers robust, high-confidence hit lists in a fraction of the time needed for manual analysis and greatly facilitates visual inspection of the associated NMR spectra. These features enable considerably higher throughput, the assessment of larger libraries, and shorter turn-around times.

INTRODUCTION

Because of its high information content and robustness, nuclear magnetic resonance (NMR) spectroscopy has been a widely used and established technique for hit finding and hit validation as well as detailed characterization and optimization of protein−ligand interactions since the mid-1990s.[1−6] The goal of these campaigns is to identify specific binders to a target and to utilize them as starting points for chemistry optimization, to reveal new pockets, or to map the available interaction space. In particular, NMR experiments are routinely employed in fragment-based lead discovery because they are ideally suited to detect binding of compounds to proteins in the typical micromolar to millimolar K_D range.[7,8] Additionally, since the NMR spectrum of the compound is directly observed in the screening sample, NMR spectroscopy provides "in-built" quality control, enabling a structure consistency check, concentration measurement, and binding assessment all from the same sample.

Ligand-observed NMR binding experiments are most commonly used for primary fragment screening, as there is no requirement for labeled protein, no restriction on the size of the receptor, and the ability to detect binding at compound concentrations far below the K_D. These characteristics extend the applicability of NMR spectroscopy to a wider target and ligand space, typically investigated by the pharmaceutical industry. Among such experiments, saturation transfer difference (STD), relaxation-edited experiments such as T1ρ and Carr−Purcell−Meiboom−Gill (CPMG), and water-ligand observed via gradient spectroscopy (WaterLOGSY)[9−13] are frequently used to detect binding of small molecules. The output of these experiments is usually one-dimensional NMR spectra in which molecules interacting with the protein are revealed by characteristic changes in the intensity or sign of the respective compound NMR signals. When assessing compound

Received: January 5, 2016
Published: March 10, 2016

DOI: 10.1021/acs.jmedchem.6b00019
J. Med. Chem. 2016, 59, 3303−3310

mixtures, the identity of the compounds is determined by comparing the NMR signals to the one-dimensional reference spectra of the individual compounds.

When proton-detected experiments are used in primary screening campaigns, compounds are typically measured in mixtures of 6−12 molecules, which are grouped together in such a way that there is minimal peak overlap in the ^{1}H spectrum. This reduces the measurement time and allows efficient screening of a typical midsize fragment library, consisting of 500−2000 compounds in 50−300 samples, over a period of several days. Following the mixture screening, compounds considered as primary hits are often remeasured as single compounds using the same set of experiments. This confirms binding at lower total compound concentration and sorts out potential false positives that may arise from undesired compound−compound interactions.

In addition to the established proton-detected experiments described above, fluorine-detected ligand-observed NMR experiments have been developed and implemented in recent years and have gained significant popularity.[14−18] This is mainly due to the ability to detect ^{19}F with very high sensitivity, permitting experiments to be conducted at compound concentrations in the low-micromolar range, the simplicity of the recorded spectra, and the very large chemical shift range (>200 ppm vs 10 ppm observed in proton experiments). All of this allows the generation of mixtures comprising a large number of compounds (typically 10−30) without any signal overlap. In addition, because of the large chemical shift anisotropy, the fluorine signal is very sensitive to binding events, which enables the identification of weakly binding compounds from fragment pools of ^{19}F-containing fragments.

Despite the advantages and proven performance of NMR-based hit generation, the manual analysis of the resulting large number of spectra and the spectral complexity often observed represent a significant impediment in hit finding and expansion campaigns. Visual inspection of hundreds to thousands of data sets from mixture screening and single-compound confirmation experiments is tedious, time-consuming, and prone to errors. A software tool that addresses these challenges and can streamline this procedure starting from the raw data to generate a report and visualization of suggested hits would facilitate NMR-based screening considerably.

In this paper, we report a novel computational tool that processes and analyzes ligand-observed NMR binding data in a fully automated fashion. The performance of the program is demonstrated using three typical data sets from a total of 248 samples containing 1846 compounds: an ^{19}F-detected mixture screen interrogating an in-house library of fluorinated fragments, a primary ^{1}H-detected fragment screen employing T1ρ and STD experiments on mixtures, and ^{1}H-detected single-compound confirmation experiments.

RESULTS AND DISCUSSION

The general strategy to determine compound binding on the basis of ligand-observed experiments is described in the Experimental Section and illustrated in Figure S1 in the Supporting Information. Throughout the paper, names written in *italics* refer to names used by the automatic analysis, while names in "quotation marks" refer to the manual analysis. The automatic analysis can handle various types of experiments with the associated spectra, namely, *reference* spectra of individual library compounds and *T1rho*, *STD*, and *CPMG* experiments to assess ligand binding. For peak matching between the *reference* spectra and the spectra recorded on the screening sample, it utilizes a so-called *scout* spectrum; most commonly a regular ^{1}H spectrum is used, but alternatively, an STD off-resonance or T1ρ_{short} spectrum can be used. Spectra of samples without and with protein are classified as *blank* and *protein*, respectively. The overall workflow is shown in Figure 1, and details of the individual processes are discussed below.

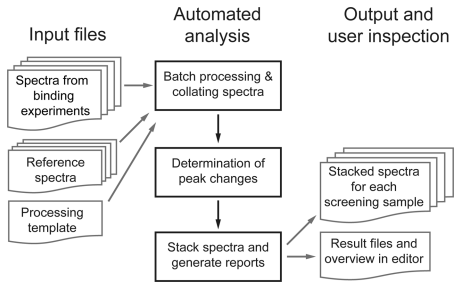


Figure 1. Workflow of the automatic procedure. Raw data of the screening experiments and reference spectra of the individual compounds are used as input. After data preparation (processing, collating) and analysis, results are prepared in a list format as well as displayed in an interactive editor for user verification.

NMR Data Preparation. Since NMR spectra of the library compounds and screening samples are often acquired on different instruments and under different conditions (e.g., receiver gain, number of scans, different buffer or temperature), a series of steps have been implemented to prepare all of the spectra for the subsequent handling and to allow a consistent analysis.

