Supplementary information

What is under the hump?

Mass Spectrometry based analysis of complex mixtures in processed food – Lessons from the characterisation of black tea thearubigins, coffee melanoidines and caramel

Nikolai Kuhnert ^{a*}, Farnoosh Dairpoosh ^a, Ghada Yassin ^a, Agnieszka Golon^a and Rakesh Jaiswal^a

A representative selection of molecular formulas for a total of 52 terpenes (268) and 99 commonly occurring dietary phenolic compounds and polyphenols (148) were collected. For each list, the O/C and H/C ratio were calculated and the values were scatter plotted. The elemental ratio boundaries were defined for terpenes and polyphenols using statistical software JMP 8.2 in order to show the area where these class of compounds have the most intensity in a scatter plot matrix (Figure 59, 60). The software enables to draw an ellipse that contains the specified mass of points. The density ellipsoid is a graphical indicator of the correlation between two variables and it is computed from the bivariate normal distribution fit to the *X* and *Y* variables. The bivariate normal density is a function of the means and standard deviations of the *X* and *Y* variables and the correlation between them. These ellipses are both density contours and confidence curves. As confidence curves, they show where a given percentage of the data is expected to lie, assuming the bivariate normal distribution.



Figure 59- Van Krevelen representation with density ellipse for polyphenols

Figure 60- Van Krevelen representation with density ellipse for terpenes

Figure 61 shows the two density ellipses developed for the polyphenols with sugar moities and without sugar moities separately, overlap.



Figure 61- Density ellipse for the phenolic and polyphenolic compounds;

(a): density ellipse for the polyphenols with sugar moieties

(b): density ellipse for the polyphenols without sugar moieties

(c): (a) and (b) overlap

The obtained intervals for polyphenols and terpenes were added to the Van Krevelen representation (see Figure 62).



Figure 62- The improved Van Krevelen representation for all classes of compounds including polyphenols and terpenes

With sufficient elemental ratio data, boundaries for H/C and O/C ratios can be defined for individual classes of naturally occurring compounds: while any data point that lies within the elemental ratio boundaries defined for a certain class of compounds might correspond to this class of compound, it does not inevitably do so. However, if a data point falls outside the elemental ratio boundaries, it certainly does not belong to this class of compounds (265).

The presence of a datapoint within the elemental ratio boundaries defined in the Van Krevelen diagram, allows tentative assignment of a compound to a given class of natural products. In principle, data analysis can be expanded to allow correct choice of molecular formulas, if more than one molecular formula suggestion is given in the computation routine. If a choice of data points exsists, it is suggested here that the Van Krevelen diagram allows selection of a likely class of natural products in particular, if one data point lies within elemental ratio boundaries and others do not. Therefore this technique adds a welcome addition to Fiehns seven golden rules in correct assignment of molecular formulas.

References:

268- Dawidar, A. M.; Metwally, M. A.; Mogib, M.; Ayyad, S. N. Terpenoids, in *Chemistry of Natural Products*, first edition, I.S.B. N. 977-19-5462-8

148- Sakakibara, H.; Honda, Y.; Nakagawa, S.; Ashida, H.; Kanazawa, K. Simultaneous Determination of All Polyphenols in Vegetables, Fruits, and Teas. *J. Agric. Food. Chem.* **2003**, *51*, 571-581.

265- Wu, Z.; Rodgers, R. P.; Marshall, A. G. Two- and Three-Dimensional Van Krevelen Diagrams: A Graphical Analysis Complementary to the Kendrick Mass Plot for Sorting Elemental Compositions of Complex Organic Mixtures based on Ultrahigh-Resolution Broadband Fourier Transform Ion Cyclotron Resonance Mass Measurements. *Anal. Chem.* **2004**, *76*, 2511-2516.