Technical Note

Chemical Proportionality within Molecular Networks

Daniel Petras[1,2,#,*], Andrés Mauricio Caraballo-Rodríguez[1,2,#], Alan K. Jarmusch[1,2,3], Carlos Molina-Santiago[4], Julia M. Gauglitz[1,2], Emily C. Gentry[1,2], Pedro Belda-Ferre[5], Diego Romero[4], Shirley M. Tsunoda[1], Pieter C. Dorrestein[1,2], Mingxun Wang[1,2*]

[1] Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, San Diego, CA, USA

[2]Collaborative Mass Spectrometry Innovation Center, University of California San Diego, La Jolla, San Diego, CA, USA

[3]Immunity, Inflammation, and Disease Laboratory, National Institute of Environmental Health Sciences (NIEHS), Research Triangle Park, NC, USA

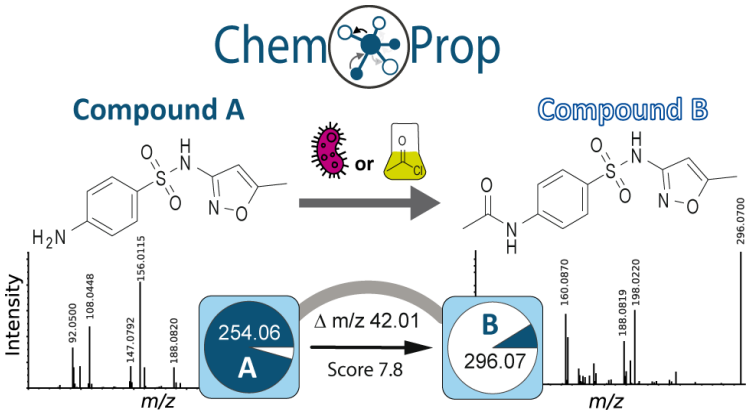
[4]Instituto de Hortofruticultura Subtropical y Mediterránea “La Mayora”, Universidad de Málaga-Consejo Superior de Investigaciones Científicas (IHSMUMA-CSIC), Departamento de Microbiología, Universidad de Málaga, Bulevar Louis Pasteur 31 (Campus Universitario de Teatinos), 29071 Málaga, Spain.

[5]Department of Pediatrics, University of California San Diego, La Jolla, San Diego, CA, USA

[#]These authors contributed equally

*Correspondence should be addressed to Mingxun Wang (miw023@ucsd.edu) for questions regarding software development and infrastructure and to Daniel Petras (functionalmetabolomics@gmail.com) for questions regarding concept and experimental validation.

Abstract: Molecular networking of non-targeted tandem mass spectrometry data connects structurally related molecules based on similar fragmentation spectra. Here we report the Chemical Proportionality (ChemProp) contextualization of molecular networks. ChemProp scores the changes of abundance between two connected nodes over sequential data series (*e.g.* temporal or spatial relationships) which can be displayed as a direction within the network to prioritize potential biological and chemical transformations or proportional changes of (biosynthetically) related compounds. We tested the ChemProp workflow on a ground truth data set of defined mixture and highlighted the utility of the tool to prioritize specific molecules within biological samples, including bacterial transformations of bile acids, human drug metabolism and bacterial natural products biosynthesis. The ChemProp workflow is freely available through the Global Natural Products Social Molecular Networking (GNPS) environment.



Key Words: Metabolomics; Non-targeted Metabolomics; Tandem-Mass Spectrometry; Molecular Networking; Biochemical Transformation; Data analysis

Introduction

To understand the metabolism of a given biological system, the identification of metabolites and their dynamical changes through (bio)chemical transformation is fundamentally important. Many metabolomics studies, that make use of non-targeted tandem mass spectrometry (MS/MS), are performed in a longitudinal or spatial fashion.[1,2] From such data, one can hypothesize the extent of putative (bio)chemical modifications by correlating changes in peak area that are associated with temporal or spatial patterns.

There are numerous challenges in the interpretation of non-targeted mass spectrometry data. Two fundamental challenges are the annotation of MS/MS spectra and providing meaningful interpretation of the biological role of the numerous compounds detected. Molecular networking in the GNPS web platform (gnps.ucsd.edu) aims to tackle the former challenge in annotation by connecting similar compounds based on their MS/MS spectra, which reflects similarities in chemical structure.[3–5] By doing so, molecular networking provides a framework of chemical-structural similarity in non-targeted MS/MS data upon which additional information can be displayed such as relative metabolite abundance.[6] In addressing the latter challenge of identifying potential (metabolic) transformations, several approaches have been described. A paired mass distance (PMD) approach was developed to link biochemical reactions available in databases,[7] such as KEGG,[8] through prediction of chemical transformations based on mass differences. Meta-mass shift analysis is focused on all the mass differences within molecular networks,[9] irrespective of whether a known metabolite has been mapped onto biochemical pathways. The use of commonly observed delta masses for modification searches is also used in proteomic studies, where the mass difference between two observed peptides arise via genetic changes, or via post-translational and chemical modifications.[10]

