Send Orders for Reprints to reprints@benthamscience.ae

Current Bioinformatics, 2015, *10*, 639-657

Multiscale Mapping of AIDS in U.S. Countries *vs* Anti-HIV Drugs Activity with Complex Networks and Information Indices

Diana María Herrera-Ibatá[*,1], Ricardo Alfredo Orbegozo-Medina[2] and Humberto González-Díaz[3,4]

[1]*Department of Information and Communication Technologies, University of A Coruña UDC, 15071, A Coruña, Spain*

[2]*Department of Microbiology and Parasitology, University of Santiago de Compostela (USC), 15782, Santiago de Compostela, A Coruña, Spain*

[3]*Department of Organic Chemistry II, Faculty of Science and Technology, University of the Basque Country UPV/EHU, 48940, Leioa, Spain*

[4]*IKERBASQUE, Basque Foundation for Science, 48011, Bilbao, Spain*

D.M. Herrera-Ibatá

Abstract: In this work, we reviewed different aspects about the epidemiology, drugs, targets, chem-bioinformatics, and systems biology methods, related to AIDS/HIV. Next, we developed a new model to predict complex networks of the prevalence of AIDS in U.S. counties taking into consideration the values of Gini coefficients of social income inequality. We also used activity/structure data of anti-HIV drugs in preclinical assays. First, we trained different Artificial Neural Networks (ANNs) using as input Markov and Symmetry information indices of social networks and of molecular graphs. We obtained the data about AIDS prevalence and Gini coefficient from the AIDSVu database of Emory University and the data about anti-HIV drugs from ChEMBL database. We used Box-Jenkins operators to measure the shift with respect to average behavior of counties from states and drugs from reference compounds assayed in a given protocol, target, or organism. To train/validate the model and predict the complex network we needed to analyze 43,249 data points including values of AIDS prevalence in 2310 counties in U.S. *vs* ChEMBL results for 21,582 unique drugs, 9 viral or human protein targets, 4856 protocols, and 10 possible experimental measures. The best model found was a Linear Neural Network (LNN) with Accuracy, Specificity, Sensitivity, and AUROC above 0.72-0.73 in training and external validation series. The new linear equation was shown to be useful to generate complex network maps of drug activity *vs* AIDS/HIV epidemiology in U.S. at county level.

Keywords: Anti-HIV drugs, AIDS in U.S. at county level, Gini coefficient, Multiscale models, Box-Jenkins moving average operators, Shannon Entropy, indices of neighborhood symmetry.

INTRODUCTION

AIDS/HIV Epidemiology and Chemotherapy

Human immunodeficiency virus (HIV) is a retrovirus belonging to the family of lentiviruses that causes AIDS. Retroviruses [1] can use their RNA and host DNA to make viral DNA, and are known for their long incubation periods. There are two types of HIV: HIV type 1 and HIV type 2. Worldwide, the predominant virus is HIV-1, the HIV-2 is less common and appears to progress more slowly [2]. In AIDS, the immune system is severely affected, and the body is susceptible to a variety of infections, such as bacteria, parasites, viruses, etc., which can cause fatal diseases in people with the syndrome [3]. Since the virus was discovered in the 1980s, many progresses have been made in the management of HIV/AIDS [4]. Antiretroviral treatment which can stop the replication of HIV, has been one of the most important discoveries [5].

*Address correspondence to this author at the Department of Information and Communication Technologies, University of A Coruña UDC, 15071, A Coruña, Spain; E-mail: diana.herrera@udc.es

Despite progresses, HIV [5] remains a public health challenge. After thirty years in the AIDS epidemic, there are over 34 million people living with HIV [6], and still 2.5 million new infections and 1.7 million deaths each year.