Batch Processing. For each type of spectrum, the program performs processing either with default parameters (from the original data set, including apodization function, zero-filling, Fourier transform, phase correction, and baseline correction) or with a user-defined processing template.

Spectral Alignment. Chemical shift reference offset and local peak shifts are two major factors that lead to peak misalignment. A systematic offset observed for one set of spectra can be corrected by calibrating all of the experiments to a chosen peak from an internal reference molecule, e.g., the 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) peak (Figure S2 in the Supporting Information). For ^{19}F data sets, which often contain no internal reference substance, alignment of the spectra can be achieved using a global spectral alignment tool that is implemented in the software. This tool accurately determines and minimizes the offset between two spectra by computing their pairwise fast Fourier transform cross-correlation function.[19] Local peak misalignment between the *reference* and screening spectra is also common, in particular for ^{19}F experiments, and the automatic analysis copes with these problems by allowing a tolerance when matching or integrating peaks for comparison (details are provided in the subsequent sections).

Spectral Normalization. The intensities of the spectra usually need to be normalized for two reasons: First, while the on- and off-resonance STD spectral pair have very comparable intensities, spectra of the T1ρ pair (corresponding to short and long spin-lock times) can differ significantly in signal-to-noise ratio, making it necessary to normalize them before peaks can be compared quantitatively. Therefore, the region of a solvent peak (typically the dimethyl sulfoxide peak) is integrated and normalized to a specified value (Figure S3 in the Supporting

DOI: 10.1021/acs.jmedchem.6b00019
J. Med. Chem. 2016, 59, 3303−3310

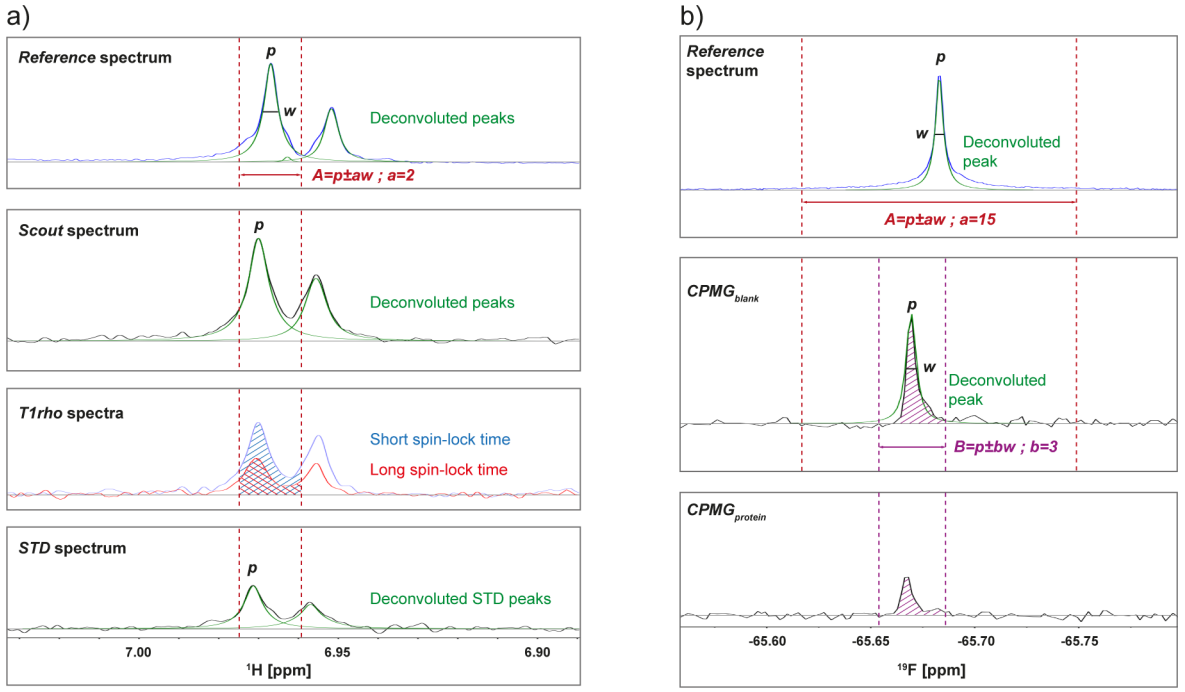


Figure 2. Illustration of the automatic analysis procedure. (a) For ^{1}H-detected experiments, the analysis starts from a *reference* peak picked by global spectral deconvolution (GSD) (top panel; deconvoluted GSD peaks are shown in green). Region *A* is determined by the peak center (p) and peak width at half height (w) as $p \pm 2w$ (represented by two vertical dotted red lines). Peaks are also picked by GSD in the *scout* spectrum (second panel). Peaks are then matched between the *reference* and *scout* spectra within the defined region *A*. For *T1rho* spectra (third panel), both spectra of the pair are sum-integrated within region *A* to calculate the peak change (integral areas are colored in blue and red). This process is repeated for all of the *reference* peaks, and the average peak change is used to judge the binding status of the corresponding compound. If the difference spectrum is used to analyze STD experiments (as shown in the bottom panel), the peaks in both the *reference* and difference spectra are picked using GSD and matched using the same region *A* as a tolerance. (b) For ^{19}F-detected experiments, peaks picked by GSD in the *reference* spectrum (top panel) are matched to those peaks picked by GSD in the *blank* spectrum (middle panel) within region *A*. The peak of the *blank* spectrum is then matched to the *protein* spectrum (bottom panel) using the narrower region *B* outlined by magenta lines. The same region *B* is utilized to sum-integrate the peaks in the *blank* and *protein* spectra (areas colored in magenta).

Information). This ensures that changes in peak integrals can be correctly quantified in the automatic analysis. Second, the *reference* spectra are normalized to allow proper visualization when displayed with the screening spectra. This is achieved by setting the largest peak within the region of interest (ROI) or a designated peak (e.g., the DSS peak) to the same intensity in all of the spectra.

