CHEMBIOCHEM

Supporting Information

Development and Validation of 2D Difference Intensity Analysis for Chemical Library Screening by Protein-Detected NMR Spectroscopy

[John M. Egner](http://orcid.org/0000-0003-2572-1368)^{[+](http://orcid.org/0000-0003-2572-1368)}, Davin R. Jensen⁺, Michael D. Olp, Nolan W. Kennedy, Brian F. Volkman, Francis C. Peterson, Ass. Prof. Brian C. Smith, and [R. Blake Hill](http://orcid.org/0000-0001-6228-3905)*^[a]

cbic_201700386_sm_miscellaneous_information.pdf

Supplemental material content

Supplemental Figure 1: Comparison of carbon- and nitrogen-based methods for hit detection against CXCL12.

Supplemental Figure 2: Evaluation of baseline noise intensity levels between fragment mixture samples show random behavior.

Supplemental Figure 3: Classification of hits against targets.

Supplemental Figure 4: Difference intensity analysis applied to CXCL12 "empty" region and individual fragment mixture screening spectra.

Supplemental Figure 5: FTMap analysis of Fis1 solution structure ensemble.

Supplemental Figure 6: Comparison of carbon- and nitrogen-based methods for hit detection against Fis1.

Supplemental Figure 7: Size distribution analysis by Dynamic Light Scattering suggests all 6 interfering fragments induce protein aggregation.

Supplemental Figure 8: Deconvoluted mass spectra of Fis1 with and without 6 interfering compounds revealed 2-methylbenzene-1,4-diol directly modifies Fis1.

Supplemental Figure 9: Comparison of estimated false positive and negative rates using PCA and DIA against all targets screened.

Supplemental Figure 10: PCA score plots for CXCL12, BRD4(BD2), and Fis1 excluding aggregation-causing fragment mixtures.

Supplemental Figure 11: PCA score plots from 2D difference spectra for CXCL12, BRD4(BD2), and Fis1 excluding aggregation-causing fragment mixtures.

Supplemental Figure 12: Schematic of compound library handling.

Supplemental Methods and References: Recombinant Fis1 and BRD4(BD2) purifications.

Supplemental Table 1: CXCL12 Positive and Negative summations from difference intensity analysis.

Supplemental Table 2: BRD4(BD2) Positive and Negative summations from difference intensity analysis.

Supplemental Table 3: Fis1 Positive and Negative summations from difference intensity analysis.

Supplemental Table 4: Compounds that induced NMR line broadening, not indicative of binding events against Fis1.

Supplemental Table 5: Aggregator-prone fragments identified in the Zenobia library using Aggregation Advisor computational filter.

Supplemental Figures

Supplemental Figure 1: Comparison of carbon- and nitrogen-based methods for hit detection against CXCL12. (**A**) Nitrogen-based screening methods allow for detectable chemical shift perturbations in response to binding, whereas (**B**) carbon-based screening methods did not. ¹H-¹⁵N SOFAST-HMQC spectral overlays (A) and ¹H-¹³C HSQC spectral overlays (**B**) of indicated multiplexed samples for CXCL12 are shown.

Supplemental Figure 2: Evaluation of baseline noise intensity levels between fragment mixture samples show random behavior. (**A**) Using NMRDraw[25], baseplane noise level was measured for each individual CXCL12 fragment mixture sample. The dotted red line indicates the mean of the baseplane noise levels. (**B**) Using the computed mean from **A**, individual z-scores for each CXCL12 fragment mixture sample was calculated. The z-score equation used for calculation was as follows: $z = (x - \mu) / s$, where $x =$ measured noise intensity, $\mu =$ mean noise intensity, and s = standard deviation. (**C**) An analysis of variance (ANOVA) was used to determine if baseplane noise level intensity differed between CXCL12 fragment mixture spectra; ANOVA fitting residuals are shown. The ANOVA results indicate no statistical significance between spectral baseplane noise levels of fragment mixture samples (p = 0.0579).

