

# Fractionation-assisted NMR-based metabolomics of *Grindelia squarrosa*

By Jason McFarlane

May 26, 2017

## Abstract

To identify compounds, a nuclear magnetic resonance spectroscopy (NMR)-based metabolomics method was used on an extract of *Grindelia squarrosa*. The extract was separated based on polarity, one of the parameters used to evaluate the potential for a new bioactive compound to be suitable for oral administration. The seven most polar fractions—out of eleven total fractions—were analyzed by NMR spectroscopy using an HSQC pulse sequence. The resulting peaks were queried against the Biological Magnetic Resonance Databank. The database returned 597 compounds, which were selected to include only those compounds that had greater than 50% of the peaks matching. After selection, 88 compounds were reported, including three bioactive compounds: agmatine, 4-guanidinobutyric acid, and dopamine.

## Introduction

A large part of modern drug development is based on compounds isolated from natural sources such as plants and marine animals. *Grindelia squarrosa* is a plant that has traditionally been used as an extract of the flowers and leaves to remove mucus from bronchitis patients,<sup>1</sup> and has been shown to have some antimicrobial activity.<sup>2</sup> *G. squarrosa* has a resinous coat to the flower, presumably as a form of protection from pests; other compounds that have a role in protecting plants against pests have been shown to have drug activity, such as morphine.<sup>3</sup> These traits make *G. squarrosa* a potential candidate for novel drug compounds.

Unfortunately, traditional methods of assaying an extract against a specific target—be it a cancer cell line, receptor or microorganism—and separating the components and “chasing” the active fraction until the active constituent is isolated has several disadvantages. First of all, the process is quite inefficient (as many compounds of interest are naturally present only in very small concentration), requiring multiple assays and very large amounts of sample and solvent to extract useable concentrations.<sup>4</sup> Additionally, testing an extract of unknown composition against a complex biological target results in a situation that relies more on chance than scientific insight.<sup>5</sup> Finally, as the identity of the active constituent is only identified after the isolation is done, there is a large rate of rediscovering known compounds.

One method that is gaining popularity in natural product research is metabolomics. Metabolomics is the study of the secondary metabolites produced by an organism.<sup>6</sup> This approach is advantageous in drug research as it allows the researcher to quickly discard compounds that have been previously identified and prioritize those compounds that have novel structures.<sup>5</sup> Metabolomics also allows a library of compounds to be compiled for testing against multiple targets. Two main techniques are used for metabolomics. Liquid Chromatography-Mass Spectrometry (LC-MS) is the most common, due to high sensitivity and ease of automation. However, Nuclear Magnetic Resonance spectroscopy (NMR) is also used; the advantages of NMR are the wealth of structural information gleaned as well as the ability to detect any organic compound. In brief, NMR detects the frequency of spinning nuclei in a magnetic field. That frequency is dependent on the strength of the magnetic field, which in turn is dependent on the strength of the NMR instrument, the type of nuclei (isotope and element), and the electrical shielding produced by the electrons surrounding the nucleus. This frequency produces a diagnostic spectrum of peaks for a particular molecule. For organic molecules, hydrogens produce a strong signal and are present in all organic molecules. The other nuclei that can be detected by NMR that is ubiquitous in organic molecules is carbon-13; however, NMR spectra based on carbon-13 have low sensitivity due to the low natural abundance of carbon-13, typically about 1%.

One of the most influential works in the natural products field is Lipinski's Rule of Five. Lipinski's Rule of Five, simply put, is a set of guidelines that predict favorable pharmacokinetic properties of a compound that has potential as an oral drug.<sup>7</sup> It was developed in order to offset the number of compounds discovered through high-throughput screening that, though they show

intrinsic activity, are unable to be absorbed by the body. Lipinski's Rule of Five states that if a drug has more than 5 hydrogen bond donors, a molecular weight greater than 500 amu, a logP greater than 5 (where P is the 1-octanol-water partition coefficient), or more than 10 hydrogen bond acceptors, it will most likely exhibit poor pharmacokinetic parameters. This allows potential drug compounds to be eliminated without expensive in vivo testing. Additionally, as logP is a property that can be determined independent of structure, it can be used as a filter to increase the presence of potential drug compounds in an extract in a metabolomics driven approach to natural product discovery.<sup>8</sup>