Regions of Interest. The user can define one or more ROIs for which the analysis is performed. This allows the selection of specific regions of the spectrum to be analyzed (e.g., the aliphatic or aromatic region depending on the type of ligand being assessed) and the exclusion of solvent and buffer peaks, which should not be taken into account for data analysis.

Method for Automated Binding Detection Based on Changes in Peak Signals. The principles applied in manual assessment of compound binding by ligand-observed NMR spectroscopy (Figure S1) form the foundation of the algorithms used in the automated analysis.

For ^{1}H-detected experiments, the analysis consists of the following steps (Figure 2a):

Step 1: The peaks in the *reference* spectra are picked using the global spectral deconvolution (GSD) algorithm.[20] GSD performs a complete deconvolution of the frequency-domain spectrum and returns a reliable list of peaks and their parameters (including chemical shift, peak intensity, peak integral, line width at half height, and peak shape). Since no assignment procedure to connect peaks with specific protons is needed, a multiplet is treated as several individual peaks. To avoid the use of spectral noise or small signals from impurities for data analysis, peaks with intensities lower than 10% (default value, adjustable) of the maximum peak intensity in the ROI are filtered out.

Step 2: In order to determine whether a compound is present in the screening sample, the program tries to match the *reference* spectrum with the *scout* spectrum. Peaks are picked in the *scout* experiment using GSD and are matched to peaks picked in the *reference* spectrum using a tolerance defined as

$$A = p \pm aw \tag{1}$$

where p is the center of the *reference* peak, w is the peak width at half height, and a is a factor used to compensate for potential peak misalignment. The parameter a is defined by the user and is usually set between 2 and 10 for proton spectra. If the percentage of matched peaks for a given ligand is below a user-defined threshold (typically 40%), the compound is classified as *missing*, and no further analysis is performed on it. Correctly identified peaks are then subjected to subsequent analysis.

Step 3. For each *reference* peak, an integration region *A* is determined according to eq 1 (Figure 2a). The spectral pairs of screening experiments are integrated within region *A* using the conventional summation method, resulting in two integrals, I_1 and I_2. The change in peak integral, ΔI, is then calculated as follows:

DOI: 10.1021/acs.jmedchem.6b00019
J. Med. Chem. 2016, 59, 3303−3310

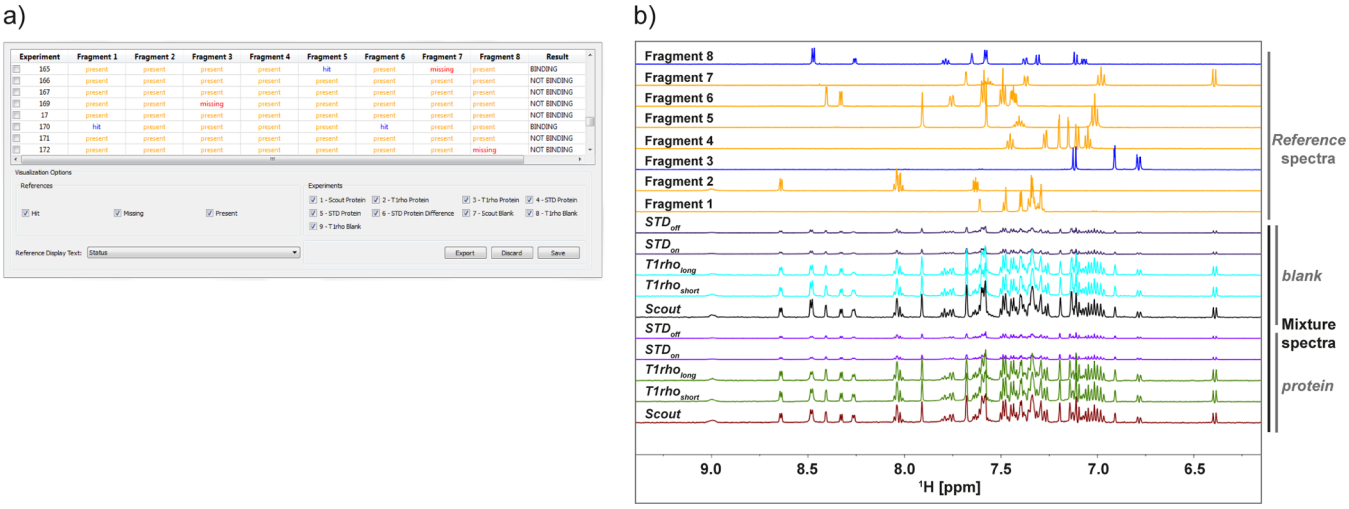


Figure 3. Visualization of the analysis results. (a) The *Results Editor* displays the status of all of the ligands (columns "Fragment 1" to "Fragment 8") in all of the samples (column "Experiment"). The status is color-coded as red, orange, or blue for *missing*, *present*, or *hit*, respectively. Pull-down menus allow the display of other results such as the average/maximum change of peak integrals, percent of matched peaks, and name of each ligand. Double clicking on an "Experiment" field displays the corresponding spectra document that contains the single and stacked views for visual inspection. (b) The stacked view contains all of the relevant spectra of the binding experiments as well as the *reference* spectra, which are displayed in the same color scheme as used in the *Results Editor*. Individual spectra are labeled on the left side for clarity. The result for each compound can be interactively modified in the *Results Editor* as well as the spectral overview.

$$\Delta I = \begin{cases} \dfrac{I_1 - I_2}{I_1} & \text{if } I_1 > I_2 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

This process is repeated for all of the *reference* peaks in the ROI. The average peak change, $\Delta\bar{I}$, is calculated as follows:

$$\Delta\bar{I} = \sum_{i=1}^{n} \Delta I_i / n \quad (3)$$

where n is the number of *reference* peaks.

