



## NMR Metabolomics Protocols for Drug Discovery

Fatema Bhinderwala and Robert Powers

### Abstract

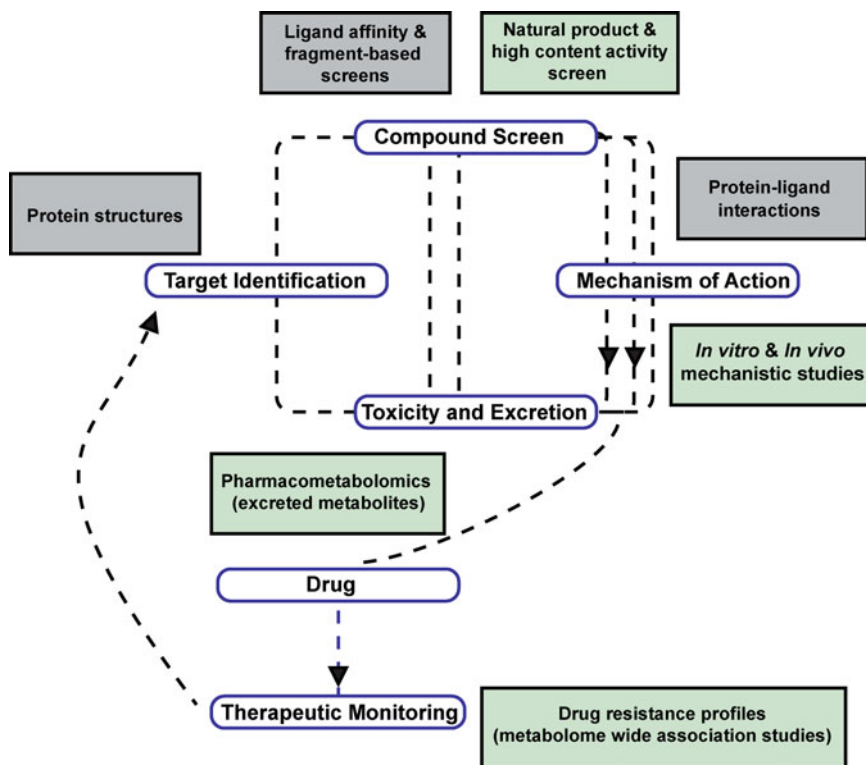
Drug discovery is an extremely difficult and challenging endeavor with a very high failure rate. The task of identifying a drug that is safe, selective, and effective is a daunting proposition because disease biology is complex and highly variable across patients. Metabolomics enables the discovery of disease biomarkers, which provides insights into the molecular and metabolic basis of disease and may be used to assess treatment prognosis and outcome. In this regard, metabolomics has evolved to become an important component of the drug discovery process to resolve efficacy and toxicity issues and as a tool for precision medicine. A detailed description of an experimental protocol is presented that outlines the application of NMR metabolomics to the drug discovery pipeline. This includes (1) target identification by understanding the metabolic dysregulation in diseases, (2) predicting the mechanism of action of newly discovered or existing drug therapies, (3) and using metabolomics to screen a chemical lead to assess biological activity. Unlike other OMICS approaches, the metabolome is “fragile” and may be negatively impacted by improper sample collection, storage, and extraction procedures. Similarly, biologically irrelevant conclusions may result from incorrect data collection, preprocessing or processing procedures, or the erroneous use of univariate and multivariate statistical methods. These critical concerns are also addressed in the protocol.

**Key words** Metabolomics, Chemometrics, NMR, Drug discovery

---

### 1 Introduction

Nuclear magnetic resonance (NMR) spectroscopy is an instrument of choice for drug discovery and design (*see* Fig. 1) [1, 2]. For decades, NMR has provided structural details for druggable protein targets and is routinely used for investigating protein-drug interactions and conducting high-throughput ligand-affinity screens. In this regard, NMR is able to obtain atomic-level resolution of protein-ligand structures in near physiological conditions that results in better quality chemical leads [3, 4]. Accordingly, NMR has enhanced the success rate of the drug discovery process and has made numerous contributions to the development of new therapeutics [5–7]. In addition to its role in structural biology, NMR has also made significant contributions to the field of metabolomics, especially in the application of metabolomics to drug discovery [8–12].



**Fig. 1** A schematic diagram outlines the overall drug discovery process and highlights the role of NMR- and NMR-based metabolomics at every step in the process. Traditional drug discovery roles for NMR are boxed in grey, while NMR metabolomics applications are boxed in green

A traditional view of a disease is as a genome or proteome level dysfunction, but there is a growing appreciation for the role of metabolomic perturbations in a variety of human diseases [13]. For example, an alteration in glucose metabolism is a well-known hallmark of cancer (i.e., Warburg effect), but other metabolic changes are also associated with cancer, such as glutaminolysis, redox imbalance, and flux changes in the tricarboxylic acid (TCA) cycle or the pentose phosphate pathway (PPP) [14]. Similarly, antibiotic resistance in *Staphylococcus aureus* has been shown to result from a re-direction of carbon flow into metabolic processes that counteract antibiotic activity [15, 16]. Another example is mitochondrial metabolism, especially changes in NAD levels, which is an indicator of aging- and age-related disorders [17]. The NAD-sirtuin pathway regulates neuronal protection and is an important process that prevents axon degradation, which is a primary cause of Parkinson's disease and Alzheimer's. Thus, lowered NAD<sup>+</sup> levels is a crucial metabolic observation in neurodegeneration that is being manipulated for disease prevention [15, 17]. Thus, the characterization of metabolite levels in any biological system coupled with the ability to trace the metabolic

fate of individual atoms has the potential to beneficially impact a variety of biomedical needs [12]. Specifically, metabolomics is an invaluable tool of drug discovery, since metabolic changes may be used to identify therapeutic targets, verify *in vivo* drug efficacy, decipher mechanisms of action, and identify potential toxic side effects.

The drug discovery pipeline is an extremely resource intensive endeavor. The process is time-consuming and may require upward of 12 years before a new drug obtains FDA approval [18] and costly with an average expense of \$2.6 billion dollars per approved drug [19]. The process is also prone to a very high failure rate where only a few lead compounds that emerge from traditional medicinal chemistry efforts can be transferred to clinical trials [20]. Furthermore, very few drug leads make it past phase I clinical trials, and most compounds eventually fail in phase III [21]. On average, less than one in ten compounds receive FDA approval and are launched as new drug entities [22]. Of course, FDA-approved drugs are still routinely pulled from the market or encounter safety concerns when broadly prescribed [23]. This low overall success rate is a result of numerous problems routinely encountered throughout the drug discovery process [24, 25]. These problems are multifactorial, complex in nature, disease specific, and, in some cases, poorly understood; they have also been extensively discussed and reviewed in the literature [26–31]. Accordingly, it is well appreciated that a single technological advancement is not going to resolve the entirety of the challenges facing drug discovery research, but a new approach may lead to better chemical leads, to a reduction in toxicity issues, or to an improvement in selectivity and efficacy. Thus, incorporating NMR metabolomics into a standard drug discovery pipeline may lead to an improved success rate, but it won't solve all of the field's problems.

Metabolomics has the potential to identify therapeutic targets in the drug discovery process. First, metabolomics may improve our fundamental understanding of cellular physiology, disease development, or disease progression, which may lead to the discovery of new therapeutic targets. In addition, metabolic profiles resulting from a disease state or drug resistance may identify cellular or metabolic processes that are critical to the system. In this manner, specific proteins associated with these upregulated or downregulated pathways may be targeted in either a single drug therapy or as part of a multidrug treatment. The underlying assumption is that the perturbed metabolic processes are required to maintain the drug resistance or disease state. Accordingly, inhibiting these proteins may reverse the drug resistance or eliminate the disease. Of course, the metabolic profile of a disease state may also be used to monitor a response to treatment [32]. The return of the metabolic profile to a healthy state would be indicative of a positive therapeutic response. Similarly, a lack of a response would suggest

an ineffective treatment, and, conversely, a drastic change in the metabolic profile may be indicative of a toxic response. Thus, in total, metabolomics may beneficially impact all the steps of the drug discovery process, from target identification to therapeutic monitoring.

Cancer biology has pioneered the idea of metabolic dysregulation as the source of disease pathology. Warburg aerobic glycolysis and glutaminolysis are essential features of cancer pathogenesis and are known to occur in most tumors [33]. These metabolic processes are potentially linked to oncogene and tumor suppressor regulation [34]. Thus, new therapeutic targets to improve cancer management may be based on modulating these altered metabolic processes. For example, a new experimental treatment for cancer is 4-phenylbutyrate [35], which is a sacrificial metabolite that induces an alternative metabolic pathway to overcome a defective urea cycle and repair a nitrogen imbalance. Presumably, 4-phenylbutyrate may indirectly address a dysregulated glutaminolysis by facilitating the excretion of glutamine in the urine [36]. Specifically, 4-phenylbutyrate is converted to 4-phenylacetate, which then reacts with glutamine to form the excretion product phenylacetylglutamine.

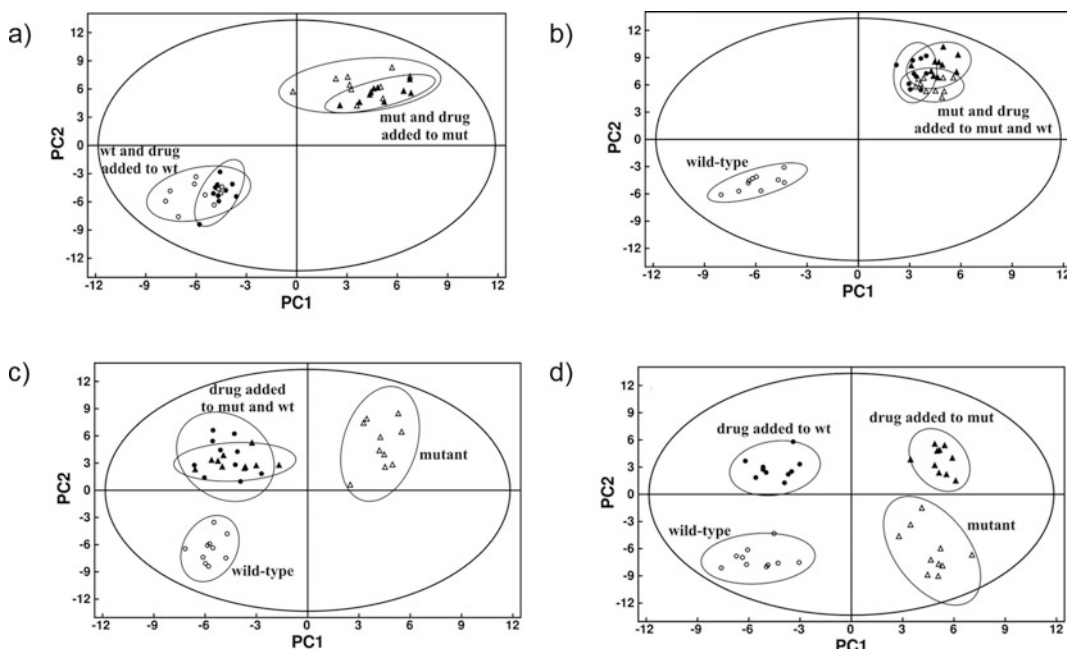
In a similar manner, ongoing efforts to develop novel cancer treatments by targeting metabolic dysregulation (e.g., glycolysis, glutaminolysis, PPP) has highlighted the therapeutic importance of metabolic enzymes such as pyruvate dehydrogenase kinase, lactate dehydrogenase A, glutaminase 1, thioredoxin reductase, and gamma-glutamylcysteine synthetase [37, 38]. In fact, inhibitors developed against all of these metabolic targets are currently at various stages of preclinical and clinical trials. Indeed, anti-metabolite compounds and enzyme inhibitors are among the most commonly prescribed drugs. For example, gemcitabine, a first-line treatment for pancreatic cancer, is a nucleotide (metabolite) analog that interferes with DNA synthesis [34]. Notably, gemcitabine resistance results from a metabolic shift that increases the cellular production of deoxycytidine triphosphate to function as a competitive inhibitor of gemcitabine [34]. Similarly, azathioprine is a purine analog that disrupts RNA and DNA synthesis and is used to treat autoimmune diseases like Crohn's disease and rheumatoid arthritis [39]. Sulfanilamide antibiotics inhibit bacterial enzymes that require para-aminobenzoic acid (PABA) for synthesizing folic acid, a coenzyme in DNA synthesis. Treatments for inborn errors of metabolism are other successful examples of employing metabolite analogs as a therapy [40]. Clearly, a metabolic view of disease pathology has beneficially contributed to existing drugs and the creation of next-generation therapies and has the potential of significantly improving patient outcomes [41].

As illustrated above, NMR-based metabolomics is an asset for drug discovery since it helps characterize metabolite level

perturbations and altered metabolic pathways that may be used to identify protein targets for designing new drugs. NMR is also routinely used as a high-throughput screen (HTS) to identify and evolve chemical leads against these protein targets. NMR ligand-affinity screens such as SAR-by-NMR, MS/NMR, multi-step NMR, SHAPES, and FAST-NMR that make use of saturation transfer difference (STD), WaterLOGSY,  $^{19}\text{F}$  NMR, HSQC chemical shift perturbations (CSP), transfer NOEs, and fragment-based libraries are established means of identifying and validating chemical leads through experimentally observed protein-ligand interactions [42–47]. Consequently, combining NMR HTS techniques with metabolomics to implement a high content drug screen is a logical next step. One such example is the NMR-based metabolomics screen of a library of 56 kinase inhibitors targeting eEF-2 kinase and NF- $\kappa$ B [48]. Tiziani et al. monitored global metabolic changes resulting from compound treatment to a panel of cancer cell lines (e.g., HeLa, CCRF-CEM, SKOV-3, and acute myeloid leukemia (AML) primary cells) in a 96-well format. SDS-induced cell lysates were then used to collect one-dimensional (1D)  $^1\text{H}$  CPMG and two-dimensional (2D)  $^1\text{H}$  JRES NMR spectra followed by multivariate statistical analysis of the entire spectral dataset. The resulting scores and backscaled loadings plot calculated relative to controls were used to identify active lead compounds based on the magnitude of the metabolic response. Notably, approximately 100 samples were screened in less than 24 h.

NMR metabolomics combined with traditional medicinal chemistry can, thus, be used to validate *in vitro* efficacy, selectivity, and specificity while avoiding off-target effects (*see* Fig. 2). Metabolic perturbations resulting from a drug treatment may be correlated to the inhibition of the desired protein target and/or as a result of unexpected off-target effects [49]. If the lead compound inhibits additional protein targets beside the specifically designed drug target, then the lack of selectivity may help predict and anticipate toxicity issues in future clinical trials. Thus, medicinal chemistry techniques can then be used to evolve the chemical lead in an iterative fashion using insights and feedback from the metabolomics experiments. The goal is to maintain the metabolic profile associated with inhibiting the desired protein target while diminishing or eliminating metabolic changes resulting from unwanted off-target effects. This global approach of monitoring all possible changes to the system in response to a drug's activity would be difficult, if not impossible, to replicate in an *in vitro* setting. Even if the specific off-target proteins could be identified, it is not feasible to reproduce the synergistic and simultaneous impact on multiple cellular processes.

In a similar manner, an *in vivo* mechanism of action for a new chemical lead may be inferred by comparing its induced metabolic profile against metabolic profiles for drugs with known mechanisms



**Fig. 2** Cartoon illustrations of hypothetical PCA scores plot for the following scenarios: (a) inactive compound; (b) active and selective inhibitor; (c) active, nonselective inhibition of target and secondary protein; and (d) active, nonselective preferential inhibition of secondary protein. Labels correspond to wild-type cells (wt) and mutant cells (mut) (Reproduced with permission from ref. 51. Copyright © 2006 American Chemical Society)

of action [50]. Compounds that share the same mechanisms of action would be expected to have similar impacts on the metabolome. Technological advances have reduced NMR analyses times to less than a few minutes per sample [51]. Thus, like other secondary assays routinely used in a drug discovery pipeline, metabolomics can be applied in a high-throughput fashion to interrogate an entire library of chemical leads. Accordingly, metabolomics is a cost-effective means of filtering out problematic lead compounds prior to expensive clinical trials.

Mechanistic understanding of *in vivo* drug activity is an important and, typically, challenging stage of the drug discovery process. Knowledge of *in vitro* and *in vivo* mechanisms of action improves the overall success rate and will beneficially enhance the drug design effort [50]. Unfortunately, elucidating a mechanism of action for a specific chemical lead may be an arduous, time-intensive, and resource-intensive task. Multiplying this effort across a large screening library is impractical. Instead, predicting the *in vivo* mechanism of action for new chemical leads by leveraging known entities is a valuable alternative to resolving this problem. Furthermore, the availability of a set of drugs with a diverse range of biological activities for such a comparison will increase the likelihood of successfully assigning a mechanism of action. In addition,

identifying compounds that target various biological processes and have a range of mechanisms of action may also assist in disease management and in combating the emergence of resistance. This may be accomplished through a multidrug therapy that combines compounds with distinct modes of action. Halouska et al. [50] demonstrated the general approach by comparing the metabolic profile of *Mycobacterium smegmatis* following treatment with a range of antibiotics with known mechanisms of action and some promising chemical leads from high-throughput cell-based screens. 1D  $^1\text{H}$  NMR spectra combined with multivariate analysis were used to characterize the relative global metabolomes. As expected, antibiotics with similar modes of action clustered together in the orthogonal projections to latent structures (OPLS) scores plot. The HTS chemical leads were clustered together with known cell wall inhibitors implying a similar mechanism of action, which was supported by literature reports. It is also possible to invert the protocol and use a drug with a known mechanism of action to modulate a metabolic pathway predicted to be important to disease progression of drug resistance. In this manner, if the targeted metabolic pathway is reversed following drug treatment and the desired therapeutic outcome is achieved (i.e., cell death), then an *in vivo* mechanism of action is likely indicated. A recent example of this process involved a vancomycin-intermediate resistant strain (VISA) of *Staphylococcus aureus* that exhibited an increase in purine biosynthesis relative to a vancomycin-susceptible strain (VSSA) [15]. Treating VISA strains with a combination drug therapy consisting of a sub-lethal dose of vancomycin and 6-mercaptopurine (i.e., purine substrate analog) led to a statistically significant increase in staphylococcal killing. In addition to reversing the acquired vancomycin resistance, the metabolomics results also indicate (as expected) that the *in vivo* mechanism of action for 6-mercaptopurine is an inhibitor of purine biosynthesis.