One application of Lipinski's Rule of Five is to use reverse phase column chromatography to separate compounds based on logP. Camp et al, 2012, tested various solid phase extraction cartridges to optimize the retention of these components.<sup>8</sup> They verified the resulting extract by high pressure liquid chromatography (HPLC) which correlates logP and retention time. Building on this method, Grkovic et al, 2014, used solid phase extraction to compile NMR fingerprints of fractionated natural product extracts.<sup>9</sup> They then identified unique fractions from the 1 dimensional NMR spectra of all of the 220 fractions; when tested, this led to the discovery of a new agent to explain the cellular mechanism of Parkinson's disease. Both solid phase extraction and column chromatography rely on the same principles—polar compounds are more attracted to the polar mobile phase and non-polar compound are attracted to the non-polar stationary phase leading to complete retention in the case of solid phase extraction and separation in the case of column chromatography—and indeed, non-polar solid phase extraction and reverse phase column chromatography use the same packing material (C-18 bound to a silica gel support).

Using NMR-based metabolomics, the chemical shifts of the different molecules are used to query the Biological Magnetic resonance databank (BMRB).<sup>10</sup> Unfortunately, due to the large number of compounds, too many peaks overlap in a 1-dimensional NMR spectrum, so 2-dimensional spectra will be obtained to separate this overlap by resolving peaks produced by hydrogens based on their carbon neighbors.

## Methods

The entire above ground portion of *G. squarrosa* was collected from the northern exposure of Kenna Cartwright Park in Kamloops, British Columbia. The plant matter was air-dried over a three-week period, then 0.75 g of leafy matter and 2.97 g flower buds (including both flowers in bloom and flowers that had not yet opened) were collected for a total mass of 3.72 g plant matter. This plant matter was ground, then transferred to an Erlenmeyer Flask. 100 mL of ACS grade hexanes were used to rinse the mortar and pestle used into the Erlenmeyer Flask, after which the flask was sealed and left to sit for 24 hours.

The contents of the Erlenmeyer were filtered through coarse filter paper and the hexanes extract discarded. The fibrous plant matter was returned to the Erlenmeyer, and 100 mL of ACS grade acetone was added. The flask was sealed and the plant material extracted for 24 hours. The acetone extract was filtered into a 250-mL round bottomed flask (RBF) and evaporated to dryness by rotary evaporation (rotovap), producing a deep green oil. The plant material was extracted again with 100 mL of HPLC grade methanol for 24 hours. This final extract was filtered into a different 250-mL RBF and evaporated to dryness, producing a light green-yellow oil.

The dried acetone extract was resuspended in four 5 mL portions of methanol, which were transferred to the RBF containing the dried methanol extract before being transferred to a vial. This vial was stored at 4°C for one week and then dried by nitrogen blow down, producing a viscous, dark green oil. This oil was resuspended in 10 mL of methanol by ultra-sonication. The suspension was transferred to a 50-mL RBF. 2 mL of preparative C-18 125 Å 55-105 µm Waters silica solid phase was added to the RBF and the mixture was rotovapped to produce a thick green paste, adhering the metabolites in the extract to the surface of the packing material. This paste was transferred to the surface of a 3 cm x 10 cm C-18 column.

The metabolites were eluted in with a stepwise eluent gradient from 35:65 methanol : water to 100% methanol as outlined in recent research by Camp et al.<sup>8</sup> Though the research by Camp et al was done using a High Pressure Liquid Chromatography column, the eluent composition over the 20 minute run was mimicked in the steps of the gradients used in this experiment. 100 mL each of 35, 45, 55, 65, 70, 75, 80 and 90% methanol—with the remainder being made up of deionized water— was used to elute the compounds from the column. Then 800 mL of 100% methanol was passed through the column until the dark green band—believed to be chlorophyll—was eluted.

