Phytochemistry 93 (2013) 182–191

Contents lists available at SciVerse ScienceDirect

Phytochemistry

journal homepage: www.elsevier.com/locate/phytochem

Classification of flavonoid compounds by using entropy of information theory

Gloria Castellano [a,*], Juan Luis González-Santander [a], Ana Lara [a], Francisco Torrens [b]

[a] Facultad de Ciencias Experimentales, Universidad Católica de, Valéncia San Vicente Mártir, Guillem de Castro-94, E-46001 Valencia, Spain
[b] Institut Universitari de Ciéncia Molecular, Universitat de València, Edificid'Instituts de Paterna, P.O. Box 22085, E-46071 Valencia, Spain

ARTICLE INFO

Article history:
Received 13 June 2012
Received in revised form 8 March 2013
Available online 3 May 2013

Keywords:
Information entropy
Molecular classification
Antioxidant activity
Flavonoid
Polyphenol

ABSTRACT

A total of 74 flavonoid compounds are classified into a periodic table by using an algorithm based on the entropy of information theory. Seven features in hierarchical order are used to classify structurally the flavonoids. From these features, the first three mark the group or column, while the last four are used to indicate the row or period in a table of periodic classification. Those flavonoids in the same group and period are suggested to show maximum similarity in properties. Furthermore, those with only the same group will present moderate similarity. In this report, the flavonoid compounds in the table, whose experimental data in bioactivity and antioxidant properties have been previously published, are related.

1. introduction

The aim of the present report is to develop the learning potentialities of our program of molecular structural classification, which is applied to flavonoids. The molecules with similar chemical structures are described. We have extended the study of general approaches to the processing of structural information. In this way, the antioxidant activity of any flavonoid can be predicted, depending on its molecular structure as compared with those that have been already classified.

Flavonoids are polyphenolic compounds that are secondary metabolites exclusively from vegetal origin. All flavonoids arise from the combined biosynthesis of the shikimic acid and acetate–malonate pathways. There is much interest in the biological effects of flavonoids because they exhibit antioxidant, anti-inflammatory, anti-tumor, anti-viral and anti-allergical properties (Croft, 1998; Dumay et al., 2004; Khadem and Marles, 2010; Lacikova et al., 2009; Pietta, 2000; Yang et al., 2001). Moreover, flavonoids may play an important role in atherosclerosis prevention and therapy (Yi et al., 2001).

Flavonoids (*cf.* Fig. 1) are C15 compounds composed of two phenolic rings connected by a three-carbon unit.

These compounds can act as antioxidants by a number of potential pathways. The most important one is likely to be by free-radical scavenging, in which the polyphenol can break the chain reaction of free radicals (Pietta, 2000). The resulting radical formed on the polyphenol has to be stable so as to prevent it from acting as a chain-propagating radical (Croft, 1998). Another pathway of apparent antioxidant action of the flavonoids is chelation of catalytic metal ions, as copper or iron, that may prevent their involvement in Fenton-type and Haber–Weiss reactions, which can generate hydroxyl radicals. The ability of flavonoids to react with metal ions may also render them pro-oxidant: reduce Cu^{2+} to Cu^{+} and hence allow the formation of initiating radicals (Cao et al., 1997; Halliwell et al., 1995).

The radical-scavenging antioxidants inhibit the free-radical-mediated oxidation of lipids, proteins and DNA, which is involved in diseases. Phenolic compounds act as antioxidants by inhibiting enzymes involved in radical generation (Anouar et al., 2009; Dewick, 2002; Fukumoto and Mazza, 2000).

Among the several methods designed to measure the activity of compounds, there are such as the lipid peroxidation inhibition capacity (LPIC) assay (Zhang et al., 2007), 1,1-diphenyl-2-picrylhydrazyl (DPPH) scavenging (Brand-Williams et al., 1995), cyclic voltammetry (Simic et al., 2007) and using the induction period method (Kadoma and Fujisawa, 2008). The antioxidants act synergistically together with other antioxidants *in vivo* (Omata et al., 2008).