Metabolomics can also provide valuable information during a clinical trial. Metabolomics can assist with therapeutic monitoring and drug dosage adjustments and by differentiating responders from non-responders. Simply, specific metabolic profiles may be attributed to each group, which then can be used to classify individual patients by monitoring metabolite changes in common biofluids such as urine, serum, or fecal material. Another aspect of a clinical trial that has benefited from metabolomics is monitoring patient compliance. In this regard, patient non-compliance is readily identified by the presence of distinct and/or strong signals in the NMR spectrum that are not normally present, such as metabolites from drug metabolism or from consuming caffeine, ethanol, or nicotine. As a result of these clear benefits to drug discovery, metabolomics continues to be adopted by the pharmaceutical and biotechnology industry. A prime example is COMET, the Consortium for Metabonomic Toxicity [52], which brought together pharmaceutical

companies to utilize large datasets of NMR metabolomics analysis of urine and serum samples to predict drug toxicity.

To address the expanding interest in NMR metabolomics, we present a detailed, step-by-step description of a metabolomics workflow with a specific attention to details important to its application to a drug discovery and design process. The protocol includes methods relevant to understanding the mechanism of action of novel drug leads in comparison with existing therapeutics. Furthermore, we describe the application of metabolomics to a wide variety of sample sources commonly encountered in drug discovery programs. This includes bacterial and mammalian cell cultures, human and animal tissue samples, and different biofluids (e.g., urine, serum, and fecal material). The entire process from cell culturing and sample collection to biological interpretation is described. Importantly, the protocol addresses critical concerns regarding the proper (1) preparation and handling of metabolomics samples; (2) preprocessing, processing, and acquisition of spectral data; (3) metabolite identification; and (4) statistical analysis.

Unlike other OMICS techniques, the composition of the metabolome can easily change from the processing, handling, and storage of samples [53]. Metabolites are chemically labile or unstable, are subject to thermal degradation or oxidation, and are readily transformed by residual enzymatic activity. For example, the oxidation of cysteine and glutathione or the conversion of ATP to AMP (because of a labile phosphate group) is a common occurrence in metabolomics samples [54]. These biologically irrelevant sample variations may distort our understanding of a system and lead to erroneous conclusions. Accordingly, metabolomics requires a robust, efficient, and fast workflow to yield reproducible and reliable data. Univariate and multivariate statistical analyses are essential components of any metabolomics study and provide the means to interpret the dataset and identify the key group variances (i.e., metabolite changes) [55]. Nevertheless, issues such as the incorrect application of statistical techniques, the subsequent lack of proper model validation, or the over-interpretation of models are common occurrences in the field of metabolomics [56]. Accordingly, the protocol presented herein addresses these issues by providing instructions to avoid common mistakes.

---

## 2 Materials

All samples should be prepared using nanopure water ( $\text{H}_2\text{O}$ ) and/or deuterated water ( $\text{D}_2\text{O}$ ). All safety regulations should be followed for the proper handling and disposal of chemicals and biological samples used in the study.



## 2.1 Laboratory Equipment

1. Bruker AVANCE III HD 700 MHz NMR spectrometer equipped with a 5 mm quadruple resonance QCI-P cryoprobe ( $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ , and  $^{31}\text{P}$ ) with  $z$ -axis gradients, an automatic tune and match system (ATM), and a SampleJet automated sample changer system with Bruker ICON-NMR software (Bruker Biospin, Billerica, MA).
2. BSL-2 biosafety level grade hood (e.g., Biological Safety Cabinet, LF BSC class 2 type A, Thermo Fischer Scientific, Waltham, MA).
3. Nanopure ultra water system (Barnstead Inc., Dubuque, IA).
4. Lab Armor bead bath (Chemglass Life Sciences, Vineland NJ).
5. Incubator capable of maintaining physiological temperature and proper carbon dioxide levels (e.g., HERA CELL VIOS 250i  $\text{CO}_2$  Incubator, Thermo Fischer Scientific, Waltham, MA).
6. pH meter and probe.
7. Refrigerated centrifuge capable of speeds up to  $21,100 \times g$  (e.g., SORVALL micro 21R centrifuge, Thermo Fischer Scientific, Waltham, MA).
8. Speed Vac for solvent removal (e.g., SAVANT SC210A Speed-Vac concentrator, Thermo Fischer Scientific, Waltham, MA).
9. Freeze dryer to remove water (e.g., FreeZone 4.5, LAB-CONCO, Kansas City, MO).
10. 1  $\mu\text{L}$  to 1000  $\mu\text{L}$  pipettes
11. FastPrep-96 homogenizer (MP Biomedicals, Santa Ana, CA).
12. ACCU-SCOPE 3030ph microscope (Commac, NY).
13. Cryogenic storage container (Taylor Wharton, Theodor, Al).
14.  $-80^\circ\text{C}$  freezer.
15. Flask shaker incubator (up to 250 rpm).
16. Sonic Dismembrator (Sonicator).

## 2.2 Disposable Supplies

1. 1  $\mu\text{L}$  to 1 mL pipette tips.
2. 10 mL aspirating pipettes.
3. 15 mL Falcon tubes.
4. 2 mL Eppendorf tubes.
5. 1 mL screw-cap microcentrifuge tubes.

## 2.3 Isotopically Labeled Solvents and Reagents (See Notes 1 and 2)

1. Deuterium oxide ( $\text{D}_2\text{O}$ , 99.8 atom% D).
2. 3-(trimethylsilyl) propionic-2,2,3,3- $\text{d}_4$  acid sodium salt (TMSP- $\text{d}_4$ , 99.8 atom% D).
3. Dimethyl sulfoxide- $\text{d}_6$  ( $\text{DMSO-}\text{d}_6$ , 99.8 atom% D).

4.  $^{13}\text{C}_6$ -glucose (99%  $^{13}\text{C}$ ).
5.  $^{13}\text{C}_2$ -acetate (99%  $^{13}\text{C}$ ).
6. Other potential  $^{13}\text{C}$ -carbon labeled or  $^{15}\text{N}$ -nitrogen labeled reagents.

## 2.4 Buffers

1. Wash buffer, phosphate-buffered saline (PBS) at pH 7.4: 137 mM NaCl, 2.7 mM KCl, 10 mM  $\text{Na}_2\text{HPO}_4$ , and 2 mM  $\text{KH}_2\text{PO}_4$ . To prepare 1 L PBS buffer at pH 7.4, add 8.0 g of NaCl, 0.2 g of KCl, 2.68 g of  $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$ , and 0.24 g of  $\text{KH}_2\text{PO}_4$  to a final volume of 1 L of Nanopure water.
2. NMR buffer: 50 mM phosphate buffer at pH 7.2 (uncorrected, *see Note 3*) in 600  $\mu\text{L}$  of 99.8%  $\text{D}_2\text{O}$ . Add 50  $\mu\text{M}$  (1D NMR experiment) or 500  $\mu\text{M}$  (2D NMR experiment) TMS- $\text{d}_4$  as an internal chemical shift reference.

## 2.5 Cell Media (See Note 4)

1. Dulbecco's modified eagle medium (DMEM) and Roswell Park Memorial Institute (RPMI) medium with the addition of an appropriate amount of fetal bovine serum (5–20% v/v) and/or supplemental glucose (1–25 mM) are commonly used to culture mammalian cells.
2. Bacterial cells are typically cultured in complex media, such as Luria-Bertani (LB) media for *E. coli*, Middlebrook's albumin dextrose catalase media (MADC) for *Mycobacterium smegmatis*, or M9 minimal media.

## 2.6 Software and Databases

1. Bruker ICON-NMR software for automated NMR data acquisition (Bruker Biospin, Billerica, MA).
2. MVAPACK metabolomics toolkit for processing and analyzing chemometric data (<http://bionmr.unl.edu/mvpack.php>) [57].
3. PCA/PLS-DA utilities for quantifying separation in PCA, PLS-DA, and OPLS-DA scores plots (<http://bionmr.unl.edu/pca-utils.php>) [58].
4. NMRPipe software for processing and visualizing NMR data (<https://www.ibbr.umd.edu/nmrpipe/install.html>) [59].
5. NMRViewJ software for processing and visualizing NMR data (One Moon Scientific, Inc. Westfield, NJ; <https://nmrfx.org/>) [60].
6. R statistical package (<https://www.r-project.org/>) [44].
7. Chenomx (Chenomx, Inc., Edmonton, AB, Canada) software for automated metabolite assignment and quantification from 1D  $^1\text{H}$  NMR spectra (<https://www.chenomx.com/>).
8. MetaboAnalyst software for the statistical, functional, and integrative analysis of metabolomics data (<http://www.metaboanalyst.ca/>) [61].

9. Human Metabolomics Database (HMDB) of reference NMR data for known metabolites (<http://www.hmdb.ca/>) [62].
10. Biological Magnetic Resonance Data Bank (BMRB) of reference NMR data for known metabolites <http://www.bmrbl.wisc.edu/metabolomics/>) [63].
11. Non-uniform schedule (NUS) generator (<http://bionmr.unl.edu/dgs-gensched.php>) for NUS NMR data acquisition [64].
12. Small Molecule Pathway Database (SMPDB) (<http://www.smpdb.ca>) for modeling metabolic pathway changes from list of metabolite changes [65].

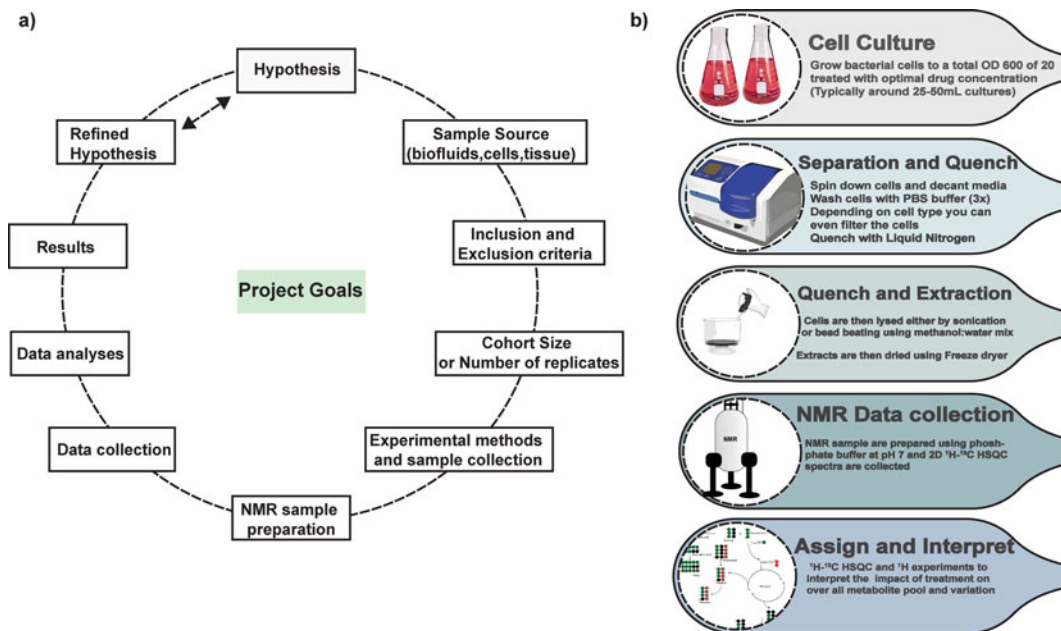
---

### 3 Methods

Extreme consistency is key to the success of any metabolomics experiment. Accordingly, it is essential to apply *all* experimental protocols and parameters uniformly to *all* groups in the study. Any variation in the preparation, handling, or storage of the sample may perturb the metabolome and lead to erroneous results. For example, different sample processing times, samples handled by different investigators, and cell cultures grown in different incubators or shakers; different sources or lot numbers of chemicals, media, or disposables; and different buffer/media preparations (even if the same recipe was prepared by the same individual) are all likely to induce a biologically irrelevant bias in the samples *and needs to be avoided*. In effect, the *only* difference between groups should be the intended variable (e.g., wild type versus mutant; treated versus untreated, healthy versus disease, etc.). In general, the defined variable(s) is expected to induce a measurable change in the metabolome that is detectable by NMR. For example, a proper dosage for a drug used in a metabolomics project needs to be experimentally determined. The drug dosage needs to be high enough to induce a detectable metabolic response, but not too high to induce a generic cell death phenotype (*see Note 5*).

#### 3.1 Experimental Design

A scientific hypothesis needs to be clearly defined in order to properly design a metabolomics experiment (*see Fig. 3*). In other words, simply characterizing the metabolome of a cell, tissue, or biofluid is likely to fail to provide any scientific insight or value without a logically designed comparative analysis. In this regard, a metabolomics experiment is fundamentally a comparison between two or more groups (e.g., control versus drug treated), where a difference or similarity in a metabolome is expected to answer a scientific question. The next critical step is to identify an appropriate source for the metabolome or the type of biological samples to be compared (i.e., established or primary cell lines, tissue samples from an animal model, human biofluids, etc.). The choice of system



**Fig. 3 (a)** A summary of the design and planning steps in a typical metabolomics-based drug discovery project. **(b)** A workflow for a typical cell-based NMR metabolomics experiment focused on understanding the mechanism of action for a chemical lead

will then determine the number of biological replicates that can be practically obtained, which, in turn, will define the statistical significance of any observed results. Given a testable hypothesis and the type of biological samples, there are essentially an endless number of experimental variables that can be manipulated, where each permutation defines a new group. Practically speaking, it would be impossible to account for every possible permutation in a single experimental design. Instead, the identification of which groups to use is fundamentally driven by the scientific question the study is designed to answer (*see* Fig. 3) Thus, a general set of protocols is presented for a generic metabolomics experiment that can be applied to a drug discovery project. The protocol provides a means to determine a drug's mechanism of action and to monitor a patient's response to treatment.

The simplest experimental design for a metabolomics drug discovery project will consist of two groups: untreated samples and drug-treated samples. Nevertheless, it is typical to significantly expand the experimental design beyond just two groups (*see* Figs. 2 and 3). Common groups include:

1. Controls (e.g., wild-type cells or healthy subjects)
2. Control with treatment (positive control, e.g., wild-type cells or healthy animals treated with a drug lead)

3. Experimental model (e.g., animal disease model or treatment naïve patients)
4. Experimental model with a drug treatment (e.g., animal disease model or patients treated with a drug lead)
5. Negative control (e.g., mutant cells with no treatment, patients with alternative disease)
6. Negative control with a drug treatment (e.g., mutant cell lines treated with a drug lead)

The metabolomics study may consist of any combination of biological samples originating from cell cultures, tissue samples, and/or biofluids. Human or animal biofluids may consist of urine, serum, plasma, blood, fecal material, saliva, sweat, condensed breath, cerebrospinal fluid (CSF), or any other body fluid or excretion. Similarly, tissue samples usually consist of organs, such as the heart, lungs, brain, liver, or kidneys from sacrificed animals. Tissue samples from human subjects typically originate from biopsies (i.e., muscle, adipose, etc.), surgeries (i.e., tumors), or tissue banks. Each biofluid, tissue type, or time point would constitute a different group. Of course, there are numerous experimental protocols that could also be varied, such as different media, nutrients, temperature, diets, etc.

Besides sample type and experimental protocols, there are other variations that may be applied to form additional groups. For example, different groups may be formed by varying the drug dosage or by using a placebo. Of course, multiple drugs could be tested in a single study, and group membership would be defined by what drug a cell culture or an individual receives. In addition, the number of drug treatments received by a cohort could vary. Time could be another variable in the experimental design (e.g., when samples are collected and when treatments are administered). Similarly, different groups may consist of different mutant or wild-type cell lines or strains. For example, a mutant cell line may correspond to the drug protein target being inactivated (i.e., negative control) through a variety of genetic mutations or gene knockout methods. Alternatively, the mutant cell lines could be various human isolates with variable levels of resistance or susceptibility to the tested drug. The cell lines could just be different types of bacteria (e.g., *E. coli*, *S. aureus*, etc.) or different types of cancers (e.g., pancreatic, breast, etc.). Similarly, animal models or human cohorts may have different stages or severity of the disease or even different diseases. Furthermore, multiple biofluid samples (i.e., both urine and serum) may be collected from the same animal or human participant at multiple time points during the study. Similarly, more than one tissue or organ sample may be harvested from each sacrificed animal.