AIDS Prevalence in U.S. at County Level

There are several databases with epidemiological data about AIDS prevalence. For instance, researchers at the Rollins School of Public Health at Emory University compiled state and county level information for AIDS prevalence in the U.S. in the database AIDSVu (http://aidsvu.org/about-aidsvu/). This information on AIDSVu comes from the U.S. Centers for Disease Control and Prevention's (CDC) national surveillance database. AIDSVu is an interactive map that is available online and it shows the prevalence of HIV in the United States. The values used in this study were the rate of adults/adolescents living with an HIV diagnosis in 2010 *per* 100,000 populations (see Fig. **1**). The county-level HIV data is estimated for persons aged 13 and older living with an HIV infection diagnosis. All race groups are non-Hispanic, and the Hispanic/Latino ethnicity is inclusive of all races. The total number of counties is $n_a = 2310$.

2212-392X/15 $58.00+.00 © 2015 Bentham Science Publishers

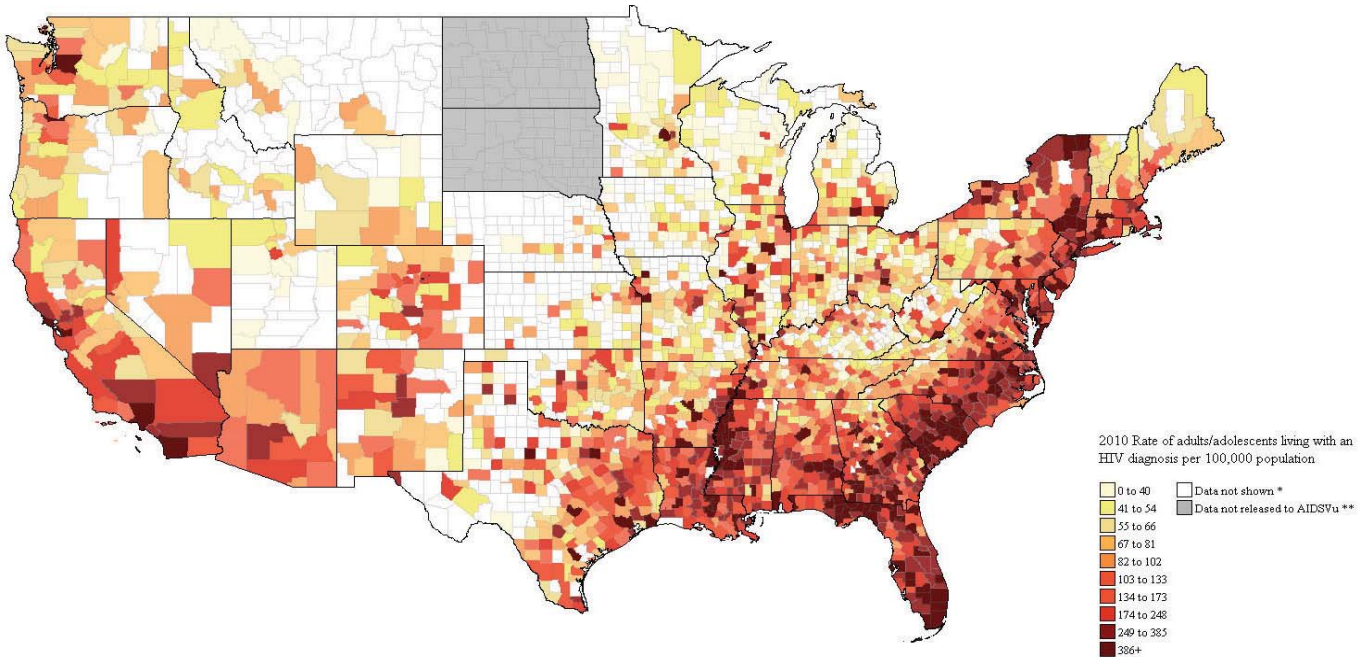


Fig. (1). AIDSVu map of HIV rate in U.S. at county level 2010.

