

[g005]

## PERIODIC CLASSIFICATION OF HUMAN IMMUNODEFICIENCY VIRUS INHIBITORS

Francisco Torrens and Gloria Castellano

<sup>1</sup> Institut Universitari de Ciència Molecular, Universitat de València, Edifici d'Instituts de Paterna, P. O. Box 22085, E-46071 València, Spain.

**ABSTRACT.** Classification algorithms are proposed based on *information entropy* and applied to 13 human immunodeficiency virus type 1 inhibitors. A number of results are compatible with the data suffering combinatorial explosion. After the *equipartition conjecture*, entropy production is most uniformly distributed. In ddI the formula is  $N_4O_3S_0P_0X_0$ ,  $X = F, Cl$ ; it is selected as a *reference* <11111>. In most cases (ddI, ddC, d4T, novel proposed ligand) the formula is  $N_{3-4}O_3S_0P_0X_0$ , while in 3TC the formula is  $N_3O_3S_1P_0X_0$ . The analysis compares well with other classification taken as *good*.

### Introduction

*Ab initio* theoretical calculations, molecular dynamics simulations and docking are useful tools for investigating important biological complexes.<sup>1-12</sup> At least three anti-human immunodeficiency virus type 1 (HIV-1) drugs, for combination therapy, have become the standard treatment of acquired immunodeficiency syndrome (AIDS) drugs that have been licensed for clinical use, or are subjected to advanced clinical trials, belong to one of the following three classes: (1) nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs/NtRTIs) [abacavir (ABC), emtricitabine [(-)FTC], zidovudine (AZT), didanosine (ddI), zalcitabine (ddC), stavudine (d4T), lamivudine (3TC), tenofovir disoproxil fumarate], (2) non-nucleoside reverse transcriptase inhibitors (NNRTIs) [emivirine, efavirenz, nevirapine, delavirdine], and (3) protease inhibitors (PIs) [lopinavir, nelfinavir, ritonavir, amprenavir, saquinavir, indinavir]. Various other events in the HIV replicative cycle can be considered as potential targets for chemotherapeutic intervention: (1) viral entry *via* blockade of the viral coreceptors CXCR4 [bicyclam (AMD3100) derivatives] and CCR5 (TAK-799 derivatives), (2) viral adsorption *via* binding to the viral envelope glycoprotein gp120 (polysulphates, polysulphonates, polycarboxylates, polyoxometalates, polynucleotides and negatively charged albumins), (3) viral assembly and disassembly *via* NCp7 Zn finger-targeted agents [2,2'-dithiobisbenzamides (DIBAs) and azodicarbonamide (ADA)], (4) virus-cell fusion *via* binding to the viral envelope glycoprotein gp41 (T-1249), (5) proviral deoxyribonucleic acid (DNA) integration *via* integrase inhibitors, *e.g.* 4-aryl-2,4-dioxobutanoic acid derivatives, as well as (6) viral messenger ribonucleic acid (mRNA) transcription *via* inhibitors of the transcription (transactivation) process (flavopiridol, fluoroquinolones). In addition, new NRTIs, NNRTIs and PIs have been developed that possess, respectively: (1) improved metabolic characteristics, *e.g.*,

phosphoramidate and cyclosaligenyl pronucleotides bypassing the first phosphorylation step of NRTIs, (2) increased activity [*second* or *third* generation NNRTIs (TMC-125, DPC-093)] and (3) different, non-peptidic scaffold, *e.g.* cyclic urea (mozenavir) and 4-hydroxy-2-pyrone (triplanavir).