Supplemental Figure 4: Total intensity analysis of CXCL12 screening spectra (before subtracting DMSO control spectrum to generate DIA). The error estimated in the intensities is 6% from five consecutive HSQCs on a control sample.

Supplemental Figure 5: FTMap analysis of Fis1 solution structure ensemble. The strongest ranked hot spot resulting from FTMap analysis on each Fis1 solution states 1-20 is shown. Each hot spot is colored separately and highlighted on Fis1 solution state 1 for comparison. The number of probes (CS score) from each ensemble state is tabulated.

Supplemental Figure 6: Comparison of carbon- and nitrogen-based methods for hit detection against Fis1. (**A**) Nitrogen-based screening methods allow for detectable chemical shift perturbations in response to binding, whereas (**B**) carbon-based screening methods did not. ¹H-¹⁵N SOFAST-HMQC spectral overlays (A) and ¹H-¹³C HSQC spectral overlays (**B**) of indicated multiplexed samples for Fis1 are shown.

A)

Supplemental Figure 9: Comparison of estimated false positive and negative rates using PCA and DIA against all targets screened. (**A**) The confusion matrices for each target and analytical method (PCA or DIA) were computed. To be considered an actual hit, the initial fragment mixture had to contain chemical shit perturbations that localized to a putative binding pocket (i.e. unique structural region). Note, we used equal thresholding levels between PCA and DIA at greater than or less than 68% confidence interval (CI) or 1σ (**B**) The confusion matrices shown in **A** were used to approximate estimated false positive and negative rates $[33]$, $[34]$ based on the following equations. The estimated false positive rate (FPR) and false negative rate (FNR) were calculated by the following: FPR = \sum false positive hit / \sum actual hit condition, where false positive hits were classified as a hit by PCA or DIA thresholding, but found to be an actual non-hit; FNR = \overline{p} false negative hit / \overline{p} actual non-hit condition, where false negative hits were classified as

a non-hit by PCA or DIA thresholding, but found to be an actual hit. Note, we did not collect individual spectra of any individual protein target against all fragments in our library and therefore, the reported false positive and negative rates are estimations.

Supplemental Figure 10: PCA score plots for CXCL12, BRD4(BD2), and Fis1 excluding aggregation-causing fragment mixtures. The PCA score plots are annotated with 68% (blue ellipse) and 95% (red ellipse) confidence intervals of datasets for (**A**) CXCL12 SOFAST-HMQC, (**B**) BRD4(BD2) HSQC, and (**C**) Fis1 SOFAST-HMQC spectra. For each target, fragment mixtures classified as potential hits are highlighted in red. For CXCL12 dataset, principal components were calculated after removing multiplexed samples #20, 21, 22, 41, 42, and 48. For BRD4(BD2) dataset, principal components were calculated after removing multiplexed samples #09, 14, 16, 18-22, 48, and 59. For Fis1 dataset, principal components were calculated after removing multiplexed samples #09, 21, 22, 48, and 59. PCA plots using R or Bruker software gave similar, but not identical results indicating differences in the methods each software package uses for calculating principal components.

Supplemental Figure 11: PCA score plots from 2D difference spectra for CXCL12, BRD4(BD2), and Fis1 excluding aggregation-causing fragment mixtures. The PCA score plots are annotated with 68% (blue ellipse) and 95% (red ellipse) confidence intervals of datasets for (**A**) CXCL12 SOFAST-HMQC, (**B**) BRD4(BD2) HSQC, and (**C**) Fis1 SOFAST-HMQC spectra. For each target, fragment mixtures classified as potential hits are highlighted in red. For CXCL12 dataset, principal components were calculated after removing multiplexed samples #20, 21, 22, 41, 42, and 48. For BRD4(BD2) dataset, principal components were calculated after removing multiplexed samples #09, 14, 16, 18- 22, 48, and 59. For Fis1 dataset, principal components were calculated after removing multiplexed samples #09, 21-22, 48, and 59.