71-15 mL aliquots were collected. These aliquots were pooled according to Table 1, giving 11 fractions that were evaporated to dryness by rotovap. The fractions with the lowest logP (fractions 1-7) were resuspended in 0.5 mL methanol-d<sub>4</sub> spiked with tetramethylsilane as an internal standard and centrifuged to remove any insoluble components. Each of the selected samples were analyzed by NMR spectroscopy, using an HSQC pulse sequence (Number of Scans=64, Time Domain Data points=2048 hydrogen x 512 carbon-13, hydrogen offset 4.7=ppm, carbon-13 offset=85.0 ppm, Spectral Width=10.9920 ppm hydrogen x 172.9149 ppm

carbon-13). HSQC probes interactions between neighboring hydrogens and carbons on the same molecule and produces a relatively simple 2-dimensional plot, with the intensity of the peaks projecting into the 3<sup>rd</sup> dimension. A list of peaks was extracted from the processed spectra using the peak picking program integrated into Topspin v2.4. The list of peaks was queried against the BMRB, which returned a list of compounds. The list of compounds from each fraction was combined and filtered to remove compounds that had less than 50% of the expected peaks present in the experimental spectra. As well, any duplicates were removed.

*Table 1. The manner in which the aliquots were pooled to produce the fractions later analyzed by NMR spectroscopy.*

<b>Fraction number</b>	<b>Subfraction range</b>	<b>Mass (g)</b>
<b>1</b>	1-8	0.2227
<b>2</b>	9-14	0.0138
<b>3</b>	15-20	0.0144
<b>4</b>	21-28	0.0552
<b>5</b>	29-34	0.0233
<b>6</b>	35-45	0.0414
<b>7</b>	46-50	0.0281
<b>8</b>	51-55	0.0222
<b>9</b>	56-60	0.0106
<b>10</b>	61-65	0.0049
<b>11</b>	66-71	0.0099

## Results and Discussion

Querying the database produced 597 unique compounds which were filtered to 88 compounds identified with relative certainty, reported in Table 2. This is greater than the number of compounds returned by the database before fractionation with 477 compounds, showing that fractionation allowed for compounds that may have been obscured by other, more concentrated, compounds in the NMR spectra to be identified. In addition, because not all of the fractions were analyzed, this increase in compounds between fractionated and crude extracts is probably much more significant. Compounds of note for their biological activity are: agmatine (antidiabetic and neuropathic pain reduction<sup>11</sup>), 4-Guanidinobutyric acid (gastric lesions inhibitor<sup>12</sup>), and dopamine (neurotransmitter).

*Table 2. The compounds returned from the BMRB database. The peak match is the percentage of expected peaks from each compound—from the spectrum of the pure compound—that were actually present in the experimental spectra of the mixture. Only those compounds that had greater than 50% of their peaks present in the NMR spectrum are reported.*

Compound	Peak Match
3-Ureidopropionic acid	1
Acetyl phosphate	1
Adipic acid	1
Citrate	1
cyclohexane	1
Methanol	1
Methylmalonic acid	1
Pyruvic acid	1
scyllo-Inositol	1
sulfoacetic acid	1
thioacetamide	1
Glutaric acid	0.88
Xylitol	0.86
Agmatine	0.8
D-(+)-Threo-isocitric acid	0.8
4-Methyl-2-oxovaleric acid	0.78
O-Phospho-L-serine	0.78
D-Saccharate	0.75
propanol	0.75
4-methylvaleric acid	0.73
Pimelic acid	0.73
isovaleric-acid	0.71
suberic acid	0.69
2-ketobutyric acid	0.67
3-Methyl-2-oxobutinoic acid	0.67

D-Citrulline	0.67
DL-alpha-Glycerol phosphate	0.67
L-Dihydroorotic acid	0.67
L-Homoserine	0.67
L-Serine	0.67
meso-Erythritol	0.67
O,O-diethyl thiophosphate	0.67
16-hydroxyhexadecanoic acid	0.62
N(alpha)-Acetyl-DL-ornithine	0.62
(R)-2-hydroxybutyric acid	0.61
1-Hexadecanol	0.61
DL-Serine	0.6
i-Erythritol	0.6
Isethionic acid	0.6
L-Ascorbate	0.6
D-Glucuronate	0.58
1,3-Propanediol	0.57
2-Aminoethyl dihydrogen phosphate	0.57
L-Citrulline	0.57
L-Isoleucine	0.57
Stearic Acid	0.57
Valeric acid	0.56
L(+)-Selenomethionine	0.55
1-Octanol	0.54
DL-2-Aminobutyric acid	0.54
N-Acetyl-D-glucosamine 1-phosphate	0.53
Nepsilon-Acetyl-L-lysine	0.53
L-Arginine L-Glutamate	0.52
(R)-Lactate	0.5
1,3-Diaminopropane	0.5
2-chloroethanol	0.5
2-Isopropylmalic acid	0.5
2,2-dimethylsuccinic acid	0.5
3-butyn-1-ol	0.5
3,5-dichlorocatechol	0.5
4-Chlorophenol	0.5
4-Guanidinobutyric acid	0.5
4-Hydroxy-benzoic acid	0.5
Ala-Ala	0.5
Alanine	0.5
alpha-Ketoglutaric acid	0.5