In cases of large local peak misalignment, i.e., a peak is shifted completely out of the integration region A in the spectral pair of the binding experiment, noise in both spectra will be integrated, producing a meaningless value of ΔI. Since this would compromise the analysis, the situation is addressed by discarding all ΔI values for which the maximum intensity in either spectrum does not exceed 10% (default value, adjustable) of the maximum peak intensity found in the ROI.

Step 4: If a *blank* experiment (e.g., STD or T1ρ pairs acquired on samples without the protein) is used as a control, the average peak change is calculated according to eqs 2 and 3 and is subtracted from the average peak change for the corresponding experiment with protein added.

Step 5: For each ligand that is not *missing*, the determined average peak change $\Delta\bar{I}$ is compared with a threshold T to classify the ligand as follows:

If $\Delta\bar{I} < T$, the ligand is not a hit and is called *present*
If $\Delta\bar{I} \geq T$, the ligand is called a *hit*

T is an empirical threshold whose value can be changed on the basis of experience. Typically, T values of 2−20% and 10−40% are used for STD and T1ρ experiments, respectively.

When the difference spectrum of the STD experiment is used instead of the STD pair, the difference peaks are picked by GSD, and their peak centers are matched with the *reference* peaks within the tolerance defined in eq 1 (bottom panel in Figure 2a). Next, the percentage of matched peaks for each *reference* spectrum is calculated and used to judge whether a compound is a hit. It should be noted that important information, i.e., the quantification of the signal change, is lost when the difference spectrum is used. In addition, the difference spectra tend to be noisy (low signal-to-noise ratio), and the peaks can be hard to distinguish from the noise unless the observed changes are significant. Therefore, whenever possible, the analysis of off/on-resonance spectral pairs instead of their difference spectra is preferred in the automatic mode.

For ^{19}F-detected experiments, the analysis consists of the following steps (Figure 2b):

Step 1: The peaks are picked using GSD in all of the spectra (*reference* and screening spectra). Similar to the ^{1}H case, the user can define a threshold to filter out peaks representing spectral noise or small signals from impurities.

Step 2: For each *reference* spectrum, the picked *reference* peaks are matched to the closest GSD peaks in the first screening spectrum (*blank*) within a tolerance as defined in eq 1. A larger tolerance (e.g., a = 15) has to be applied to ^{19}F peaks because peak misalignments between the *reference* and screening spectra can be greater than those of ^{1}H peaks as a result of the higher sensitivity of fluorine nuclei to their environment. If the percentage of matched peaks for this *reference* spectrum is below a user-defined value (typically 60%, guaranteeing that at least one peak is matched), the corresponding compound is classified as *missing*, and further analysis is not pursued.

Step 3: For each peak in the *blank* spectrum that is matched to a *reference* peak in step 2, a narrower integration region B is determined using eq 4:

$$B = p \pm bw \quad (4)$$

where p is the center of the peak in the *blank* spectrum and w is the peak width at half height. The parameter b is automatically calculated as a/5, with the upper and lower limits for b being 4 and 2, respectively. This region is then used to sum-integrate the peaks in both the *blank* and *protein* spectra. Utilizing two different factors a and b ensures that a larger region (A) is used to match peaks between the *reference* and *blank* spectrum but a

DOI: 10.1021/acs.jmedchem.6b00019
J. Med. Chem. 2016, 59, 3303−3310

smaller region (B) is employed for the peak integration of the screening experiments. This accounts for the observation that there is generally less local peak misalignment between the *blank* and *protein* spectra than between the *reference* and *blank* spectra. At the same time, the smaller integration area reduces the possibility of introducing errors in case there are other peaks in the vicinity.

Step 4: The peak change, ΔI, is calculated according to eq 2 using the values determined in step 3.

Step 5: The average peak change, $\Delta\bar{I}$, is then calculated for the whole *reference* using eq 3, and it is utilized for hit calling using the same criteria as described for the ^{1}H case.

For both ^{1}H and ^{19}F analyses, the program offers a one-window setup where all of the user-defined parameters are easily accessible.

Presentation of Analysis Results for User Verification. Upon completion of the automatic processing and analysis, an interactive table called the *Results Editor* is displayed (Figure 3a). This interface lists the status of all of the ligands in all of the samples, coded in the same colors as in the corresponding *reference* spectra in the stacked view. It also acts as a navigator, since double clicking on an experiment number opens a file saved for each mixture showing the stacked view of all of the relevant spectra (Figure 3b and Figure 4a).

Both the *Results Editor* and the stacked spectra allow the user to review and interactively edit the results conveniently. Moreover, a log file is generated that contains all of the analysis details, such as those for peak picking, peak integration, and calculation of the compound peak changes, which can be useful if the user wants to review the details of the analysis.

Performance Evaluation on NMR Screening Data. Going beyond method development and using sample data for initial testing, we evaluated the performance of the computational tool by analyzing several large in-house sets of fragment binding data, as they are routinely acquired in fragment-based chemistry campaigns. To ensure an unbiased evaluation, the training and test data sets were completely separate. Setting up the automatic analysis, the parameters used by the program are by-and-large transferrable from one data set to another. Critical parameters to adjust for a particular data set are the tolerance factors for peak matching and integration (parameter a for both ^{1}H and ^{19}F) and the thresholds applied for hit calling, which can be determined by performing a small test run with a fraction of the complete data set.

First, we tested an ^{19}F-detected data set of 19 mixtures containing up to 30 fluorinated fragments in each sample. We are particularly interested in analyzing such data quickly and efficiently since such a library and approach can be used as a first, rapid ligandability assessment and to identify early hits in a new campaign. Next, we analyzed ^{1}H-detected T1ρ and STD spectra from two data sets; one consisted of experiments performed on 55 compounds measured as singles, and the other comprised 174 samples containing compound mixtures. For each of the three sets, the data were initially analyzed manually in the course of project support and then analyzed automatically. For the purpose of comparing the outcomes, we used the result of the manual analysis as a benchmark. Compounds identified as "hit" in the manual analysis but determined as *present* by the automatic analysis are termed "false negatives". Conversely, compounds picked up as *hit* by the automatic process but called "no hit" when the data were assessed manually are termed "false positives". While we decided to employ this convention to compare the results, we point out that there might have been mistakes in our manual analysis due to peak overlap or other adverse effects in the spectra where drawing a clear conclusion proved to be challenging.