We developed a chemical proportionality (ChemProp) approach, integrated with feature-based molecular networking[3] to address the challenge in identifying related metabolites in non-targeted MS/MS data. ChemProp aims to find two or more structurally related molecules that have a proportional relationship to each other between sequential data series (*e.g.* time or space). For example, a (bio)chemical reaction that causes the mass difference (*m/z*) of 14.016 could result both from a methylation or a demethylation reaction, but current methodology does not highlight or in any way indicate that these changes are associated to spatial or temporal data. ChemProp scores the peak area changes of connected nodes in a molecular network across a sequential data frame by comparing their proportions. The ChemProp scoring can be used to guide the formulation of hypotheses regarding the direction of the change, that can be indicated directly within a molecular network or used on a dataset level to explore pattern changes between all connected compounds (**Figure 1**).

In this technical note, we present illustrative examples of insights gained via ChemProp in the case of a defined mixture (*i.e.* ground truth dataset) that illustrates an acetylation reaction of sulfamethoxazole, biological datasets of bacterial transformations of bile acids, human drug metabolism of omeprazole and proportional changes in the biosynthesis of bacterial natural products.

Experimental Concept

The concept of ChemProp is illustrated in Figure 1a and b. In order to establish a proportion based directionality of potential transformations or pattern change of structurally related compounds, ChemProp relies on the following assumptions. The first is that a given reactant and product or two otherwise structurally related compounds are connected in a molecular network through their MS/MS (chemical) similarity. And second, that the abundance of the initial compound would decrease over time/space and the abundance of the new compound would increase.

To obtain this information, ChemProp makes use of Feature-based Molecular Networking (FBMN)[6] and the peak areas of a given feature pair (connected in the networking, e.g. Compound A and B in Figure 1a) across a sequential data series. The ChemProp score is derived via the log-ratio of the proportional value of feature pairs at one time point vs. the proportional value at a sequential time point.

In the hypothetical example in **Figure 1a**, this would correspond to the log-ratio of A_1/B_1 by A_2/B_2. In practice, samples 1 and 2 can be time points in a longitudinal study, but also data points in a spatial study, or other experimental designs such as two treatment groups. Note, a constant (k = 1.0 e^{-10}) is added to each value to avoid any zero values before calculating the ratio. A positive ChemProp score indicates a forward change (A -> B), whereas a negative ChemProp score indicates a reversed change (B ->A).

Figure 1b showcases different examples of relations that would be captured with high scores as well as challenging relations that would result in low scores. The magnitude of the change in abundance is thereby reflected in the absolute changes in proportions and represented as a ChemProp score (the higher the score, the higher the ratio). As a default cut-off we recommend to use a ChemProp score of 2, which would represent a 10-fold change in the feature pair. However, an optimal cut-off is compound and study dependent.

The input required to perform the ChemProp workflow is a FBMN GNPS task ID from a job that includes metadata indicating the sequential order of samples. We recommend using the ReDU metadata template,[16] that is validated to be compatible with ChemProp. The output of the ChemProp workflow consists of a .graphML file which can be directly loaded into network visualization software such as Cytoscape. The .graphML is a summary file that contains the delta mass as well as ChemProp and cosine scores of the connected nodes. The sign of the ChemProp score can be used to map the directionality in the form of arrows in Cytoscape. In addition to the .graphML network file, ChemProp also provides a tabulated output of node connections, delta masses and ChemProp scores as a .csv file that can be used for further statistical analysis of global transformations within datasets.