Betaine	0.5
Citraconic acid	0.5
Creatine	0.5
Creatine phosphate	0.5
Dihydroxyacetone	0.5
DL-threo-beta-Methylaspartate	0.5
dopamine	0.5
epichlorohydrin	0.5
Ethanolamine	0.5
Glycogen	0.5
Glycolaldehyde	0.5
Guaiacol	0.5
Isobutyric acid	0.5
L-(-) Arabitol	0.5
L-Glutamic acid	0.5
L-Glutamine	0.5
n-acetylglycine	0.5
N,N-Dimethylglycine	0.5
p-fluorobenzoic acid	0.5
Propionic acid	0.5
Succinic acid	0.5
syringic acid	0.5

Two large issues are inherent in this technique. One is that the BMRB provides no statistical measurement of the validity of peak assignment. This means that beyond instituting arbitrary peak match cut offs—as I did at 0.5—there is no way to filter the results for accuracy beyond physically examining each spectrum individually, which for the total 597 unique compounds returned by the database would be prohibitively time consuming. The peak match itself is inadequate as measure of assignment accuracy. The peak match would be skewed towards compound with simple substructures. For example, citrate—shown in Figure 1—has only two carbon-hydrogen pairs to produce a signal in the HSQC spectrum. In addition, due to symmetry, the signal those two pairs produce are at an identical ppm (the units of the spectrum axes). In order for the BMRB to return a peak match of 1, all it need to detect is the single signal produced by citrate. This issue is present in many of the compounds shown in Table 2

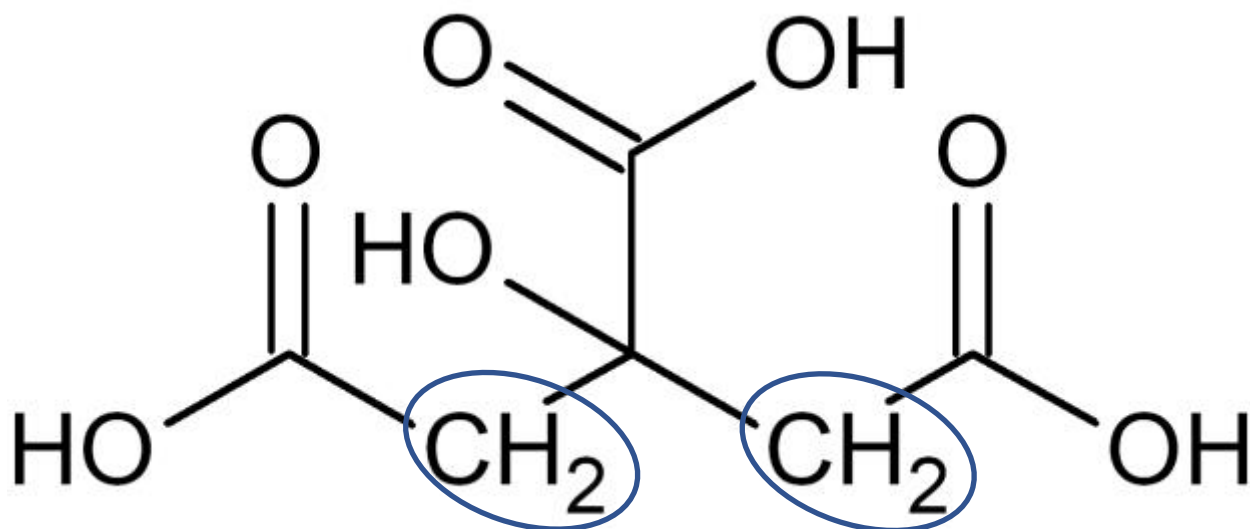


Figure 1. The structure of citrate. The circled atoms are those that produce a signal in the HSQC spectra.