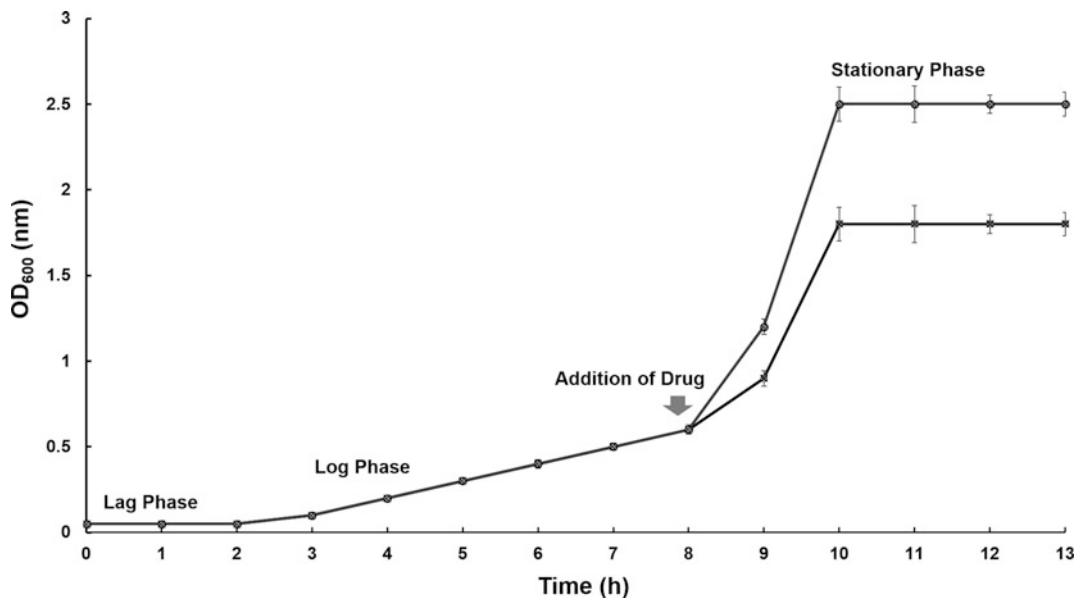
Another important decision regarding the experimental design is the number of biological replicates required per group. In

general, more biological replicates led to a better statistical significance in differentiating between the various groups. But of course, there are practical limits to the number of cell cultures, animals, or human cohorts that can be prepared or recruited for any study. Facility and equipment capacities, the availability of personnel, and cost are all major factors that greatly limit the number of samples or cohorts that are possible. There is also an obvious multiplication factor based on the number of desired groups and the number of replicates per group. Accordingly, there may be a tendency to maximize the number of groups at the expense of the number of replicates per group. This is inadvisable since it would likely lead to statistically insignificant conclusions. Instead, the number of groups should be reduced to achieve an acceptable minimal number of biological replicates per group. In fact, sample size calculation for metabolomics studies is complex and not well defined, where recent efforts suggest sample sizes may need to be >20 per group [66, 67]. In the case of metabolomics studies involving human cohorts, the need for replicates is even greater and more challenging to estimate, but sample sizes >60–100 cohorts per group are typical targets.

One final factor to consider for the design of a metabolomics study involving human cohorts is the inclusion and exclusion criteria. Simply, what individual characteristics, traits, or behavior are needed for the goals of the study? Conversely, what are potential confounding factors that may confuse or obscure the desired outcome of the study and should be avoided? The specific inclusion and exclusion criteria are likely unique to each study; and there are no general guidelines to apply. Criteria commonly considered include age, body mass index, diet, comorbidities, ethnicity, gender, health test panel (e.g., blood pressure, lipid panel, metabolic panel, urinalysis, etc.), physical activity, race, and the use of alcohol, recreational drugs, or tobacco (*see* Table 1).

**Table 1**  
**A list of typical exclusion criteria for a metabolomics clinical study**

<i>Exclusion criteria</i>
<ul style="list-style-type: none"> <li>• Acute illness within 2 weeks</li> <li>• Comorbidity (e.g., chronic lung disease, depression, etc.)</li> <li>• Excessive intake of alcohol</li> <li>• Obesity</li> <li>• Pregnant women or menstruating women</li> <li>• Recreational drug use</li> <li>• Risk of transmitting disease through urine (e.g., hepatitis B or C)</li> <li>• Significant abnormalities in clinical chemistry of hematology</li> <li>• Smoker</li> </ul>



**Fig. 4** An example of bacterial growth curve for a cell line treated with a drug at an OD<sub>600</sub> of 0.6. The cellular metabolome should be extracted during the log phase, which captures the majority of metabolic variations

### 3.2 Determining Drug Dosage for Cultured Cells

The reported minimum inhibitory concentration (MIC) for a drug against a specific cell strain is a good starting point for determining the optimal drug dosage for a metabolomics study. Typically, drug concentrations that range between 1 and 24 times the reported MIC or MIC<sub>50</sub> are tested to identify an optimal drug dosage. The goal is to identify a drug dosage that stresses the cells, but does not induce cell death. This is accomplished by identifying a drug dosage that results in a decrease in cell growth by approximately 50% (MIC<sub>50</sub>, *see* Fig. 4).

Along with dosage, the time point for administering the drug needs to be experimentally determined. First, enough biomass needs to be harvested in order to extract a detectable metabolome. Second, the timing of the drug dosage should induce the desired perturbation in the metabolome. For example, a particular metabolic process may only be active during a specific phase of the growth curve. Typically, a drug treatment is administered during the exponential or log phase for a bacterial culture. The cells are allowed to grow for at least one generation or one doubling time before harvesting.

1. Approximately 1–20 mM of a stock solution is prepared for each compound to be tested by dissolving the required amount of the compound into 0.5 mL of ethanol or DMSO-*d*<sub>6</sub> (*see* Note 6).
2. Bacterial cells are typically grown at 37 °C with shaking at 200 rpm in 50 mL of complex or minimal media in a 125 mL Erlenmeyer flask. Mammalian cells are typically grown in a

6-well plate with 3 mL of media at 37 °C and 5% percent CO<sub>2</sub> in an incubator. A total of 21 cell cultures are prepared.

3. The bacterial cells are allowed to grow until an average optical density at 600 nm (OD<sub>600</sub>) of 0.6 is achieved (*see* Fig. 4). Mammalian cells are grown to approximately 40% confluency.
4. Each cell culture receives a different aliquot of the stock drug solution such that the final drug concentration ranges from 0 to 24 times the MIC value. Each drug dosage is made in triplicate for a total of 21 cell cultures (*see* **Note 7**).
5. Following the addition of the drug leads, the cells are allowed to grow for an additional amount of time. This is typically the cell doubling time.
6. For bacterial cells, the optical density (OD<sub>600</sub>) is recorded at the end of the additional cell growth time. The growth rate inhibition is calculated by comparing the optical density of the treated cells to the untreated cells. Typically, untreated cells will reach an OD<sub>600</sub> of 1.2 (i.e., a doubling from an initial OD<sub>600</sub> of 0.6), while the OD<sub>600</sub> of drug treated cells will be lower. For mammalian cells, the cells are visualized under a microscope to measure the percent confluency and assess the amount of detached cells [53, 68, 69]. Typically, untreated cells should reach a confluency of approximately 80% (i.e., doubling from an initial 40% confluency).
7. The desired dosage corresponds to the drug concentration that inhibits cell growth by approximately 50% relative to untreated cells (*see* **Note 5**).
8. For mammalian cells, an additional consideration is the length of time following the drug treatment that is required before the cells can be harvested. It is important to provide a significant amount of time to allow a perturbation in the metabolome to occur, but it should be short enough to prevent a significant accumulation of dead cells. No dead cells are a preferred outcome. After the proper drug dosage is identified, repeat experimental **steps 2–5** with the following modifications:
  - (a) A constant drug dosage is used for each replicate cell culture. Each cell culture replicate is a different time point instead of a different drug dosage.
  - (b) The time after drug treatment that the cells are harvested are varied from zero (i.e., control) to an end point of approximately 12–24 h.
9. For each time point, the percentage of live and dead cells is determined with a cell counter. The optimal time point for harvesting the cells would be the latest time point following the drug treatment where the vast majority of the cells are still alive.



### **3.3 Determining Drug Dosage for Cohorts**

How a drug is administered (e.g., oral, intravenous, intraperitoneal, etc.), how often the drug is administered, how the drug is formulated, and the drug's dosage will all impact how the drug is adsorbed, distributed, metabolized, and excreted (ADME) in an animal or human [70, 71]. This, in turn, will determine both the peak and steady-state concentration of the drug and the resulting efficacy and toxicity of the treatment. Determining the proper drug dosage and the route of administering a drug lead for an animal model or a clinical trial is, thus, a complex question that is not easily answered. In fact, the entire field of pharmacokinetics and pharmacodynamics is dedicated to addressing these very challenging questions [72]. Available information, such as a lethal dose (LD50) from cell-based assays or reported dosages for similar compounds from the scientific literature, may provide initial guidance. Unfortunately, there is not a simple algorithm to convert an LD50 to an animal dosage. Instead, a detailed efficacy and toxicity study needs to be conducted to determine the proper drug dosage prior to embarking on a metabolomics study.

Drug dosages are typically administered according to a species' body-mass (e.g., mg/kg). Accordingly, it may appear trivial to translate an animal drug dosage to a human clinical trial by simply scaling the dosage based on relative body masses. Unfortunately, this approach is incorrect since pharmacokinetics and toxicity issues vary widely across species. Instead, a variety of factors, such as differences in allometry, metabolic rates, and physiological, need to be considered [73, 74]. Thus, the protocol for determining the proper dose and means of administering a drug candidate is beyond the intended scope of this chapter.

### **3.4 Collecting and Storage of Biofluids from Human Cohorts (See Table 1)**

1. Human volunteers in a metabolomics study must sign an informed consent form before participating.
2. Participants in a metabolomics study are typically required to complete a questionnaire that confidentially collects information regarding the individual's age, sex, race/ethnicity, residing state, diet (e.g., vegetarian, vegan, etc.), dietary restrictions (e.g., no peanuts, no gluten, low carb, low calorie, no lactose, etc.), and a list of daily medications or supplements.
3. Biofluid collection should be done after 12 h of fasting with no fish consumed for 24 h [75].
4. Daily supplements and over-the-counter medications should be stopped at least 24 h prior to collecting the biofluid(s).
5. Each type of biofluid requires specific and unique collection procedures. For example, urine specimens are best collected midstream and in the morning and can be collected by the participant. Conversely, blood needs to be collected by a medical professional in a clinical or medical facility. A detailed

discussion of common protocols employed for collecting biofluids from human and animal cohorts is beyond the scope of this chapter, but has been previously described in detail [19, 53, 68, 69, 76–78].

6. After collecting a biofluid, the samples should be processed immediately, kept on ice at all times, and stored at  $-80^{\circ}\text{C}$ .

### 3.5 Preparation of Mammalian Cells

1. Cell culture media and supplements are obtained from commercial vendors (*see* Subheading 2.5). The cell culture media may be supplemented with isotopically labeled nutrients (e.g.,  $^{13}\text{C}$ -glucose,  $^{13}\text{C}$ -acetate,  $^{15}\text{N}$ -ammonium chloride, etc.) as dictated by the goals of the experiment (*see* Subheading 3.7).
2. Cell culture procedures must follow published guidelines for the cell type to avoid misidentification and contamination [79].
3. A minimum target of ten replicate cultures is prepared per group (*see* Note 8).
4. Mammalian cells are typically grown in a  $100\text{ mm}^2$  dish with 10 mL of media at  $37^{\circ}\text{C}$  and 5% percent  $\text{CO}_2$  in an incubator. Equally distribute replicates from each group between incubators if more than one incubator is required (*see* Note 9).
5. The cells are grown until reaching approximately 80–90% confluency.
6. The drug is added to the cell culture for the treated groups using the dosage determined from Subheading 3.3. A blank solvent bolus is added to all non-treatment groups.
7. The drug is typically added to the cell culture after 80–90% confluency. But, the drug may also be added at the start of the cell culture or at other time points as required.
8. The cells are harvested after a pre-determined incubation time with the drug.

### 3.6 Preparation of Bacterial Cell Cultures

1. Perform **steps 1–3** from Subheading 3.5.
2. The drug is added to the cell culture for the treated groups using the dosage determined from Subheading 3.3. A blank solvent bolus is added to all non-treatment groups.
3. Typically, the drug is added at the start of the cell culture, but it may be added at various time points along the growth curve depending on the specific goals of the study or the proposed mechanism of action of the drug.
4. Start with a 25 mL culture of cells with an  $\text{OD}_{600}$  of 0.05 from glycerol frozen stocks.
5. For any type of bacterial cell, it is important to obtain a final  $\text{OD}_{600}$  of 20 for the entire culture volume. For example, if the total volume is 20 mL, then the minimum  $\text{OD}_{600}$  should be

1.0 per mL. This will ensure that a high-quality NMR spectrum will be obtained from each cell lysate in a reasonable amount of time.

6. Cell cultures are harvested at a specific time point of the growth curve (e.g., exponential growth, stationary phase, etc.). Importantly, all cell cultures must be harvested at the identical phase of the growth curve, not necessarily the same OD<sub>600</sub> value.
7. All replicate cell cultures are normalized based on the final OD<sub>600</sub> value. The total volume for each individual cell culture is adjusted so that each cell culture has the same effective OD<sub>600</sub> value.

### **3.7 Isotopically Labeled Metabolomics Samples**

1. Identify a <sup>13</sup>C-, <sup>15</sup>N-, or other isotopically labeled compound or nutrient to label the metabolome and act as a tracer. The tracer should be associated with the metabolic pathway of interest and expected to be affected by the experimental treatment. <sup>13</sup>C<sub>6</sub>-glucose is a common choice for a tracer since it highlights central carbon metabolism (e.g., glycolysis and TCA cycle).
2. The culture media should not contain the selected nutrient or metabolite.
3. The culture media is supplemented with an appropriate molar concentration of the tracer metabolite (*see Notes 4 and 10*). The concentration of the <sup>15</sup>N- or <sup>13</sup>C-labeled metabolite needs to be high enough ( $\geq 1\text{--}5\ \mu\text{M}$ ) to be detected by NMR. Accordingly, the culture media should be supplemented with approximately 2.5–4 g/L (0.5–25 mM) of a major nutrient source like <sup>13</sup>C<sub>6</sub>-glucose or ~10–100 mg/L of a targeted metabolite like <sup>13</sup>C-D-alanine (~1 mM). Multiple labeled isotopes may be used together if the type of label (<sup>13</sup>C or <sup>15</sup>N) differs between the compounds.
4. In general, the isotopically labeled metabolite is added to the culture medium at the start of the cell culture. Of course, it may be added at different time points to emphasize a specific cellular process or timed according to the drug treatment.

### **3.8 Extracting Water-Soluble Metabolites from Mammalian Cells**

Samples should be stored at  $-80\ ^\circ\text{C}$ , but, ideally, samples should be immediately analyzed. All samples should be kept on ice or at  $4\ ^\circ\text{C}$  during the extraction process and/or handling. There are a few critical aspects to keep note of throughout the process that can affect the quality of the resultant data (1) speed, (2) consistency, (3) random processing of samples, and (4) the efficient removal of all biomolecules and cell debris [80]. Quick and prompt sample processing and preparation are ideal to avoid extended temperature variations and chances of environmental contamination and degradation of metabolites. For example, protein degradation due to

improper shipping processes may lead to alterations in amino acid levels in biofluids. Furthermore, metabolites can chemically degrade or transform within milliseconds due to enzymatic activity, oxidation, chemical instability, or any number of other chemical processes [81]. Thus, rapid inactivation and removal of all macromolecular cell debris (usually through methanol/ethanol precipitation) that may bind or transform a metabolite are necessary steps of the protocol (*see Note 11*).

1. The extraction process for intracellular metabolites begins after the cells have grown to the desired time point. In addition to the cell extract, the media should also be analyzed for metabolomics changes as many metabolites get exchanged or effluxed outside of the cell.
2. Pre-chill the extraction solvents methanol and methanol:water mixture on dry ice.
3. Collect 1 mL of the media and immediately freeze the samples with liquid nitrogen and store at  $-80\text{ }^{\circ}\text{C}$  until ready for NMR sample preparation.
4. For the cell extracts, wash the cells still attached to the plates twice with 5 mL of PBS to remove debris. Discard the wash.
5. Lyse and quench cells with 1 mL of pre-chilled methanol at  $-20\text{ }^{\circ}\text{C}$ . Incubate cells at  $-80\text{ }^{\circ}\text{C}$  for 15 min. This process should detach the cells from the plate. Confirm cell detachment using a microscope and repeat lyse and quenching step if necessary.
6. Using a cell scraper, detach and transfer the cell suspension in methanol to a 2 mL microcentrifuge tube.
7. Centrifuge the 2 mL microcentrifuge tube for 5 min at  $15,000 \times g$  and  $4\text{ }^{\circ}\text{C}$  to pellet the cell debris.
8. Collect the supernatant and transfer to a new 2 mL microcentrifuge tube. Again, keep all samples on ice throughout the process.
9. Perform a second extraction to improve the overall efficiency. Add 0.5 mL of an 80%/20% mixture of methanol/water pre-chilled at  $-20\text{ }^{\circ}\text{C}$  to the cell pellet.
10. Centrifuge the suspension of cell debris with 80%/20% methanol/water for 5 min at  $15,000 \times g$  at  $4\text{ }^{\circ}\text{C}$  to pellet the cell debris.
11. Collect the supernatant and transfer to the 2 mL microcentrifuge tube containing the original methanol extract, thereby combining the two extractions.
12. For a third time, extract the pellet by adding 0.5 mL of ice cold water.

13. Centrifuge the cell suspension for 5 min at  $15,000 \times g$  at  $4\text{ }^{\circ}\text{C}$  to pellet the remaining cell debris.
14. Transfer the supernatant to the original microcentrifuge tube. Combine all extractions together in a 2 mL microcentrifuge tube.
15. Repeat **steps 1–14** for each replicate and for each group either in a randomized fashion or at the same time depending on the total sample number and your centrifuge capacity (*see Note 9*).
16. Use a SpeedVac or a rotary evaporator to remove the methanol. Chill the condenser with liquid nitrogen if using a rotary evaporator.
17. Following evaporation of the methanol solvent, poke a pin hole using a thumbtack and flash-freeze the samples in liquid nitrogen.
18. Freeze dry the samples using a lyophilizer; the pin hole helps establish low pressure in each sample tube.
19. Store samples in a  $-80\text{ }^{\circ}\text{C}$  freezer or proceed to preparing the NMR samples (*see Subheading 3.11*).