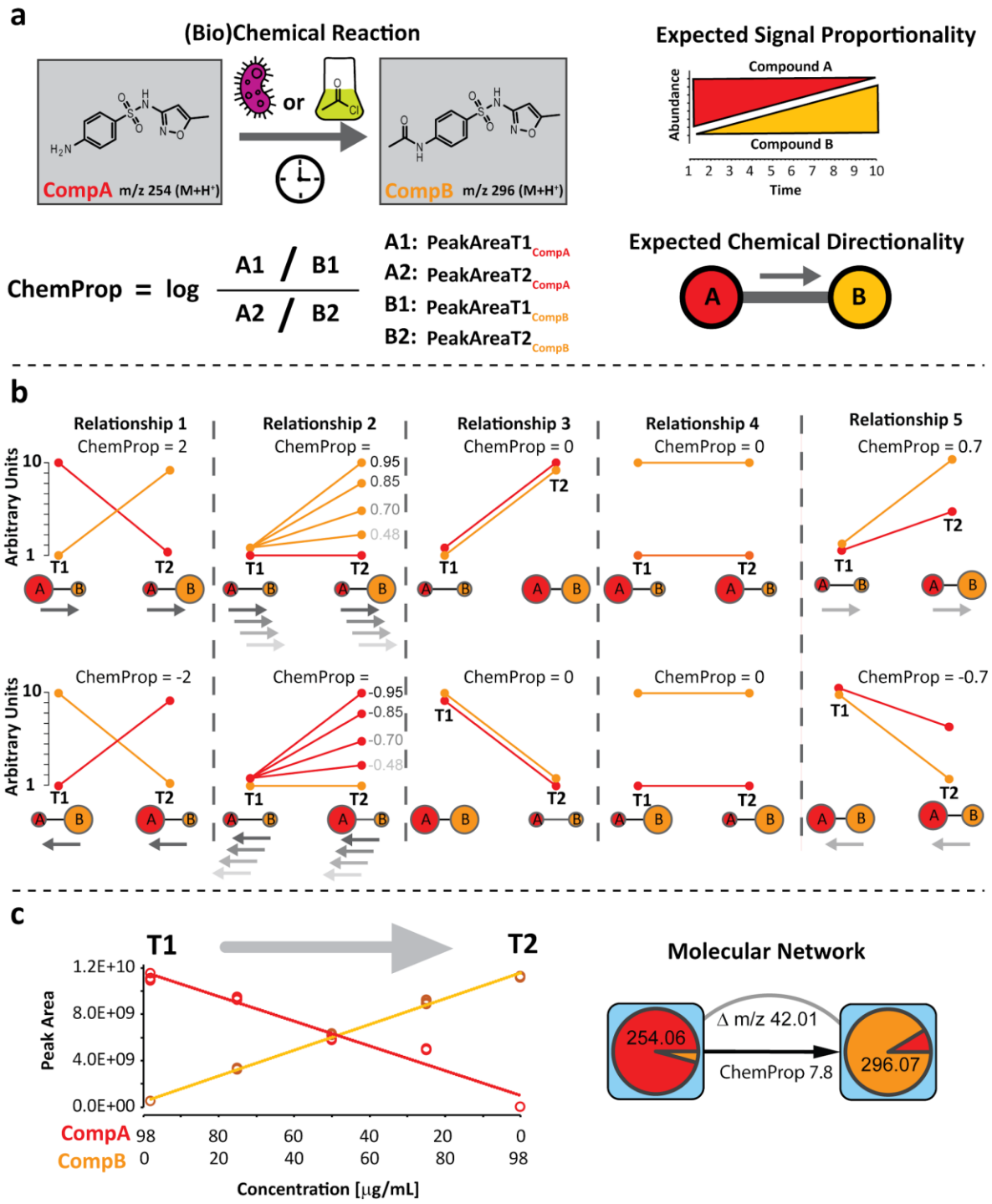


Figure 1. ChemProp concept, expected scenarios and ground truth experiment.
(a) Concept illustrated with the example of the (bio)chemical formation of N-acetyl sulfamethoxazole. (b) Plausible scenarios captured by the ChemProp approach. (c) Observed abundance changes (Y axis) of sulfamethoxazole and acetyl-sulfamethoxazole in a demonstration of proof-of-concept data (concentration X axis). If T1 (98 µg/mL; 2 µg/mL) and T2 (2 µg/mL; 98 µg/mL) are considered as two time points, the resulting chemical directionality in the molecular network indicates an acetylation reaction (+ *m/z* 42.01).

Results and Discussion

To evaluate the ChemProp approach, defined mixtures of sulfamethoxazole and N-acetyl-sulfamethoxazole in a dilution series with linear changing proportion were created and analyzed. The resulting peak areas and molecular network are shown in **Figure 1c**. The defined mixture mimics an acetylation reaction with linear conversion of reactants to products over time which represents a common metabolic phase II reaction and microbial resistance strategy (excluding reaction kinetics).[11,12] Looking at the experimentally derived peak areas of sulfamethoxazole and N-acetyl-sulfamethoxazole, an expected anti-correlation was observed. The maximum ChemProp score was 7.80 between concentration point 1 and 7, which are considered as T1 (98 ug/mL A; 2 ug/mL B) and T2 (2 ug/mL A; 98 ug/mL B), representing the largest differences in the mock acetylation reaction.

To test ChemProp with real-life samples, we applied the workflow to three publicly available MS/MS datasets. All datasets were from studies with temporal sampling. In order to explore all datasets for putative transformations and identify patterns of frequent reactions (mass shifts), we plotted the delta masses from all molecular networks from data sets against their particular ChemProp score. The global transformation patterns for all four data sets can be evaluated between the datasets and are shown in **Figure 2a**. The results reveal that different time dependent (biological) changes are distinct to each experiment. For instance, the anaerobic culturing of *Clostridium cadaveris* showed more frequently negative ChemProp scores, indicating mass losses (e.g. demethylation or dehydration). ChemProp can thus give insights into catabolic, or anabolic behavior, and can highlight the frequency of particular mass shifts/modifications.