Supplemental Figure 12: Schematic of compound library handling. For multiplexing compounds, 10 µL of each compound (352 each at 200 mM DMSO- d_6) was multiplexed from mother plates 1-4 into 59 wells of a 96-well v-bottom plate with each well containing six chemical fragments at 33.3 mM. A 60th well contained DMSO only. This step was performed manually resulting in a "mother matrix plate" with 60 wells containing a total of 60 μ L of fragments (or DMSO) in DMSO-d₆. Then, 7.2 μ L of each fragment mixture (or DMSO) from the "mother matrix plate" was transferred to single-use "daughter matrix plates" 96-well v-bottom plates. Mother and daughter plates were sealed, stored at -80 °C, and available for single-use screening purposes.

Supplemental Methods

Recombinant Fis1 purification

Recombinant ¹⁵N-Fis1¹⁻¹²⁵ was expressed in *E. coli* BL21 pRep4 cells. Cells were grown in 1 L minimal media (¹⁵N ammonium chloride, 8 g d-dextrose, solubilized multi-vitamin (CVS pharmacy), 120 μ L 250 mM calcium chloride, and 2 mL 1 M magnesium sulfate) at 37 °C to an OD₆₀₀ \approx 1.5-2.0, induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and given ~18 h for protein expression at 18 °C. Cells were harvested by centrifugation at 5,000 × *g* for 30 minutes, resuspended in Buffer A (20 mM HEPES pH 7.4, 175 mM NaCl, 30 mM imidazole, 0.02% w/v sodium azide), and stored at -80 °C until cell lysis. Prior to cell lysis, a crushed half-tablet of protease inhibitor cocktail (Roche) was added and then lysed by an emulsiflex (Avestin) at 15,000 psi. Cell lysates were cleared by centrifugation at 15,000 × *g* for 45 min. Cleared supernatant was applied onto a His60 Ni superflow resin (5 mL column, Clontech) and thoroughly washed with Buffer A. His₆-SUMO-Fis1¹⁻¹²⁵ was eluted with Buffer B (20 mM HEPES pH 7.4, 175 mM NaCl, 500 mM imidazole, 0.02% w/v sodium azide). His₆-SUMO tag was cleaved using ubiquitin-like-protease-1 (ULP1) during overnight dialysis in buffer C (20 mM HEPES pH 7.4, 175 mM NaCl, 0.1% v/v β-mercaptoethanol) at 4 °C. Cleaved ¹⁵N-Fis1¹⁻¹²⁵ was separated from affinity tag and ULP1 by loading onto Ni-agarose resin and washing with Buffer A. Recombinant $15N-Fis1^{1-125}$ was further purified by size exclusion chromatography on a Superdex S-75 column (GE healthcare) pre-equilibrated with Buffer D (20 mM HEPES pH 7.4, 175 mM NaCl, 2 mM DTT, 0.02% w/v sodium azide) and driven by an AKTA FPLC. Peak fractions were collected and concentrated to ~1 mM using Vivaspin 20 concentrators (GE healthcare) with a molecular weight cutoff of 3 kDa and stored at 4 °C. Typical yields of ¹⁵N Fis1¹⁻¹²⁵ was ~20 mg per 7 g wet cell pellet mass from 1 L minimal media growth. $13C^{15}N$ Fis1 $1-125$ was expressed and purified following same procedure as ¹⁵N Fis1, except ¹³C-glucose was used as sole carbon source.

