We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK
CITATION
INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Molecular Classification of Antitubulin Agents with Indole Ring Binding at Colchicine-Binding Site

Francisco Torrens and Gloria Castellano

Additional information is available at the end of the chapter

Abstract

Algorithms for classification and taxonomy are proposed based on criteria as *information entropy* and its production. A set of 59 antitubulin agents with trimethoxyphenyl (TMP), indole, and C=O bridge present inhibition of gastric cancer cell line MNK-45. On the basis of structure-activity relation of TMPs, derivatives are designed that are classified using seven structural parameters of different moieties. A lot of categorization methods are founded on the entropy of information. On using processes on collections of reasonable dimension, an extreme amount of outcomes occur, matching information and suffering a combinatorial increase. Notwithstanding, following the *equipartition conjecture*, an assortment factor appears among dissimilar alternatives resultant from categorization among pecking order rankings. The entropy of information allows classifying the compounds and agrees with principal component analyses. A table of periodic properties TMPs is obtained. Features denote positions R_{1-4} on the benzo and $X–R_{5/6}$ on the pyridine ring in indole cycle. Inhibitors in the same group are suggested to present similar properties; those in the same group and period will present maximum resemblance.

Keywords: periodic law, periodic property, periodic table, information entropy, equipartition conjecture, anticancer activity

1. Introduction

Experimentally, antitubulin analogues were synthesized/tested for antitubulin activity, revealing ligand-interaction principles with tubulin/related bioactivity [1–13]. Molecular modeling studies of antitubulin agents were performed to aid in the design of better antitubulin inhibitors [14–16]. In computer-aided drug design studies, comparative molecular field analysis

(CoMFA) combined with docking calculations was applied to protein-ligand-binding complexes [17–21]. A class of antitubulin agents, binding at colchicine (COL) site with an indole ring, was developed and underwent examinations for binding, antitubulin polymerization, and/or anticancer effects. The discovered properties are helpful for better-inhibitor design. Half inhibitory concentrations (IC_{50}) were collected for the inhibition of gastric cancer cell MKN-45, for 59 COL-like compounds with indole and trimethoxyphenyl (TMP) rings (**Figure 1**), which bind at COL site [22]. The IC_{50} were measured for 24 compounds and reviewed for others: 71 compounds were collected. Trial CoMFA calculations for all gave a low leave-one-out determination coefficient q^2~0.2. Examination of functional groups showed that three ones are much more bulky than the others. Functional groups of eight are much different from others. Compounds were excluded leaving 59 substances in CoMFA calculation. With data, three-dimensional (3D)-quantitative structure-activity relationship (SAR) (QSAR) examination was performed with CoMFA [23], combined with docking calculations for compounds to illustrate correlation of functional group variations with anticancer effect. An approach was employed to examine QSAR for a number of other protein-ligand-binding complexes. Functional-group substitutions locate at sites around indole ring, i.e., $R_{1–6}$ functional-group sites. Comparative QSAR modeling of 2-phenylindole-3-carbaldehyde derivatives was performed as potential antimitotic agents [24]. The KIT kinase mutants showed unique mechanisms of drug resistance to imatinib and sunitinib in gastrointestinal stromal tumor patients [25]. Gene expression profiling of gastric cancer was reported [26]. Natural product COL, obtained from *Colchicum autumnale*, is a bioactive alkaloid used in the treatment of a number of diseases [27]. It received considerable attention in the basic study of neoplasia by its capacity for interrupting *mitosis*, ending the process in metaphase [28]. The COL acts as an inhibitor of the polymerization of tubulin (a protein that contains eight Trp units) [29]. It was used as a probe to understand *microtubule* role in cells because of its big affinity to tubulin, in which structure presents a binding site (*colchicine domain*) [30, 31]. Tubulin is a target for cancer treatment: a number of drugs were developed to target it [32]. Binding with it, ligands interfere with its polymerization dynamics and exhibit an antitumor effect. In addition to developed drugs (*viz*. taxol, vibrestine), which bind with it at taxol/vibrestine-binding sites, COL presents a tubulin binding site and showed anticancer effects although with significant toxicity. Developing COL-like compounds with lesser toxicity represented an effort in finding better ligands to target tubulin at COL-binding site [33, 34]. A simple computerized algorithm useful for establishing a relation
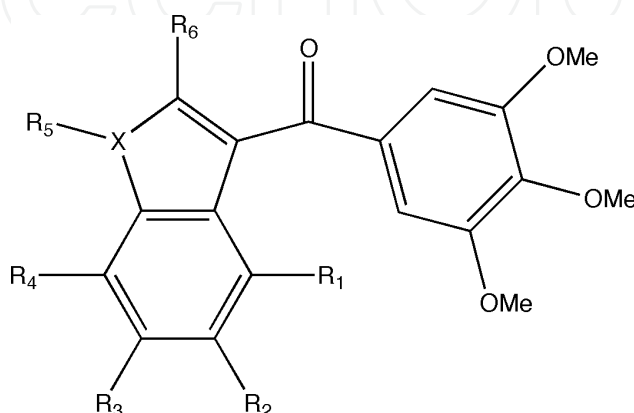

Figure 1. General structure motifs: Trimethoxyphenyl (TMP) ring/indole ring/C=O bridge.

between chemical structures [35, 36] was proposed. The preliminary idea results the entropy of information for configuration detection. The entropy of information results was expressed based on a *similarity matrix* among a pair of chemical entities. Because the entropy of information results feebly discerning for categorization reasons, the more influential concepts of *entropy production* and *equipartition conjecture* result were presented in [37]. In previous articles, the classifications by periodic properties of local anesthetics [38–40], inhibitors of human immuno-deficiency virus [41–43], and anticancer drugs [44, 45] were analyzed. The goal of the current account is expanding the promises of knowledge of the algorithm and, as compounds are unaffectedly explained by a changeable-dimension prearranged model, learning universal methods in the dispensation of prearranged information. Next goal presents a periodic classification of TMPs. A further objective is to perform a validation of the periodic table (PT) with an external property not used in the development of PT.

2. Computational method

The key problem in classification studies is to define *similarity indices* when several criteria of comparison are involved. The primary stage in counting resemblance for TMPs records the majority of the significant moieties. The *vector of properties* $\bar{i} = <i_1, i_2, \ldots i_k, \ldots>$ should be linked to each TMP i, whose parts match with dissimilar characteristic groups in the molecule, in a pecking order consistent the predictable significance of pharmacological potency. Whether moiety *m-th* results more important than portion *k-th* then $m < k$. The parts i_k are values "1" or "0", consistent if an alike portion of rank k is present in TMP i contrasted to the recommendation one. The examination comprises two regions of structure variation in TMP molecules: positions R_{1-4} on the benzo and locations X and $R_{5/6}$ on the pyridine ring in the indole cycle. The TMPs are inhibitory to gastric cancer cell line MKN-45. The *structural elements* of a TMP molecule can be *ranked* according to their contribution to MKN-45 inhibition in the order: $R_1 > R_4 > R_2 > X > R_5 > R_3 > R_6$. Index $i_1 = 1$ denotes $R_1 = H$ ($i_1 = 0$, otherwise), $i_2 = 1$ means $R_4 = H$, $i_3 = 1$ signifies $R_2 = H$, $i_4 = 1$ stands for $X = N$, $i_5 = 1$ indicates $R_5 = H$, $i_6 = 1$ represents $R_3 = OMe$, and $i_7 = 1$ implies $R_6 = CH_2–OH$. In TMP 42, $R_1 = R_4 = R_2 = R_5 = H$, $X = N$, $R_3 = OMe$ and $R_6 = CH_2–OH$; obviously its associated vector is <1,111,111>. The TMP 42 was selected as *reference* because of its greatest MNK-45 inhibition. Vectors were associated with 59 TMPs with gastric anticancer activities. Vector <1,111,110> is associated with TMP 1 since $R_1 = R_4 = R_2 = R_5 = R_6 = H$, $X = N$ and $R_3 = OMe$. Mean by r_{ij} ($0 \leq r_{ij} \leq 1$) the similarity index of a pair of TMPs linked to vectors $\bar{i}$ and $\bar{j}$, in that order. The relationship of similarity results is typified by a *similarity matrix* $\mathbf{R} = [r_{ij}]$. The similarity index among a pair of TMPs $\bar{i} = <i_1, i_2, \ldots i_k \ldots>$ and $\bar{j} = <j_1, j_2, \ldots j_k \ldots>$ is described by:

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

where $0 \leq a_k \leq 1$ and $t_k = 1$ whether $i_k = j_k$ except $t_k = 0$ whether $i_k \neq j_k$. The definition allocates a weight $(a_k)^k$ to whichever feature concerned about the explanation of molecule i or j. The MNK-45 gastric cancer inhibition data reported by Lin *et al.* were used for the present classification

study. The *grouping algorithm* applies the *stabilized* similarity matrix obtained *via* the *max-min composition rule o* as described by:

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

where $\mathbf{R} = [r_{ij}]$ and $\mathbf{S} = [s_{ij}]$ that result in matrices of the same kind and $(\mathbf{R}o\mathbf{S})_{ij}$, entry (i,j)-*th* of matrix $\mathbf{R}o\mathbf{S}$ [46–49]. On using composition rule max-min iteratively, so that $\mathbf{R}(n+1) = \mathbf{R}(n) \, o \, \mathbf{R}$, an integer n results fulfilling: $\mathbf{R}(n) = \mathbf{R}(n+1) = \ldots$ The resultant matrix $\mathbf{R}(n)$ is named as *stabilized similarity matrix*. The importance of stabilization stretches out in the categorization procedure, and stabilization generates a separation in displaced divisions. From the present on, it results implicitly that the stabilized similarity matrix is applied and named as $\mathbf{R}(n) = [r_{ij}(n)]$. The *grouping rule* is as follows: i and j results allocated in the same division whether $r_{ij}(n) \geq b$. The grouping of i ($\hat{i}$) results in the collection of molecules j that fulfills the grouping rule $r_{ij}(n) \geq b$. The matrix of clusters results in

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

where s means whichever indicator of a molecule fitting in class $\hat{i}$ (likewise for t and $\hat{j}$). Rule (3) denotes discovering the main similarity index among molecules of a pair of divisions. In information theory, *information entropy h* measures the surprise that the source emitting the sequences can give [50, 51]. We consider the utilization of a qualitative mark assay to decide the attendance of Fe in a sample of water. With no sample in the past, the analyst has to start with the pair of results supposing: 0 (Fe not present) and 1 (Fe there), which are equiprobable with likelihood 1/2. As up to a pair of elements are there in the sample solution (e.g., Fe, Ni or both), there are four achievable results neither from (0, 0) to the two being there (1, 1) *via* on a par likelihood $1/2^2$. Which of the four options goes is decided by a pair of assays, each one with a pair of clear conditions. Likewise, with three metals, there are eight options, every one with a likelihood $1/2^3$: three assays are necessary. The following configuration clearly connects uncertainty to information necessary to solve it. The amount of options results stated to the power of 2. The power to which 2 is lifted to provide the amount of occurrences N results in the logarithm to base 2 of that amount. Both information and uncertainty are described in terms of the logarithm to base 2 of the amount of achievable analytical results: $\log_2 N$. The initial uncertainty is defined in terms of the probability of the occurrence of every outcome; e.g., the definition is as follows: $I = H = \log_2 N = \log_2 1/p = -\log_2 p$, where I denotes the information held in the reply provided that there were N options, H, the first uncertainty coming from the necessity of taking into account the N options and p, the likelihood of each result whether or not all N occurrences are evenly probable to occur. The equation can be extended to the case in which the likelihood of each result does not result the same; whether it is identified from historical experiment is proven by some metals that result in more probability than other ones, the expression results are corrected so that the logarithms of the particular likelihood appropriately weighted result in: $H = -\Sigma \, p_i \log_2 p_i$, where: $\Sigma \, p_i = 1$. Take into account the first case but at present, historical experiment proved that 90% of the samples had no Fe. The amount of uncertainty results is computed as: $H = -(0.9 \log_2 0.9 + 0.1 \log_2 0.1) = 0.469$ bits. For a particular

case happening with probability p, the amount of astonishment results is proportional to $-\ln p$. Extending the outcome to a random variable X (that is able to present N achievable values x_1, ..., x_N with probabilities p_1, ..., p_N), the astonishing mean is obtained when finding out the value of X results $-\Sigma\, p_i \ln p_i$. The entropy of information is linked to similarity matrix $\mathbf{R}$ results:

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

Mean is obtained by C_b, the collection of divisions and $\widehat{\mathbf{R}}_b$, the similarity matrix at the classification level b. The entropy of information fulfills the following features. (1) $h(\mathbf{R}) = 0$ whether $r_{ij} = 0$ or $r_{ij} = 1$. (2) $h(\mathbf{R})$ results maximum whether $r_{ij} = 0.5$, i.e., as the ambiguity is maximum. (3) $h\left(\widehat{\mathbf{R}}_b\right) \leq h(\mathbf{R})$ for whichever b, i.e., categorization directs to a deficit of entropy. (4) $h\left(\widehat{\mathbf{R}}_{b_1}\right) \leq h\left(\widehat{\mathbf{R}}_{b_2}\right)$ if $b_1 < b_2$, i.e., entropy is a monotone function of grouping level b. In the categorization procedure, each *hierarchical tree* matches to a reliance of the entropy of information on the classification level, and a plot h–b is obtained. The *equipartition conjecture of entropy production* of Tondeur and Kvaalen results is suggested as an assortment principle, between dissimilar alternatives coming from categorization between pecking order rankings. Consistent with the conjecture, for a provided custody, the top arrangement of a dendrogram results in which the production of entropy results is mainly dispersed regularly, i.e., neighboring a type of equipartition. It is gone on at this point similarly *via information entropy* in its place of thermodynamic entropy. Equipartition entails a linear relationship, i.e., a steady production of entropy of information all along the extent of b, so that the *equipartition line* results are explained by:

$$h_{\mathrm{eqp}} = h_{\max} b \tag{5}$$

As the categorization results are disconnected, a mean of stating equipartition is a usual staircase function. The most excellent alternative results decided the one minimizing the addition of the square differences:

$$SS = \sum_{b_i}\left(h - h_{\mathrm{eqp}}\right)^2 \tag{6}$$

Learning procedures alike the ones met in *stochastic methods* are the results as applied in [52]. Taking into account a provided classification as *good* or perfect from practice or experience, which matches to a *reference* similarity matrix $\mathbf{S} = [s_{ij}]$ obtained for equivalent weights $a_1 = a_2 = \ldots = a$ and any amount of fabricated features. Then, take into account identical collection of molecules as in the good categorization and the real features. The similarity index r_{ij} results calculated with Eq. (1) provided matrix $\mathbf{R}$. The amount of features for $\mathbf{R}$ and $\mathbf{S}$ can vary. The learning process lies in attempting to get categorization outcomes for $\mathbf{R}$ as near as

likely to the *good* categorization. The primary weight a_1 results obtained constant and just the next weights $a_2, a_3, \ldots$ result exposed to random changes. A novel similarity matrix results *via* Eq. (1) and the novel weights. The distance among the classifications typified by **R** and **S** results is provided by:

$$D = -\sum_{ij}\left(1 - r_{ij}\right)\ln\frac{1 - r_{ij}}{1 - s_{ij}} - \sum_{ij}r_{ij}\ln\frac{r_{ij}}{s_{ij}} \quad \forall 0 \leq r_{ij}, s_{ij} \leq 1 \tag{7}$$

The definition was suggested by Kullback to measure the distance between two probability distributions, which is an amount of the distance among matrices **R** and **S** [53]. As for each matrix a matching categorization exists, the pair of categorizations result contrasted by distance, which results a non-negative amount that approximates zero as the similarity among **R** and **S** rises. The outcome of the procedure results a collection of weights permitting proper categorization. The algorithm was utilized in the production of complicated dendrograms *via* the entropy of information [54]. Our program MolClas is an easy, dependable, effective, and quick process for molecular categorization, founded on the conjecture of the equipartition of the production of the entropy of information consistent with Eqs. (1)–(7). It reads the amount of features and molecular indices. It permits the optimization of the coefficients. It not obligatorily reads the initial coefficients and the amount of iteration cycles. The correlation matrix results are computed by the algorithm or read from input. Code MolClas permits the alteration of the correlation matrix from [−1, 1] to [0, 1]. The program computes the similarity matrix of the features in symmetric storage mode, computes categorizations, checks whether categorizations result is dissimilar, computes distances among categorizations, computes the similarity matrices of categorizations, works out the entropy of information of categorizations, optimizes coefficients, carries out single/complete-linkage hierarchical cluster analyses, and charts classification plots. It was written not only to analyze the equipartition conjecture of entropy production but also to explore the world of molecular classification. Code MolClas is different from other program MolClass as referred in the literature [55]. While MolClas classifies molecules based on hierarchical dichotomic (Boolean) descriptors, MolClass discovers SARs from molecular patterns (*fingerprints*) extracted from experimental datasets and needs to interrogate big databases (PubChem, ChEMBL, ChemBank). Code MolClas is available at Internet (torrens@uv. es) and is free for academic use.

3. Calculation results and discussion

Matrix of Pearson correlation coefficients results computed among couples of vector properties $<i_1, i_2, i_3, i_4, i_5, i_6, i_7>$ for 59 TMPs. Pearson correlations result displayed in the partial correlation diagram, which encloses high ($r \geq 0.75$), medium ($0.50 \leq r < 0.75$), low ($0.25 \leq r < 0.50$), and *zero* ($r < 0.25$) partial correlations. Couples of inhibitors with superior partial associations present a vector property alike. Notwithstanding, the outcomes have to be gotten with concern since the TMP with steady vector <1,111,111> (Entry 42) presents zero standard deviation, producing

maximum partial correlation $r = 1$ with whichever TMP, which results an artifact. After the conjecture of equipartition, the intercorrelations are illustrated in the partial correlation diagram, which contains 1382 high (**Figure 2**, *red lines*), 109 medium (*orange*), 161 low (*yellow*), and 59 *zero* (*black*) partial correlations. Six out of 58 high partial correlations of Entry 42 were corrected; e.g., its correlations with Entries 3 and 47 are medium, its correlations with Entries 12, 15, and 43 are low, and its correlation with Entry 46 is *zero* partial correlation.