In the example of microbial transformations of bile acids, cholic acid and deoxycholic acid were added to a *Clostridium cadaveris* culture and incubated. After extraction and LC-MS/MS analysis, the ChemProp workflow was applied to identify potential bile acid transformation products. Figure 2b shows a molecular network of a subset of bile acids detected in the culture extracts. High ChemProp scores were observed between nodes of deoxycholic acid (DCA) (detected as $[M-3H_2O+H]^+$) or cholic acid (CA) (detected as $[M-H_2O+H]^+$) and leucine conjugated deoxycholic acid (Leu-DCA). Based on *a priori* knowledge that the bile acids were fed to the culture, we hypothesize that either parent bile acid could be the substrate for formation of Leu-DCA. The conversion of CA to Leu-DCA would require a conjugation to leucine and dehydroxylation. There was also alanine conjugated CA (Ala-CA) detected but the deoxycholic derivative was not observed. Looking at the abundance change in these relationships, the ChemProp score reflects the decrease of CDCA and CA over time while Ala-CA and Leu-CDCA increase (Figure S1).

Next, we applied ChemProp to a dataset from a study that investigated the metabolism of omeprazole[13], a proton pump inhibitor drug, in healthy humans. **Figure 2c** displays the molecular network component in which omeprazole, 5-hydroxyomeprazole and carboxyomeprazole (omeprazole metabolites), and a deuterated standard (*i.e.* omeprazole-d_3) were detected. Further, a phase II metabolite, hydroxyomeprazole-5-O-glucuronide was connected in the network. The largest ChemProp value observed was 2.89 between the omeprazole-omeprazole-d_3 node pair, (60 to 120 min time interval). We observed that omeprazole-d_3 remains constant (as it was spiked into each sample), while the level of omeprazole increased from 60 to 120 min (Figure S2). While not offering any biological insight, this observation supports the intended measure of ChemProp. On the other hand, the CYP2C19 transformation of omeprazole to 5-hydroxyomeprazole, a

primary metabolite of omeprazole which is of biological relevance, had a ChemProp value of 0.45. While the absolute ChemProp value was smaller than that of other test cases within the study, the value was sufficient to be prioritized as interesting and was found to reflect the combined effect of pharmacokinetic absorption from the orally dosed (single dosage) omeprazole into the blood plasma, as well as the subsequent metabolism by CYP2C19 in the intestine and liver. Combined absorption and metabolism is reflected in the omeprazole-carboxyomeprazole node pair. The hydroxyomeprazole-5-O-glucuronide and 5-hydroxyomeprazole node pair had a greater ChemProp value (2.02) during the 60 to 120 min interval which reflects the lesser signal of glucuronide metabolite as its formation is dependent on the 5-hydroxyomeprazole precursor (Figure S2).

Lastly, we applied the ChemProp approach to explore the production of bacterial natural products. Over a culturing time course of *Bacillus subtilis, the* production of different surfactin derivatives, a class of well-known and characteristic lipopeptides [14], changed in time dependent fashion (**Figure 2d** and **Figure S3**). From a biological perspective, we expected a change of surfactin levels over time as they have been described in the context of motility in *B. subtilis*[15]. We observed two clusters of surfactin derivatives with similar ChemProp patterns. Notably, while the first group of surfactins A-E (Figure 2d and Figure S3, left side of network) did not show strong variance over the time course (low ChemProp score), the second cluster of surfactin derivatives (*m/z* 1036.69 and *m/z* 1067.73, right side of the network), started to increase after 24h of cultivation (high ChemProp score, relationship 2). According to these proportional changes, the network shows high ChemProp scores within the two groups of compounds (2.4-3.1) and low scores within these groups (0.04-0.5). The observed mass differences with the networks were Δ *m/z* 14.02 and Δ *m/z* 28.04 which correspond to methylations or variations in the amino acids incorporated during the biosynthesis of these bacterial metabolites[16]. However, less frequently observed mass shifts occurred between the two derivative groups (left and right side of network) with Δ *m/z* 31.04 and Δ *m/z* 45.05 (Figure 3d) that are differentially produced over the time course.