### **3.9 Extracting Water-Soluble Metabolites from Gram-Negative Bacteria**

1. Spin down the suspension culture at  $3200 \times g$  at  $4\text{ }^{\circ}\text{C}$  for 20 min to obtain a pellet.
2. Aspirate the media, keeping 1 mL of the media for NMR analysis.
3. Wash the pellet with ice cold PBS before extraction.
4. Add 1.5 mL of a 1:1 mixture of methanol and water to the pellet and transfer to a 2 mL microcentrifuge tube.
5. Pipet to mix the pellet with a 1:1 mixture of methanol and water pre-chilled to  $-20\text{ }^{\circ}\text{C}$ .
6. Lyse the cells using a handheld sonicator by pulsing for six 10 s bursts. Take precautions to avoid any sample heating during the sonication process.
7. Incubate the microcentrifuge tubes on ice for 2 min to extract the metabolome.
8. Centrifuge at  $1000 \times g$  for 10 min at  $4\text{ }^{\circ}\text{C}$  to remove cell debris.
9. Collect the supernatant (0.75 mL) and transfer to a new 2 mL microcentrifuge tube.
10. Repeat the metabolome extraction by adding 0.75 mL of 1:1 mixture of methanol and water pre-chilled to  $-20\text{ }^{\circ}\text{C}$  to the cell pellet.
11. Repeat **steps 6–8** and combine the supernatant with the previous extract.
12. Repeat **steps 1–11** for all replicate samples and groups (*see Note 9*).

13. Use a SpeedVac or a rotary evaporator to remove the methanol.
14. Flash-freeze the samples in liquid nitrogen.
15. Remove the water and bring to dryness using a lyophilizer.
16. Store samples in a  $-80\text{ }^{\circ}\text{C}$  freezer or proceed to preparing the NMR samples (*see* Subheading 3.11).

### 3.10 *Extracting Water-Soluble Metabolites from Gram-Positive Bacteria*

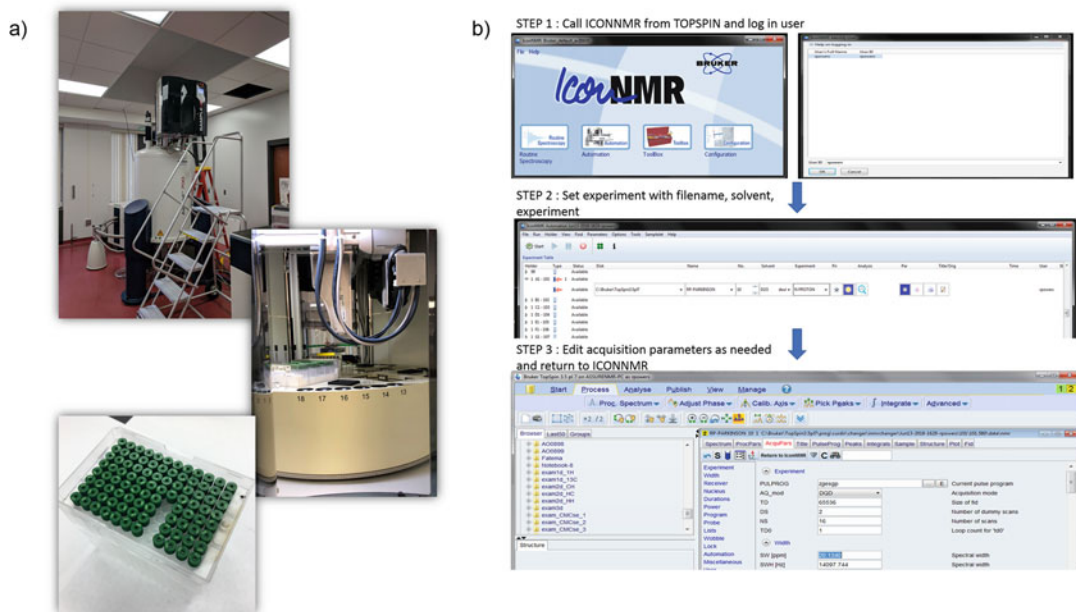
1. Perform **steps 1–5** from Subheading 3.9.
2. Lyse the cells using Matrix B lysing tubes using a FastPrep system. Each sample is bead beaten for 45 s burst at 1200 rpm for three rounds. Take precautions to avoid any sample heating during the process.
3. Perform **steps 7–16** from Subheading 3.9.

### 3.11 *Preparation of NMR Samples from Cell Lysates or Tissue Extracts*

1. For 1D NMR experiments, lyophilized cell-free lysates or tissue extracts are suspended in 550  $\mu\text{L}$  of 100% 50 mM  $\text{D}_2\text{O}$  phosphate buffer at pH 7.2 (uncorrected, *see* **Note 3**) with 50  $\mu\text{M}$  TMSP- $\text{d}_4$  as a chemical shift and concentration reference.
2. For 2D NMR experiments, lyophilized cell-free lysates or tissue extracts are suspended in 550  $\mu\text{L}$  of 100% 50 mM  $\text{D}_2\text{O}$  phosphate buffer at pH 7.2 (uncorrected, *see* **Note 3**) with 500  $\mu\text{M}$  TMSP- $\text{d}_4$  as a chemical shift and concentration reference.
3. Centrifuge the sample at  $14,000 \times g$  for 10 min to remove any particulates.
4. The sample is transferred to a 4" 5 mM SampleJet NMR tube with a RAININ XLS pipette (*see* **Note 12**).
5. Repeat **steps 1–4** for each replicate and for each group (*see* **Note 9**).
6. Each sample is added to a 96-well plate SampleJet configuration equilibrated to  $4\text{ }^{\circ}\text{C}$  to prevent metabolite degradation (*see* Fig. 5).

### 3.12 *Preparation of NMR Samples from Urine*

1. Thaw the urine samples stored at  $-80\text{ }^{\circ}\text{C}$  on ice.
2. Samples are centrifuged at  $15,000 \times g$  for 10 min at  $4\text{ }^{\circ}\text{C}$  to pellet any particulates in the sample.
3. 500  $\mu\text{L}$  of urine is transferred into a new Eppendorf tube and mixed with 50  $\mu\text{L}$  of 50 mM phosphate buffer in 99.8%  $\text{D}_2\text{O}$  at pH 7.2 (uncorrected, *see* **Note 3**). 50  $\mu\text{M}$  (for 1D  $^1\text{H}$  NMR) or 500  $\mu\text{M}$  (for 2D NMR) TMSP- $\text{d}_4$  are added to each sample as a chemical shift and concentration reference.
4. The sample is transferred to a 4" 5 mM SampleJet NMR tube with a RAININ XLS pipette (*see* **Note 12**).
5. Repeat **steps 1–7** for each replicate and for each group (*see* **Note 9**).



**Fig. 5** High-throughput sample preparation and data collection. **(a)** Images illustrating the loading of replicate metabolomics samples into the 96-well plate SampleJet configuration. **(b)** ICONNMR screenshots illustrating the stepwise workflow for setting up a high-throughput 1D  $^1\text{H}$  NMR metabolomics screen

- Each sample is added to a 96-well plate SampleJet configuration equilibrated to  $4^\circ\text{C}$  to prevent metabolite degradation (*see* Fig. 5).

### 3.13 Preparation of NMR Samples from Serum

- Thaw the serum samples stored at  $-80^\circ\text{C}$  on ice.
- 500  $\mu\text{L}$  of serum sample is transferred into a new Eppendorf tube and mixed with 1 mL of methanol to precipitate proteins.
- Samples are centrifuged at  $15,000 \times g$  for 10 min at  $4^\circ\text{C}$ .
- Transfer the supernatants into new microcentrifuge tubes.
- Remove the methanol by rotary evaporation using a SpeedVac.
- The samples are flash-frozen in liquid nitrogen and then dried by lyophilization.
- The dried samples are reconstituted into 550  $\mu\text{L}$  of 50 mM phosphate buffer in 99.8%  $\text{D}_2\text{O}$  at pH 7.2 (uncorrected, *see* Note 3). 50  $\mu\text{M}$  (for 1D  $^1\text{H}$  NMR) or 500  $\mu\text{M}$  (for 2D NMR) TMSP- $\text{d}_4$  are added to each sample as a chemical shift and concentration reference.
- The sample is transferred to a 4" 5 mM SampleJet NMR tube with a RAININ XLS pipette (*see* Note 12).
- Repeat steps 1–8 for each replicate and for each group (*see* Note 9).

- Each sample is added to a 96-well plate SampleJet configuration equilibrated to 4 °C to prevent metabolite degradation (*see* Fig. 5).

### 3.14 Preparation of NMR Samples from Fecal Matter

- Thaw the fecal samples stored at –80 °C on ice.
- Weigh 3–50 mg of the fecal material in a microcentrifuge tube containing matrix D type lysing beads.
- Add 1 mL of 1:1 water:methanol mixture to extract the metabolome and precipitate proteins.
- Using a FastPrep, bead beat the samples for 30 s at 1200 rpm thrice.
- Centrifuge at 15,000 × *g* for 10 min at 4 °C.
- Collect the supernatants in a separate microcentrifuge tube.
- Repeat **steps 3–6** to re-extract the pellet and combine both supernatants.
- Remove the methanol by rotary evaporation using a SpeedVac and water by lyophilization.
- Reconstitute in 550 µL of 50 mM phosphate buffer in 99.8% D<sub>2</sub>O at pH 7.2 (uncorrected, *see* **Note 3**). 50 µM (for 1D NMR) or 500 µM (for 2D NMR) of TMSP-d<sub>4</sub> will be added to each sample as a chemical shift and concentration reference.
- The sample is transferred to a 4" 5 mM SampleJet NMR tube with a RAININ XLS pipette (*see* **Note 12**).
- Repeat **steps 1–10** for each replicate and for each group (*see* **Note 9**).
- Each sample is added to a 96-well plate SampleJet configuration equilibrated to 4 °C to prevent metabolite degradation (*see* Fig. 5).

### 3.15 NMR Data Collection

All NMR experiments are conducted at 298 K using a Bruker AVANCE III HD 700 MHz spectrometer equipped with a 5 mm quadruple resonance QCI-P cryoprobe (<sup>1</sup>H, <sup>13</sup>C, <sup>15</sup>N, and <sup>31</sup>P) with *z*-axis gradients. An automatic tune and match system (ATM) and a SampleJet automated sample changer system with Bruker ICON-NMR software were used to automate the NMR data collection (*see* Fig. 5).

#### 3.15.1 1D <sup>1</sup>H NMR

- Load the NMR metabolomics samples into the SampleJet automated sample changer system (*see* Fig. 5 and **Note 9**). Check that the SampleJet is in the correct mode (i.e., 5 mm tubes).
- Log into an account on the spectrometer workstation and start the Topspin software.

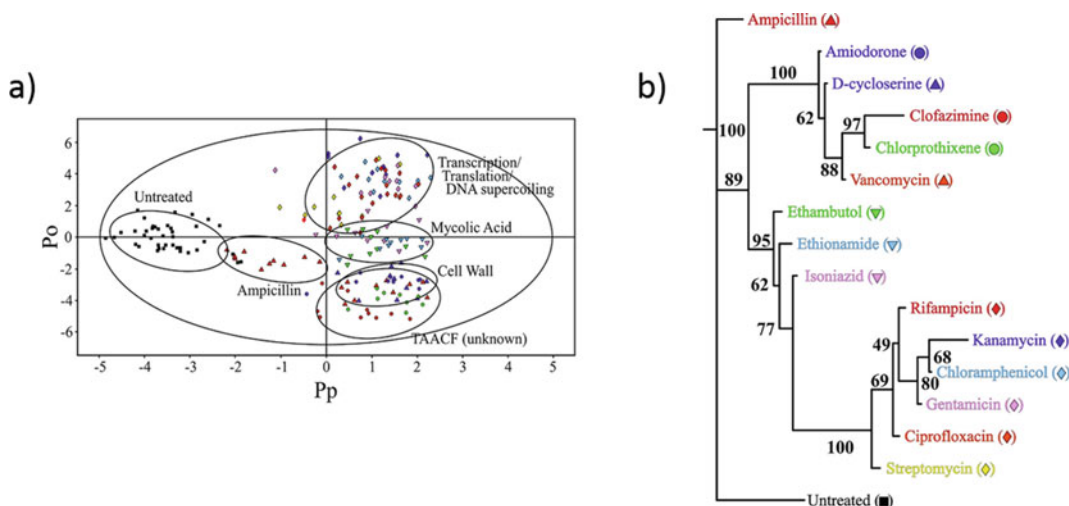


3. The first NMR sample is lowered into the magnet using the Bruker command, *sv 101*, where 101 corresponds to sample one in rack one.
4. The spectrometer is locked onto the D<sub>2</sub>O solvent frequency using the Bruker command, *lock D<sub>2</sub>O* (*see Note 13*).
5. The NMR sample is shimmed for optimal signal and suppression of the water signal by typing the Bruker command *topshim*. This will initiate an automated gradient shimming procedure, which may take a few min to complete (*see Note 13*).
6. The sample is automatically tuned and matched using the ATM system by typing the Bruker command, *atma*.
7. The 90° pulse length (μs) is determined by measuring a null spectrum with an approximate 360° pulse using the Bruker *zg* pulse sequence (*see Note 14*).
8. A 1D <sup>1</sup>H NMR spectrum is obtained for each sample using a standard excitation sculpting water suppression pulse program (Bruker *zgesgp* pulse sequence) that provides optimal suppression of the residual water signal while maintaining a flat baseline (*see Note 15*).
9. Typical experimental parameters for a 1D <sup>1</sup>H NMR spectrum obtained on a Bruker 700 MHz spectrometer with a cryoprobe correspond to 128 scans, 16 dummy scans, 32,768 data points, a spectral width of 11,160.7 Hz, and a relaxation delay of 1.5 (*see Note 16*).
10. Automated data collection of the entire set of metabolomics samples is accomplished using Bruker ICONNMR 5 (*see Fig. 5*).
11. The sample filename, solvent, pulse program, and temperature parameters are all defined in Bruker ICONNMR 5 (*see Notes 15–19*).
12. Collect the 1D <sup>1</sup>H NMR spectrum for each replicate and each group (*see Note 9*).
13. The data is processed initially with Topspin to verify spectral quality but exported for further analysis.

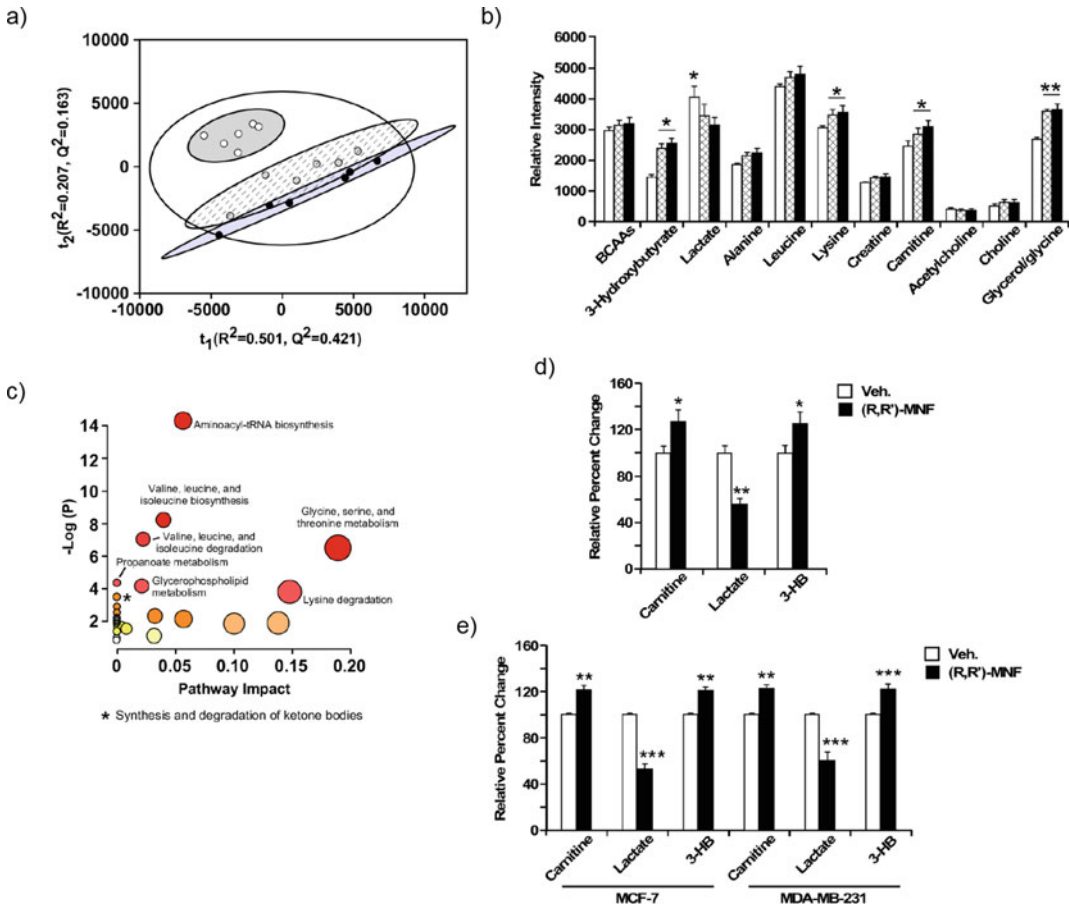
### 3.15.2 2D <sup>1</sup>H–<sup>13</sup>C-HSQC NMR (*See Note 20*)

1. Follow **steps 1–7** from Subheading **3.15.1**.
2. Using ICONNMR 5, the sample filename, solvent, pulse program, and temperature parameters are adjusted (*see Notes 16–19*).
3. The ICONNMR setup is similar to a 1D <sup>1</sup>H NMR data collection as shown in *Fig. 5*.
4. A standard 2D <sup>1</sup>H–<sup>13</sup>C-HSQC experiment (Bruker *hsqcetgpsisp2* pulse program) is used to determine the <sup>1</sup>H–<sup>13</sup>C chemical shift correlations for all <sup>13</sup>C-labeled metabolites in the metabolomics sample (*see Note 20*).