^{19}F-Detected Mixture Screening. Out of the 551 compounds evaluated in the screen, the manual and automatic analyses came to the same conclusion for 537 compounds, representing an overall agreement of 97% and demonstrating excellent performance (Table 1). An example of the screening data and the numerical analysis is shown in Figure 4b.

Table 1. Results of Comparisons between Manual and Automatic Data Analyses

category	manual	automatic
^{19}F-Detected Mixture Screen: Overall Agreement 97%		
compounds evaluated	551	551
outcome "no hit"/*present* or *missing*	487	491
outcome "hit"/*hit*	64	60
overlapping "hit"/*hit*	56/64 (88%)	56/60 (93%)
"false positives"		4/60 (7%)
"false negatives"		8/64 (13%)
^{1}H-Detected Single Compound Confirmation: Overall Agreement 91%		
compounds evaluated	55	55
outcome "no hit"/*present* or *missing*	24	21
outcome "hit"/*hit*	31	34
overlapping "hit"/*hit*	30/31 (97%)	30/34 (88%)
"false positives"		4/34 (12%)
"false negatives"		1/31 (3%)
^{1}H-Detected Mixture Screen: Overall Agreement 93%		
compounds evaluated	1240	1240
outcome "no hit"/*present* or *missing*	1190	1132
outcome "hit"/*hit*	50	108
overlapping "hit"/*hit*	34/50 (68%)	34/108 (31%)
"false positives"		74/108 (69%)
"false negatives"		16/50 (32%)

Eight of the 64 "hits" were not picked up by the automatic procedure (13% "false negative" rate), whereas four compounds (out of 60) were classified as *hits* by the automatic analysis but not identified as "hits" manually (7% "false positive" rate). To understand these different conclusions, we hypothesized that the misclassified compounds display values of signal reduction close to the chosen hit-calling threshold of 40% (signal reduction upon protein addition). Indeed, an assessment of where both "false positive" and "false negative" compounds are located in a correlation plot showed that most of them are within ±10% of the threshold (Figure S4a in the Supporting Information). One way to address this is to inspect the automatically generated results list and manually assess compounds that are close to the threshold. Molecules of interest can then be selected to be included in the follow-up work. We also evaluated the large discrepancy observed for outliers and found that it can be attributed to peak inconsistencies between the *reference* and $CPMG_{blank}$ spectra as well as partial peak overlap (Figure S4b,c).

^{1}H-Detected Single-Compound Confirmation Experiments. Sets of single-compound confirmation experiments are typically generated in a step subsequent to mixture screening as well as during analoguing and iterative chemistry follow-up cycles. The presence of only one compound in the

DOI: 10.1021/acs.jmedchem.6b00019
J. Med. Chem. 2016, 59, 3303−3310

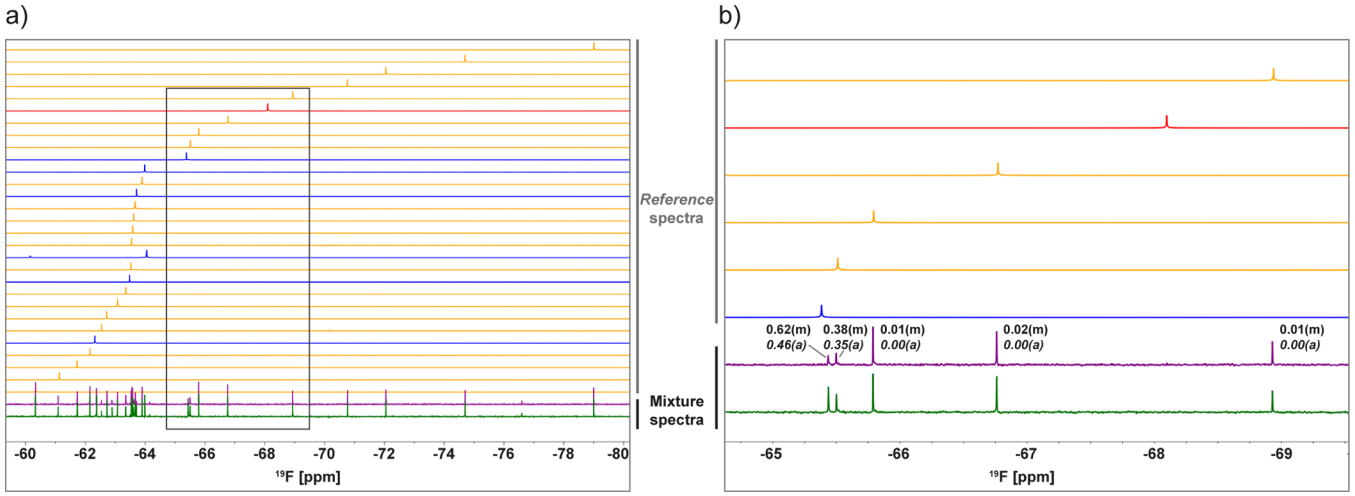


Figure 4. Results of ^{19}F mixture screening. (a) Stacked view with the mixture screening spectra without protein (green) and with protein (purple) shown at the bottom and all of the ^{19}F *reference* spectra of the individual compounds in the mixture shown above. The *reference* spectra are ordered by decreasing chemical shift values from left to right and colored according to the scheme described in Figure 3, with *hits* shown in blue. (b) Zoom into the region of the stacked view marked by the box shown in panel (a), showing only *reference* spectra for the compounds in this area. The respective signal reduction values are indicated next to the mixture peaks, with the values measured in the manual (m) and automatic (*a*) analyses written in regular and *italic* font, respectively. For all of the compounds shown, the manual and automatic analyses were in agreement.

sample results in moderate complexity of the spectrum and generally little overlap of NMR signals. However, large numbers of samples often make this step very time-consuming. For the 55 compounds analyzed in our test set, the overall agreement of the manual and automatic analyses is 91% (Table 1). The number of overlapping hits is 30, representing 97% (of the 31 manual "hits"), and resulting in a very low "false negative" rate of 3% (one compound out of the 31 manual "hits"). The "false positive" rate was also low at 12% (four compounds out of the 34 *hits*).