The advent of so many new compounds, other than those that have been formally approved for the treatment of HIV infections, will undoubtedly improve the prognosis of patients with AIDS and AIDS-associated diseases. Nucleoside analogues constitute a family of biological molecules (ddI, d4T, ddC and T3C), which play an important role in the transcription process of HIV. The normal nucleoside substrates, used by reverse transcriptase (RT) to synthesize DNA, are mimicked by these nucleoside analogues, which lacked a 3'-OH group and, consequently, act as chain terminators when incorporated into DNA by RT. Although these nucleoside analogues show good activity as inhibitors of HIV, their long-term usefulness is limited by toxicities. Resistance and mutation are also problems. The development of better drugs requires a better understanding of how the drugs work, the mechanism of drug resistance and the interaction with the receptor, as well as the stability of the drugs inside the active site. A novel HIV RT inhibitor ligand was proposed, which indicated the highest docking scores and more hydrogen-bond interactions with the residues of the RT active site.<sup>13</sup> A simple computerized algorithm, useful for establishing a relationship between chemical structures and their biological activities or significance, is proposed and exemplified here. The starting point is to use an informational or configurational entropy for pattern recognition purposes. The entropy is formulated on the basis of a *matrix of similarity* between two chemical or biochemical species. As entropy is weakly discriminating for classification purposes, the more powerful concept of *entropy production* and its *equipartition conjecture* are introduced.<sup>14</sup> In earlier publications the periodic classification of local anaesthetics was analyzed.<sup>15</sup> The aim of the present report is to develop the learning potentialities of the code and, since molecules are more naturally described *via* a varying size structured representation, the study of general approaches to the processing of structured information.

Table 1 contains the vectors associated with 13 HIV-1 inhibitors of various types: non-nucleoside reverse transcriptase inhibitors (NNRTIs), nucleoside reverse transcriptase inhibitors (NRTIs), nucleotide reverse transcriptase inhibitors (NtRTIs) and protease inhibitors (PIs).

Table 1. Vector properties of human immunodeficiency virus type 1 inhibitors.

Non-nucleoside reverse transcriptase inhibitors (NNRTIs)			
1. efavirenz	<00110>	3. delavirdine	<00111>
2. nevirapine	<10111>		
Nucleoside reverse transcriptase inhibitors (NRTIs)			
4. didanosine (ddI)	<11111>	8. novel proposed ligand	<11111>
5. zalcitabine (ddC)	<01111>	9. zidovudine (AZT)	<00111>
6. stavudine (d4T)	<01111>	10. abacavir (ABC)	<00111>
7. lamivudine (3TC)	<01011>		

Nucleotide reverse transcriptase inhibitor (NtRTIs)			
11. tenofovir disoproxil	<00101>		
Protease inhibitors (PIs)			
12. amprenavir	<00011>	13. lopinavir	<10111>

## Results and Discussion

Many HIV-1 inhibitors fit the following general scheme: (base derivative)–(furan ring). The base is often a guanine (Gua) or cytosine (Cys); the furan normally contains one O atom. The intercorrelations in the matrix of Pearson correlation coefficients are illustrated in the partial correlation diagram, which contains 28 high ( $r \geq 0.75$ ), 24 medium ( $0.50 \leq r < 0.75$ ) and 12 low ( $0.25 \leq r < 0.50$ ) partial correlations. Inhibitors with high partial correlations show a similar vector property. However, the results should be taken with care, because both compounds with constant <11111> vector (Entries 4, 8) show null standard deviation, causing high partial correlations ( $r = 1$ ) with any inhibitor, which is an artifact. With the equipartition conjecture the intercorrelations are illustrated in the partial correlation diagram, which contains 20 high, 22 medium and 18 low partial correlations. Notice that 22 out of the 24 high partial correlations have been corrected. The grouping rule for  $0.88 \leq b_1 \leq 0.93$  allows:

$$C_{b_1} = (1,3,9,10)(2,13)(4,8)(5,6)(7)(11)(12)$$

The seven classes show an entropy  $h(\hat{\mathbf{R}}_{b_1}) = 24.22$ . Dendrogram and radial tree<sup>20–22</sup> separate the same classes. The NRTI ddI and novel proposed ligand inhibitors are grouped into the same class, and NRTI ddC and d4T. Inhibitors belonging to the same class appear highly correlated in partial correlation diagrams, in agreement with previous results for Entries 4–8. At level  $0.82 \leq b_2 \leq 0.84$  classification is:

$$C_{b_2} = (1,3,9,10,11)(2,13)(4,8)(5,6,7)(12)$$

Five classes result in this case; the entropy decreases to  $h(\hat{\mathbf{R}}_{b_2}) = 11.90$ . Both dendrogram and radial tree matching to  $\langle i_1, i_2, i_3, i_4, i_5 \rangle$  and  $C_{b_2}$  separate the same five classes, in agreement with both partial correlation diagrams, dendrogram, binary tree and previous results obtained for Entries 4–8. A high degree of similarity is found for Entries 2–13, 3–9–10, 4–8 and 5–6. Again NRTI ddI and novel proposed ligand are grouped into the same class and NRTI ddC, d4T and 3TC. The lower-level  $b_2$  classification process shows lower entropy and may be more parsimonious. The  $b_2$  may have greater signal-to-noise ratio than  $b_1$  classification. Entries 4–8 belong to the same class at any grouping level  $b$ . A comparative analysis of the set of 1–13 classes is in agreement with previous results obtained for Entries 4–8. Once more NRTI ddI and novel proposed ligand are grouped into the same class and NRTI ddC, d4T and 3TC. The inclusion in the radial tree (*cf.* Figure 1) is in agreement with partial correlation