The Human immunodeficiency virus has prompted many researchers worldwide to discover new compounds and/or molecular or cellular targets to become useful against the disease. There are different datasets with experimental outcomes of the interactions of anti-HIV compounds with their respective targets. ChEMBL (https://www.ebi.ac.uk/chembl/) [7-9] is one of the biggest; ChEMBL is an open large-scale bioactivity database containing huge information largely manually extracted from the medicinal chemistry literature. Information regarding the compounds tested (including their structures), the biological or physicochemical assays performed on these and the targets of these assays are recorded in a structured form, there are now >1.3 million distinct compound structures and 12 million bioactivity data points. The data are mapped to >9000 targets, of which 2827 are human protein targets [9]. Specifically, ChEMBL contains > 43,000 outcomes for assays of anti-HIV compounds. CHEMBL contains a general data set of potential anti-HIV compounds composed of >8,000 multiplexing assay endpoints (results of multiple assays) [7, 8]. The dataset used to train and validate our model includes N = 43,249 statistical cases formed by N_d = 21,582 unique drugs; which have been assayed each one in at least one out of 10 possible standard type measures determined in at least one out of 4856 different assays (experimental protocols reported as different in ChEMBL). Each assay involves, in turn, at least one out of 9 non-molecular or protein targets expressed in the tissues, cells, or viral particles of at least one out of 5 different organisms (including human cells lines). In this work, we reviewed and tabulated a number of compounds, targets, biological activities present in ChEMBL for the HIV (see Table **1**).

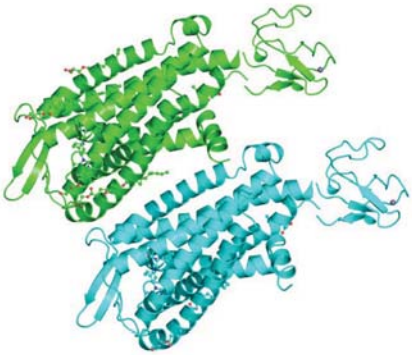
HIV Targets

Some of the more important targets are proteins present in the virus or in the host. The spikes projecting from the surface of HIV-1 are composed of the Envelope glycoprotein (Env), Env facilitates HIV entry by a process of direct fusion between the virion membrane and the target cell [10]. Envelope glycoprotein consists of two noncovalently associated subunits derived by proteolytic cleavage of the gp160 biosynthetic precursor: the external subunit gp120, which is responsible for binding to specific target cell receptors [10]. The fusion of the HIV-1 Envelope glycoprotein and target cell membranes is initiated by binding of the viral Envelope surface subunit gp120 to the CD4 receptor on the surface of the $CD4^+$ T cells [11]. This interaction creates a high affinity binding site for a chemokine coreceptor like CXCR4 and/or CCR5, necessary for HIV-1 entry into the target cell and subsequent infection [12]. The C-C chemokine receptor type 3 and C-C chemokine receptor type 2 are alternatives coreceptors with CD4 for HIV-1 infection [13]. This fusion introduces the contents of the virion into the cytoplasm of the cell, setting the stage for reverse transcription and thus to convert their RNA genomes into DNA.

The reverse transcription is an essential step in retroviral replication [14]. Once the reverse transcription has occurred, the integrase enzyme facilitates the incorporation of HIV-1 proviral DNA into the host cell genome and catalyses a function vital to viral replication [15]. After, the HIV protease plays a crucial role in the viral life cycle by processing polyproteins into structural and functional proteins essential for viral maturation [16, 17]. In Fig. (**2**), you can see some viral and human targets involved in the antiretroviral therapy. The 3D structural models of targets for anti-HIV inhibitors illustrated in this figure were downloaded from PDBe (http://www.ebi.ac.uk/pdbe/). The proteins illustrated in Fig. (**2**) were Human CXCR4 (file: 3oe8) [18], Human CCR5 (file: 4mbs)[19], HIV-1 Reverse Transcriptase (file:4mfb) [20], HIV integrase (file:3vqe) [21], HIV protease (file:4he9) [22].

Table 1. Review of ChEMBL outcomes for assays of anti-HIV drugs.