The second issue that is grounded in the very nature of databases is that no new compounds can be discovered through simply querying a database. This is because the database is populated with known compounds. This problem is compounded by small database sizes, as NMR-based metabolomics is an emerging field. Querying a database that is too small not only cannot return unknown compounds, but known compounds remain unidentified as well.

## Conclusions and Future Work

477 compounds were identified in the crude *G. squarrosa* extract and 597 compounds were identified in the fractions of greatest polarity, 1-7. Filtering these compounds by 50% peak match resulted in 88 compounds, three of which are known to have biological activity.

In order to address the issues with NMR-based metabolomics, reconstructed spectra should be made from all the identified compounds. This is done by overlaying all of the individual compounds' HSQC spectra on one plot, and comparing this plot to the experimental spectra. Matlab is a matrix manipulation software that can do this sort of reconstruction and overlaying of compounds.<sup>13</sup> The overlay allows for visual confirmation of the BMRB's peak assignment. Peaks that are present in the reconstruction, but not the original spectra suggest that the compounds responsible for those peaks were assigned by the database in error, and should be excluded from the list of peaks. Peaks that appear in the experimental spectra but not in the reconstructed spectra are due to novel compounds.

Further investigation should be done on those unassigned peaks. Different NMR pulse programs probe different interactions in molecules. For example, a COSY spectrum looks at hydrogens that are separated by up to three bonds. This allows for the substructure of the molecule to be parsed by "walking" along the neighboring hydrogens. COSY spectra are more complex than HSQC, and fewer databases are configured in such a way as to allow COSY peaks to be directly inputted, but identifying an unknown peak in the HSQC spectrum and using the COSY spectra to see which other peaks belong to the same molecule allows a single molecule to be queried against a database as if it were a pure compound.

## Literature Cited

1. Stuhr, E. T. The distribution, abundance and uses of wild drug plants in Oregon and Southern California. *Econ. Bot.* **1**, 57–68 (1947).
2. Hassan, H. M., Jiang, Z.-H., Asmussen, C., McDonald, E. & Qin, W. Antibacterial activity of northern Ontario medicinal plant extracts. *Can. J. Plant Sci.* **94**, 417–424 (2013).
3. Morimoto, S. *et al.* Morphine metabolism in the opium poppy and its possible physiological function: biochemical characterization of the morphine metabolite, bismorphine. *J. Biol. Chem.* (2001). doi:10.1074/jbc.M107105200
4. Bendaikha, S., Gadaut, M., Harakat, D. & Magid, A. Acylated flavonol glycosides from the flower of *Elaeagnus angustifolia* L. *Phytochemistry* **103**, 129–136 (2014).
5. Kurita, K. L. & Linington, R. G. Connecting phenotype and chemotype: high-content discovery strategies for natural products research. *J. Nat. Prod.* **78**, 587–596 (2015).
6. Dona, A. C. *et al.* A guide to the identification of metabolites in NMR-based metabonomics/metabolomics experiments. *Comput. Struct. Biotechnol. J.* **14**, 135–153 (2016).
7. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **23**, 3–25 (1997).
8. Camp, D., Davis, R. A., Campitelli, M., Ebdon, J. & Quinn, R. J. Drug-like Properties: Guiding Principles for the Design of Natural Product Libraries. *J. Nat. Prod.* **75**, 72–81 (2012).
9. Grkovic, T. *et al.* NMR Fingerprints of the Drug-like Natural-Product Space Identify Iotrochotazine A: A Chemical Probe to Study Parkinson's Disease. *Angew. Chem. Int. Ed.* **53**, 6070–6074 (2014).
10. Ulrich, E. L. *et al.* BioMagResBank. *Nucleic Acids Res.* **36**, D402–408 (2008).
11. Piletz, J. E. *et al.* Agmatine: clinical applications after 100 years in translation. *Drug Discov. Today* **18**, 880–893 (2013).
12. Hwang, I. Y. & Jeong, C. S. Inhibitory Effects of 4-Guanidinobutyric Acid against Gastric Lesions. *Biomol. Ther.* **20**, 239–244 (2012).
13. Bingol, K., Bruschweiler-Li, L., Li, D.-W. & Bruschweiler, R. Customized Metabolomics Database for the Analysis of NMR 1H–1H TOCSY and 13C–1H HSQC-TOCSY Spectra of Complex Mixtures. *Anal. Chem.* **86**, 5494–5501 (2014).