diagrams, dendrograms, binary trees and previous results for Entries 4–8. Moreover, the classification presents lower bias and greater precision, with lower divergence with regard to the original distribution.

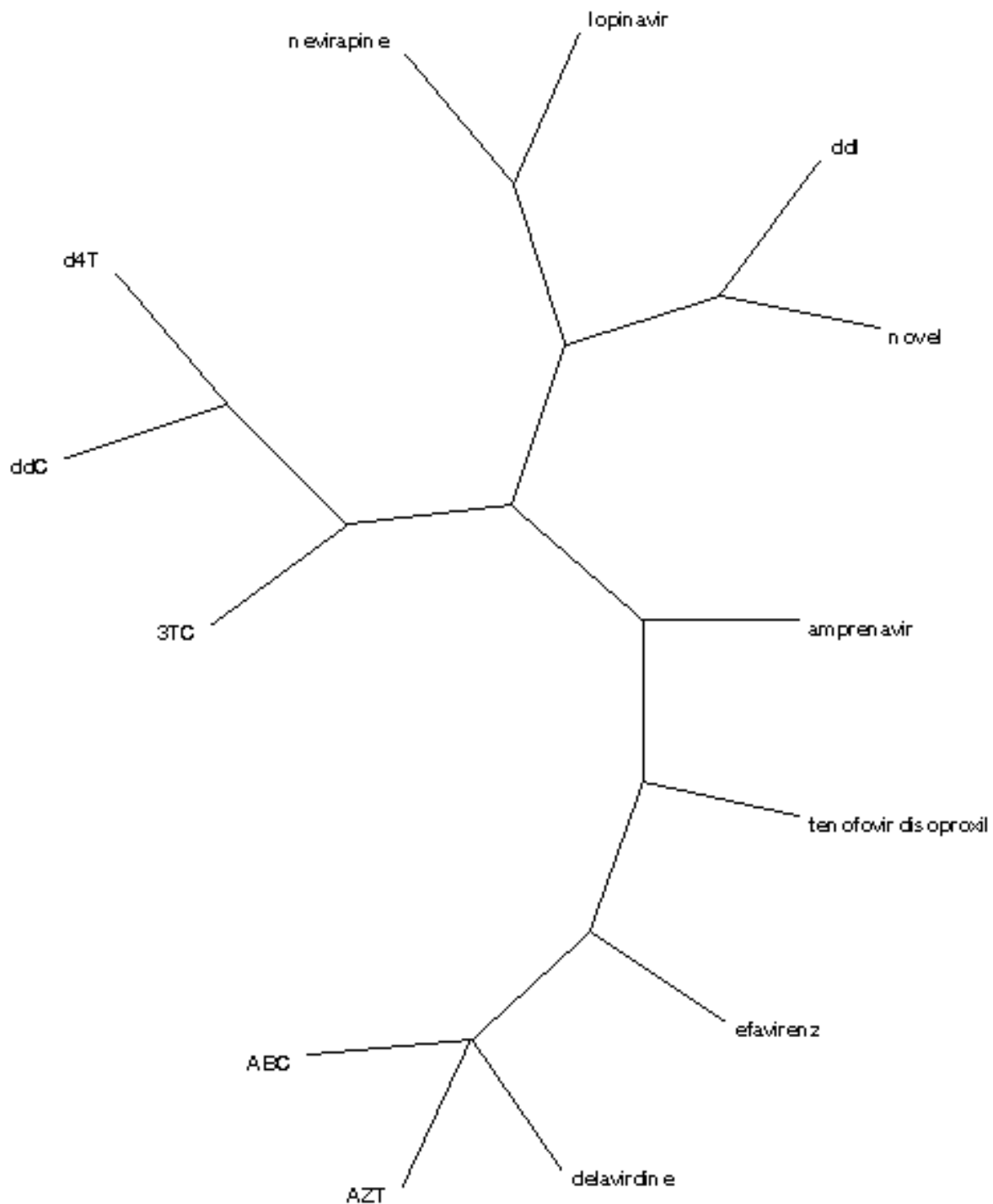


Figure 1. Radial tree for human immunodeficiency virus type 1 inhibitors.