The grouping rule in the case with equal weights $a_k = 0.5$ for $b_1 = 0.97$ allows the classes:

C–b_1 = (1,5–8,10,11,13,16,17,26–28,41,42,44,45,48,58,59),(2,4,9,18,19,49),(3),(12),

(14,20–25,29–33,35,50–55),(15,43),(34,36–40,56,57),(46),(47)

The nine groupings are obtained with associated entropy h–**R**–b_1 = 39.44. The *dendrogram* (binary tree) matching with <$i_1,i_2,i_3,i_4,i_5,i_6,i_7$ > and C–b_1 is calculated [56–58]; it provides a binary taxonomy that separates the same nine classes: from top to bottom, the data bifurcate into groupings 3, 4, 8, 9, 1, 2, 5, 6, and 7 with 1, 1, 1, 1, 20, 6, 19, 2, and 8 TMPs, respectively [59]. The TMPs 42, 26, etc. with the greatest inhibitory activity are grouped into the same class.
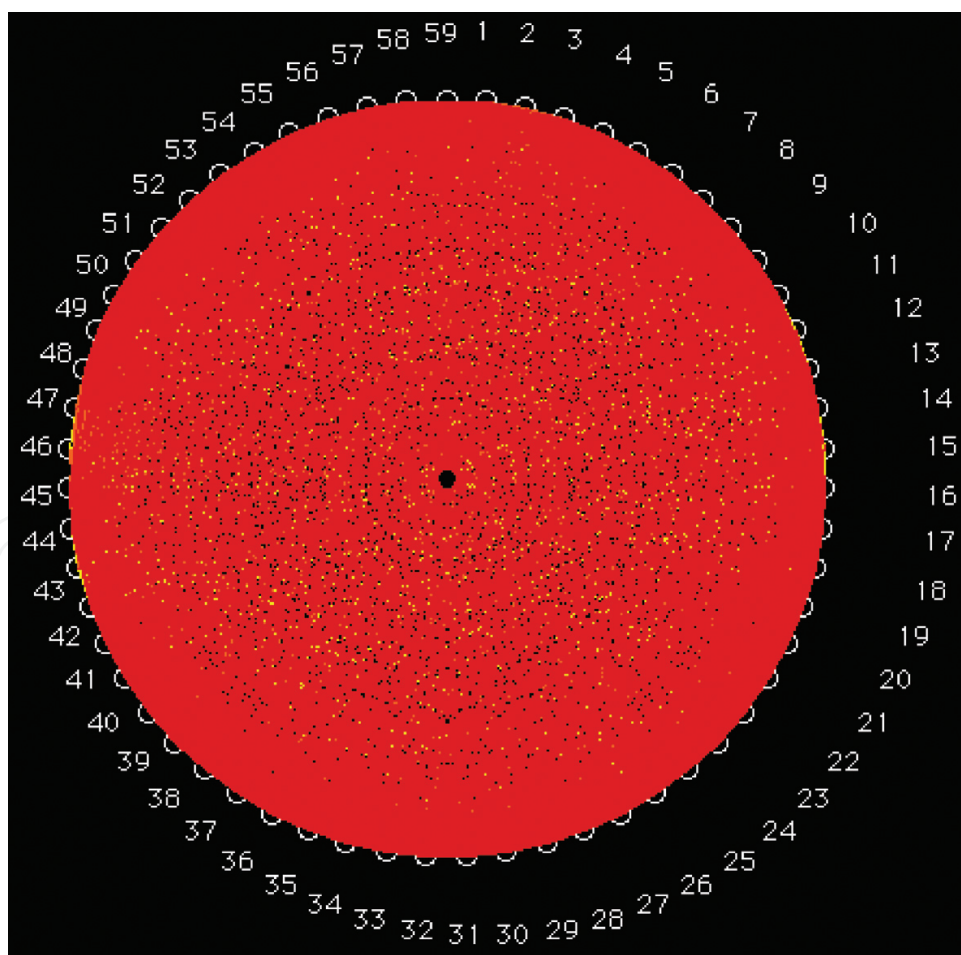

Figure 2. Partial correlation diagram: High (*red*), medium (*orange*), and low (*yellow*) correlations.

The TMPs in the same grouping appear highly correlated in the partial correlation diagram. At level b_2 with $b_2 = 0.86$, the set of classes results in:

C–b_2 = (1,4–8,10,11,13,14,16–42,44,45,48–59),(2,9),(3,47),(12,15),(43),(46).

Six classes result and entropy decays to h–**R**–b_2 = 16.18. Dendrogram matching to <$i_1,i_2,i_3,i_4,i_5,i_6,i_7$> and C–$b_2$ divides the same six classes: from top to bottom data bifurcate into classes 5, 6, 1, 2, 3, and 4 with 1, 1, 51, 2, 2, and 2 TMPs, respectively. Again, TMPs with the greatest inhibitory potency belong to the same class. The TMPs in the same class appear highly correlated in the partial correlation diagram and dendrogram. An analysis of set containing 1–59 classes was performed, in agreement with partial correlation diagram and dendrograms. In view of partial correlation diagram and dendrograms, we split data into seven classes: (1,26–28,41,42,45,58,59), (5–8,10,11,13,16,17,44,48), (14,20–25,29–33,35,50–55), (34,36–40,56,57), (2,4,9,18,19,49), (3,47), and (12,15,43,46). **Figure 3** displays corresponding tree. Again, TMPs with the greatest activity correspond to the same class.

The illustration of the classification above in a radial tree (**Figure 4**) shows the same classes, in qualitative agreement with the partial correlation diagram and dendrograms. Once more, TMPs with the greatest potency are included in the same grouping.

Program SplitsTree analyzes cluster analysis (CA) data [60]. Based on *split decomposition*, it takes a *distance matrix* and produces a graph that represents the relations between taxa. For ideal data, graph is a tree, whereas less ideal data cause a tree-like network, which is interpreted as possible evidence for different and conflicting data. As split decomposition does not attempt to force data on to a tree, it gives a good indication of how *tree*-like are given data. Splits graph for 59 TMPs in (**Figure 5**) shows that most TMP groups collapse: (1,2,4–11,13,16–19,26–28,41,42,44,45,48,49,58,59), (3,47), (12,15,43), (14,20–25,29–33,35,50–55), and (34,36–40,56,57); classes 1, 2, and 5 coincide. No conflicting relation appears between TMPs. Splits graph is in partial agreement with partial correlation diagram, dendrograms, and radial tree.