5. Typical experimental parameters for a 2D  $^1\text{H}$ - $^{13}\text{C}$ -HSQC NMR spectrum obtained on a Bruker 700 MHz spectrometer with a cryoprobe correspond to 128 scans, 32 dummy scans, and a 1.0 s relaxation delay. The spectrum is collected with 2k data points and a spectrum width of 4734 Hz in the direct dimension and 64 data points and a spectrum width of 18,864 Hz in the indirect dimension (*see Note 21*).
6. Implementation of fast NMR methods that includes non-uniform sampling significantly decreases data acquisition time and/or increases spectral resolution but may introduce artifacts (*see Note 22*).
7. Collect the 2D  $^1\text{H}$ - $^{13}\text{C}$ -HSQC NMR spectrum for each replicate and each group (*see Note 9*).
8. Process the data initially with Topspin to verify spectral quality but exported for further analysis (*see Figs. 6 and 7*).



**Fig. 6** A representative example of bacterial cells treated with existing drug therapies to decipher the in vivo mechanism of action of unknown drugs using 1D  $^1\text{H}$  NMR-based metabolomics. **(a)** 2D OPLS-DA scores plot demonstrating the clustering pattern for 12 antibiotics with known biological targets and 3 compounds of unknown in vivo activity: untreated cells (■ black), chloramphenicol (◆ teal), ciprofloxacin (◆ orange), gentamicin (◆ pink), kanamycin (◆ purple), rifampicin (◆ red), streptomycin (◆ yellow), ethambutol (▼ green), ethionamide (▼ blue), isoniazid (▼ pink), ampicillin (▲ red), D-cycloserine (▲ purple), vancomycin (▲ orange), amiodorone (● purple), chlorprothixene (● green), and clofazimine (● red)-treated *M. smegmatis* cells. The ellipses correspond to the 95% confidence limits from a normal distribution for each cluster. The untreated *M. smegmatis* cells (■ black) were designated the control class, and the remainder of the cells was designated as treated. The OPLS-DA used one predictive component and six orthogonal components to yield a  $R^2_X$  of 0.715,  $R^2_Y$  of 0.803, and  $Q^2$  of 0.671. **(b)** Metabolomics tree diagram was determined from the OPLS-DA scores plot. The coloring scheme and associated symbol for each compound in the tree diagram correlate with colored symbols in the OPLS-DA scores plot. The bootstrap numbers for each node are indicated on the tree diagram (Reproduced with permission from ref. 52. Copyright © 2012 American Chemical Society)



**Fig. 7** Metabolomics analysis on the effect of (R,R')-MNF in PANC-1 cells. **(a)** PCA scores resulting from modeling of the 1D  $^1\text{H}$  NMR data matrix from untreated PANC-1 cells (open symbols) and PANC-1 cells treated with 0.5  $\mu\text{M}$  (R,R')-MNF (gray symbols) or 1  $\mu\text{M}$  (R,R')-MNF (black symbols). A statistically significant degree of separation is observed between treated and untreated groups. The ellipses correspond to 95% confidence intervals for a normal distribution. Each principal component is labeled with the corresponding  $R^2$  and  $Q^2$  values. **(b)** Bar graph of 1D  $^1\text{H}$  NMR peak intensities (relative metabolite concentrations) resulting from the analysis of PANC-1 cellular extracts after a 1-h incubation with 0.5  $\mu\text{M}$  (hatched bars) or 1  $\mu\text{M}$  (filled bars) of (R,R')-MNF versus untreated cells (open bars). Metabolites were identified from the backscaled loadings plots as the major contributors to the group separations in the OPLS models. Benjamini-Hochberg corrected Student's  $t$ -test  $p$ -values from pairwise comparisons are indicated. An asterisk indicates significance between the control and 1  $\mu\text{M}$  of (R,R')-MNF, while an asterisk with underlying bar indicates significance between the control and both treated groups. **(c)** Metabolic pathway analysis of the identified metabolites found in the 1D  $^1\text{H}$  NMR spectra of the PANC-1 extracts after 1-h incubation with 0.5  $\mu\text{M}$  or 1  $\mu\text{M}$  of (R,R')-MNF. Each circle represents a matched pathway and is colored according to its  $p$  values from the pathway enrichment analysis. Statistically significant pathways ( $p < 0.05$ ) are labeled with their common name. **(d)** and **(e)** Targeted metabolomics on the effect of 1  $\mu\text{M}$  (R,R')-MNF on intracellular concentrations of carnitine,  $\text{L}$ -lactate, and 3-hydroxybutyrate in PANC-1 cells **(d)** as well as MCF-7 and MDA-MB-231 breast tumor cells **(e)**. \*  $p \leq 0.05$ ; \*\*  $p \leq 0.01$ ; \*\*\*  $p \leq 0.001$  (Reproduced with permission from M. Bernier, J. Catazaro, N. S. Singh, Q. Shi, M. Wang, Q. Yang, A. Wnorowski, K. Jozwiak, R. Powers, and I. W. Wainer\* (2017) "GPR55 Receptor Antagonists Decrease Glycolytic Activity in PANC-1 Pancreatic Cancer Cell Lines and Tumor Xenografts", *International Journal of Cancer*, 141(10):2131–2142 Copyright © 2017 John Wiley and Sons)

### 3.16 NMR Data Processing (See Note 23)

All NMR data is processed and analyzed with our MVAPACK software [57], PCA/PLS-DA utilities [58], NMRPipe [59], and NMRViewJ [60]. See example processing scripts at <http://bionmr.unl.edu/wiki/Scripts>. Please see the NMRPipe and nmrDraw tutorial (<https://spin.niddk.nih.gov/NMRPipe/doc1/>) and NMRViewJ documentation (<http://docs.nmrfox.org/>) for more details.

#### 3.16.1 1D <sup>1</sup>H NMR (See Figs. 6 and 8)

1. Apply a 1.0 Hz exponential apodization function to the FID.
2. Fourier transform the FID.
3. The resulting NMR spectrum is automatically simultaneously phased corrected and normalized with the phase-scatter correction algorithm [82].
4. Reference the NMR spectrum to the peak of TMSP-d<sub>4</sub> (0.0 ppm).
5. Remove the noise and solvent regions manually.

#### 3.16.2 2D <sup>1</sup>H-<sup>13</sup>C-HSQC NMR (See Figs. 7 and 8)

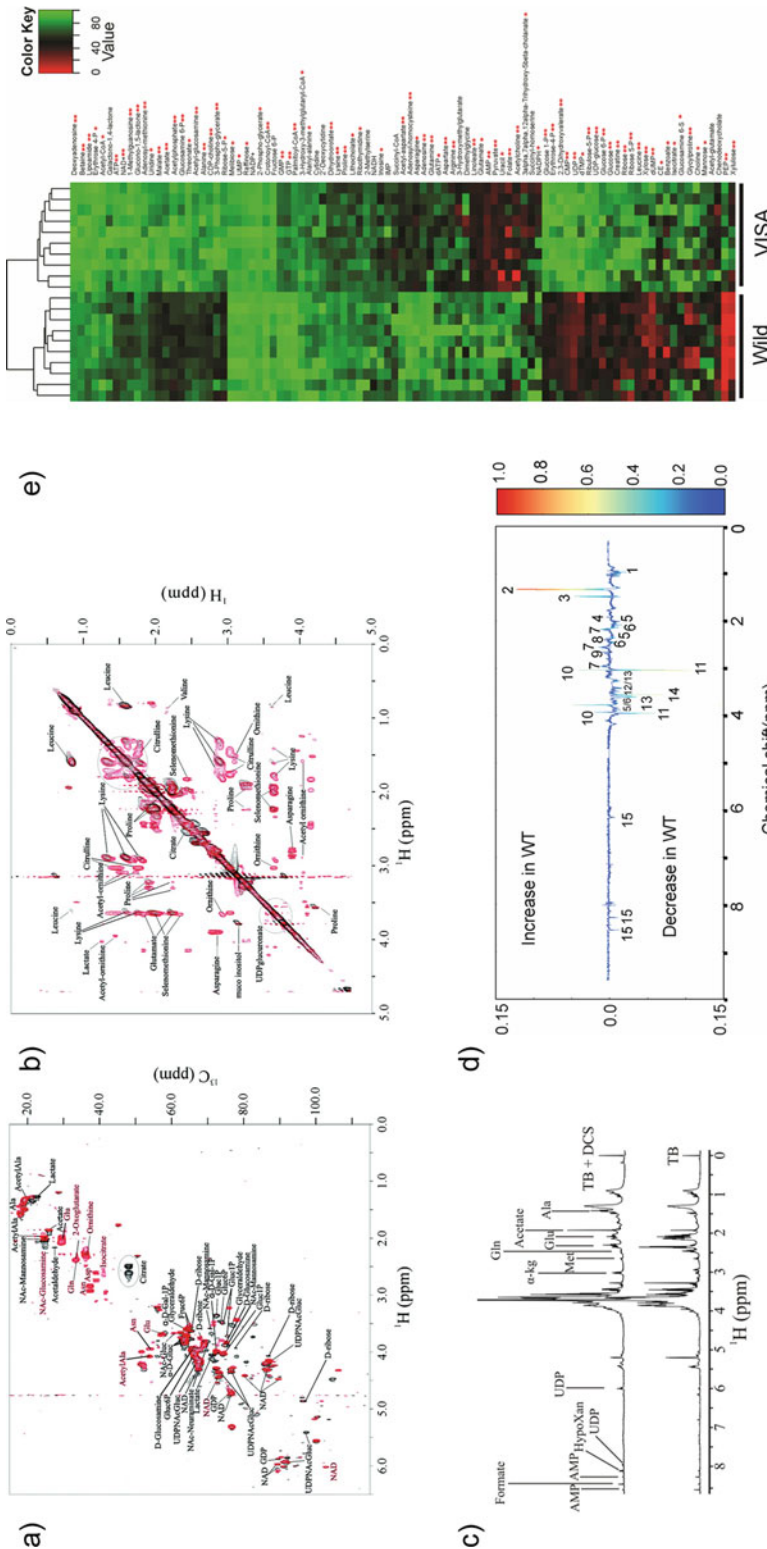
1. Apply a sine-bell apodization function to the  $t_2$  dimension.
2. Zero fill the  $t_2$  dimension three times.
3. Fourier transform the  $t_2$  dimension, manually phase correct, and delete the imaginary data.
4. Transpose the data matrix.
5. Apply a sine-bell apodization function to the  $t_1$  dimension.
6. Zero fill the  $t_1$  dimension three times.
7. Fourier transform the  $t_1$  dimension and phase correct manually.
8. Reference the NMR spectrum in both dimensions to the peak of TMSP-d<sub>4</sub> (0.0 ppm).

### 3.17 NMR Data Preprocessing for Multivariate Modeling

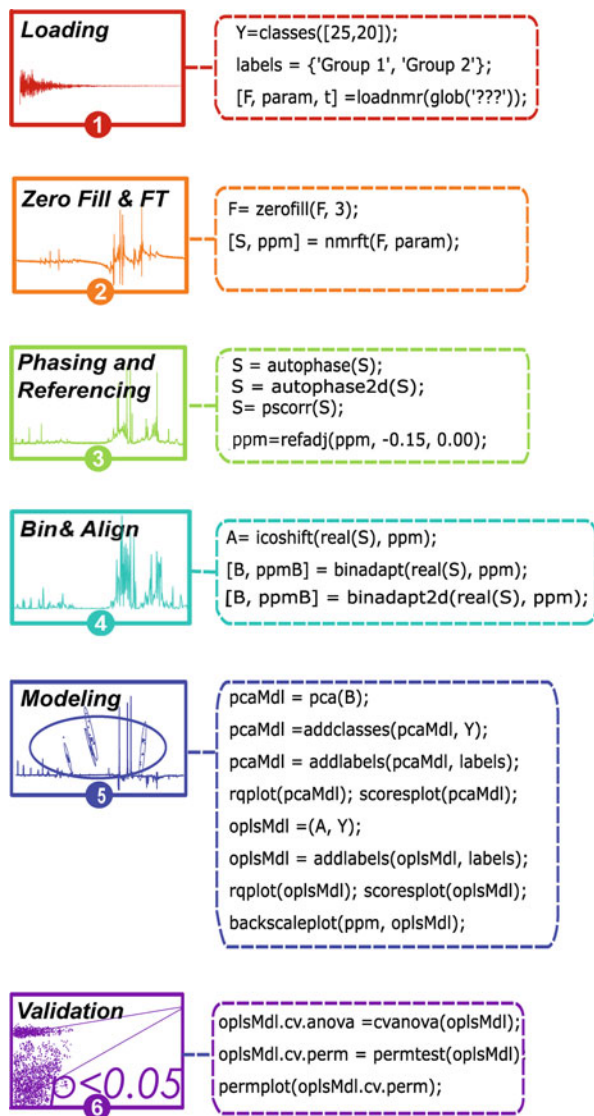
To obtain an accurate and reliable multivariate statistical model, it is essential that the dataset is properly preprocessed to remove normal systematic variations resulting from biological variability, instrument instability, and inconsistency in sample handling and preparation. Key preprocessing steps include (1) alignment, (2) normalization, (3) binning, and (4) scaling, which is illustrated in Fig. 9. All NMR datasets are processed with our MVAPACK software [82] and our PCA/PLS-DA utilities [58]. See example MVAPACK scripts at <http://bionmr.unl.edu/wiki/Scripts>.

#### 3.17.1 1D <sup>1</sup>H NMR

1. Spectra may first be normalized based on either the total cell count or the total protein concentration using the BCA (bicinchoninic acid) protein estimation assay. The BCA assay uses parallel dishes that are similarly treated.
2. Normalize the spectra with the PSC algorithm [82].



**Fig. 8** Examples of a typical (a) 2D  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectrum and (b) 2D  $^1\text{H}$ - $^1\text{H}$  TOCSY spectrum acquired from bacterial cells (Reproduced with permission from Zhang, B.; Halouska, S.; Schiaffo, C. E.; Sadykov, M. R.; Somerville, G. A.; Powers, R. NMR analysis of a stress response metabolic signaling network. *J. Proteome Res.*, 2011, 10, 3743–3754. Copyright © 2011 American Chemical Society). (c) 1D  $^1\text{H}$  NMR spectrum acquired from bacterial cells (Reproduced with permission from S. Halouska, R. J. Fenton, D. K. Zinniel, D. D. Marshall, R. G. Barletta\* and R. Powers\* (2014) “Metabolomics Analysis Identifies D-Alanine- $\text{D}$ -alanine Ligase as the Primary Lethal Target of  $\text{D}$ -cycloserine in *Mycobacteria*”, *J. Proteome Res.*, 13(2):1065-1076. Copyright © 2014, American Chemical Society). (d) Back-scaled loadings plot for 1D  $^1\text{H}$  NMR spectra showing 15 different metabolites that vary between groups (Reproduced with permission from ref. 15. Copyright © 2017 American Society for Microbiology). (e) Heatmap showing differences between the vancomycin-susceptible *Staphylococcus aureus* (VSSA) strains and *S. aureus* with intermediate-level resistance to vancomycin (VISA) metabolome. The heatmap showing differences between the VSSA or VISA strains and normalized peak intensities of the VSSA or VISA strain, and hierarchical clustering is depicted by the dendrogram. The color scale ranges from 0 (red; less intense) to 1 (green; intense). Statistical significance is denoted by a single asterisk if determined using Student’s t test at the 95% confidence level or by two asterisks if determined by Student’s t test with a Bonferroni correction for multiple hypothesis testing (Reproduced with permission from ref. 15. Copyright © 2017 American Society for Microbiology)



**Fig. 9** Example of an MVAPACK processing script for a 1D  $^1\text{H}$  NMR dataset. The numbered steps in the flow diagram correspond to the numbered lines in the processing script

3. Align and/or bin the spectra. For principal component analysis (PCA), use the following parameters:
  - (a) Globally align the spectral data to the peak of TMS- $\text{d}_4$  at 0.0 ppm.
  - (b) Regionally align the spectral data using the icoshift algorithm [83].
  - (c) Bin the spectral data using the adaptive, intelligent binning algorithm [84].

4. For orthogonal projection to latent structures (OPLS), use the following parameters:
  - (a) Globally align the spectral data to the peak of TMS- $d_4$  at 0.0 ppm.
  - (b) Regionally align the spectral data using the icoshift algorithm [83].
  - (c) The spectral data are not binned. Instead, use the full-resolution spectral data to build the model.
5. Manually remove the solvent peaks and noise regions.
6. Scale the dataset using Pareto scaling.
7. Generate the PCA or OPLS model from the data matrix.

### 3.17.2 2D $^1H$ - $^{13}C$ -HSQC NMR

1. Spectra may first be normalized based on either the total cell count or the total protein concentration as explained above.
2. Normalize the spectral data using standard normal variate normalization.
3. Bin the spectral data using a generalized adaptive, intelligent binning algorithm [84].
4. Pareto scale the data.
5. Generate a PCA or OPLS model from the data matrix.