Program SplitsTree analyzes cluster analysis (CA) data.<sup>23</sup> Its *split decomposition* takes a *distance matrix* and produces a graph representing the relationships between taxa. For ideal data this graph is a tree whereas less ideal data will give rise to a tree-like network. As split decomposition does not force data onto a tree, it can provide a good indication of how *tree-like* given data are. The splits graph for the

13 HIV-1 inhibitors of Table 1 (*cf.* Figure 2) reveals no conflicting relationship. The splits graph is in general agreement with partial correlation diagrams, dendrograms and binary trees (Figure 1).

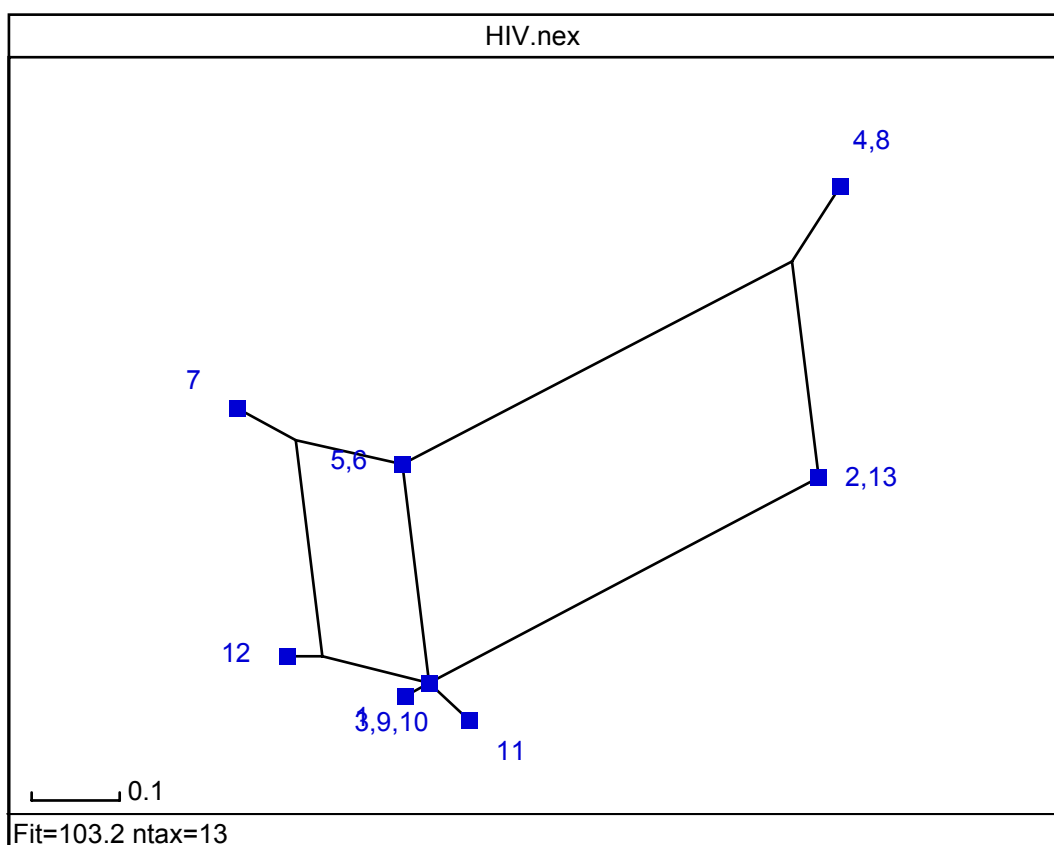


Figure 2. Splits graph for the human immunodeficiency virus type 1 inhibitors.