^{1}H-Detected Mixture Screening. In this data set, a total of 1240 fragments in 174 mixtures (each containing up to eight compounds) were analyzed. This type of data is by far the most difficult and time-consuming to analyze manually, mainly because of the high complexity and resulting peak overlap but also because of the large amount of data usually generated in such primary screening campaigns. To assess whether the peak matching performs robustly on such data, we first analyzed the percentage of peaks matched between the *reference* and *scout* spectra. Of the 1240 compounds, we found that 1154 compounds (93%) have 40% or more of their peaks matched and that 50% of the compounds (622) have 70% or more of their peaks matched, showing good performance. Regarding the evaluation of binding, the manual and automatic analyses came to the same conclusion for 1150 compounds, i.e., the "hit"/*hit* or "no hit"/*present* or *missing* classifications were identical (Table 1). This represents an overall agreement of 93%, a promising result considering the factors described above. In the manual analysis, 50 fragments were identified as "hits", whereas the automatic analysis classified 108 compounds as *hits*. There was an overlap of 34 compounds between the manual and automatic analyses, representing an agreement of 68%. The "false positive" rate of 69% is relatively high, but it is possible that there are bona fide hits in this set. The setup described above facilitates the inspection of results, enabling the user to conveniently re-evaluate such compounds in a quick and easy manner. When we evaluated the "false positives" in more detail, we determined that the majority of compounds (over one-third) read out incorrectly because some of their peaks overlap with those of real hits and that approximately one-third of the compounds show signal changes just above the threshold for hit calling. However, we were also able to rescue 10 compounds that were missed by the manual analysis. The "false negative" rate for this data set was determined to be 32%, representing 16 of the 50 manually found "hits". Since it is possible that the automatic analysis missed real hits in this case, we assessed these compounds in more detail to understand the reason for the observed disconnect. In the majority of cases, the manual hit calling was based on a single ligand peak signal that showed a change above the threshold in the binding experiments, although other peaks of the same ligand were either less affected or difficult to assess because of signal overlap. In our experience, such compounds often do not validate well in subsequent experiments, making the loss of such molecules acceptable. Since the automatic analysis is based on the average change across all of the compound peaks, this value was determined to be under or close to the threshold and therefore not picked as a *hit*.

CONCLUSIONS

Although ligand-observed NMR experiments are well-established and extensively employed to assess the binding of ligands to receptors, currently no software is generally available to automatically analyze such data. Many laboratories rely on in-house-written scripts and software modules to aid the analysis, which usually only perform one particular step of the process and need continued modifications or are not maintained over time. This study describes a robust and streamlined data handling platform that performs all of the analysis steps in a fully automated fashion while also enabling convenient user-based validation of the outcome. The common difficulties encountered in real-world data analysis, such as peak misalignment, low signal-to-noise ratio, and interference from impurities and noise have been largely addressed in the algorithms to improve the robustness of the analysis. The automatic analysis provides confident hit lists that overlap with the results obtained by manual analysis. Comparing the overall agreement of the automatic and manual analyses, we found that

DOI: 10.1021/acs.jmedchem.6b00019
J. Med. Chem. 2016, 59, 3303−3310

the program works reliably on all data sets with an agreement of more than 90% across all of the analyzed compounds. The automatic analysis performs particularly well at analyzing single-compound experiments as well as ^{19}F-based mixture screening data. For the analysis of ^{1}H-detected data of compound mixtures, the comparison yields a lower overlap of identified hits. This is mainly due to the significant peak overlap of the individual compounds and to the higher intrinsic complexity of such data. The outlined analysis process is currently based on a defined integration region and thus performs particularly well when peaks are isolated but can introduce various degrees of error when peaks of different ligands overlap. This limitation is currently addressed by peak averaging, but this is certainly an area that will benefit from further development and improvements, e.g., implementation of weighted averages (where isolated peaks obtain a higher weighing factor) and logic-based assessment of peak changes (where the program would check for consistency of changes across all ligand peaks). In our experience, however, the automatic analysis still provides robust hit lists and enables the identification of otherwise-missed compounds in an efficient manner using the visualization and verification tool. Importantly, the observed discrepancies are found for compounds clustered around the threshold value used in hit calling, where the manual analysis is also more subjective and difficult. Rather than conducting an a priori analysis of large data sets, the user is given the option to use a highly enriched starting point, to subsequently qualify the outcome, and finally to check questionable cases. The data organization and display including all of the reference spectra are of invaluable help for validating results and for making decisions in cases where compound signals are close to each other.

Another important aspect of the automated analysis is the significant time savings. Automatic analyses of data sets such as the ones used in this study take between 30 min (^{19}F mixtures and ^{1}H single-compound data sets) and less than 3 h (^{1}H mixtures data set) when run on a standard PC laptop, whereas approximately half a day to multiple days of user time is required to perform manual analyses with libraries of the sizes presented here.

The features described in this article enable a more streamlined incorporation of NMR-based binding assessment in hit generation. It becomes more feasible to assess larger compound libraries to cover a more diverse chemical space, to screen focused libraries or analogue sets, and to implement short turnaround times in iterative chemistry cycles in compound optimization.

EXPERIMENTAL SECTION

Computational Methods. The program consists of several scripts that were developed using the programming language ECMAScript (JavaScript). It is called "Mnova Screen" and is compatible with the use as a special-functionality plugin within the Mnova software package. It can be run with or without the graphical user interface. Mnova is a general purpose, vendor-format-independent 1D and 2D NMR data processing and analysis program and is compatible with Windows, Macintosh, and Linux operating systems as well as mobile devices.[21]

Sample Preparation and NMR Experiments. All of the spectra were recorded on Bruker NMR spectrometers at 600 MHz ^{1}H Larmor frequency equipped with 5 mm QCI cryogenic probes at 296 K. Samples contained a volume of 170 μL in 3 mm NMR tubes, and a Bruker SampleJet was used as an autosampler during the measurements.