Recombinant BRD4(BD2) purification

Recombinant BRD4(BD2) was transformed in BL21(DE3) *E. coli* and expressed in 6× Medium P minimal media in the presence of 50 μ g/ml kanamycin to an OD $_{600}$ of 0.6-0.8. Protein expression was induced overnight with 0.1 mM IPTG at 18 °C. Cells were harvested by centrifugation at 5,000 × *g* and re-suspended in 30 mL of lysis buffer (50 mM HEPES pH 7.5 at 20 °C; 500 mM NaCl; 5% v/v glycerol, 2.5 mM Imidazole) supplemented with protease inhibitors (0.3 nM Leupeptin, 1 nM E-64, 1 nM Leupeptin, 1 nM Bestatin, 1 nM Pepstatin and 100 nM PMSF). Cells were disrupted by sonication and lysates were cleared by centrifugation for 30 min at 30,000 × *g*. The lysates were then applied to Ni-NTA resin and rocked for 1-3 h at 4 °C. Ni-NTA resin was applied to a column and washed twice with 50 mL of lysis buffer and protein was eluted using a step elution of imidazole in lysis buffer (5 mL of 50, 100, 150, 200 and 250 mM Imidazole). Fractions were monitored by SDS-PAGE and those containing recombinant protein were dialyzed overnight with his $₆$ -tagged TEV protease into lysis</sub> buffer. The dialyzed sample was then re-applied to the Ni-NTA column to remove the TEV protease and free his₆ tags. The flow-through was then concentrated to a final volume of 1 mL and applied to an Enrich SEC 70 10×300 mm column (Bio-rad) to exchange the buffer to 100 mM sodium phosphate (pH 6.5 at 20 °C), 50 mM NaCl, and 5 mM DTT. Samples were concentrated, flash frozen in liquid nitrogen, and stored at -80 °C until used.

Supplemental References

[1] Zhang, Z.; Marshall, A. G. A Universal Algorithm for Fast and Automated Charge State Deconvolution of Electrospray Mass-to-Charge Ratio Spectra. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 225–233.

Supplemental Tables

Supplemental Table 1: CXCL12 Positive and Negative summations from difference intensity analysis. Raw values are ranked ordered based on summed positive intensity (Pos. sum), summed negative intensity (Neg. sum), and negative:positive ratio (Neg./Pos. ratio). Red and blue line indicate average ± 1 SD threshold for positive and negative intensities, respectively. Average \pm 1 SD for CXCL12 difference intensity analysis is $-(4.7 \pm 0.2) \times 10^6$. Exp. = multiplexed experiment number, Pos. sum = positive intensity summation, Neg. sum = negative intensity summation, Neg./Pos. ratio = Negative intensity summation / Positive intensity summation ratio.

Supplemental Table 2: BRD4(BD2) Positive and Negative summations from difference intensity analysis. Raw values are ranked ordered based on summed positive intensity (Pos. sum), summed negative intensity (Neg. sum), and negative:positive ratio (Neg./Pos. ratio). Red and blue line indicate average ± 1 SD threshold for positive and negative intensities, respectively. Average \pm 1 SD for BRD4(BD2) difference intensity analysis is $-(1.9 \pm 0.6) \times 10^7$. Exp. = multiplexed experiment number, Pos. sum = positive intensity summation, Neg. sum = negative intensity summation, Neg./Pos. ratio = Negative intensity summation / Positive intensity summation ratio.

Supplemental Table 3: Fis1 Positive and Negative summations from difference intensity analysis. Raw values are ranked ordered based on summed positive intensity (Pos. sum), summed negative intensity (Neg. sum), and negative:positive ratio (Neg./Pos. ratio). Red and blue line indicate average ± 1 SD threshold for positive and negative intensities, respectively. Average \pm 1 SD for Fis1 difference intensity analysis is $-(3.4 \pm 0.1) \times 10^6$. Exp. = multiplexed experiment number, Pos. sum = positive intensity summation, Neg. sum = negative intensity summation, Neg./Pos. ratio = Negative intensity summation / Positive intensity summation ratio.

Supplemental Table 4: Compounds that induced NMR line broadening, not indicative of binding events against Fis1. Six compounds (1 mM) – from 18 possible compounds contained in fragment mixtures #21, 22, and 48 – were individually screened against 50 µM 15N-Fis1 and resulted in peak intensity loss and broadening. Only ninhydrin was found to induce line broadening for all protein targets screened; CXCL12, BRD4(BD2), and Fis1.