A principal component analysis (PCA)<sup>24</sup> has been carried out for the HIV-1 inhibitors. Factors  $F_1$ – $F_5$  show that  $F_1$  explains 30% of the variance (70% error),  $F_{1-2}$ , 55% of variance (45% error) and  $F_{1-3}$ , 76% of variance (24% error). For  $F_1$  and  $F_4$ ,  $i_2$  has the greatest weight in the profile; however,  $F_1$  cannot be reduced to three variables  $\{i_1, i_2, i_4\}$  without a 19% error. For  $F_2$ ,  $i_3$  has the greatest weight; notwithstanding,  $F_2$  cannot be reduced to three variables  $\{i_1, i_2, i_3\}$  without a 4% error. For  $F_3$ ,  $i_4$  and  $i_5$  have the greatest weight; furthermore,  $F_3$  can be reduced to two variables  $\{i_4, i_5\}$  with a 0% error. For  $F_5$ ,  $i_1$  has the greatest weight; nevertheless,  $F_5$  cannot be reduced to three variables  $\{i_1, i_3, i_4\}$  without a 22% error. The  $F_{1-5}$  can be considered as linear combinations of  $\{i_1, i_2, i_4\}$ ,  $\{i_1, i_2, i_3\}$ ,  $\{i_4, i_5\}$ ,  $\{i_1, i_2, i_3\}$  and  $\{i_1, i_3, i_4\}$  with 19%, 4%, 0%, 18% and 22% errors, respectively. In plot  $F_2$ – $F_1$  (*cf.* Figure 3) those HIV-1 inhibitors with the same vector property appear superposed. Inhibitors 9 and 10 (NRTI) come out placed over 3 (NNRTI), 11 (NtRTI), over 1 (NNRTI) and 13 (PI), over 2 (NNRTI). Four classes of inhibitors are clearly distinguished in agreement with classical classification: class NNRTI (3 units, *left top*), class NRTI (7 units, *right*), class NtRTI (1 unit, *left*) and class PI (2 units, *low*). The classification is in agreement with partial correlation diagrams, dendrograms, binary trees and splits graph (Figures 1–2).

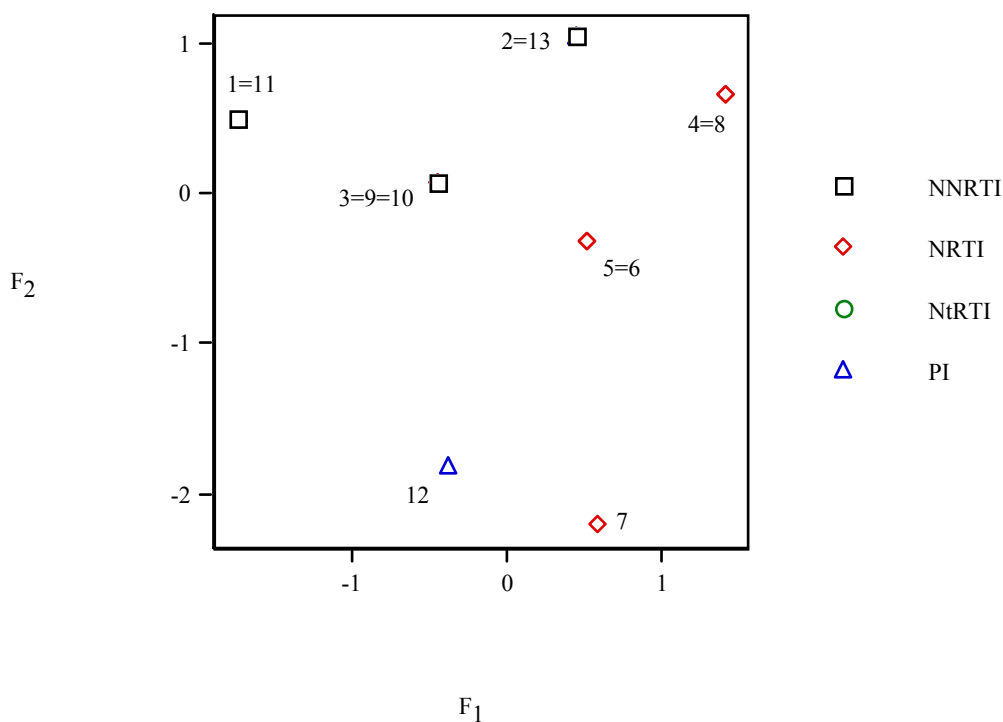


Figure 3. Principal component analysis  $F_2$  vs.  $F_1$  plot for the HIV-1 inhibitors.