Looking at the MS/MS spectra of the differential expressed variants (*m/z* 1036.69 and *m/z* 1067.73) both compounds contain a shared MS/MS fragment ion of *m/z* 685 $[M+H]^+$ (y_6+H_2O) that suggests amino acid composition (e.g. Leu/Ile-Leu/Ile-Val-Asp-Leu/Ile-Leu/Ile) (Figure S4), other fragments such as *m/z* 356 and *m/z* 370 indicate chemical variations, likely present at the □-hydroxy fatty acid chain, which we speculate could indicate the requirement of a different enzyme and/or signaling pathway that changed over the time course of *B. subtillis* growth.

This example highlights that ChemProp can be used to identify directionality of actual (biochemical) reactions and to highlight potential pattern changes of structurally /biochemically related compounds.

For both scenarios, it is important to note that the proportionality approach should be considered as a prioritization and hypothesis generating strategy that complements chemical information provided by the feature-based molecular networking workflow.

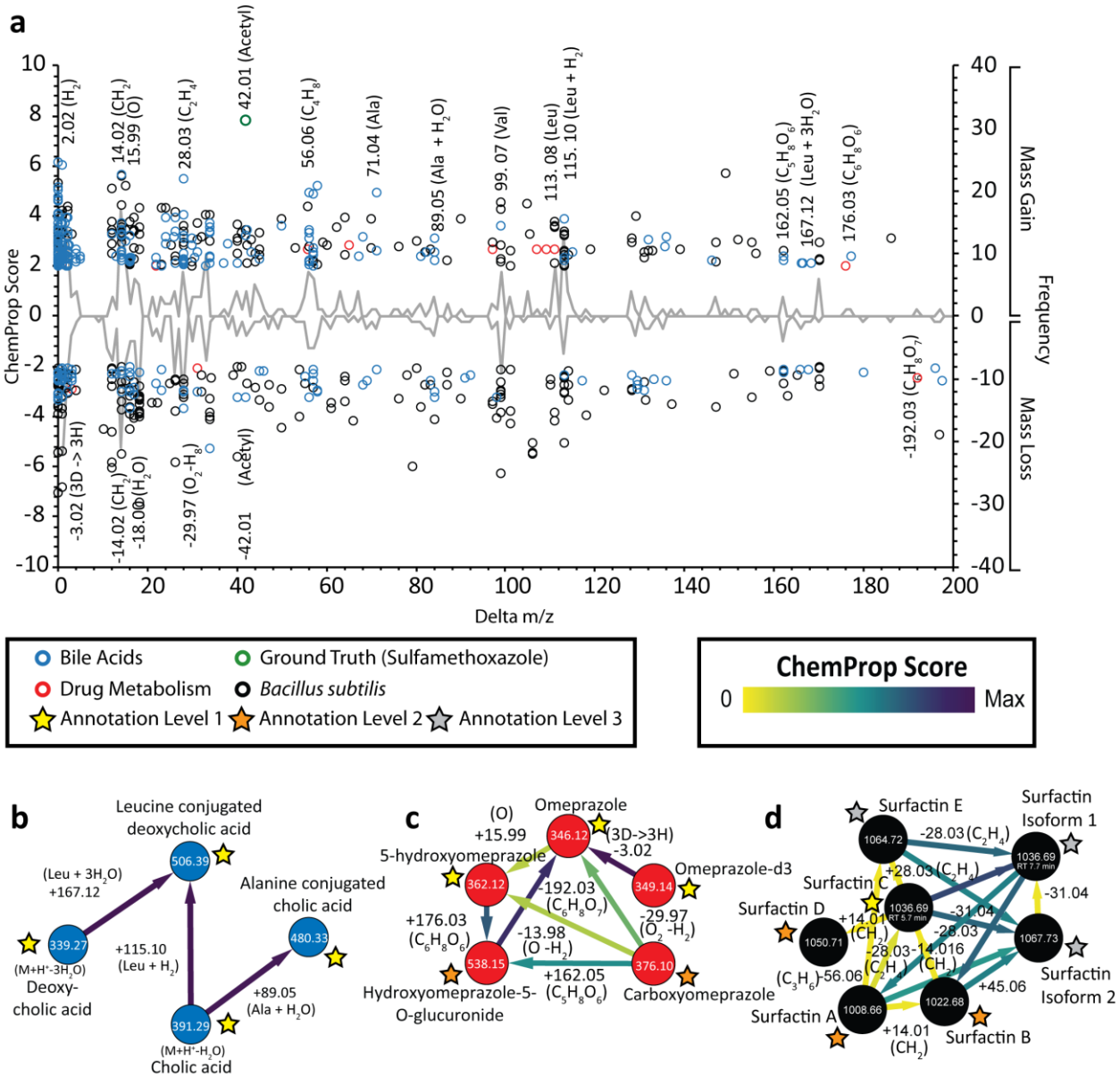


Figure 2: Global analysis and examples of captured directionality from test datasets.
(a) Summary of delta mass shifts captured from complex datasets including drug metabolism, bile acids, and fungal interaction. (b) Molecular network of bile acid modifications, highlighting conjugations with common amino acids. This example corresponds to relationships 2 and 3 (shown in Figure 1). (c) Network of detected features from a drug metabolism dataset involving omeprazole and its hydroxylated modification mediated by cytochrome P450. This example of drug metabolism corresponds to relationship 2 (shown in Figure 1). (d) Network of surfactins produced by *Bacillus subtilis* over a time course experiment. This example corresponds to relationships 2, 3 and 4 (shown in Figure 1).

