**Multidimensional Stoichiometric Compound Classification (MSCC)**

Publication Information:

Albert Rivas-Ubach, Yina Liu, Thomas S. Bianchi, Nikola Tolić, Christer Jansson, Ljiljana Paša-Tolić. (2018). Moving beyond the van Krevelen diagram: A new stoichiometric approach for compound classification in organisms. Analytical Chemsitry. DOI: 10.1021/acs.analchem.8b00529

The R script assigns one of the compounds considered in the MSCC constraints (Lipids, Protein, Amino-Sugars, Carbohydrates, Phytochemical compounds (oxy-aromatic)) to a series of elements and elemental stoichiometry ratios.

The CSV file to be run with the R code and assign compounds to different variables (metabolic features detected by High-Resolution mass spectrometry) must contain at least the following information (in columns) for each variable with no specific order (more columns can be added since they will not interfere with the R code):

“O.C” <- column for O:C ratio column

“H.C” <- H:C ratio column

“N.C” <- N:C ratio column

“P.C” <- P:C ratio column

“N.P” <- N:P ratio column

“O” <- O column

“N” <- N column

“P” <- P column

“S” <- S column

“Mass” <- exact mass column

Molecular formulas can be assigned to a high-resolution mass spectrometry feature by using software such as Formularity (Nikola et al., 2017 in Analytical Chemistry, DOI: 10.1021/acs.analchem.7b03318)