### 3.18 Statistical Analysis

Datasets are analyzed with our MVAPACK software [57], PCA/PLS-DA utilities [58], and R [85]. See example MVAPACK and R scripts at <http://bionmr.unl.edu/wiki/Scripts>.

Improper application of statistical methods is a serious issue in the field of metabolomics, which results from some common misconceptions. For example, validation of supervised multivariate statistical methods are routinely lacking. Instead, supervised OPLS and PLS models need to be validated using a cross-validation score like CV-ANOVA [86] and/or response permutation testing [87].  $R^2$  and  $Q^2$  values from these supervised models provide a measure of model fitness from the original data and a measure of consistency between the original and predicted data, respectively.  $R^2$  and  $Q^2$  values do not provide for model validation without a proper standard of comparison.

Metabolomics datasets are often described using PCA or PLS/OPLS models. It is important to note that PCA finds the largest source of variance in the dataset irrespective of the intent of the study. Separation of clusters observed in a PCA model may not have anything to do with a drug treatment if there are other major sources of variance in the data. Thus, a loadings plot generated from a PCA model represents all of the variation in the dataset, not just the expected changes induced by the drug treatment. Supervised methods, like PLS and OPLS, address this issue by aggressively forcing group separation based on the defined group

membership. So, by design, group separation is almost always achieved in a supervised model and should not be interpreted as a novel discovery in itself. Again, group separation may occur irrespective of the validity of the model. This happens because PLS and OPLS models are easily over-fitted, especially for metabolomics datasets since the number of variables (e.g., metabolites) are typically larger than the number of replicates. This forced separation of groups also contributes to a common misconception in the metabolomics community, the incorrect belief that OPLS or PLS outperforms PCA in identifying group separation and the underlying dataset variances. Remember, PCA finds the largest source of variance in the dataset, so if there is no variance in a PCA model, smaller group-specific variations are also not present. Thus, if PCA fails to identify group separation, it is unlikely that PLS or OPLS will yield a valid model despite the appearance of group separation [87]. Simply, PCA, PLS, and OPLS generate different models by extracting different information to achieve different goals.

PLS and OPLS appear to provide a similar model. In fact, a comparison of PLS and OPLS scores plots generated from the same dataset may suggest the only difference is a relative rotation of the group-defined ellipses. Nevertheless, this apparently subtle change highlights a critical difference. PLS intermingles both group-independent and group-dependent variance, whereas OPLS places group-independent variance (e.g., confounding factors such as differences in diet, age, race, etc.) orthogonal to group-dependent variance. Therefore, a metabolite identified as a major contributor to an OPLS model is strictly the result of the defined group difference. Accordingly, we strongly recommend using OPLS instead of PLS.

### 3.18.1 Univariate Analysis

1. Determine relative metabolite abundances from NMR spectral peak heights and/or peak volumes (*see* Fig. 8).
2. Normalize relative metabolite abundances on a per spectrum basis. One common approach (standard normal variate) is to convert the absolute peak intensities (arbitrary units) to a *Z*-score:

$$Z = \frac{I_i - \bar{I}}{\sigma} \quad (1)$$

where  $\bar{I}$  is the average peak intensity for the spectrum,  $I_i$  is the intensity of peak  $i$ , and  $\sigma$  is the standard deviation of peak intensities. Peak intensities may also be normalized to the total number of cells, to the total protein concentration (*see* Subheading 3.17.1), to the average spectral noise, or to an internal standard (*see* Note 24). Relative metabolite abundances may also be converted to fold changes:



$$F = \frac{I_i}{I_o} \quad (2)$$

where  $I_i$  is the normalized peak intensity of metabolite  $i$  from a treated spectrum and  $I_o$  is the normalized peak intensity of metabolite  $i$  from the control or untreated spectrum.

3. A standard Student's  $t$ -test is commonly used to determine statistical significance only for a pairwise comparison of metabolite changes based on either fold changes or normalized peak intensities (*see* **Note 25**). A statistically significant difference is typically identified by a  $p$ -value  $< 0.05$  (*see* Fig. 8).
4. A Student's  $t$ -test is insufficient for the multiple comparisons that are common to a metabolomics study [87, 88]. In order to identify the set of metabolites that exhibit a statistically significant change, a multiple hypothesis test correction method such as a Benjamini-Hochberg [89] or a Bonferroni [90] correction must be applied (*see* **Notes 25** and **26**).
5. A heatmap with hierarchical clustering is commonly generated from the fold changes or normalized peak intensities using R (*see* example R script at <http://bionmr.unl.edu/wiki/Scripts>). The heatmap may contain relative metabolite abundances for each individual replicate in the study or simply the replicate-averages for each group (*see* **Note 27**).

### 3.18.2 Multivariate Analysis

1. Generate a PCA and or OPLS model from the data matrix.
2. Fractions of explained variation ( $R^2_X$  and  $R^2_Y$ ) are computed during PCA or OPLS model training.
3. Internally cross-validate the PCA or OPLS model using seven-fold Monte Carlo cross-validation [91] to compute  $Q^2$  values (*see* **Note 28**).
4. For an OPLS model, compare the  $Q^2$  value against a distribution of null model  $Q^2$  values in 1000 rounds of response permutation testing [87]. Group membership is randomly reassigned to generate the set of null models. A  $p$ -value is calculated from a comparison of the true  $Q^2$  value to the set of null model  $Q^2$  values (*see* **Note 28**).
5. Further validate the model using CV-ANOVA significance testing, which is used to calculate another model  $p$ -value [86] (*see* **Note 28**).
6. Scores plots (*see* Fig. 6), backscaled loadings plots (*see* Fig. 8), S-plots, and/or SUS-plots are often generated from OPLS models.
7. PCA/PLS-DA utilities [58] are used to define group membership by drawing an ellipse per group onto the scores plots (*see* Figs. 6 and 7). Each ellipse corresponds to 95% confidence interval for a normal distribution. The PCA/PLS-DA utilities

also generate a metabolomics tree diagram that identifies the statistical significance ( $p$ -value and/or bootstrap value) and the relative similarity of each group in the scores plot (*see* Fig. 6). The  $p$ -value or bootstrap number from the pairwise comparison is labeled at each node in the tree.

### 3.19 Data Analysis: Metabolite Assignment from 1D $^1\text{H}$ NMR Data

NMR data are analyzed with NMRPipe [59], NMRViewJ [60], and Chenomx. See example scripts at <http://bionmr.unl.edu/wiki/Scripts>. Please see the NMRPipe and nmrDraw tutorial (<https://spin.niddk.nih.gov/NMRPipe/doc1/>) and NMRViewJ documentation (<http://docs.nmrfx.org/>) for more details.

1. Perform the identification of metabolites in a 1D  $^1\text{H}$  NMR spectrum with software programs such as Chenomx. Chenomx matches the experimental 1D  $^1\text{H}$  NMR spectrum to a database of 1D  $^1\text{H}$  NMR spectra of known metabolites. Chenomx attempts to explain or describe the experimental NMR spectrum by combining or summing as many of the individual reference metabolite NMR spectra as needed. In addition to metabolite identification, Chenomx also provides an estimate of the metabolite concentration (*see* Note 29).
2. Upload the 1D  $^1\text{H}$  NMR spectrum for processing. The NMR spectra can be batch processed or processed one at a time.
3. Phase the 1D  $^1\text{H}$  NMR spectrum.
4. Calibrate and reference the 1D  $^1\text{H}$  NMR spectrum to TMS- $\text{d}_4$ , using the known concentration of TMS- $\text{d}_4$ .
5. Send the properly phased and calibrated 1D  $^1\text{H}$  NMR spectrum to the Chenomx profiler where the spectrum is compared against the metabolite library.
6. Chenomx will overlay a 1D  $^1\text{H}$  NMR reference spectrum for each metabolite identified in the experimental 1D  $^1\text{H}$  NMR spectrum. The spectral overlay needs to be manually adjusted to optimize the alignment of the experimental 1D  $^1\text{H}$  NMR spectrum with the reference spectrum. Figure 8 shows an example of a labeled 1D  $^1\text{H}$  NMR spectrum.

### 3.20 Data Analysis: Metabolite Assignment from 2D $^1\text{H}$ - $^{13}\text{C}$ -HSQC NMR Data

NMR data are analyzed with NMRPipe [59] and NMRViewJ [60]. See example scripts at <http://bionmr.unl.edu/wiki/Scripts>. Please see the NMRPipe and nmrDraw tutorial (<https://spin.niddk.nih.gov/NMRPipe/doc1/>) and NMRViewJ documentation (<http://docs.nmrfx.org/>) for more details.

#### 3.20.1 NMRPipe Processing to Obtain .ft2 and .nv Files

1. The data files from ICONNMR can be used directly by NMRPipe to process the 2D  $^1\text{H}$ - $^{13}\text{C}$ -HSQC spectra.

2. On a Linux workstation, open a terminal and go to the directory that contains the NMR data. Type *bruker* to start the NMRpipe software.
3. Read in the experimental parameters file by clicking *Read Parameters* and verify that all the parameters have been correctly updated. Confirm that the mode of data collection has been set to *echo-antiecho* if the NMR spectrum was collected with the *hsqcetgpgpsisp2* pulse program.
4. Click *Update Script* to save an NMRPipe processing script *fid.com* file in the working directory.
5. Type *./fid.com* to start the NMRPipe processing script.
6. When the NMRPipe processing has finished, type *nmrDraw* to view the processed NMR spectrum. Please see the NMRPipe and *nmrDraw* tutorial (<https://spin.niddk.nih.gov/NMRPipe/doc1/>) for detailed instructions.
7. Phase the NMR spectrum in NMRpipe and note the  $p_0$  and  $p_1$  values for both the  $^1\text{H}$  and  $^{13}\text{C}$  dimensions.
8. Edit the NMRPipe processing script *hsqcproc.com* and replace the parameters associated with the NMRPipe phase correction command, *ps*, with the  $p_0$  and  $p_1$  values obtained from **step 7**.
9. Type *./hsqcproc.com* to start the NMRPipe processing script.
10. Repeat **steps 3–9** for each 2D  $^1\text{H}$ - $^{13}\text{C}$ -HSQC NMR spectrum in the dataset. This produces a set *.ft2* files. One *.ft2* file is created for each 2D  $^1\text{H}$ - $^{13}\text{C}$ -HSQC NMR spectrum collected for each replicate from each group.
11. Copy all of the *.ft2* files into a new folder and use the NMRPipe script *addnmr.com* to generate NMRviewJ files from the *.ft2* files. A *.nv* file will be generated for each individual spectrum (*.ft2* file) with a numerically incremented root name of *Final\_*. In addition, the script will combine all of the NMR spectra together into a single file called *results.nv*. The script will also generate the text file, *rate.txt*, which lists all of the individual *.nv* files (*Final\_*).

3.20.2 Peak Picking  
and Peak Integration of 2D  
 $^1\text{H}$ - $^{13}\text{C}$ -HSQC Spectra  
in NMRviewJ

1. Type *nmrviewj* to start NMRviewJ. Refer to NMRViewJ documentation (<http://docs.nmrfx.org/>) for more details.
2. From the *Dataset* toolbar in the *main* window, use the *Open and Draw Datasets* function to select the *result.nv* file.
3. Right click and select *attributes* to open the *attributes* window.
4. In the *attributes* window, select the *PeakPick* tab.
5. In the blank *Lists* field in the *attribute* window, type a filename (i.e., *lists*) for the new peak pick list. Click the *Pick* button. The software will automatically peak pick the displayed spectrum and populate *lists* with the peak ID number, chemical shifts, and intensity.

6. Choose *Show Peak Table* from the *Peak* toolbar on the *main* window. A *peak table* window will open that lists the peak ID, peak intensity, and the peak chemical shifts.
7. Manually edit the peak list and remove solvent peaks, noise peaks, or other spectral artifacts. Peaks are deleted from the peak table by using the delete function in the *PeakPick* tab in the *attributes* window along with the *spectrum display* window. In the *spectrum display* window, use the mouse to position the two cursors around any peak or spectral region to form a box. Then, click the *Delete* button under the *PeakPick* tab in the *attributes* window to remove the peak(s).
8. After the peak table has been completely edited, on the *peak table* window, choose the *Edit* tab and select *Compress and Degap*. Answer *yes* to the pop-up question. This will finalize changes to the peak list and prevent any further edits.
9. On the *peak table* window, choose the *Edit* tab and then select *Save Table*. A file browser window will open in order to choose a name and location to save the new peak list file. The saved peak pick file can be viewed and edited in Excel.
10. In order to obtain peak intensities across the entire set of NMR spectra in the dataset, click on the *Analysis* tab on the main window and select *Rate Analysis*. A setup window for the *Rate Analysis* will open.
11. In the *Rate Analysis* setup window:
  - (a) Set the Prefix for matrix numbers field to Final\_.
  - (b) Set the *Peaklist* field to *lists* (defined in Subheading 3.20.2, step 5).
  - (c) Make sure *Auto fit* is checked.
  - (d) Use all other default settings.
  - (e) Click *Load time file*.
  - (f) In the file browser window, select *rate.txt* (created in Subheading 3.20.1, step 11).
  - (g) Click *Measure All*. The software automatically populates the table in the *Rate Analysis* setup window with all of the peak intensities across the entire NMR dataset.
  - (h) Click *Save Table*. In the file browser window, save the peak intensities table to a new filename (i.e., *intensities*).
12. The peak list (i.e., *list*) and the peak intensities (i.e., *intensities*) files are merged in Microsoft Excel using the common peak ID column. The *ppm1* ( $^1\text{H}$  ppm) and *ppm2* ( $^{13}\text{C}$  ppm) columns are added to the peak intensities columns to generate a complete matrix of NMR peaks and intensities across the entire dataset.
13. The merged Excel file is saved to a new filename.

3.20.3 Metabolite  
Assignments from 2D  
 $^1\text{H}$ - $^{13}\text{C}$ -HSQC Peak Lists

1. The complete list of peaks obtained from the NMRviewJ analysis is searched using NMR metabolomics databases such as HMDB [62], BMRB [63], or other databases (see **Note 30**).
2. On the HMDB homepage, choose the *Search* tab and select *2D NMR Search*.
3. From the *Spectra Library* pull-down menu, choose  $^{13}\text{C}$  HSQC.
4. Cut and paste the 2D  $^1\text{H}$ - $^{13}\text{C}$ -HSQC peak lists into the *Peak List* field. One set of  $^1\text{H}$  and  $^{13}\text{C}$  chemical shifts, respectively, per line. Chemical shift values should only be separated by white space.
5. Set the  $^1\text{H}$  chemical shift error tolerance to 0.05 ppm (*X-axis Peak Tolerance*  $\pm$  field) and the  $^{13}\text{C}$  chemical shift error tolerance to 0.10 ppm (*Y-axis Peak Tolerance*  $\pm$  field).
6. Click the *Search* button. Depending on the size of the peak list, the software will return a ranked-order list of possible metabolites based on the number of chemical shift matches to reference spectrum.
7. Manually curate the list of potential metabolite assignments based on the number of chemical shift assignments, the quality of the spectral overlap (i.e., chemical shift match), the number of other metabolites in the same metabolic pathway, and the biological system (i.e., is it a reasonable or possible metabolite for the organism).
8. Obtain additional NMR (e.g., HMBC, HSQC-TOCSY) and/or MS spectral data to confirm or refute the assignment.
9. An assigned 2D  $^1\text{H}$ - $^{13}\text{C}$ -HSQC spectrum is shown in Fig. 8.

---

## 4 Notes

1. NMR isotope labeled compounds are non-radioactive and do not require any special handling. It is imperative you wear typical personal protective gear like gloves and eye protection and follow standard safety protocols.
2. Always keep deuterated solvents in a desiccator or dry box and keep them sealed or wrapped in parafilm while in storage.
3. Typically, metabolomics sample from cell lysates is prepared in 100%  $\text{D}_2\text{O}$ , and using a standard pH probe may not report pH accurately. Standard corrections are  $\text{pD} = \text{pH} + 0.4$  although Rubinson et al. [92] suggest that phosphate buffer pH is not significantly altered to warrant corrections.
4. The choice of culture medium is dependent on the specific needs of the cell type and the specific goals of the experiment. There are a variety of cell culture media available either through

commercial sources or from published recipes. Also, it may be possible to use different media for the same cell type. Again, the choice of medium will be dictated by the specific needs of the experiment. Accordingly, only a few representative culture media are listed as illustrative examples, and these are not intended to be an exhaustive list.

5. For the success of a metabolomics experiment, cell survival is an essential requirement. This presents a challenge when investigating the efficacy, side effects, mechanism of action, or resistance of a drug that impacts cell survival or is designed to kill cells. Similarly, cell mutations or experimental conditions (e.g., nutrient depletion, high temperature, anaerobic growth, etc.) that may lead to an increase in cell death are also potential problems. The goal is to identify a drug dosage and/or experimental condition that only stress the cells, but does not result in *any* cell death (*see* Subheading 3.3). In this regard, the observed metabolomic changes will only be a result of the cell's immediate response to the biological activity of the drug or the experimental condition. Conversely, if the metabolome is harvested from a mixture of live and dead cells, then it will be difficult, if not impossible, to deconvolute metabolome changes resulting from the stressor or cell death. A correct drug dosage or experimental condition is usually determined by acquiring a series of growth curves in which the drug or stressor is titrated over a wide range of values.
6. The choice of solvent and concentration is determined by the compound's solubility. The goal is to maximize the concentration of the stock solution so as to minimize the amount of solvent that needs to be added to each cell culture. The maximal amount of solvent added to a cell culture is 5–10  $\mu\text{L}$ .
7. The following range of drug concentrations is typically used: 0 $\times$ , 1 $\times$ , 2 $\times$ , 4 $\times$ , 8 $\times$ , 16 $\times$ , and 24 $\times$  the MIC value. The amount of solvent (i.e., DMSO- $d_6$  or ethanol) added to any cell culture should be 5–10  $\mu\text{L}$  or less.
8. The number of replicates per group will have a significant impact on the quality of the study and the statistical validity of the outcomes. In general, it is best to maximize the number of replicates per group, with a typical target of ten replicates per group. A variety of experimental considerations may impact the number of replicates that are practical for a given study. For example, a large number of groups may require a reduction in the number of replicates per group. Another consideration is the impact of the number of replicates on the quality of the metabolomics samples. Sacrificing quality for a greater number of replicates will not likely lead to a successful outcome. Conversely, a limited number of replicates <4 per group will likely provide meaningless results.