## Experimental Procedures

The key problem in classification studies is to define *similarity indices*, when several criteria of comparison are involved. The first step in quantifying the concept of similarity, for molecules of HIV-1 inhibitors, is to list the most important portions of such molecules. Furthermore, the *vector of properties*  $\vec{i} = \langle i_1, i_2, \dots, i_k, \dots \rangle$  should be associated with each inhibitor  $i$ , whose components correspond to different characteristic groups in the inhibitor molecule, in a hierarchical order according to the expected importance of their pharmacological potency. If the  $m$ -th portion of the molecule is pharmacologically more significant for the inhibitory effect than the  $k$ -th portion, then  $m < k$ . The components  $i_k$  are “1” or “0”, according to whether a similar (or identical) portion of rank  $k$  is present or absent in inhibitor  $i$ , compared with the reference inhibitor. The analysis includes such chemical compounds that fit the following general scheme: (base derivative)–(furan ring), since these are the most numerous and have the widest range of uses among the species used in practice of inhibition. The base portion is often a guanine (Gua) or cytosine (Cys) derivative; the furan ring normally contains one O heteroatom. In didanosine (ddI) the base is a Gua derivative, and the furan contains only one O heteroatom. It is assumed that the *structural elements* of an inhibitor molecule can be *ranked*, according to their contribution to inhibitory activity, in the following order of decreasing importance: number of N atoms > number of O atoms > number of S atoms > number of P atoms > number of halogen atoms. The ddI molecule contains four N, three O, no S, no P and no halogen (X = F, Cl) heteroatoms ( $N_4O_3S_0P_0X_0$ ). In some inhibitors the base is a Gua (ddI, novel proposed ligand), in some others, a Cys derivative (ddC,

d4T, 3TC). In most inhibitors the furan ring contains only one O heteroatom (ddI, ddC, d4T, novel proposed ligand,  $N_{3-4}O_3S_0P_0X_0$ ), while in 3TC the furan ring includes one O and one S heteroatoms ( $N_3O_3S_1P_0X_0$ ). In the NRTI inhibitor ddI the base is a Gua derivative, and the furan contains only one O heteroatom. The molecule contains four N, three O, no S, no P and no halogen heteroatoms ( $N_4O_3S_0P_0X_0$ ). Obviously its associated vector is  $\langle 11111 \rangle$ . In this study, ddI was selected as a *reference* HIV-1 inhibitor, because of the good docking scores with the receptor RT. This improves the quality of the classification for those inhibitors similar to ddI. The selection as reference of an inhibitor dissimilar to ddI, *e.g.* tenofovir disoproxil, would not improve the quality of the classification for those inhibitors similar to ddI.

Vector  $\langle 00110 \rangle$  is associated with efavirenz since it contains one N, two O, no S, no P and four halogen atoms. Let us denote by  $r_{ij}$  ( $0 \leq r_{ij} \leq 1$ ) the similarity index of two inhibitors associated with vectors  $\bar{i}$  and  $\bar{j}$ , respectively. The relation of similitude is characterized by a *similarity matrix*  $\mathbf{R} = [r_{ij}]$ . The similarity index between two inhibitors  $\bar{i} = \langle i_1, i_2, \dots, i_k, \dots \rangle$  and  $\bar{j} = \langle j_1, j_2, \dots, j_k, \dots \rangle$  is defined as:

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

(1)

where  $0 \leq a_k \leq 1$  and  $t_k = 1$  if  $i_k = j_k$ , but  $t_k = 0$  if  $i_k \neq j_k$ . The definition assigns a weight  $(a_k)^k$  to any property involved in the description of molecule  $i$  or  $j$ . The *grouping algorithm* uses the *stabilized* matrix of similarity, obtained by applying the *max-min composition rule*  $\circ$  defined by:

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

where  $\mathbf{R} = [r_{ij}]$  and  $\mathbf{S} = [s_{ij}]$  are matrices of the same type, and  $(\mathbf{R} \circ \mathbf{S})_{ij}$ , element  $(i, j)$ -th of matrix  $\mathbf{R} \circ \mathbf{S}$ . When applying this rule iteratively so that  $\mathbf{R}(n+1) = \mathbf{R}(n) \circ \mathbf{R}$ , there exists an integer  $n$  such that:  $\mathbf{R}(n) = \mathbf{R}(n+1) = \dots$ . The matrix  $\mathbf{R}(n)$  is called the *stabilized similarity matrix*. Its importance lies in the fact that in the classification it will generate a partition into disjoint classes. It is used and designated by  $\mathbf{R}(n) = [r_{ij}(n)]$ . The *grouping rule* is the following:  $i$  and  $j$  are assigned to the same class if  $r_{ij}(n) \geq b$ . The class of  $i$  noted  $i'$  is the set of species  $j$  that satisfies rule  $r_{ij}(n) \geq b$ . The matrix of classes is:

$$\hat{\mathbf{R}}(n) = \left[ \hat{r}_{ij} \right] = \max_{s,t} (r_{st}) \quad (s \in i', t \in j') \quad (3)$$

where  $s$  stands for any index of a species belonging to the class  $i'$  (similarly for  $t$  and  $j'$ ). Rule (3) means finding the largest similarity index between species of two different classes. In information theory, the *information entropy*  $h$  measures the surprise that the source emitting the sequences can give.<sup>16,17</sup> For a single event occurring with probability  $p$  the degree of surprise is proportional to  $-\ln p$ . Generalizing the result to a random variable  $X$  (which can take  $N$  possible values  $x_1, \dots, x_N$  with

probabilities  $p_1, \dots, p_N$ ), the average surprise received on learning the value of  $X$  is  $-\sum p_i \ln p_i$ . The information entropy associated with similarity matrix  $\mathbf{R}$  is:

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

Denote by  $C_b$  the classes set and by  $\mathbf{R}_b$  the similarity matrix at grouping level  $b$ . The information entropy satisfies the following properties. (1)  $h(\mathbf{R}) = 0$  if  $r_{ij} = 0$  or  $r_{ij} = 1$ . (2)  $h(\mathbf{R})$  is maximum if  $r_{ij} = 0.5$  (when the imprecision is maximum). (3)  $h(\mathbf{R}_{b_1}) \leq h(\mathbf{R}_{b_2})$  for any  $b$ , i.e. classification leads to a loss of entropy. (4)  $h(\mathbf{R}_{b_1}) \leq h(\mathbf{R}_{b_2})$  if  $b_1 < b_2$  (the entropy is a monotone function of the grouping level  $b$ ). In the classification, each *hierarchical tree* corresponds to an entropy dependence on grouping level, and an  $h$ - $b$  diagram can be obtained. The *equipartition conjecture of entropy production* is proposed, as a selection criterion among different variants resulting from classification among hierarchical trees. For a given charge or duty, the best configuration is that in which entropy production is most uniformly distributed. Equipartition implies a linear dependence, so that the *equipartition line* is described by:

$$h_{\text{eqp}} = h_{\text{max}} b \quad (5)$$

Since the classification is discrete, a way of expressing equipartition would be a regular staircase function. The best variant is chosen to be that minimizing the sum of squares of the deviations:

$$SS = \sum_{b_i} (h - h_{\text{eqp}})^2 \quad (6)$$

*Learning procedures* are implemented as follows.<sup>18</sup> Consider a given partition into classes as *good* or *ideal* from practical or empirical observations, which corresponds to a *reference* similarity matrix  $\mathbf{S} = [s_{ij}]$  obtained for equal weights  $a_1 = a_2 = \dots = a$  and 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 degree  $r_{ij}$  is then computed with Equation (1) giving matrix  $\mathbf{R}$ . The number of properties for  $\mathbf{R}$  and  $\mathbf{S}$  may differ. The learning procedure consists in trying to find classification results for  $\mathbf{R}$ , as close as possible to the *good* classification. The first weight  $a_1$  is taken constant, and the following weights  $a_2, a_3, \dots$ , subjected to random variations. A new similarity matrix is obtained using Equation (1) and the new weights. The distance between the partitions into classes characterized by  $\mathbf{R}$  and  $\mathbf{S}$  is given by:

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

The result of the algorithm is a set of weights allowing adequate classification. The procedure has been applied in the synthesis of complex flowsheets using of information entropy.<sup>19</sup>

From the present results and discussion the following conclusions can be drawn.