One-dimensional ^{1}H and ^{19}F reference spectra of individual library fragments were measured at a nominal concentration of 1 mM in 50 mM d-Tris (pH 7.6), 100 mM sodium chloride, 20% (v/v) D_2O, and 0.111 mM DSS (corresponding to a signal intensity of a compound present at 1 mM, used as an internal concentration reference). The compounds were obtained as stock solutions dissolved in deuterated dimethyl sulfoxide (DMSO-d_6), and the final DMSO concentration in the samples was 2% (v/v).

^{1}H-observed single-compound screening samples were measured in buffer containing 50 mM d-Tris (pH 7.4), 150 mM sodium chloride, 10% (v/v) D_2O, 5 mM dithiothreitol (d-DTT), and 0.022 mM DSS (corresponding to 200 μM signal intensity). The compound concentration was 200 μM, and the final DMSO-d_6 concentration was 0.4% (v/v). The conditions for ^{1}H-detected mixture samples were identical, except that the samples were measured in mixtures of eight compounds at 200 μM each, which corresponds to a total compound concentration of 1.6 mM and a final DMSO-d_6 concentration of 3.2% (v/v). NMR experiments were performed in the absence and presence of 10 μM target protein (MW = 31.7 kDa) and included a 1D ^{1}H spectrum, two T1ρ spectra with 10 and 200 ms spin-lock duration (each recorded with 128 transients and 8192 points), and STD spectra (acquired with 4 s saturation at −20 000 and 0 Hz offset for the off- and on-resonance experiments, respectively, and recorded with 32 transients and 32 768 points).

^{19}F-detected NMR screening samples contained up to 30 compounds, with each compound being present at a final concentration of 10 μM. The buffer contained 10 mM sodium phosphate (pH 7), 40 mM potassium chloride, and 0.1 mM EDTA, and the final DMSO-d_6 concentration was 0.4% (v/v). ^{19}F spectra measured on screening samples were recorded with 512 transients and 16 384 points, applying a CPMG filter with a total length of 250 ms as well as proton decoupling during acquisition. After the blank spectra (compound mixtures in buffer) were measured, the target protein (MW = 24 kDa) was added to the samples to a final concentration of 5 μM before remeasuring.

Analysis of NMR Binding Experiments. Analysis of ligand-observed NMR binding experiments is based on identifying various changes in compound peak signals upon addition of the protein target.

For ^{1}H-detected experiments, we describe the analysis of T1ρ and STD experiments (Figure S1a). T1ρ experiments are measured as a pair comprising short (10 ms, T1ρ_{short}) and long (200 ms, T1ρ_{long}) spin-lock durations. If peaks experience a signal decrease from T1ρ_{short} to T1ρ_{long} above a set threshold in the sample containing protein, the corresponding compound is qualified as a hit. Peaks that do not change significantly from T1ρ_{short} to T1ρ_{long} indicate that the corresponding ligand does not interact with the target. Next, the affected peaks need to be matched with the corresponding compound, and this is achieved by stacking and comparing the individual library compound spectra (reference spectra) with the mixture spectra. The analysis of STD experiments follows the same concept when pairs of off- and on-resonance spectra are used. However, in manual analyses it is often preferred to look at the difference spectrum and compare the difference peaks with the reference spectra to identify which compounds are hits. In addition, the ligand-observed NMR binding experiments described above are often measured in the absence of protein as well. The signal reduction, if any, is then subtracted from the value determined in the protein sample to account for potential compound aggregation.

For ^{19}F-detected experiments, a CPMG experiment is performed without and with protein. Peak changes between the two samples are assessed, and if a peak decreases from CPMG$_{blank}$ to CPMG$_{protein}$ by more than a set threshold, the compound is determined as a hit (Figure S1b). Conversely, no or small peak changes indicate that the compound does not bind to the target and therefore is not a hit. As above, the identity of the compound is revealed by comparing the spectra of the screening samples with the reference spectra of the individual compounds.

During the manual analysis, spectra were processed using Topspin (Bruker) and then overlaid with the individual reference spectra using an in-house script (Novartis). Spectra were evaluated by eye in

DOI: 10.1021/acs.jmedchem.6b00019
J. Med. Chem. 2016, 59, 3303−3310

Topspin, and hits were determined on the basis of individual scaling of spectra (to compare ^{19}F CPMG and ^{1}H T1ρ spectra) or by assessing the difference spectrum in STD experiments. For the comparative study described in this article, the aromatic region of the ^{1}H spectrum (5−10 ppm) was assessed in both the manual and automatic analyses. The entire spectral window (−84 to −54 ppm) was assessed for ^{19}F spectra.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jmedchem.6b00019.

Figures S1−S4 (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*Address: MestreLab Research, 1420 Timber Glen, Escondido, CA 92027, USA. E-mail: chen.peng@mestrelab.com. Phone: +1 760 317 5715.

*E-mail: andreas.lingel@novartis.com. Phone: +1 510 879 9410.

Present Address

§A.B.: Novartis Institutes for BioMedical Research, Basel, Switzerland.

Author Contributions

∥C.P. and A.F. contributed equally. The manuscript was written through contributions of all authors. A.L., A.F., and C.P. wrote the manuscript; C.P., M.P., C.C., and S.D. developed the software tool; A.L., A.B., and A.F. carried out the experimental work and manual analyses; and A.F. and A.L. performed the comparison between the manual and automated analyses. All of the authors have approved the final version of the manuscript.

Notes

The authors declare the following competing financial interest(s): C.P., M.P., C.C., and S.D. are employees of MestreLab Research, and A.F., A.B., and A.L. are employees of Novartis.