The main structural features of flavonoids have to pay attention to some structural requirements for efficient antioxidant potency: an *ortho*-dihydroxy (catechol) structure on the B ring, which confers greater stability to aroxyl radicals, possibly through hydrogen bonding, and which participates in electron delocalization; 2,3-double bond in conjugation with a 4-keto function provides electron delocalization from the B ring; hydroxyl groups at positions 3 and 5 provide hydrogen bonding to the keto group (Bors et al., 1990; Croft, 1998; Patra and Chua, 2011). Even, the occurrence of

* Corresponding author. Tel.: +34 963544431; fax: +34 963543274.
E-mail address: gloria.castellano@ucv.es (G. Castellano).

0031-9422/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.phytochem.2013.03.024

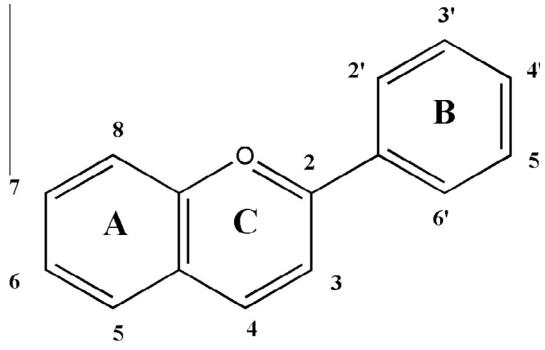

Fig. 1. Basic structure of flavonoids.

substituents with positive mesomeric and inductive electronic effects is also needed. The presence of certain hydroxyl groups on the flavonoid rings enhances antioxidant activity; glycosylation of flavonoids diminishes their activity when compared to the corresponding aglycones; alkyl and mesomeric methoxy groups possessing positive inductive effects show such function, and O-methylation may reflect steric effects that perturb planarity and decrease antioxidant activity and hydrophobicity. Flavonoids with gallate or cinnamate moiety reflect the contribution from gallic or cinnamic acid, and increase antioxidant action (Rice-Evans et al., 1996). Conjugation between the A- and B rings permits a resonance effect of the aromatic nucleus that lends stability to the flavonoid radical.

In an earlier publication, it was examined the molecular classification of 33 phenolic compounds derived from cinnamic and benzoic acids (Castellano et al., 2012). In the present report, 74 flavonoids: anthocyanidins, flavanes, flavones, flavonols and flavanones have been clustered.

2. Calculation

The computational method was applied to polyphenols (Castellano et al., 2012). First step in quantifying the concept of similarity for molecules of flavonoid compounds is to list the most important moieties with respect to the antioxidant activity. Properties vector $\bar{i} = \langle i_1, i_2, \ldots i_k, \ldots \rangle$ should be associated with each flavonoid i, whose components correspond to a number of characteristic groups in the molecule, in a hierarchical order according to the expected importance of their antioxidant potency. Components i_k are "1" or "0", according to experimental conclusions of antioxidant power to structural variations in the flavonoid: index $i_1 = 1$ denotes either a cinnamic or benzoic ester group at 3-position on ring C (Fig. 1), $i_2 = 1$ signifies an O-dihydroxy (catechol) structure on the B ring for electron delocalization, $i_3 = 1$ indicates the presence of two hydroxyl groups at positions 3 and 5, $i_4 = 1$ means the presence of a 2,3-double bond in conjugation with a 4-keto function, providing electron delocalization from the B ring , $i_5 = 1$ represents both a 2,3-double bond and 4,5-double bond that cause electron delocalization from the B ring, $i_6 = 1$ shows the absence of alkoxyl and glycoxyl groups on ring B, and $i_7 = 1$ indicates the absence of alkoxyl and glycoxyl groups on ring A (Table 1).

Let us denote by r_{ij} $(0 \leqslant r_{ij} \leqslant 1)$ the similarity index of two phenolic compounds associated with the $\bar{i}$ and $\bar{j}$ vectors, respectively. The relation of similitude is characterized by a similarity matrix $\bar{\bar{R}} = [r_{ij}]$. The similarity index between two phenolic compounds $\bar{i} = \langle i_1, i_2, \ldots i_k, \ldots \rangle$ and $\bar{j} = \langle j_1, j_2, \ldots j_k, \ldots \rangle$ is defined as:

$$r_{ij} = \sum_k t_k (a_k)^k \quad (k = 1, 2, \ldots) \tag{1}$$

where $0 \leqslant a_k \leqslant 1$ and $t_k = 1$ if $i_k = j_k$, but $t_k = 0$ if $i_k \neq j_k$. This definition assigns a weight $(a_k)^k$ to any property involved in the description of molecule i or j. The hierarchical order of the seven structural features is expressed by their corresponding weights. For instance, for all $a_k = 0.5$ these weights are 0.5, 0.25, 0.125, 0.0625, 0.03125, 0.015625 and 0.0078125.