9. In order to avoid the introduction of bias, always randomize sample handling steps. If the samples are always processed in the order of the sample number, then a time bias will be induced across the entire dataset. Instead, if the order is constantly changed at each step, the processing time and any impact on the metabolome will be randomized, which in turn will minimize or eliminate any bias.
10. The experimental conditions for optimal labeling should be experimentally verified for every cell line or cell type used for a metabolomics experiment. Factors to consider include the composition of the culture media and the  $^{13}\text{C}$ -carbon source (e.g., glucose, pyruvate, or glutamine) that are used for the cell culture. For mammalian cells, a time course experiment should be conducted where cells are cultured from 1 to 48 h to assess the rate of carbon consumption. For bacterial cells, a time course experiment should be conducted where cells are cultured for one to three doubling times to assess the rate of carbon consumption. Examples of media used for  $^{13}\text{C}$ -carbon labeling of mammalian cell metabolome are Dulbecco's modified Eagle medium (DMEM) and Roswell Park Memorial Institute (RPMI) media. Examples of media used for  $^{13}\text{C}$ -carbon labeling of bacterial cell metabolome are Middlebrook's albumin dextrose catalase (MADC) media or M9 minimal media.
11. Removal of proteins and other biomolecules by methanol or ethanol precipitation is preferred over mechanical filtration methods or the application of Carr-Purcell-Meiboom-Gill (CPMG) NMR  $T_2$  filtering techniques. Filtering techniques are known to impart biologically irrelevant group differences, due to metabolite-macromolecule binding [87].
12. Smaller diameter NMR tubes of 3 mm (160  $\mu\text{L}$ ) or 1.7 mm (45  $\mu\text{L}$ ) may be needed if the available metabolomics sample is limited. Filling of these smaller diameter NMR tubes may be improved by a liquid handling robot, such as a Gilson 215 Liquid Handler. In addition, the NMR acquisition parameters will likely need to be adjusted to account for the lower sensitivity due to the lower number of nuclei in the samples.
13. Topshim requires the sample to contain either a  $\text{D}_2\text{O}$  or  $\text{H}_2\text{O}$  solvent. It is advisable to create a shim file with a parameter set that produces an optimal set of shims for your sample type. Read in a shim file using the Bruker command *rsb* and select the appropriate Topshim shim file. If you are doing this for the first time, complete the command *topshim*; if you are not satisfied with the shim performance, use command *topshim tuneb tunea* to obtain an improved set of shims. Write the shim set parameters with the Bruker command *wsb* and save it to a new file name for future reference.

14. The 90° pulse length is commonly measured by incrementing the P1 pulse in the *zgpg* pulse program by 1 μs or smaller increments and by plotting the relative peak heights or intensities. A maximum peak height should be observed at the pulse length corresponding to the 90° pulse. Conversely, a minimum or null spectrum should be observed at the pulse length corresponding to the 360° pulse length. In practice, a more accurate measure of the 90° pulse is obtained by measuring the 360° pulse length and dividing by four to obtain the 90° pulse length. A typical 90° pulse length for a metabolomics sample ranges from approximately 8–13 μs or longer. Among other factors, the relative salt concentration of the metabolomics sample affects the 90° pulse, in which a higher salt concentration results in a longer 90° pulse. Other factors also contribute to the observed 90° pulse, so it is always necessary to experimentally determine the 90° pulse for each sample or set of samples.
15. Excitation sculpting parameters (*zgesgp*)—32,768 data points (TD), SW = 12.02 ppm, O1P (transmitter offset) = 4.70 ppm, D1 = 1 s, NS (number of scans) = 128, DS (dummy/steady state scans) = 16, P1 = 9.5–13.5 μs, SPNAM (shaped pulse for water suppression) = SINC1.1000 at 26.39 dB or 0.00228 W.
16. The NMR data acquisition parameters need to be adjusted to compensate for differences in the field strength and sensitivity of the NMR spectrometer actually used for the data collection. Specifically, the number of scans, the number of data points, the sweep width (13.79 ppm, <sup>1</sup>H frequency range), and the frequency offset (centered on water peak at 4.70 ppm) need to be adjusted according to the type and configuration of the NMR spectrometer used for the study.
17. For high-throughput NMR data collection, please refer to the Bruker ICONNMR manual to explore various configuration options. For example, composite experiments allow for the collection of multiple 1D and 2D experiments for the same metabolomics sample. An experimental set consisting of a 1D <sup>1</sup>H, a 2D <sup>1</sup>H–<sup>13</sup>C HSQC, and a 2D <sup>1</sup>H–<sup>13</sup>C HMBC experiment may be subsequently collected for the same sample before moving to the next sample in the queue.
18. It is imperative that NMR data are collected at the same temperature for a queue of metabolomics samples. ICONNMR assists this by allowing for a temperature delay when a large number of samples are in the SampleJet queue. For example, a 15–60-s delay may be inserted prior to data acquisition to allow each sample to equilibrate to the probe temperature. We recommend a 60-s delay for both pre- and post-sample insertion to prevent any temperature variation.



19. Parameters to check before you queue experiments in ICONNMR for 1D  $^1\text{H}$  NMR are number of scans *ns*, number of dummy scans *ds*,  $90^\circ$  pulse *p1*, delay *d1*, sweep width *sw*, receiver gain *rg*, experiment temperature *te*, and automation setup *aunm*. We recommend using *au\_zgonly* as the automation setup. This will collect all samples at the same receiver gain, which will avoid peak intensity variation across the dataset.
20. NMR metabolomics studies are beginning to make use of HMBC, TOCSY, HSQC-TOCSY, 2D J-Resolved spectra, or other experiments along with the traditional 2D  $^1\text{H}$ - $^{13}\text{C}$  HSQC approach. Heteronuclear isotope labels like  $^{15}\text{N}$ ,  $^{31}\text{P}$ , and other isotope-labeled metabolites may be detected in addition to  $^1\text{H}$ - and  $^{13}\text{C}$ -labeled metabolites [93]. Accordingly, experimental parameters, data processing and preprocessing methods, and data analysis techniques all need to be adjusted to accommodate the specifics of each NMR experiment.
21. 2D  $^1\text{H}$ - $^{13}\text{C}$ -HSQC parameters (*hsqcetgppisp2*)—1024 data points in F2 and F1, non-uniform sampling at 25%, O1P = 4.7 ppm, O2P (offset for  $^{13}\text{C}$ ) = 75 ppm, NS = 64, DS = 16, d1 = 2, P1 = 10–13  $\mu\text{s}$  depending on salinity, CPDPRG2 = *garp* (decoupling program), PCPD2 = 55  $\mu\text{s}$  at PLW12 = 4.09 W.
22. Non-uniform sampling of 2D  $^1\text{H}$ - $^{13}\text{C}$  HSQC data can be performed on metabolomics samples albeit with some caution. From our experience we recommend no lower than 20% sparsity using the burst augmented scheduler available at <http://bionmr.unl.edu/dgs-gensched.php>. Download the sampling schedules as a text file for Topspin. Similarly, processing scripts need to be adapted for NUS datasets and a quick guide to that can be found online at <http://bionmr.unl.edu/wiki/NUS>.
23. A minimalistic approach to the processing of NMR data is optimal for a metabolomics analysis based on multivariate statistics models such as PCA and OPLS. The resulting multivariate statistical model is dependent on the choice of processing and preprocessing protocols. Baseline corrections and alignment, type of weighting (apodization) function, the type of spectral alignment or referencing, the resulting phase correction or phase correction algorithm, the number of zero-fills or the application of linear prediction, or any other data manipulation method will affect the outcome of the statistical model. Accordingly, it is best to avoid data processing steps since it is difficult to ascertain if the data processing induced a biologically irrelevant bias to the data or improved the model.
24. For NMR, relative peak intensities are averaged across all replicates per group and for each NMR peak assigned to the

metabolite. Most metabolites will have more than one peak in an NMR spectrum, and commonly all NMR peaks are incorporated into an average relative peak intensity. If a metabolite is differentially labeled due to contributions from different pathways, it is then appropriate to show single resonances for the metabolite. NMR peaks may need to be scaled by the number of attached hydrogens, since peak intensity is proportional to the number of nuclei.

25. There are a wide variety of statistical tests available to analyze the data. Some examples are the Mann-Whitney  $U$  test [94], Welch's  $t$ -test [95], Hotelling's  $t$ -squared statistic [96], and one-way analysis of variance [87], among others. The proper choice of a statistical test depends on several factors. A good introduction to what dictates the choice of statistical tests can be found in *A Biologist's Guide To Statistical Thinking And Analysis* [95].
26. In effect, the uncertainty in each pairwise comparison (as determined by the Student's  $t$ -test) is compounded with the addition of each metabolite to a set. The actual  $p$  value for a set of metabolites is defined as:

$$p = 1 - (1 - \alpha)^m \quad (3)$$

where  $m$  is the number of hypotheses (metabolites) and  $\alpha$  is typically defined as 0.05. Accordingly, a set of ten metabolites becomes an insignificant  $p = 0.401$  even though each individual metabolite is statistically significant based on a pairwise Student's  $t$ -test with a  $p < 0.05$ .

27. It is more informative to display all replicates from each group in the heatmap. Hierarchical clustering of each replicate is indicative of the relative group separation and provides further confirmation of an observed group separation from a PCA, PLS, or OPLS scores plot.
28. Typical  $R^2$  values  $> Q^2$  values, and  $Q^2$  values  $> 0.4$ . While  $p$ -values  $< 0.05$  are typically acceptable for a supervised PLS-DA or OPLS model, usually, high-quality PLS/OPLS models from metabolomics datasets yield  $p$ -values  $\ll 0.001$ .
29. NMR field strength and sample pH are important aspects of any chemical shift database matching. Thus, we strongly recommend performing a chemical shift match with the correct database (e.g., same NMR field strength and pH) while using Chenomx or similar software.
30. BMRB, HMDB, and other such databases operate in a similar fashion. Simply upload a peak list with a set of chemical shift tolerances and obtain a list of potential matches.

## Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant Number (1660921). This work was supported in part by funding from the Redox Biology Center (P30 GM103335, NIGMS) and the Nebraska Center for Integrated Biomolecular Communication (P20 GM113126, NIGMS). The research was performed in facilities renovated with the support from the National Institutes of Health (RR015468-01). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

## References

1. Hajduk PJ, Greer J (2007) A decade of fragment-based drug design: strategic advances and lessons learned. *Nat Rev Drug Discov* 6:211. <https://doi.org/10.1038/nrd2220>
2. Pellecchia M, Sem DS, Wüthrich K (2002) Nmr in drug discovery. *Nat Rev Drug Discov* 1:211. <https://doi.org/10.1038/nrd748>
3. Pellecchia M, Bertini I, Cowburn D, Dalvit C, Giralt E, Jahnke W, James TL, Homans SW, Kessler H, Luchinat C, Meyer B, Oschkinat H, Peng J, Schwalbe H, Siegal G (2008) Perspectives on NMR in drug discovery: a technique comes of age. *Nat Rev Drug Discov* 7 (9):738–745
4. Hughes JP, Rees S, Kalindjian SB, Philpott KL (2011) Principles of early drug discovery. *Br J Pharmacol* 162(6):1239–1249. <https://doi.org/10.1111/j.1476-5381.2010.01127.x>
5. Chessari G, Woodhead AJ (2009) From fragment to clinical candidate—a historical perspective. *Drug Discov Today* 14 (13):668–675. <https://doi.org/10.1016/j.drudis.2009.04.007>
6. Harner MJ, Frank AO, Fesik SW (2013) Fragment-based drug discovery using NMR spectroscopy. *J Biomol NMR* 56(2):65–75. <https://doi.org/10.1007/s10858-013-9740-z>
7. Huth JR, Mendoza R, Olejniczak ET, Johnson RW, Cothron DA, Liu Y, Lerner CG, Chen J, Hajduk PJ (2005) ALARM NMR: a rapid and robust experimental method to detect reactive false positives in biochemical screens. *J Am Chem Soc* 127(1):217–224. <https://doi.org/10.1021/ja0455547>
8. Powers R (2014) The current state of drug discovery and a potential role for NMR metabolomics. *J Med Chem* 57(14):5860–5870. <https://doi.org/10.1021/jm401803b>
9. Powers R (2009) NMR metabolomics and drug discovery. *Magn Reson Chem* 47(S1): S2–S11. <https://doi.org/10.1002/mrc.2461>
10. Robertson DG, Reilly MD, Baker JD (2007) Metabonomics in pharmaceutical discovery and development. *J Proteome Res* 6 (2):526–539. <https://doi.org/10.1021/pr060535c>
11. Wishart DS (2008) Applications of metabolomics in drug discovery and development. *Drugs R&D* 9(5):307–322. <https://doi.org/10.2165/00126839-200809050-00002>
12. Wishart DS (2016) Emerging applications of metabolomics in drug discovery and precision medicine. *Nat Rev Drug Discov* 15:473. <https://doi.org/10.1038/nrd.2016.32>
13. Zhang L, Hatzakis E, Patterson AD (2016) NMR-based metabolomics and its application in drug metabolism and cancer research. *Curr Pharmacol Rep* 2(5):231–240. <https://doi.org/10.1007/s40495-016-0067-9>
14. Wu W, Zhao S (2013) Metabolic changes in cancer: beyond the Warburg effect. *Acta Biochim Biophys Sin* 45(1):18–26. <https://doi.org/10.1093/abbs/gms104>
15. Gardner SG, Marshall DD, Daum RS, Powers R, Somerville GA (2018) Metabolic mitigation of *Staphylococcus aureus* Vancomycin intermediate-level susceptibility. *Antimicrob Agents Chemother* 62(1). <https://doi.org/10.1128/aac.01608-17>
16. Gaupp R, Lei S, Reed JM, Peisker H, Boyle-Vavra S, Bayer AS, Bischoff M, Herrmann M, Daum RS, Powers R, Somerville GA (2015) *Staphylococcus aureus* metabolic adaptations during the transition from a daptomycin susceptibility phenotype to a daptomycin nonsusceptibility phenotype. *Antimicrob Agents Chemother* 59(7):4226–4238. <https://doi.org/10.1128/AAC.00160-15>