Conclusion

Molecular Networking aims to enhance chemical insight from non-targeted MS/MS experiments by connecting spectrally-related and thus structurally-related compounds. The ChemProp approach facilitates the prioritization of relative changes of connected nodes within molecular networks over sequential data series (*e.g.* time or space). ChemProp thus enhances one's' ability to formulate hypotheses from non-targeted LC-MS/MS data with respect to mass changes in a biological context, such as microbial modifications, drug metabolism, and changes in biosynthesis patterns. The proportionality approach can be used to suggest directionality of (bio)chemical reactions in time-courses, spatial mapping, or treatment/control experiments and in a broader sense, to highlight abundance pattern changes among related compounds.

Data and Software availability

There are two portions of the ChemProp infrastructure, ChemProp GNPS/ProteoSAFe workflow and the ChemProp Results Exploration Dashboard. The citable source code is available at 10.5281/zenodo.4081635 and active development is at GitHub: https://github.com/CCMS-UCSD/GNPS_Workflows/tree/master/chemprop.

The ChemProp workflow is available through the GNPS environment and github at: https://gnps.ucsd.edu/ProteoSAFe/index.jsp?params=%7B%22workflow%22%3A%20%22CHEMPROP%22%7D.

Detailed instructions, including a step-to-step tutorial, for the use of ChemProp is available through the GNPS documentation: https://ccms-ucsd.github.io/GNPSDocumentation/chemdir/.

All raw and centroid MS/MS data used in this manuscript can be downloaded from the MassIVE repository under the following accession numbers: MSV000085688, MSV000084681, MSV000082493, MSV000082402.

Acknowledgements

A.M.C.R. and P.C.D. were supported by the National Sciences Foundation grant IOS-1656481 and National Institutes of Health Award 1DP2GM137413-01. D.R. was founded by ERC Starting Grant (BacBio 637971) and Ministerio de Ciencia e Innovación (PID2019-107724GB-I00). We thank Dr. Krista Longnecker and all other participants of the ChemProp online Workshop for their feedback or helpful discussion.

Author contributions

DP and AMCR conceived the concept of ChemProp. MW implemented ChemProp into the GNPS environment. DP, AMCR, AKJ, CMS, DR, JMG, ECG and PBF performed sample preparation, MS/MS experiments, analyzed data and validated the ChemProp workflow. SMT provided data and analysis for the omeprazole example. PCD provided guidance, feedback and infrastructure. All authors wrote and edited the manuscript.

Ethics declarations

Mingxun Wang is a co-founder of Ometa labs LLC. Pieter C. Dorrestein is a scientific advisor for Sirenas LLC, Cybele Microbiome and Galileo and a scientific advisor and co-founder of Ometa labs LLC and Enveda with approval by UCSD.

References

(1) Stewart, C. J.; Embleton, N. D.; Marrs, E. C. L.; Smith, D. P.; Fofanova, T.; Nelson, A.; Skeath, T.; Perry, J. D.; Petrosino, J. F.; Berrington, J. E.; Cummings, S. P. Longitudinal Development of the Gut Microbiome and Metabolome in Preterm Neonates with Late Onset Sepsis and Healthy Controls. *Microbiome* **2017**, *5* (1), 75. https://doi.org/10.1186/s40168-017-0295-1.

(2) Petras, D.; Jarmusch, A. K.; Dorrestein, P. C. From Single Cells to Our Planet—Recent Advances in Using Mass Spectrometry for Spatially Resolved Metabolomics. *Curr. Opin. Chem. Biol.* **2017**, *36*, 24–31. https://doi.org/10.1016/j.cbpa.2016.12.018.

(3) Watrous, J.; Roach, P.; Alexandrov, T.; Heath, B. S.; Yang, J. Y.; Kersten, R. D.; van der Voort, M.; Pogliano, K.; Gross, H.; Raaijmakers, J. M.; Moore, B. S.; Laskin, J.; Bandeira, N.; Dorrestein, P. C. Mass Spectral Molecular Networking of Living Microbial Colonies. *Proc Natl Acad Sci U A* **2012**, *109* (26), E1743-52. https://doi.org/10.1073/pnas.1203689109.