Usually in quantitative structure-property relationships (QSPRs), the information archive encloses fewer than 100 molecules and thousands of *X*-variables. There are a lot of *X*-variables that nobody is able to find out by *inspection* configurations, tendencies, groupings, etc. in the molecules. *Principal component analysis* (PCA) results a method helpful to *summarize* the knowledge enclosed in the **X**-matrix and place it comprehensible [61–66]. The PCA acts by decomposing the **X**-matrix as the product of two matrices **P** and **T**. The *loading matrix* (**P**), with knowledge concerning the variables, encloses some vectors [*principal components* (PCs)], in which results are obtained as linear combinations of the first *X*-variables. In the *score matrix* (**T**), with knowledge about the molecules, each molecule result is expressed by projections on to PCs instead of original variables: **X** = **TP′** + **E**. Knowledge not enclosed in the matrices stays as *unexplained* X-*variance* in a *residual matrix* (**E**). Each PC$_i$ results a novel coordinate stated as a linear combination of the first characteristics x$_j$: PC$_i$ = $\Sigma_j b_{ij} x_j$. The novel coordinates PC$_i$ result *scores* or *factors* whereas the coefficients b_{ij} result the *loadings*. The scores are sorted consistently with the knowledge regarding the entire variability between molecules. The *score-score plots* present the places of the molecules in the novel coordinate scheme, whereas the *loading-loading*
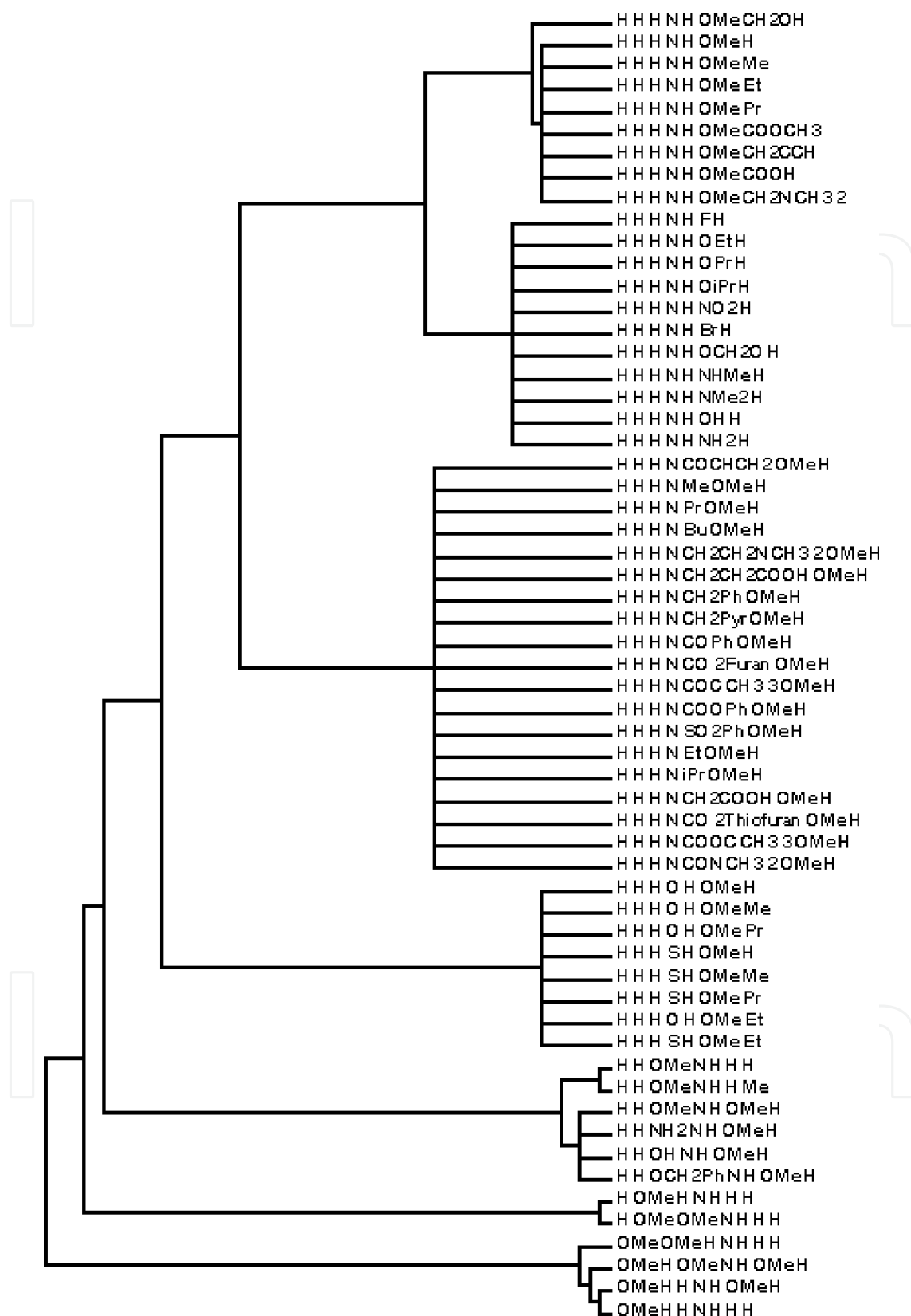
Figure 3. Dendrogram of TMP ring/indole ring/C=O bridge as MKN-45 inhibitors.
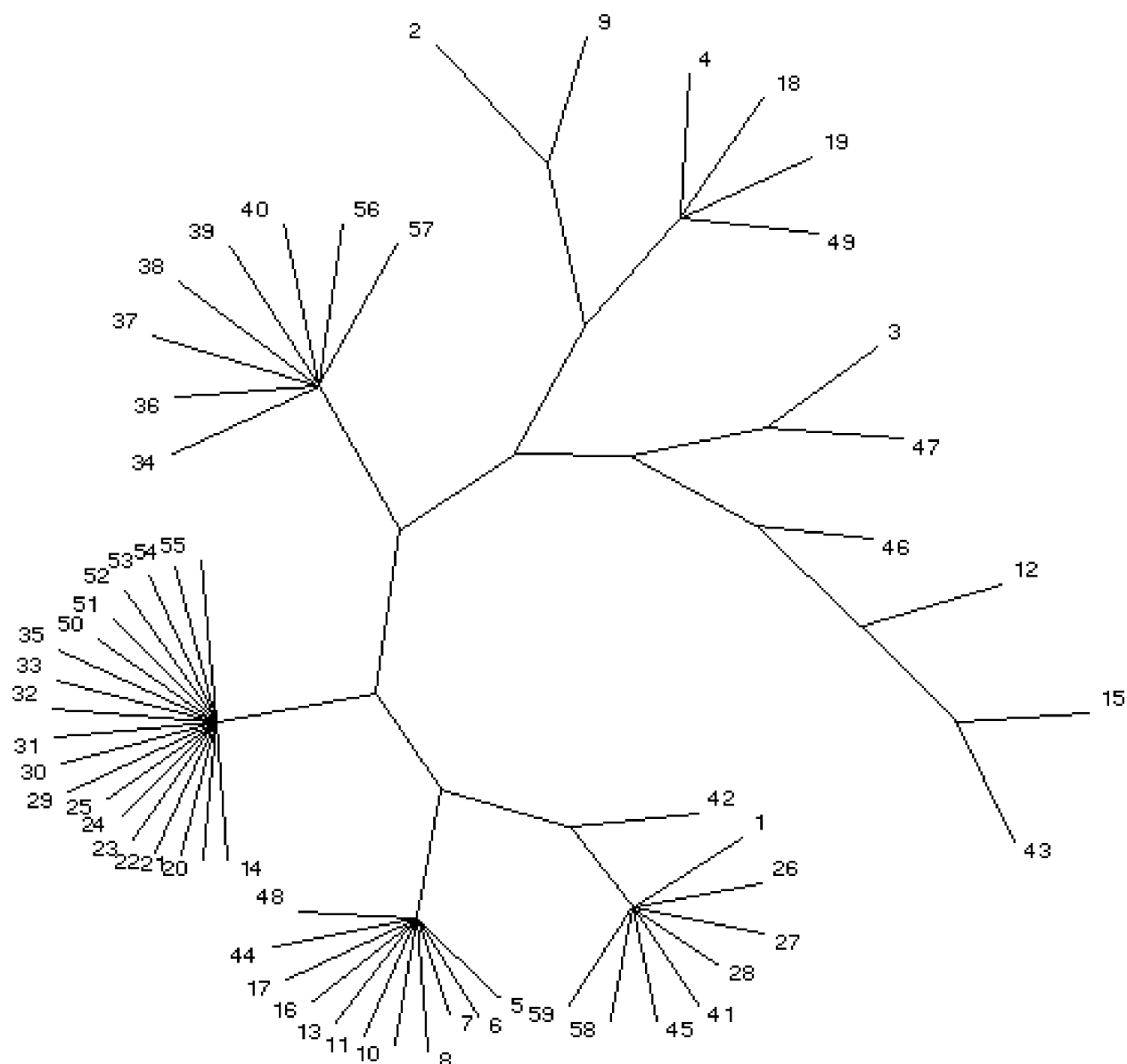
Figure 4. Radial tree of TMP ring/indole ring/C=O bridge as MKN-45 inhibitors.

plots display the position of the properties that correspond to the molecules in the novel coordinate scheme. The PCs show a pair of features. (1) The PCs result taken out in decreasing sequence of significance: the first PC encloses more knowledge than the second one, the second more than the third one, and so on. (2) Each PC results orthogonal to each other: no correlation exists between information contained in different PCs. A PCA was performed for TMPs. The importance of PCA factors F_{1-7} for $\{i_1,i_2,i_3,i_4,i_5,i_6,i_7\}$ was calculated. In particular, the use of the first factor F_1 explains 27% of the variability of data (73% error), the combined application of the first two factors $F_{1/2}$ accounts for 45% of variance (55% error), the utilization of the first three factors F_{1-3} justifies 60% of variability (40% error), etc. Factor loadings of PCA were computed. Profile of PCA F_1–F_2 for vector property was calculated. For F_1, variable i_6 shows the maximum weight in the profile; notwithstanding, F_1 is not able to be downgraded to two variables $\{i_5,i_6\}$ devoid of a 48% error. For F_2, variable i_4 presents the maximum weight and F_2
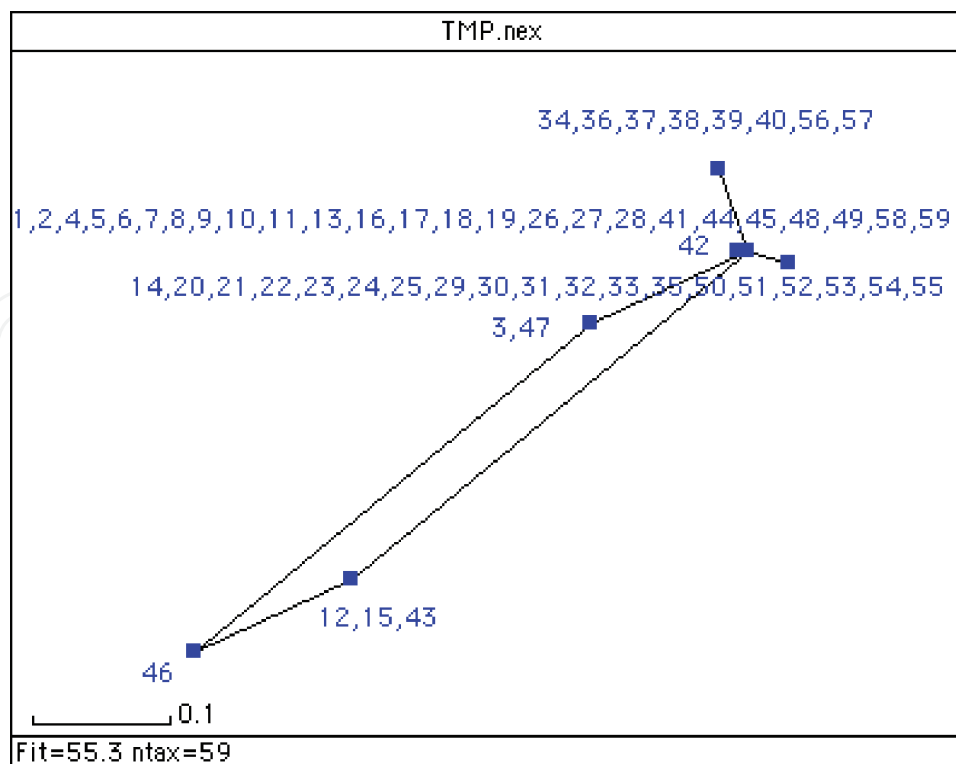
Figure 5. Splits graph of TMP ring/indole ring/C=O bridge as MKN-45 inhibitors.