Example of a hypothetic dataset:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | **Elemental and Stoichiometric Information (Required)** | | | | | | | | | | | | Sample Information (e.g. HRMS peak intensity values) | | | | | |
| Feature ID | Elemental Formula | Formula Type | **Mass** | **C** | **H** | **O** | **N** | **P** | **S** | **O.C** | **H.C** | **N.C** | **P.C** | **N.P** | Sample A1 | Sample A2 | Sample A3 | Sample B1 | Sample B2 | Sample B3 |
| 1 | C19H33O5N9S | CHONS | 499.2325 | 19 | 33 | 5 | 9 | 0 | 1 | 0.26 | 1.74 | 0.47 | 0.00 | 0.00 | 4010617 | 81524.69 | 1963844 | 194415.8 | 165433.9 | 234049.5 |
| 2 | C23H40O6N10S | CHONS | 584.2853 | 23 | 40 | 6 | 10 | 0 | 1 | 0.26 | 1.74 | 0.43 | 0.00 | 0.00 | 13767.63 | 127648.5 | 36546.04 | 644199.4 | 688459.8 | 266972.5 |
| 3 | C22H37O13N6P | CHONP | 594.2497 | 22 | 37 | 13 | 6 | 1 | 0 | 0.59 | 1.68 | 0.27 | 0.05 | 6.00 | 4489.683 | 5635.086 | 5276.031 | 10172.77 | 160151.2 | 122011.4 |
| 4 | C29H50O7N6 | CHON | 594.3741 | 29 | 50 | 7 | 6 | 0 | 0 | 0.24 | 1.72 | 0.21 | 0.00 | 0.00 | 334464.9 | 306728.8 | 768420.6 | 458695.1 | 991765 | 647233.8 |
| 5 | C29H48O | CHO | 412.3705 | 29 | 48 | 1 | 0 | 0 | 0 | 0.03 | 1.66 | 0.00 | 0.00 | 0.00 | 4096455 | 13353856 | 4783002 | 3930514 | 3643681 | 6525600 |
| 6 | C25H38O4 | CHO | 402.277 | 25 | 38 | 4 | 0 | 0 | 0 | 0.16 | 1.52 | 0.00 | 0.00 | 0.00 | 12387915 | 20738031 | 13838732 | 19831475 | 8440247 | 12537276 |
| 7 | C21H36O9N4S | CHONS | 520.2203 | 21 | 36 | 9 | 4 | 0 | 1 | 0.43 | 1.71 | 0.19 | 0.00 | 0.00 | 2463422 | 877958.9 | 4909353 | 2520199 | 2214910 | 2387963 |
| 8 | C24H43O3N3 | CHON | 421.3304 | 24 | 43 | 3 | 3 | 0 | 0 | 0.13 | 1.79 | 0.13 | 0.00 | 0.00 | 238469 | 491862.1 | 423312 | 851000.8 | 62098.04 | 794108 |
| 9 | C15H10O13S2 | CHOS | 461.9563 | 15 | 10 | 13 | 0 | 0 | 2 | 0.87 | 0.67 | 0.00 | 0.00 | 0.00 | 281164.4 | 406600.1 | 353023.7 | 326417.8 | 920016.4 |  |
| 10 | C6H9O2N3 | CHON | 155.0695 | 6 | 9 | 2 | 3 | 0 | 0 | 0.33 | 1.50 | 0.50 | 0.00 | 0.00 | 61411371 | 63992262 | 51359312 | 54900324 | 65758004 | 89868102 |
| 11 | C26H34O8N6S | CHONS | 590.2159 | 26 | 34 | 8 | 6 | 0 | 1 | 0.31 | 1.31 | 0.23 | 0.00 | 0.00 | 2361216 | 1672936 | 1528812 | 819687.5 | 1338865 | 734678.8 |
| 12 | C23H34O6N4 | CHON | 462.2478 | 23 | 34 | 6 | 4 | 0 | 0 | 0.26 | 1.48 | 0.17 | 0.00 | 0.00 | 12890.59 | 11951.13 | 21134.02 | 44777.59 | 34437.82 | 15504.69 |
| 13 | C24H39O6N8PS | CHONPS | 568.2791 | 24 | 39 | 6 | 8 | 1 | 1 | 0.25 | 1.63 | 0.33 | 0.04 | 8.00 | 9371.055 | 10850.6 | 2433.84 | 2355.149 | 7869.067 | 56467.08 |
| 14 | C18H29O9N5S | CHONS | 491.1686 | 18 | 29 | 9 | 5 | 0 | 1 | 0.50 | 1.61 | 0.28 | 0.00 | 0.00 | 366989.5 | 37494.8 | 231979.3 | 81608.06 | 46769.41 | 21508.33 |
| 15 | C31H64O3 | CHO | 484.4855 | 31 | 64 | 3 | 0 | 0 | 0 | 0.10 | 2.06 | 0.00 | 0.00 | 0.00 | 3114134 | 4450268 | 2003169 | 3446747 | 2766780 | 2538759 |
| 16 | C23H51O5N2P | CHONP | 466.3536 | 23 | 51 | 5 | 2 | 1 | 0 | 0.22 | 2.22 | 0.09 | 0.04 | 2.00 | 3158513 | 4520448 | 1367594 | 8373131 | 7762882 | 4809900 |
| 17 | C21H33O7N5S | CHONS | 499.2101 | 21 | 33 | 7 | 5 | 0 | 1 | 0.33 | 1.57 | 0.24 | 0.00 | 0.00 | 2142696 | 999184.6 | 1823207 | 3131098 | 1340149 | 1483401 |
| 18 | C12H23O10NS2 | CHONS | 405.0763 | 12 | 23 | 10 | 1 | 0 | 2 | 0.83 | 1.92 | 0.08 | 0.00 | 0.00 | 1330840 | 1611011 | 1587143 | 1218215 | 1703107 | 455192.8 |
| 19 | C24H33O6N7 | CHON | 515.2492 | 24 | 33 | 6 | 7 | 0 | 0 | 0.25 | 1.38 | 0.29 | 0.00 | 0.00 | 153298.2 | 43339.43 | 147315.6 | 450367.3 | 932065 | 365217.7 |
| 20 | C13H21O8N5 | CHON | 375.139 | 13 | 21 | 8 | 5 | 0 | 0 | 0.62 | 1.62 | 0.38 | 0.00 | 0.00 | 854467.3 | 33046.17 | 2072787 | 3122.609 | 1752.768 |  |
| 21 | C7H10O7 | CHO | 206.0427 | 7 | 10 | 7 | 0 | 0 | 0 | 1.00 | 1.43 | 0.00 | 0.00 | 0.00 | 69411783 | 1.14E+08 | 81137086 | 1.18E+08 | 1.26E+08 | 1.13E+08 |
| 22 | C24H45O9N5 | CHON | 547.3217 | 24 | 45 | 9 | 5 | 0 | 0 | 0.38 | 1.88 | 0.21 | 0.00 | 0.00 | 819663.5 | 865958.9 | 674266.1 | 202725.6 | 287174.9 |  |

The R code performs 3 main steps:

1. Assign each metabolic feature to a specific compound (Lipids, Carbohydrates, Protein, etc..) according to the constraints defined in the MSCC (see Rivas-Ubach et al., 2018 in Analytical Chemistry DOI: 10.1021/acs.analchem.8b00529).

After this, R exports the same read dataset with an extra column with the compound assignation name.

1. The R code also calculate the proportions of each compound category found in the dataset and exports a table with a summary of the proportions. Not Matched compounds and Double Matched compounds are also indicated.
2. Finally, it can also perform a pie-chart plot the different calculated proportions.

The only information that has to be inserted in the R code is the directory path to READ the CSV file and the directory paths to save the output tables.