(4) Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kapono, C. A.; Luzzatto-Knaan, T.; Porto, C.; Bouslimani, A.; Melnik, A. V.; Meehan, M. J.; Liu, W.-T.; Crüsemann, M.; Boudreau, P. D.; Esquenazi, E.; Sandoval-Calderón, M.; Kersten, R. D.; Pace, L. A.; Quinn, R. A.; Duncan, K. R.; Hsu, C.-C.; Floros, D. J.; Gavilan, R. G.; Kleigrewe, K.; Northen, T.; Dutton, R. J.; Parrot, D.; Carlson, E. E.; Aigle, B.; Michelsen, C. F.; Jelsbak, L.; Sohlenkamp, C.; Pevzner, P.; Edlund, A.; McLean, J.; Piel, J.; Murphy, B. T.; Gerwick, L.; Liaw, C.-C.; Yang, Y.-L.; Humpf, H.-U.; Maansson, M.; Keyzers, R. A.; Sims, A. C.; Johnson, A. R.; Sidebottom, A. M.; Sedio, B. E.; Klitgaard, A.; Larson, C. B.; P, C. A. B.; Torres-Mendoza, D.; Gonzalez, D. J.; Silva, D. B.; Marques, L. M.; Demarque, D. P.; Pociute, E.; O'Neill, E. C.; Briand, E.; Helfrich, E. J. N.; Granatosky, E. A.; Glukhov, E.; Ryffel, F.; Houson, H.; Mohimani, H.; Kharbush, J. J.; Zeng, Y.; Vorholt, J. A.; Kurita, K. L.; Charusanti, P.; McPhail, K. L.; Nielsen, K. F.; Vuong, L.; Elfeki, M.; Traxler, M. F.; Engene, N.; Koyama, N.; Vining, O. B.; Baric, R.; Silva, R. R.; Mascuch, S. J.; Tomasi, S.; Jenkins, S.; Macherla, V.; Hoffman, T.; Agarwal, V.; Williams, P. G.; Dai, J.; Neupane, R.; Gurr, J.; Rodríguez, A. M. C.; Lamsa, A.; Zhang, C.; Dorrestein, K.; Duggan, B. M.; Almaliti, J.; Allard, P.-M.; Phapale, P.; Nothias, L.-F.; Alexandrov, T.; Litaudon, M.; Wolfender, J.-L.; Kyle, J. E.; Metz, T. O.; Peryea, T.; Nguyen, D.-T.; VanLeer, D.; Shinn, P.; Jadhav, A.; Müller, R.; Waters, K. M.; Shi, W.; Liu, X.; Zhang, L.; Knight, R.; Jensen, P. R.; Palsson, B. O.; Pogliano, K.; Linington, R. G.; Gutiérrez, M.; Lopes, N. P.; Gerwick, W. H.; Moore, B. S.; Dorrestein, P. C.; Bandeira, N. Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* **2016**, *34* (8), 828–837. https://doi.org/10.1038/nbt.3597.

(5) Aron, A. T.; Gentry, E. C.; McPhail, K. L.; Nothias, L.-F.; Nothias-Esposito, M.; Bouslimani, A.; Petras, D.; Gauglitz, J. M.; Sikora, N.; Vargas, F.; van der Hooft, J. J. J.; Ernst, M.; Kang, K. B.; Aceves, C. M.; Caraballo-Rodríguez, A. M.; Koester, I.; Weldon, K. C.; Bertrand, S.; Roullier, C.; Sun, K.; Tehan, R. M.; P, C. A. B.; Christian, M. H.; Gutiérrez, M.; Ulloa, A. M.; Tejeda Mora, J. A.; Mojica-Flores, R.; Lakey-Beitia, J.;

Vásquez-Chaves, V.; Zhang, Y.; Calderón, A. I.; Tayler, N.; Keyzers, R. A.; Tugizimana, F.; Ndlovu, N.; Aksenov, A. A.; Jarmusch, A. K.; Schmid, R.; Truman, A. W.; Bandeira, N.; Wang, M.; Dorrestein, P. C. Reproducible Molecular Networking of Untargeted Mass Spectrometry Data Using GNPS. *Nat. Protoc.* **2020**, *15* (6), 1954–1991. https://doi.org/10.1038/s41596-020-0317-5.