is able to be downgraded to two variables $\{i_4,i_5\}$ with a 5% error. For F_3, variable i_7 assigns the maximum weight and F_3 is able to be downgraded to two variables $\{i_4,i_7\}$ with a 3% error. For F_4, variable i_3 consigns the maximum weight; however, F_4 is not able to be downgraded to two variables $\{i_2,i_3\}$ devoid of a 15% error. For F_5, variable i_1 represents the maximum weight and F_5 is able to be downgraded to two variables $\{i_1,i_6\}$ with a 6% error. For F_6, variable i_2 explains the maximum weight; notwithstanding, F_6 is not able to be downgraded to two variables $\{i_1,i_2\}$ devoid of a 25% error. For F_7, variable i_5 accounts for the maximum weight; nevertheless, F_7 is not able to be downgraded to two variables $\{i_5,i_6\}$ devoid of a 36% error. In PCA F_2–F_1 scores plot (**Figure 6**), TMPs with the same vector property collapse: (1,26–28,41,45,58,59), (2,9), (4,18,19,49) (5–8,10,11,13,16,17,44,48), (14,20–25,29–33,35,50–55) and (34,36–40,56,57). Seven TMP classes are clearly distinguished: class 1 with 9 compounds ($0 < F_1 < F_2$, *right*), class 2 with 11 substances ($F_1 < F_2 \approx 0$, *middle*), class 3 with 19 molecules ($F_1 >> F_2$, *bottom right*), class 4 with 8 organics ($0 < F_1 << F_2$, *top*), class 5 (6 units, $F_1 < F_2 \approx 0$, *middle*), class 6 (2 units, $F_1 << F_2 < 0$, *left*) and class 7 (4 units, $F_1 < F_2 < 0$, *bottom*). The classification is in agreement with partial correlation diagram, dendrograms, radial tree, and splits graph.

From PCA factor loadings of TMPs, F_2–F_1 loadings plot (**Figure 7**) depicts the seven properties. In addition, as a complement to the scores plot for the loadings, it is confirmed that TMPs in class 1, located in the right side, present a contribution of R_3 = OMe situated in the same side. The TMPs in class 3 in the bottom have more pronounced contribution of X = N in the same location. Two classes of properties are clearly distinguished in the loadings plot: class 1 {R_1,R_4, R_2,R_3} ($F_1 > F_2 > 0$, *right*) and class 2 {X,R_5,R_6} ($F_1 < F_2$, *left*).
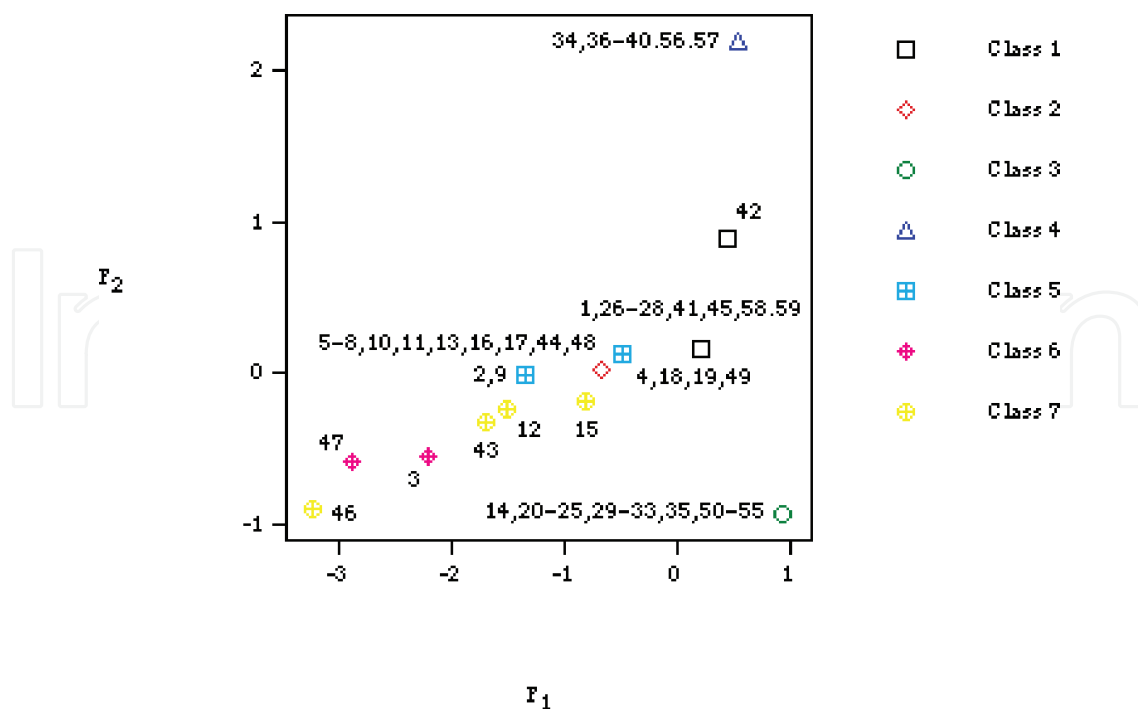
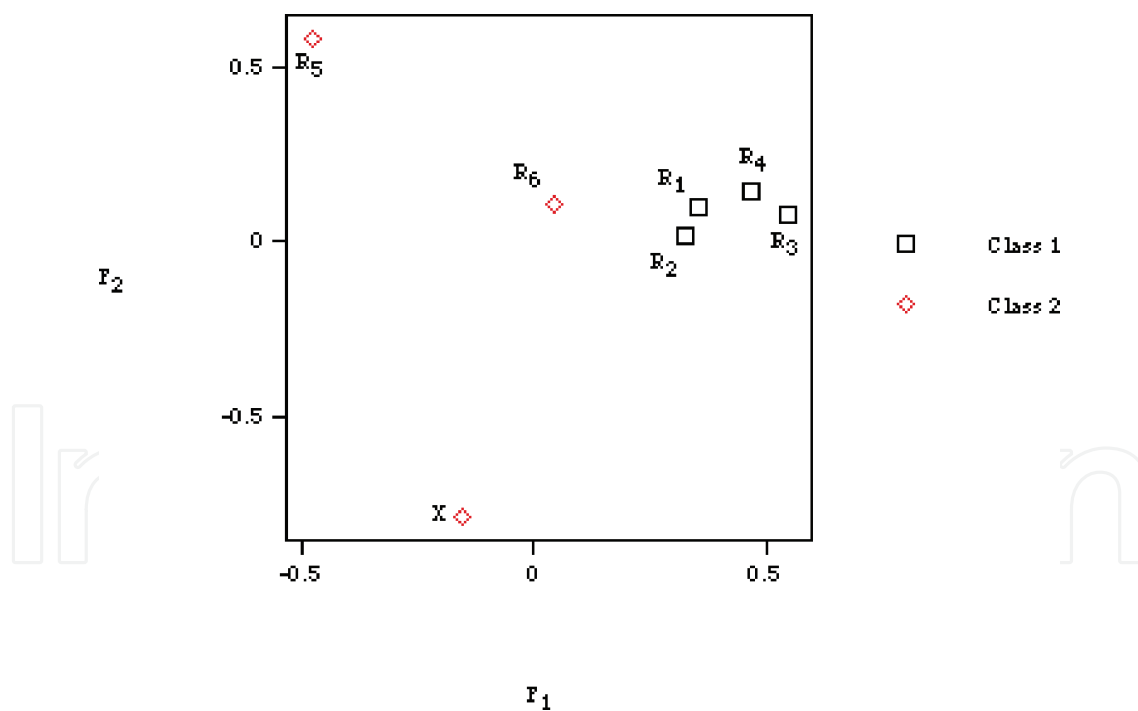**Figure 6.** Principal component analysis F_2–F_1 scores plot for TMP ring/indole ring/C=O bridge.