■ ACKNOWLEDGMENTS

A.L., A.F., and C.P. thank Xiaolu Zhang and Jasna Fejzo for early work on the collaboration. A.L. and A.F. thank Andreas O. Frank for feedback and informal discussions and Andrew Proudfoot for reading the manuscript. C.P. thanks Jiangli Yan and Wei Wang at Pfizer La Jolla for insightful discussions and exchange of ideas during the initial development stages.

■ ABBREVIATIONS USED

NMR, nuclear magnetic resonance; GSD, global spectral deconvolution; STD, saturation transfer difference; CPMG, Carr−Purcell−Meiboom−Gill; WaterLOGSY, water-ligand observed via gradient spectroscopy; DSS, 4,4-dimethyl-4-silapentane-1-sulfonic acid; DMSO, dimethyl sulfoxide; DTT, dithiothreitol; ROI, region of interest

■ REFERENCES

(1) Campos-Olivas, R. NMR Screening and Hit Validation in Fragment Based Drug Discovery. *Curr. Top. Med. Chem.* **2011**, *11*, 43−67.

(2) Coles, M.; Heller, M.; Kessler, H. NMR-Based Screening Technologies. *Drug Discovery Today* **2003**, *8*, 803−810.

(3) Lepre, C. A.; Moore, J. M.; Peng, J. W. Theory and Applications of NMR-Based Screening in Pharmaceutical Research. *Chem. Rev.* **2004**, *104*, 3641−3676.

(4) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science* **1996**, *274*, 1531−1534.

(5) Stockman, B. J.; Dalvit, C. NMR Screening Techniques in Drug Discovery and Drug Design. *Prog. Nucl. Magn. Reson. Spectrosc.* **2002**, *41*, 187−231.

(6) Betz, M.; Saxena, K.; Schwalbe, H. Biomolecular NMR: A Chaperone to Drug Discovery. *Curr. Opin. Chem. Biol.* **2006**, *10*, 219−225.

(7) Harner, M. J.; Frank, A. O.; Fesik, S. W. Fragment-Based Drug Discovery Using NMR Spectroscopy. *J. Biomol. NMR* **2013**, *56*, 65−75.

(8) Dalvit, C. NMR Methods in Fragment Screening: Theory and a Comparison with Other Biophysical Techniques. *Drug Discovery Today* **2009**, *14*, 1051−1057.

(9) Dalvit, C.; Pevarello, P.; Tatò, M.; Veronesi, M.; Vulpetti, A.; Sundström, M. Identification of Compounds with Binding Affinity to Proteins via Magnetization Transfer from Bulk Water. *J. Biomol. NMR* **2000**, *18*, 65−68.

(10) Gossert, A. D.; Henry, C.; Blommers, M. J. J.; Jahnke, W.; Fernández, C. Time Efficient Detection of Protein-Ligand Interactions with the Polarization Optimized PO-WaterLOGSY NMR Experiment. *J. Biomol. NMR* **2009**, *43*, 211−217.

(11) Hajduk, P.; Olejniczak, E.; Fesik, S. W. One-Dimensional Relaxation-and Diffusion-Edited NMR Methods for Screening Compounds That Bind to Macromolecules. *J. Am. Chem. Soc.* **1997**, *119*, 12257−12261.

(12) Mayer, M.; Meyer, B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chem., Int. Ed.* **1999**, *38*, 1784−1788.

(13) Deverell, C.; Morgan, R. E.; Strange, J. H. Studies of Chemical Exchange by Nuclear Magnetic Relaxation in the Rotating Frame. *Mol. Phys.* **1970**, *18*, 553−559.

(14) Dalvit, C.; Fagerness, P. E.; Hadden, D. T.; Sarver, R. W.; Stockman, B. J. Fluorine-NMR Experiments for High-Throughput Screening: Theoretical Aspects, Practical Considerations, and Range of Applicability. *J. Am. Chem. Soc.* **2003**, *125*, 7696−7703.

(15) Dalvit, C.; Flocco, M.; Veronesi, M.; Stockman, B. J. Fluorine-NMR Competition Binding Experiments for High-Throughput Screening of Large Compound Mixtures. *Comb. Chem. High Throughput Screening* **2002**, *5*, 605−611.

(16) Dalvit, C.; Mongelli, N.; Papeo, G.; Giordano, P.; Veronesi, M.; Moskau, D.; Kümmerle, R. Sensitivity Improvement in ^{19}F NMR-Based Screening Experiments: Theoretical Considerations and Experimental Applications. *J. Am. Chem. Soc.* **2005**, *127*, 13380−13385.

(17) Vulpetti, A.; Hommel, U.; Landrum, G.; Lewis, R.; Dalvit, C. Design and NMR-Based Screening of LEF, a Library of Chemical Fragments with Different Local Environment of Fluorine. *J. Am. Chem. Soc.* **2009**, *131*, 12949−12959.

(18) Vulpetti, A.; Dalvit, C. Design and Generation of Highly Diverse Fluorinated Fragment Libraries and Their Efficient Screening with Improved ^{19} F NMR Methodology. *ChemMedChem* **2013**, *8*, 2057−2069.

(19) Wong, J. W. H.; Durante, C.; Cartwright, H. M. Application of Fast Fourier Transform Cross-Correlation for the Alignment of Large Chromatographic and Spectral Datasets. *Anal. Chem.* **2005**, *77*, 5655−5661.

(20) Cobas, C.; Seoane, F.; Domínguez, S.; Sykora, S. A New Approach to Improving Automated Analysis of Proton NMR Spectra through Global Spectral Deconvolution (GSD). *Spectrosc. Eur.* **2011**, *23*, 26−30.

(21) Cobas, C.; Iglesias, I.; Seoane, F. NMR Data Visualization, Processing, and Analysis on Mobile Devices. *Magn. Reson. Chem.* **2015**, *53*, 558−264.

DOI: 10.1021/acs.jmedchem.6b00019
J. Med. Chem. 2016, 59, 3303−3310