(6) Nothias, L.-F.; Petras, D.; Schmid, R.; Dührkop, K.; Rainer, J.; Sarvepalli, A.; Protsyuk, I.; Ernst, M.; Tsugawa, H.; Fleischauer, M.; Aicheler, F.; Aksenov, A. A.; Alka, O.; Allard, P.-M.; Barsch, A.; Cachet, X.; Caraballo-Rodriguez, A. M.; Da Silva, R. R.; Dang, T.; Garg, N.; Gauglitz, J. M.; Gurevich, A.; Isaac, G.; Jarmusch, A. K.; Kameník, Z.; Kang, K. B.; Kessler, N.; Koester, I.; Korf, A.; Le Gouellec, A.; Ludwig, M.; H, C. M.; McCall, L.-I.; McSayles, J.; Meyer, S. W.; Mohimani, H.; Morsy, M.; Moyne, O.; Neumann, S.; Neuweger, H.; Nguyen, N. H.; Nothias-Esposito, M.; Paolini, J.; Phelan, V. V.; Pluskal, T.; Quinn, R. A.; Rogers, S.; Shrestha, B.; Tripathi, A.; van der Hooft, J. J. J.; Vargas, F.; Weldon, K. C.; Witting, M.; Yang, H.; Zhang, Z.; Zubeil, F.; Kohlbacher, O.; Böcker, S.; Alexandrov, T.; Bandeira, N.; Wang, M.; Dorrestein, P. C. Feature-Based Molecular Networking in the GNPS Analysis Environment. *Nat. Methods* **2020**, *17* (9), 905–908. https://doi.org/10.1038/s41592-020-0933-6.

(7) Yu, M.; Petrick, L. Reactomics: Using Mass Spectrometry as a Reaction Detector. *bioRxiv* **2020**, 855148. https://doi.org/10.1101/855148.

(8) Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28* (1), 27–30. https://doi.org/10.1093/nar/28.1.27.

(9) Hartmann, A. C.; Petras, D.; Quinn, R. A.; Protsyuk, I.; Archer, F. I.; Ransome, E.; Williams, G. J.; Bailey, B. A.; Vermeij, M. J. A.; Alexandrov, T.; Dorrestein, P. C.; Rohwer, F. L. Meta-Mass Shift Chemical Profiling of Metabolomes from Coral Reefs. *Proc Natl Acad Sci U A* **2017**, *114* (44), 11685–11690. https://doi.org/10.1073/pnas.1710248114.

(10) Avtonomov, D. M.; Kong, A.; Nesvizhskii, A. I. DeltaMass: Automated Detection and Visualization of Mass Shifts in Proteomic Open-Search Results. *J Proteome Res* **2019**, *18* (2), 715–720. https://doi.org/10.1021/acs.jproteome.8b00728.

(11) Pluvinage, B.; Dairou, J.; Possot, O. M.; Martins, M.; Fouet, A.; Dupret, J.-M.; Rodrigues-Lima, F. Cloning and Molecular Characterization of Three Arylamine N-Acetyltransferase Genes from Bacillus Anthracis: Identification of Unusual Enzymatic Properties and Their Contribution to Sulfamethoxazole Resistance. *Biochemistry* **2007**, *46* (23), 7069–7078.

(12) Reis, P. J. M.; Homem, V.; Alves, A.; Vilar, V. J. P.; Manaia, C. M.; Nunes, O. C. Insights on Sulfamethoxazole Bio-Transformation by Environmental Proteobacteria Isolates. *J. Hazard. Mater.* **2018**, *358*, 310–318. https://doi.org/10.1016/j.jhazmat.2018.07.012.

(13) Jarmusch, A. K.; Vrbanac, A.; Momper, J. D.; Ma, J. D.; Alhaja, M.; Liyanage, M.; Knight, R.; Dorrestein, P. C.; Tsunoda, S. M. Enhanced Characterization of Drug Metabolism and the Influence of the Intestinal Microbiome: A Pharmacokinetic, Microbiome, and Untargeted Metabolomics Study. *Clin. Transl. Sci. n/a* (n/a). https://doi.org/10.1111/cts.12785.

(14) Moro, G. V.; Almeida, R. T. R.; Napp, A. P.; Porto, C.; Pilau, E. J.; Lüdtke, D. S.; Moro, A. V.; Vainstein, M. H. Identification and Ultra-High-Performance Liquid Chromatography Coupled with High-Resolution Mass Spectrometry Characterization of Biosurfactants, Including a New Surfactin, Isolated from Oil-Contaminated Environments. *Microb. Biotechnol.* **2018**, *11* (4), 759–769. https://doi.org/10.1111/1751-7915.13276.

(15) Molina-Santiago, C.; Pearson, J. R.; Navarro, Y.; Berlanga-Clavero, M. V.; Caraballo-Rodriguez, A. M.; Petras, D.; García-Martín, M. L.; Lamon, G.; Haberstein, B.; Cazorla, F. M. The Extracellular Matrix Protects Bacillus Subtilis Colonies from Pseudomonas Invasion and Modulates Plant Co-Colonization. *Nat. Commun.* **2019**, *10* (1), 1–15.

(16) Hu, F.; Liu, Y.; Li, S. Rational Strain Improvement for Surfactin Production: Enhancing the Yield and Generating Novel Structures. *Microb. Cell Factories* **2019**, *18* (1), 1–13.