Figure 7. PCA F_2–F_1 loadings plot for TMP ring/indole ring/C=O bridge.

Instead of 59 TMPs in the $\mathfrak{R}^7$ space of seven vector properties, we consider seven properties in the $\mathfrak{R}^{59}$ space of 59 TMPs. The dendrogram for vector properties separates properties {R_1, R_4,R_2,R_3} (class 1) from {X,R_5,R_6} (class 2), in agreement with PCA loadings plot. The splits graph for properties indicates no conflicting relation between vector components, separating

properties $\{R_1, R_4, R_2, R_3\}$ (class 1) from $\{X, R_5, R_6\}$ (class 2), in agreement with PCA loadings plot and dendrogram. A PCA was performed for the vector properties. The use of only the first factor F_1 explains 51% of variance (49% error), the combined application of the first two factors $F_{1/2}$ accounts for 71% of variability (29% error), the utilization of the first three factors F_{1-3} rationalizes 82% of variance (18% error), etc. In the PCA F_2–F_1 scores plot, property R_4 appears superimposed on R_1. Two groupings of properties are distinguished: class 1 $\{R_1, R_4, R_2, R_3\}$ ($F_1 > F_2$, *right*) and class 2 $\{X, R_5, R_6\}$ ($F_1 < F_2$, *left*), in agreement with PCA loadings plot, dendrogram and splits graph. Format for PT of TMPs (**Table 1**) indicates that TMPs are categorized first by i_1, then i_2, i_3, i_4, i_5, i_6, and i_7. Vertical groups result described by $\{i_1, i_2, i_3, i_4\}$ and horizontal periods, by $\{i_5, i_6, i_7\}$. Periods of eight elements are considered; e.g., group g0011 denotes $<i_1, i_2, i_3, i_4> = <0011>$: $<0011100>$ ($R_1 \neq H$, $R_4 \neq H$, $R_2 = H$, $X = N$, $R_5 = H$, $R_3 \neq OMe$, $R_6 \neq CH_2$–OH), etc. The TMPs in the same column appear close in partial correlation diagram, dendrograms, radial tree, splits graph, and PCA scores.

The change of property P (inhibition of gastric cancer cell MKN-45) of vector $<i_1, i_2, i_3, i_4, i_5, i_6, i_7>$ is expressed in the decimal system $P = 10^6 i_1 + 10^5 i_2 + 10^4 i_3 + 10^3 i_4 + 10^2 i_5 + 10 i_6 + i_7$ *vs.* structural parameters $\{i_1, i_2, i_3, i_4, i_5, i_6, i_7\}$, for TMPs. The property was not used in the development of PT and serves to validate it. Most points appear superimposed, and lines $i_{2/6}$ on i_1 and i_7 on i_4. Results show the order of importance of parameters: $i_1 > i_2 > i_3 > i_4 > i_5 > i_6 > i_7$, in agreement with PT of properties with vertical groups defined by $\{i_1, i_2, i_3, i_4\}$ and horizontal periods by $\{i_5, i_6, i_7\}$. The variation property P of vector $<i_1, i_2, i_3, i_4, i_5, i_6, i_7>$ in base 10 *vs.* the number of group in PT, for TMPs, reveals minima corresponding to compounds with $<i_1, i_2, i_3, i_4> ca. <0011>$ (group g0011) and maxima *ca.* $<1111>$ (group g1111). Periods p010, p100, p110, and p111 represent rows 1–4, respectively. For groups 3 and 6, period p110 is superimposed on p100, and for group 8, all periods coincide. The corresponding function $P(i_1, i_2, i_3, i_4, i_5, i_6, i_7)$ indicates a series of cyclic *waves* obviously controlled by minima or maxima, which propose a periodic performance that evokes the shape of a trigonometric function. For $<i_1, i_2, i_3, i_4, i_5, i_6, i_7>$, maximum results are obviously presented. The space in $<i_1, i_2, i_3, i_4, i_5, i_6, i_7>$ elements among every couple of successive maxima is eight, which agrees with TMP collections in consecutive periods. The maxima are in similar locations in the curve and are in phase. The typical points in phase have to match with the components in similar group in PT. For maxima $<i_1, i_2, i_3, i_4, i_5, i_6, i_7>$, there is consistency among the two descriptions; notwithstanding, the constancy is not universal. The assessment of the waves presents a pair of dissimilarities: (1) periods are incomplete and (2) periods 2 and 3 are somewhat staircase like. The most characteristic points of the plot are maxima that lie about group g1111. The values of $<i_1, i_2, i_3, i_4, i_5, i_6, i_7>$ are repeated as the periodic law (PL) states. An empirical function $P(p)$ reproduces different $<i_1, i_2, i_3, i_4, i_5, i_6, i_7>$ values; a minimum of $P(p)$ presents significance just if it is contrasted with the previous $P(p-1)$ and afterward $P(p + 1)$ points, necessitating to satisfy:

$$P_{\min}(p) < P(p - 1)$$
$$P_{\min}(p) < P(p + 1)$$

(8)

Sequenced relationship (8) has to be done again at determined gaps peer to the dimension of the period and is equal to:

g0011	g0101	g0111	g1001
–OMe –OMe –H –N –H –H –H		–OMe –H –H –N –H –H –H	–H –OMe –OMe –N –H –H –H
	–OMe –H –OMe –N –H –OMe –H	–OMe –H –H –N –H –OMe –H	

g1011	g1101	g1110	g1111
			–H –H –H –N –CO–CH=CH$_2$ –OMe –H
			–H –H –H –N –Me –OMe –H
			–H –H –H –N –Pr –OMe –H
			–H –H –H –N –Bu –OMe –H
			–H –H –H –N –CH$_2$–CH$_2$–N(CH$_3$)$_2$ –OMe –H
			–H –H –H –N –CH$_2$–CH$_2$–CO–OH –OMe –H
			–H –H –H –N –CH$_2$–Ph –OMe –H
			–H –H –H –N –CH$_2$–Pyr –OMe –H
			–H –H –H –N –CO–Ph –OMe –H
			–H –H –H –N –CO–2-Furan –OMe –H
			–H –H –H –N –CO–C(CH$_3$)$_3$ –OMe –H
			–H –H –H –N –CO–O–Ph –OMe –H
			–H –H –H –N –SO$_2$–Ph –OMe –H
			–H –H –H –N –Et –OMe –H
			–H –H –H –N –i-Pr –OMe –H
			–H –H –H –N –CH$_2$–CO–OH –OMe –H
			–H –H –H –N –CO–2-Thiofuran –OMe –H
			–H –H –H –N –CO–O–C(CH$_3$)$_3$ –OMe –H
			–H –H –H –N –CO–N(CH$_3$)$_2$ –OMe –H
–H –OMe –H –N –H –H –H	–H –H –OMe –N –H –H –H		–H –H –H –N –H –F –H
	–H –H –OMe –N –H –H –Me		–H –H –H –N –H –OEt –H
			–H –H –H –N –H –OPr –H
			–H –H –H –N –H –O-i-Pr –H
			–H –H –H –N –H –NO$_2$ –H
			–H –H –H –N –H –Br –H
			–H –H –H –N –H –O–CH$_2$–O– –H
			–H –H –H –N –H –NHMe –H
			–H –H –H –N –H –N(Me)$_2$ –H
			–H –H –H –N –H –OH –H
			–H –H –H –N –H –NH$_2$ –H
	–H –H –OMe –N –H –OMe –H	–H –H –H –O –H –OMe –H	–H –H –H –N –H –OMe –H
	–H –H –NH$_2$ –N –H –OMe –H	–H –H –H –O –H –OMe –Me	–H –H –H –N –H –OMe –Me
	–H –H –OH –N –H –OMe –H	–H –H –H –O –H –OMe –Pr	–H –H –H –N –H –OMe –Et
	–H –H –O–CH$_2$–Ph –N –H –OMe –H		–H –H –H –N –H –OMe –Pr
			–H –H –H –N –H –OMe –CO–O–CH$_3$
			–H –H –H –N –H –OMe

g1011	g1101	g1110	g1111
		–H –H –H –S –H –OMe –H	–CH$_2$–C≡CH
		–H –H –H –S –H –OMe –Me	–H –H –H –N –H –OMe –CO–OH
		–H –H –H –S –H –OMe –Pr	–H –H –H –N –H –OMe
		–H –H –H –O –H –OMe –Et	–CH$_2$–N(CH$_3$)$_2$
		–H –H –H –S –H –OMe –Et	
			–H –H –H –N –H –OMe –CH$_2$–OH

Table 1. Periodic properties for 2-phenylindole-3-carbaldehyde derivatives.