The grouping algorithm uses the stabilized similarity matrix obtained by applying the max–min composition rule defined:

$$(\bar{\bar{R}} o \bar{\bar{S}})_{ij} = \max_k [\min_k (r_{ik}, s_{kj})] \tag{2}$$

where $\bar{\bar{R}} = [r_{ij}]$ and $\bar{\bar{S}} = [s_{ij}]$ are two similarity matrices and $(\bar{\bar{R}} o \bar{\bar{S}})_{ij}$ is the (i,j)th element of the matrix $\bar{\bar{R}} o \bar{\bar{S}}$ (Kaufmann, 1975). It can be shown that when applying this rule iteratively so that $\bar{\bar{R}}(n+1) = \bar{\bar{R}}(n) o \bar{\bar{R}}$, there is an integer n such that: $\bar{\bar{R}}(n) = \bar{\bar{R}}(n+1) = \ldots$ The resulting matrix $\bar{\bar{R}}(n)$ is called the stabilized similarity matrix, which importance lies in the fact that in the classification process, it will generate a partition into disjoint classes. It is used and designated by $\bar{\bar{R}}(n) = [r_{ij}(n)]$. The grouping rule follows: i and j are assigned to the same class if $r_{ij}(n) \geqslant b$, where b is the grouping level of the classification. Class of i noted $\hat{i}$ is the set of species j that satisfies the rule $r_{ij}(n) \geqslant b$. The matrix of classes is:

$$\bar{\bar{R}}(n) = [\hat{r}_{\hat{i}\hat{j}}] = \max_{s,t} (r_{st}) (s \in \hat{i}, t \in \hat{j}) \tag{3}$$

where s stands for any index of a species belonging to the class $\hat{i}$ (similarly for t and $\hat{j}$). Rule (3) means finding the largest similarity index between species of two different classes.

The information entropy associated with the matrix of similarity $\bar{\bar{R}}$ is:

$$h(\bar{\bar{R}}) = -\sum_{i,j} r_{ij} \ln r_{ij} - \sum_{i,j} (1 - r_{ij}) \ln (1 - r_{ij}) \tag{4}$$

In the classification algorithm, each hierarchical tree corresponds to a dependence of entropy on the grouping level and an h–b diagram can be obtained (White, 1989). The equipartition conjecture of entropy production is proposed as a selection criterion, among different variants, resulting from classification among hierarchical trees. According to this conjecture, the best configuration of a dendrogram is that in which entropy production is most uniformly distributed.

Learning procedures similar to stochastic methods are implemented. Consider a given partition into classes as good from practical observations. This corresponds to a reference similarity matrix $\bar{\bar{S}} = [s_{ij}]$ obtained for an arbitrary number of fictitious properties. Next, consider the same set of species as in the good classification and the actual properties. The similarity of degree r_{ij} is then computed with Eq. (1) giving the matrix $\bar{\bar{R}}$. The number of properties for $\bar{\bar{R}}$ and $\bar{\bar{S}}$ is:

$$D = -\sum_{ij} (1 - r_{ij}) \ln \frac{1 - r_{ij}}{1 - s_{ij}} - \sum_{ij} r_{ij} \ln \frac{r_{ij}}{s_{ij}} \quad \forall 0 \leqslant r_{ij}, s_{ij} \leqslant 1 \tag{5}$$

The definition was suggested by that introduced in information theory by Kullback (1959) to measure the distance between two probability distributions. Algorithm result is a set of weights allowing adequate classification. Such a procedure was applied in the synthesis of complex flow sheets using information entropy (Iordache et al., 1993).

2.1. Calculation outcomes

The matrix of Pearson correlation coefficients has been calculated between each pair of vector properties $\langle i_1, i_2, i_3, i_4, i_5, i_6, i_7 \rangle$ of the 74 flavonoids. The Pearson intercorrelations are computed for the partial correlation diagram, which contains high partial correlations ($r \geqslant 0.75$), medium partial correlations ($0.50 \leqslant r < 0.75$), low partial correlations ($0.25 \leqslant r < 0.50$) and *zero* partial correlations ($r < 0.25$). Pairs of flavonoid compounds with high partial