1. Many algorithms for classification are based on *information entropy*. For sets of moderate size an excessive number of results appear compatible with data, and the number suffers a combinatorial



explosion. However, after the *equipartition conjecture*, one has a selection criterion between different variants resulting from classification between hierarchical trees. According to the conjecture, the best configuration is the one in which the entropy production is most uniformly distributed. The method avoids the problem of other methods of continuum variables, because for the four compounds with constant <11111> vector, the null standard deviation always causes a Pearson correlation coefficient of  $r = 1$ . The lower level classification processes show lower entropy and may be more parsimonious. The good comparison of our classification results, with other taken as *good*, confirm the adequacy of the property vector selected for the molecular structures. Information entropy and principal component analyses permit classifying the inhibitors and agree. In general, the classical classes are recognized.

2. The analysis is in agreement with principal component analysis. It compares well with other classification taken as *good* based on docking, density functional, molecular dynamics, the Rule of Five, absorption, distribution, metabolism, excretion and toxicity. The analysis of the interactions of the proposed novel ligand with the reverse-transcriptase active site strongly suggests that the proposed novel ligand could be a good potential inhibitor for anti-HIV chemotherapy.

*Acknowledgements.* The authors acknowledge financial support from the Spanish MEC DGCyT (Project No. CTQ2004-07768-C02-01/BQU), EU (Program FEDER) and Generalitat Valenciana (DGEUI INF01-051, INFRA03-047 and OCYT GRUPOS03-173).

## References

- (1) Da Silva, C. H. T. P.; Carvalho, I.; Taft, C. A. *J. Comput.-Aided Mol. Design* **2005**, *19*, 83-92.
- (2) Da Silva, C. H. T. P.; del Ponte, G.; Neto, A. F.; Taft, C. A. *Bioorg. Chem.* **2005**, *33*, 274-284.
- (3) Da Silva, C. H. T. P.; Almeida, P.; Taft, C. A. *J. Mol. Model.* **2004**, *10*, 38-43.
- (4) Da Silva, C. H. T. P.; Taft, C. A. *Biophys. Chem.* **2005**, *117*, 73-77.
- (5) Da Silva, C. H. T. P.; Taft, C. A. *J. Biomol. Struct. Dynam.* **2004**, *22*, 59-63.
- (6) Arissawa, M.; Taft, C. A.; Felcman, J. *Int. J. Quantum Chem.* **2003**, *93*, 422-432.
- (7) Kuno, M.; Palangsuntikul, R.; Hannongbua, S. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1584-1590.
- (8) Sharma, B.; Kaushik, N.; Singh, K.; Kumar, S.; Pandey, V. N. *Biochemistry* **2002**, *41*, 15685-15697.
- (9) De Clercq, E. *Biochim. Biophys. Acta* **2002**, *1587*, 258-275.
- (10) Kasai, N.; Mizushina, Y.; Sugawara, F.; Sakaguchi, K. *J. Biochem.* **2002**, *132*, 819-828.
- (11) Painter, G. R.; Andrews, C. W.; Furman, P. A. *Nucleosides, Nucleotides Nucl. Acids* **2000**, *19*, 13-29.
- (12) Mlinaric, A.; Kreft, S.; Umek, A.; Strukelj, B. *Pharmazie* **2000**, *55*, 75-77.

- (13) Da Silva, C. H. T. P.; Carvalho, I.; Taft, C. A. *J. Theor. Comput. Chem.* **2006**, *5*, 579-586.
- (14) Tondeur, D.; Kvaalen, E. *Ind. Eng. Chem., Fundam.* **1987**, *26*, 50-56.
- (15) Torrens, F.; Castellano, G. *Int. J. Mol. Sci.* **2006**, *7*, 12-34.
- (16) Shannon, C. E. *Bell Syst. Tech. J.* **1948**, *27*, 379-423.
- (17) Shannon, C. E. *Bell Syst. Tech. J.* **1948**, *27*, 623-656.
- (18) White, H. *AI Expert* **1989**, *12*, 48-48.
- (19) Iordache, O.; Corriou, J. P.; Garrido-Sánchez, L.; Fonteix, C.; Tondeur, D. *Comput. Chem. Eng.* **1993**, *17*, 1101-1113.
- (20) IMSL, *Integrated Mathematical Statistical Library (IMSL)*; IMSL: Houston, 1989.
- (21) Tryon, R. C. *J. Chronic Dis.* **1939**, *20*, 511-524.
- (22) Page, R. D. M. *Program TreeView*; University of Glasgow, 2000.
- (23) Huson, D. H. *Bioinformatics* **1998**, *14*, 68-73.
- (24) Hotelling, H. *J. Educ. Psychol.* **1933**, *24*, 417-441.