$$P_{\min}(p) - P(p - 1) < 0$$
$$P(p + 1) - P_{\min}(p) > 0 \tag{9}$$

Because relationship (9) is just suitable for minima, additional universal others are wanted for all positions p; $D(p) = P(p + 1) - P(p)$ differences are computed by allocating each value to TMP p:

$$D(p) = P(p + 1) - P(p) \tag{10}$$

In the place of $D(p)$, the values of $R(p) = P(p + 1)/P(p)$ are obtained by assigning $R(p)$ to TMP p; whether PL is universal, components in similar group in equivalent locations in dissimilar periodic waves assure:

either

$$D(p) > 0 \text{ or } D(p) < 0 \tag{11}$$

either

$$R(p) > 1 \text{ or } R(p) < 1 \tag{12}$$

Notwithstanding, the outcomes demonstrate that this is not the case, so PL is not universal but with anomalies. The change of $D(p)$ *vs.* group number shows that for group 6, periods p100 and p110 collapse. It introduces lack of consistency among $<i_1, i_2, i_3, i_4, i_5, i_6, i_7>$ Cartesian and PT charts. Whether constancy were exact, every position in each period present similar sign: in general, a tendency exists in the positions to provide $D(p) > 0$ for the lower groups but not for group 8; however, the latter results should be taken with care because $D(p)$ are calculated using data from the next period. In detail, irregularities exist in which TMPs for successive periods are not always in phase. The change of $R(p)$ *vs.* group number shows that for groups 3 and 6, periods p100 and p110 collapse, and, for group 8, all periods coincide, confirming the lack of steadiness among

Cartesian and PT representations. Whether control were precise or not, every position in every period presents $R(p)$ either smaller or larger than one. A tendency exists in the positions to provide $R(p) > 1$ for the lower groups but not for group 8; however, the latter should be taken with care because $R(p)$ are calculated from the next period. Confirmed incongruities exist in which TMPs for successive periods are not always in phase.

4. Conclusion

1. Several criteria were selected to reduce analysis to manage quantity of trimethoxyphenyl, indole, carbonyl bridge antitubulins referred to structural parameters related to positions R_{1-4} on benzo, $R_{5/6}$ on pyridine, and heteroatom X in indole. Molecular *structural elements* were *ranked* according to inhibitory activity: $R_1 > R_4 > R_2 > X > R_5 > R_3 > R_6$. In compound 42, $R_1 = R_4 = R_2 = R_5 = H$, $X = N$, $R_3 = OMe$ and $R_6 = CH_3$–OH <1,111,111>, which was selected as *reference*. Many classification algorithms are based on *information entropy*. For moderate-sized sets, an excessive number of results appear compatible with data and suffer a combinatorial explosion; however, after the *equipartition conjecture*, one has a selection criterion, according to which the best configuration is that in which entropy production is most uniformly distributed. Method avoids the problem of continuum variables because for compound with constant <1,111,111> vector, null standard deviation causes Pearson correlation coefficient of one. Classification is in agreement with the analyses by principal components.

2. Code MolClas is an easy, dependable, effective, and quick process for the classification of molecules founded on the conjecture of the equipartition of the production of the entropy of information. The code was developed not just to examine the conjecture of equipartition but, in addition, to discover the world of the classification of molecules.

3. The periodic law does not convince the category of the laws of physics: (1) antitubulin inhibitory powers do not result done again; maybe their chemical nature; (2) sequence relations are done again with exemptions. The examination compels the declaration: relationships that whichever molecule p presents with its neighbor $p + 1$ are more or less done again for each period. Periodicity result is not universal; notwithstanding, if a usual order of molecules are agreed, the rule should be phenomenological. The antiproliferative potency did not generate the table of periodic classification and serves to confirm it. The examination of other antitubulin features would give an insight into the achievable generalization of the periodic table.

Acknowledgements

Author details

Francisco Torrens[1]* and Gloria Castellano[2]

*Address all correspondence to: torrens@uv.es

1 Institute for Molecular Science, University of Valencia, Valencia, Spain

2 Department of Experimental Sciences and Mathematics, Faculty of Veterinary and Experimental Sciences, Valencia Catholic University Saint Vincent Martyr, Valencia, Spain

References

[1] DeMartino G, Edler MC, LaRegina G, Coluccia A, Barbera MC, Barrow D, Nicholson RI, Chiosis G, Brancale A, Hamel E, Artico M, Silvestri R. New arylthioindoles: Potent inhibitors of tubulin polymerization. 2. Structure-activity relationships and molecular modeling studies. Journal of Medicinal Chemistry. 2006;**49**:947-954

[2] DeMartino G, LaRegina G, Coluccia A, Edler MC, Barbera MC, Brancale A, Wilcox E, Hamel E, Artico M, Silvestri R. Arylthioindoles, potent inhibitors of tubulin polymerization. Journal of Medicinal Chemistry. 2004;**47**:6120-6123

[3] Chang JY, Hsieh HP, Chang CY, Hsu KS, Chiang YF, Chen CM, Kuo CC, Liou JP. 7-Aroyl-aminoindoline-1-sulfonamides as a novel class of potent antitubulin agents. Journal of Medicinal Chemistry. 2006;**49**:6656-6659

[4] Liou JP, Chang YL, Kuo FM, Chang CW, Tseng HY, Wang CC, Yang YN, Chang JY, Lee SJ, Hsieh HP. Concise synthesis and structure-activity relationships of combretastatin A-4 analogues, 1-aroylindoles and 3-aroylindoles, as novel classes of potent antitubulin agents. Journal of Medicinal Chemistry. 2004;**47**:4247-4257

[5] Chang JY, Yang MF, Chang CY, Chen CM, Kuo CC, Liou JP. 2-amino and 2′-aminocombretastatin derivatives as potent antimitotic agents. Journal of Medicinal Chemistry. 2006;**49**:6412-6415

[6] Liou JP, Mahindroo N, Chang CW, Guo FM, Lee SWH, Tan UK, Yeh TK, Kuo CC, Chang YW, Lu PH, Tung YS, Lin KT, Chang JY, Hsieh HP. Structure−activity relationship studies of 3-aroylindoles as potent antimitotic agents. ChemMedChem. 2006;**1**:1106-1118

[7] Rappl C, Barbier P, Bourgarel-Rey V, Gregoire C, Gilli R, Carre M, Combes S, Finet JP, Peyrot V. Interaction of 4-arylcoumarin analogues of combretastatins with microtubule network of HBL100 cells and binding to tubulin. Biochemistry. 2006;**45**:9210-9218

[8] Romagnoli R, Baraldi PG, Remusat V, Carrion MD, Cara CL, Petri D, Fruttarolo F, Pavani MG, Tabrizi MA, Tolomeo M, Grimaudo S, Balzarini J, Jordan MA, Hamel E. Synthesis and biological evaluation of 2-(3′,4′,5′-trimethoxybenzoyl)-3-amino 5-aryl thiophenes as a new class of tubulin inhibitors. Journal of Medicinal Chemistry. 2006;**49**:6425-6428

[9] Nguyen TL, McGrath C, Hermone AR, Burnett JC, Zaharevitz DW, Day BW, Wipf P, Hamel E, Gussio R. A common pharmacophore for a diverse set of colchicine site inhibitors using a structure-based approach. Journal of Medicinal Chemistry. 2005;**48**:6107-6116

[10] Kim DY, Kim KH, Kim ND, Lee KY, Han CK, Yoon JH, Moon SK, Lee SS, Seong BL. Design and biological evaluation of novel tubulin inhibitors as antimitotic agents using a pharmacophore binding model with tubulin. Journal of Medicinal Chemistry. 2006;**49**: 5664-5670

[11] Liou JP, Wu ZY, Kuo CC, Chang CY, Lu PY, Chen CM, Hsieh HP, Chang JY. Discovery of 4-amino and 4-hydroxy-1-aroylindoles as potent tubulin polymerization inhibitors. Journal of Medicinal Chemistry. 2008;**51**:4351-4355

[12] Hsieh HP, Liou JP, Mahindroo N. Pharmaceutical design of antimitotic agents based on combretastatins. Current Pharmaceutical Design. 2005;**11**:1655-1677

[13] Mahindroo N, Liou JP, Chang JY, Hsieh HP. Antitubulin agents for the treatment of cancer – A medicinal chemistry update. Expert Opinion on Therapeutic Patents. 2006;**16**:647-691

[14] Ducki S, Mackenzie G, Lawrence NJ, Snyder JP. Quantitative structure-activity relationship (5D-QSAR) study of combretastatin-like analogues as inhibitors of tubulin assembly. Journal of Medicinal Chemistry. 2005;**48**:457-465

[15] Bellina F, Cauteruccio S, Monti S, Rossi R. Novel imidazole-based combretastatin A-4 analogues: Evaluation of their in vitro antitumor activity and molecular modeling study of their binding to the colchicine site of tubulin. Bioorganic & Medicinal Chemistry Letters. 2006;**16**:5757-5762

[16] Brown ML, Rieger JM, Macdonald TL. Comparative molecular field analysis of colchicine inhibition and tubulin polymerization for combretastatins binding to the colchicine binding site on beta-tubulin. Bioorganic & Medicinal Chemistry. 2000;**8**:1433-1441

[17] Pan X, Tan N, Zeng G, Han H, Huang H. 3D-QSAR and docking studies of aldehyde inhibitors of human cathepsin K. Bioorganic & Medicinal Chemistry. 2006;**14**:2771-2778