17. Verdin E (2015) NAD<sup>+</sup> in aging, metabolism, and neurodegeneration. *Science* 350 (6265):1208
18. Van Norman GA (2016) Drugs, devices, and the FDA: part 1: an overview of approval processes for drugs. *JACC Basic Transl Sci* 1 (3):170–179. <https://doi.org/10.1016/j.jacbts.2016.03.002>
19. DiMasi JA, Grabowski HG, Hansen RW (2016) Innovation in the pharmaceutical industry: new estimates of R&D costs. *J Health Econ* 47:20–33
20. Mohs RC, Greig NH (2017) Drug discovery and development: role of basic biological research. *Alzheimers Dement (N Y)* 3 (4):651–657
21. Harrison RK (2016) Phase II and phase III failures: 2013–2015. *Nat Rev Drug Discov* 15:817. <https://doi.org/10.1038/nrd.2016.184>
22. Smietana K, Siatkowski M, Moeller M (2016) Trends in clinical success rates. *Nat Rev Drug Discov* 15(6):379–380. <https://doi.org/10.1038/nrd.2016.85>
23. Downing NS, Shah ND, Aminawung JA et al (2017) Postmarket safety events among novel therapeutics approved by the us food and drug administration between 2001 and 2010. *JAMA* 317(18):1854–1863. <https://doi.org/10.1001/jama.2017.5150>
24. Settleman J, Cohen RL (2016) Communication in drug development: “translating” scientific discovery. *Cell* 164(6):1101–1104. <https://doi.org/10.1016/j.cell.2016.02.050>
25. Scannell JW, Bosley J (2016) When quality beats quantity: decision theory, drug discovery, and the reproducibility crisis. *PLoS One* 11(2): e0147215/0147211-e0147215/0147221. <https://doi.org/10.1371/journal.pone.0147215>
26. Kamb A, Wee S, Lengauer C (2007) Why is cancer drug discovery so difficult? *Nat Rev Drug Discov* 6(2):115–120. <https://doi.org/10.1038/nrd2155>
27. Swinney DC (2004) Opinion: biochemical mechanisms of drug action: what does it take for success? *Nat Rev Drug Discov* 3 (9):801–808. <https://doi.org/10.1038/nrd1500>
28. Pink R, Hudson A, Mouries M-A, Bendig M (2005) Opportunities and challenges in Antiparasitic drug discovery. *Nat Rev Drug Discov* 4(9):727–740. <https://doi.org/10.1038/nrd1824>
29. Koul A, Arnoult E, Lounis N, Guillemont J, Andries K (2011) The challenge of new drug discovery for tuberculosis. *Nature* 469 (7331):483–490. <https://doi.org/10.1038/nature09657>
30. Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, Schacht AL (2010) How to improve R&D productivity: the pharmaceutical industry’s grand challenge. *Nat Rev Drug Discov* 9(3):203–214. <https://doi.org/10.1038/nrd3078>
31. Payne DJ, Gwynn MN, Holmes DJ, Pompliano DL (2007) Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat Rev Drug Discov* 6(1):29–40. <https://doi.org/10.1038/nrd2201>
32. Gebregiworgis T, Powers R (2012) Application of NMR metabolomics to search for human disease biomarkers. *Comb Chem High Throughput Screen* 15(8):595–610
33. Tennant DA, Durán RV, Gottlieb E (2010) Targeting metabolic transformation for cancer therapy. *Nat Rev Cancer* 10:267. <https://doi.org/10.1038/nrc2817>
34. Shukla SK, Purohit V, Mehla K, Gunda V, Chaika NV, Vernucci E, King RJ, Abrego J, Goode GD, Dasgupta A, Illies AL, Gebregiworgis T, Dai B, Augustine JJ, Murthy D, Attri KS, Mashadova O, Grandgenett PM, Powers R, Ly QP, Lazenby AJ, Grem JL, Yu F, Matés JM, Asara JM, J-w K, Hankins JH, Weekes C, Hollingsworth MA, Serkova NJ, Sasson AR, Fleming JB, Oliveto JM, Lysiotis CA, Cantley LC, Berim L, Singh PK (2017) MUC1 and HIF-1 $\alpha$  signaling crosstalk induces anabolic glucose metabolism to impart gemcitabine resistance to pancreatic cancer. *Cancer Cell* 32(1):71–87. e77. <https://doi.org/10.1016/j.ccell.2017.06.004>
35. Al-Keilani MS, Al-Sawalha NA (2017) Potential of Phenylbutyrate as adjuvant chemotherapy: an overview of cellular and molecular anticancer mechanisms. *Chem Res Toxicol* 30 (10):1767–1777. <https://doi.org/10.1021/acs.chemrestox.7b00149>
36. Choi Y-K, Park K-G (2018) Targeting glutamine metabolism for cancer treatment. *Biomol Ther (Seoul)* 26(1):19–28
37. Seo M, Crochet RB, Lee Y-H (2014) Chapter 14—targeting altered metabolism—emerging cancer therapeutic strategies. In: Neidle S (ed) *Cancer drug design and discovery*, 2nd edn. Academic Press, San Diego, pp 427–448. <https://doi.org/10.1016/B978-0-12-396521-9.00014-0>
38. Jones NP, Schulze A (2012) Targeting cancer metabolism—aiming at a tumour’s sweet-spot. *Drug Discov Today* 17(5):232–241. <https://doi.org/10.1016/j.drudis.2011.12.017>
39. Copeland RA, Harpel MR, Tummino PJ (2007) Targeting enzyme inhibitors in drug

- discovery. *Expert Opin Ther Targets* 11 (7):967–978. <https://doi.org/10.1517/14728222.11.7.967>
40. Kinch MS, Umlauf S, Plummer M (2015) An analysis of FDA-approved drugs for metabolic diseases. *Drug Discov Today* 20(6):648–651
  41. Phan LM, Yeung S-CJ, Lee M-H (2014) Cancer metabolic reprogramming: importance, main features, and potentials for precise targeted anti-cancer therapies. *Cancer Biol Med* 11(1):1–19. <https://doi.org/10.7497/j.issn.2095-3941.2014.01.001>
  42. 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 (47):15292–15299. <https://doi.org/10.1021/ja0651759>
  43. Powers R, Mercier KA, Copeland JC (2008) The application of FAST-NMR for the identification of novel drug discovery targets. *Drug Discov Today* 13(3):172–179. <https://doi.org/10.1016/j.drudis.2007.11.001>
  44. Venkitakrishnan RP, Benard O, Max M, Markley JL (2012) Use of NMR saturation transfer difference spectroscopy to study ligand binding to membrane proteins. *Methods Mol Biol* 914:47–63. [https://doi.org/10.1007/978-1-62703-023-6\\_4](https://doi.org/10.1007/978-1-62703-023-6_4)
  45. 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(3):285–295
  46. Moy FJ, Haraki K, Mobilio D, Walker G, Tabei K, Tong H, Siegel MM, Powers R (2001) MS/NMR: a structure-based approach for discovering protein ligands and for drug design by coupling size exclusion chromatography, mass spectrometry, and nuclear magnetic resonance spectroscopy. *Anal Chem* 73 (3):571–581
  47. Dalvit C, Fagerness PE, Hadden DTA, Sarver RW, Stockman BJ (2003) Fluorine-NMR experiments for high-throughput screening: theoretical aspects, practical considerations, and range of applicability. *J Am Chem Soc* 125(25):7696–7703. <https://doi.org/10.1021/ja034646d>
  48. Tiziani S, Kang Y, Choi JS, Roberts W, Pateronostro G (2011) Metabolomic high-content nuclear magnetic resonance-based drug screening of a kinase inhibitor library. *Nat Commun* 2:545. <https://doi.org/10.1038/ncomms1562>. <https://www.nature.com/articles/ncomms1562#supplementary-information>
  49. Forgue P, Halouska S, Werth M, Xu K, Harris S, Powers R (2006) NMR metabolic profiling of *Aspergillus nidulans* to monitor drug and protein activity. *J Proteome Res* 5 (8):1916–1923
  50. Halouska S, Fenton RJ, Barletta RG, Powers R (2012) Predicting the in vivo mechanism of action for drug leads using NMR metabolomics. *ACS Chem Biol* 7(1):166–171. <https://doi.org/10.1021/cb200348m>
  51. Weljie AM, Newton J, Mercier P, Carlson E, Slupsky CM (2006) Targeted profiling: quantitative analysis of 1H NMR metabolomics data. *Anal Chem* 78(13):4430–4442. <https://doi.org/10.1021/ac060209g>
  52. Lindon JC, Keun HC, Ebbels TMD, Pearce JMT, Holmes E, Nicholson JK (2005) The consortium for Metabonomic toxicology (COMET): aims, activities and achievements. *Pharmacogenomics* 6(7):691–699. <https://doi.org/10.2217/14622416.6.7.691>
  53. Torell F, Bennett K, Rännar S, Lundstedt-Enkel K, Lundstedt T, Trygg J (2017) The effects of thawing on the plasma metabolome: evaluating differences between thawed plasma and multi-organ samples. *Metabolomics* 13 (6):66. <https://doi.org/10.1007/s11306-017-1196-9>
  54. Catala A, Culp-Hill R, Nemkov T, D’Alessandro A (2018) Quantitative metabolomics comparison of traditional blood draws and TAP capillary blood collection. *Metabolomics* 14 (7):100. <https://doi.org/10.1007/s11306-018-1395-z>
  55. Worley B, Powers R (2013) Multivariate analysis in metabolomics. *Curr Metabolomics* 1 (1):92–107. <https://doi.org/10.2174/2213235X11301010092>
  56. Worley B, Powers R (2016) PCA as a practical indicator of OPLS-DA model reliability. *Curr Metabolomics* 4(2):97–103. <https://doi.org/10.2174/2213235X04666160613122429>
  57. Worley B, Powers R (2014) MVAPACK: a complete data handling package for NMR metabolomics. *ACS Chem Biol* 9 (5):1138–1144. <https://doi.org/10.1021/cb4008937>
  58. Worley B, Halouska S, Powers R (2013) Utilities for Quantifying Separation in PCA/PLS-DA scores plots. *Anal Biochem* 433 (2):102–104. <https://doi.org/10.1016/j.ab.2012.10.011>
  59. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6(3):277–293
  60. Johnson BA (2004) Using NMRView to visualize and analyze the NMR spectra of

- macromolecules. In: Downing AK (ed) *Protein NMR Techniques*. Humana Press, Totowa, NJ, pp 313–352. <https://doi.org/10.1385/1-59259-809-9:313>
61. Xia J, Sinelnikov IV, Han B, Wishart DS (2015) MetaboAnalyst 3.0-making metabolomics more meaningful. *Nucleic Acids Res* 43(W1):W251–W257. <https://doi.org/10.1093/nar/gkv380>
  62. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, Djoumbou Y, Mandal R, Aziat F, Dong E, Bouatra S, Sinelnikov I, Arndt D, Xia J, Liu P, Yallou F, Bjorn Dahl T, Perez-Pineiro R, Eisner R, Allen F, Neveu V, Greiner R, Scalbert A (2013) HMDB 3.0-the human Metabolome database in 2013. *Nucleic Acids Res* 41(D1):D801–D807. <https://doi.org/10.1093/nar/gks1065>
  63. Markley JL, Anderson ME, Cui Q, Eghbalnia HR, Lewis IA, Hegeman AD, Li J, Schulte CF, Sussman MR, Westler WM, Ulrich EL, Zolnai Z (2007) New bioinformatics resources for metabolomics. World Scientific Publishing Co. Pte. Ltd, Singapore, pp 157–168. [https://doi.org/10.1142/9789812772435\\_0016](https://doi.org/10.1142/9789812772435_0016)
  64. Worley B, Powers R (2015) Deterministic multidimensional nonuniform gap sampling. *J Magn Reson* 261:19–26. <https://doi.org/10.1016/j.jmr.2015.09.016>
  65. Jewison T, Su Y, Disfany FM, Liang Y, Knox C, Maciejewski A, Poelzer J, Huynh J, Zhou Y, Arndt D, Djoumbou Y, Liu Y, Deng L, Guo AC, Han B, Pon A, Wilson M, Rafatnia S, Liu P, Wishart DS (2014) SMPDB 2.0: big improvements to the small molecule pathway database. *Nucleic Acids Res* 42(D1):D478–D484. <https://doi.org/10.1093/nar/gkt1067>
  66. Blaise BJ, Correia G, Tin A, Young JH, Vergnaud A-C, Lewis M, Pearce JTM, Elliott P, Nicholson JK, Holmes E, Ebbels TMD (2016) Power analysis and sample size determination in metabolic Phenotyping. *Anal Chem* 88(10):5179–5188. <https://doi.org/10.1021/acs.analchem.6b00188>
  67. Nyamundanda G, Gormley IC, Fan Y, Gallagher WM, Brennan L (2013) MetSizeR: selecting the optimal sample size for metabolomic studies using an analysis based approach. *BMC Bioinformatics* 14:338
  68. Bando K, Kawahara R, Kunimatsu T, Sakai J, Kimura J, Funabashi H, Seki T, Bamba T, Fukusaki E (2010) Influences of biofluid sample collection and handling procedures on GC–MS based metabolomic studies. *J Biosci Bioeng* 110(4):491–499. <https://doi.org/10.1016/j.jbiosc.2010.04.010>
  69. Emwas A-HM, Salek RM, Griffin JL, Merzaban J (2013) NMR-based metabolomics in human disease diagnosis: applications, limitations, and recommendations. *Metabolomics* 9(5):1048–1072. <https://doi.org/10.1007/s11306-013-0524-y>
  70. Lipinski CA (2001) Drug-like properties and the causes of poor solubility and poor permeability. *J Pharmacol Toxicol Methods* 44(1):235–249. [https://doi.org/10.1016/S1056-8719\(00\)00107-6](https://doi.org/10.1016/S1056-8719(00)00107-6)
  71. Li AP (2001) Screening for human ADME/Tox drug properties in drug discovery. *Drug Discov Today* 6(7):357–366. [https://doi.org/10.1016/S1359-6446\(01\)01712-3](https://doi.org/10.1016/S1359-6446(01)01712-3)
  72. Heller AA, Lockwood SY, Janes TM, Spence DM (2018) Technologies for measuring pharmacokinetic profiles. *Annu Rev Anal Chem* 11:79–100. <https://doi.org/10.1146/annurev-anchem-061417-125611>
  73. Reigner BG, Blesch KS (2002) Estimating the starting dose for entry into humans: principles and practice. *Eur J Clin Pharmacol* 57(12):835–845
  74. Nair AB, Jacob S (2016) A simple practice guide for dose conversion between animals and human. *J Basic Clin Pharm* 7(2):27–31
  75. O’Gorman A, Gibbons H, Brennan L (2013) Metabolomics in the identification of biomarkers of dietary intake. *Comput Struct Biotechnol J* 4:e201301004
  76. Bouatra S, Aziat F, Mandal R, Guo AC, Wilson MR, Knox C, Bjorn Dahl TC, Krishnamurthy R, Saleem F, Liu P, Dame ZT, Poelzer J, Huynh J, Yallou FS, Psychogios N, Dong E, Bogumil R, Roehring C, Wishart DS (2013) The human urine Metabolome. *PLoS One* 8(9):e73076. <https://doi.org/10.1371/journal.pone.0073076>
  77. Emwas A-H, Luchinat C, Turano P, Tenori L, Roy R, Salek RM, Ryan D, Merzaban JS, Kaddurah-Daouk R, Zeri AC, Nagana Gowda GA, Raftery D, Wang Y, Brennan L, Wishart DS (2015) Standardizing the experimental conditions for using urine in NMR-based metabolomic studies with a particular focus on diagnostic studies: a review. *Metabolomics* 11(4):872–894. <https://doi.org/10.1007/s11306-014-0746-7>
  78. Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, Sinelnikov I, Krishnamurthy R, Eisner R, Gautam B, Young N, Xia J, Knox C, Dong E, Huang P, Hollander Z, Pedersen TL, Smith SR, Bamforth F, Greiner R, McManus B, Newman JW, Goodfriend T, Wishart DS (2011) The human serum metabolome. *PLoS One* 6(2):e16957. <https://doi.org/10.1371/journal.pone.0016957>

79. Geraghty RJ, Capes-Davis A, Davis JM, Downard J, Freshney RI, Knezevic I, Lovell-Badge R, Masters JRW, Meredith J, Stacey GN, Thraves P, Vias M (2014) Guidelines for the use of cell lines in biomedical research. *Br J Cancer* 111:1021. <https://doi.org/10.1038/bjc.2014.166>
80. Halouska S, Zhang B, Gaupp R, Lei S, Snell E, Fenton RJ, Barletta RG, Somerville GA, Powers R (2013) Revisiting protocols for the NMR analysis of bacterial metabolomes. *J Integr OMICS* 3(2):120–137. <https://doi.org/10.5584/jiomics.v3i2.139>
81. Westerhoff HV, Chen Y-D (1984) How do enzyme activities control metabolite concentrations? *Eur J Biochem* 142(2):425–430. <https://doi.org/10.1111/j.1432-1033.1984.tb08304.x>
82. Worley B, Powers R (2014) Simultaneous phase and scatter correction for NMR datasets. *Chemom Intell Lab Syst* 131:1–6. <https://doi.org/10.1016/j.chemolab.2013.11.005>
83. Savorani F, Tomasi G, Engelsen SB (2010) Icoshift: a versatile tool for the rapid alignment of 1D NMR spectra. *J Magn Reson* 202(2):190–202. <https://doi.org/10.1016/j.jmr.2009.11.012>
84. De Meyer T, Sinnaeve D, Van Gasse B, Tshiporkova E, Rietzschel ER, De Buyzere ML, Gillebert TC, Bekaert S, Martins JC, Van Criekinge W (2008) NMR-based characterization of metabolic alterations in hypertension using an adaptive, intelligent binning algorithm. *Anal Chem* 80(10):3783–3790. <https://doi.org/10.1021/ac7025964>
85. Development Core Team R (2011) R: a language and environment for statistical computing, vol 1
86. Eriksson L, Trygg J, Wold S (2008) CV-ANOVA for significance testing of PLS and OPLS<sup>®</sup> models. *J Chemom* 22(11–12):594–600. <https://doi.org/10.1002/cem.1187>
87. Triba MN, Le Moyec L, Amathieu R, Goossens C, Bouchemal N, Nahon P, Rutledge DN, Savarin P (2015) PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters. *Mol BioSyst* 11(1):13–19. <https://doi.org/10.1039/C4MB00414K>
88. Goodacre R, Broadhurst D, Smilde AK, Kristal BS, Baker JD, Beger R, Bessant C, Connor S, Capuani G, Craig A, Ebbels T, Kell DB, Manetti C, Newton J, Paternostro G, Somorjai R, Sjöström M, Trygg J, Wulfert F (2007) Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics* 3(3):231–241. <https://doi.org/10.1007/s11306-007-0081-3>
89. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B Methodol* 57(1):289–300
90. Bland JM, Altman DG (1995) Multiple significance tests: the Bonferroni method. *BMJ* 310(6973):170. <https://doi.org/10.1136/bmj.310.6973.170>
91. Xu Q-S, Liang Y-Z (2001) Monte Carlo cross validation. *Chemom Intell Lab Syst* 56(1):1–11. [https://doi.org/10.1016/S0169-7439\(00\)00122-2](https://doi.org/10.1016/S0169-7439(00)00122-2)
92. Rubinson KA (2017) Practical corrections for p (H,D) measurements in mixed H<sub>2</sub>O/D<sub>2</sub>O biological buffers. *Anal Methods* 9(18):2744–2750. <https://doi.org/10.1039/C7AY00669A>
93. Bhinderwala F, Lonergan S, Woods J, Zhou C, Fey PD, Powers R (2018) Expanding the coverage of the Metabolome with nitrogen-based NMR. *Anal Chem* 90(7):4521–4528. <https://doi.org/10.1021/acs.analchem.7b04922>
94. Ruxton GD (2006) The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test. *Behav Ecol* 17(4):688–690. <https://doi.org/10.1093/beheco/ark016>
95. Fay MP, Proschan MA (2010) Wilcoxon–Mann–Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. *Statist Surv* 4:1–39. <https://doi.org/10.1214/09-SS051>
96. Hotelling H (1931) The economics of exhaustible resources. *J Polit Econ* 39(2):137–175