[18] Wolohan P, Reichert DE. CoMFA and docking study of novel estrogen receptor subtype selective ligands. Journal of Computer-Aided Molecular Design. 2003;**17**:313-328

[19] Liu H, Huang X, Shen J, Luo X, Li M, Xiong B, Chen G, Shen J, Yang Y, Jiang H, Chen K. Inhibitory mode of 1,5-diarylpyrazole derivatives against cyclooxygenase-2 and cyclooxygenase-1: Molecular docking and 3D QSAR analyses. Journal of Medicinal Chemistry. 2002;**45**:4816-4827

[20] Cheng F, Shen J, Luo X, Zhu W, Gu J, Ji R, Jiang H, Chen K. Molecular docking and 3-D-QSAR studies on the possible antimalarial mechanism of artemisinin analogues. Bioorganic & Medicinal Chemistry. 2002;**10**:2883-2891

[21] Du J, Qin J, Liu H, Yao X. 3D-QSAR and molecular docking studies of selective agonists for the thyroid hormone receptor beta. Journal of Molecular Graphics and Modelling. 2008;**27**:95-104

[22] Lin IH, Hsu CC, Wang SH, Hsieh HP, Sun YC. Comparative molecular field analysis of anti-tubulin agents with indole ring binding at the colchicine binding site. Journal of Theoretical and Computational Chemistry. 2010;**9**:279-291

[23] Cramer RD, III, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. Journal of the American Chemical Society 1988;**110**:5959-5967

[24] Halder AK, Adhikari N, Jha T. Comparative QSAR modelling of 2-phenylindole-3-carbaldehyde derivatives as potential antimitotic agents. Bioorganic & Medicinal Chemistry Letters. 2009;**19**:1737-1739

[25] Gajiwaja KS, Wu JC, Christensen J, Deshmukh GD, Diehl W, DiNitto JP, English JM, Greig MJ, He YA, Jacques SL, Lunney EA, McTigue M, Molina D, Quenzer T, Wells PA, Yu X, Zhang Y, Zou A, Emmett MR, Marshall AG, Zhang HM, Demetri GD. KIT kinase mutants show unique mechanisms of drug resistance to imatinib and sunitinib in gastrointestinal stromal tumor patients. Proceedings of the National Academy of Sciences of the United States of America. 2009;**106**:1542-1547

[26] Marimuthu A, Jacob HKC, Jakharia A, Subbannayya Y, Keerthikumar S, Kashyap MK, Goel R, Balakrishnan L, Dwivedi S, Pathare S, Dikshit JB, Maharudraiah J, Singh S, Kumar GSS, Vijayakumar M, Kumar KVV, Premalatha CS, Tata P, Hariharan R, Roa JC, Prasad TSK, Chaerkady R, Kumar RV, Pandey A. Gene expression profiling of gastric cancer. Journal of Proteomics and Bioinformatics. 2011;**4**(4):74-82

[27] Santavy F. Substanzen der herbstzeitlose und ihre derivate. XXII. Photochemische produkte des colchicins und einige seiner derivate. Collection of Czechoslovak Chemical Communications. 1951;**16**:665-675

[28] Andreu JM, Perez-Ramirez B, Gorbunoff MJ, Ayala D, Timasheff SN. Role of the colchicine ring a and its methoxy groups in the binding to tubulin and microtubule inhibition. Biochemistry. 1998;**37**:8356-8368

[29] Nunez J, Fellous A, Francon J, Lennon AM. Competitive inhibition of colchicine binding to tubulin by microtubule-associated proteins. Proceedings of the National Academy of Sciences of the United States of America. 1979;**76**:86-90

[30] Lee RM, Gewirtz DA. Colchicine site inhibitors of microtubule integrity as vascular disrupting agents. Drug Development Research. 2008;**69**:352-358

[31] Bhattacharyya B, Panda D, Gupta S, Banerjee M. Anti-mitotic activity of colchicine and the structural basis for its interaction with tubulin. Medicinal Research Reviews. 2008;**28**:155-183

[32] Jordan MA, Wilson L. Microtubules as a target for anticancer drugs. Nature Reviews. Cancer. 2004;**4**:253-265

[33] Brancale A, Silvestri R. Indole, a core nucleus for potent inhibitors of tubulin polymerization. Medicinal Research Reviews. 2007;**27**:209-238

[34] Sengupta S, Thomas SA. Drug target interaction of tubulin-binding drugs in cancer therapy. Expert Review of Anticancer Therapy. 2006;**6**:1433-1447

[35] Varmuza K. Pattern Recognition in Chemistry. New York: Springer; 1980

[36] Benzecri JP. L'Analyse des Données. Paris: Dunod; 1984. Vol. 1

[37] Tondeur D, Kvaalen E. Equipartition of entropy production. An optimality criterion for transfer and separation processes. Industrial & Engineering Chemistry Fundamentals. 1987;**26**:50-56

[38] Torrens F, Periodic CG. Classification of local anaesthetics (procaine analogues). International Journal of Molecular Sciences. 2006;**7**:12-34

[39] Castellano-Estornell G, Torrens-Zaragozá F. Local anaesthetics classified using chemical structural indicators. Nereis. 2009;**2009**(2):7-17

[40] Torrens F, Castellano G. Using chemical structural indicators for periodic classification of local anaesthetics. Int. J. Chemoinf. Chemical Engineer. 2011;**1**(2):15-35

[41] Torrens F, Castellano G. Table of periodic properties of human immunodeficiency virus inhibitors. International Journal of Computational Intelligence in Bioinformatics and Systems Biology. 2010;**1**:246-273

[42] Torrens F, Molecular CG. Classification of thiocarbamates with cytoprotection activity against human immunodeficiency virus. International Journal of Chemical Modeling. 2011;**3**:269-296

[43] Torrens F, Molecular CG. Classification of styrylquinolines as human immunodeficiency virus integrase inhibitors. International Journal of Chemical Modeling. 2014;**6**:347-376

[44] Torrens F, Castellano G. Modelling of complex multicellular systems: Tumour–immune cells competition. Chemistry Central Journal. 2009;**3**(Suppl. I):75–1-1

[45] Torrens F, Castellano G. Information theoretic entropy for molecular classification: Oxadiazolamines as potential therapeutic agents. Current Computer-Aided Drug Design. 2013;**9**:241-253

[46] Kaufmann A. Introduction à la Théorie des Sous-ensembles Flous. Paris: Masson; 1975. Vol. 3

[47] Cox E. The Fuzzy Systems Handbook. New York: Academic; 1994

[48] Kundu S. The min–max composition rule and its superiority over the usual max–min composition rule. Fuzzy Sets and Systems. 1998;**93**:319-329

[49] G. Lambert-Torres G, Pereira Pinto JO, Borges da Silva LE. In: Wiley Encyclopedia of Electrical and Electronics Engineering. New York: Wiley; 1999

[50] Shannon CE. A mathematical theory of communication: Part I, discrete noiseless systems. Bell System Technical Journal. 1948;**27**:379-423

[51] Shannon CE. A mathematical theory of communication. Part II, The discrete channel with noise. Bell System Technical Journal. 1948;**27**:623-656

[52] Neural WH. Network learning and statistics. AI Expert. 1989;**4**(12):48-52

[53] Kullback S. Information Theory and Statistics. New York: Wiley; 1959

[54] Iordache O, Corriou JP, Garrido-Sánchez L, Fonteix C, Tondeur D. Neural network frames. Application to biochemical kinetic diagnosis. Computers and Chemical Engineering. 1993; **17**:1101-1113

[55] Wildenhain J, FitzGerald N, Tyers M. MolClass: A web portal to interrogate diverse small molecule screen datasets with different computational models. Bioinformatics. 2012;**28**: 2200-2201

[56] Tryon RCA. Multivariate analysis of the risk of coronary heart disease in Framingham. Journal of Chronic Diseases. 1939;**20**:511-524

[57] Jarvis RA, Patrick EA. Clustering using a similarity measure based on shared nearest neighbors. IEEE Transactions on Computers. 1973;**C22**:1025-1034

[58] Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proceedings of the National Academy of Sciences of the United States of America. 1998;**95**:14863-14868

[59] Page RDM. Program TreeView. Glasgow: Universiy of Glasgow; 2000

[60] Huson DH. SplitsTree: Analyzing and visualizing evolutionary data. Bioinformatics. 1998;**14**:68-73

[61] Hotelling H. Analysis of a complex of statistical variables into principal components. Journal of Education & Psychology. 1933;**24**:417-441

[62] Kramer R. Chemometric Techniques for Quantitative Analysis. New York: Marcel Dekker; 1998

[63] Patra SK, Mandal AK, Pal MK. State of aggregation of bilirubin in aqueous solution: Principal component analysis approach. Journal of Photochemistry and Photobiology A. 1999;**122**:23-31

[64] Jolliffe IT. Principal Component Analysis. New York: Springer; 2002

[65] Xu J, Hagler A. Chemoinformatics and drug discovery. Molecules. 2002;**7**:566-600

[66] Shaw PJA. Multivariate Statistics for the Environmental Sciences. New York:Hodder-Arnold; 2003