CHEMBLID	Target	Accession	Target Type	Organism	Compounds	Activities
CHEMBL274	C-C chemokine receptor type 5	P51681	protein	*H. sapiens*	2922	4740
CHEMBL4015	C-C chemokine receptor type 2	P41597	protein	*H. sapiens*	2567	4674
CHEMBL2414	C-C chemokine receptor type 4	P51679	protein	*H. sapiens*	1335	2489
CHEMBL2107	C-X-C chemokine receptor type 4	P61073	protein	*H. sapiens*	550	967
CHEMBL3473	C-C chemokine receptor type 3	P51677	protein	*H. sapiens*	1276	1550
CHEMBL5412	C-C chemokine receptor type 2	P51683	protein	*M. musculus*	74	90
CHEMBL3676	C-C chemokine receptor type 5	P51682	protein	*M. musculus*	63	65
CHEMBL378	HIV type 1		organism	*H. sapiens*	16547	41411
CHEMBL243	HIV type 1 protease	Q72874	protein	*H. sapiens*	5503	8268
CHEMBL247	HIV type 1 reverse transcriptase	Q72547	protein	*H. sapiens*	3292	7187
CHEMBL380	HIV type 2		organism	*H. sapiens*	1703	2834
CHEMBL613758	HIV		organism	*H. sapiens*	1421	2718
CHEMBL3471	HIV type 1 integrase	Q7ZJM1	protein	*H. sapiens*	1269	2910
CHEMBL612359	HIV3		organism	*H. sapiens*	104	136
CHEMBL613498	HIV type 1		organism	*H. sapiens*	76	76
CHEMBL4609	HIV type 1 Tat protein	P04326	protein	*H. sapiens*	71	75
CHEMBL3520	Envelope polyprotein GP160	P04578	protein	HIV type 1	129	330

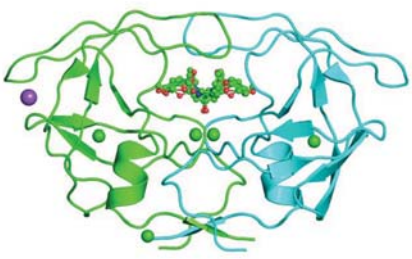


CCR5 Chemokine receptor

CXCR4 Chemokine receptor

HIV-1 Reverse Transcriptase

HIV-1 Integrase



HIV-1 Protease

Fig. (2). 3D structural models of targets for anti-HIV inhibitors downloaded from PDBe (http://www.ebi.ac.uk/pdbe/): Human CXCR4 (file: 3oe8) [18], Human CCR5 (file: 4mbs)[19], HIV-1 Reverse Transcriptase (file:4mfb) [20], HIV integrase (file:3vqe) [21], HIV protease (file:4he9) [22].

In Fig. (**3**), we illustrate the structure of some compounds in the fight against HIV. Additionally, the antiretroviral therapy includes the entry inhibitors; they can be subdivided into distinct sub-classes. The first one is the fusion inhibitors, like Enfuvirtide, that inhibits entry of HIV into the CD4 cell because they bind to glycoprotein gp41 (a protein on the viral membrane). This sub-class prevents fusion of the virus and the CD4 cell membrane [23]. The second one is the CCR5 inhibitors, like Maraviroc, it binds to the CCR5 receptor on the membrane of human cells such as CD4 cells. This binding prevents the interaction of HIV-1 gp120 and human CCR5, which is necessary for entry into the cell [11]. Another type of anti-HIV drugs are the Nucleoside/Nucleotide Reverse Transcriptase Inhibitors (NRTI's). They inhibit the viral reverse transcriptase enzyme, which is responsible for transcribing viral RNA into double stranded DNA. Some examples of this class of drugs are: Zidovudine Didanosine, Zalcitabine, Stavudine, Lamivudine, Abacavir, Tenofovir, Emcitrabine [24]. There are also Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs) that inhibit the viral reverse transcriptase enzyme. NNRTIs bind directly to the reverse transcriptase enzyme [25]. Currently there are four: Nevirapine, Delavirdine, Efavirenz, Etravirine [26]. Among other important anti-HIV compounds are the integrase inhibitors. The integration is an essential element of the HIV type 1 (HIV-1) replication cycle, allowing the transfer of virally encoded DNA into the host chromosome before replication, the Raltegravir, Elvitegravir and Dolutegravir are integrase inhibitors [27, 28].

Last, the protease inhibitors are a different type of anti-HIV compounds, the protease is necessary to form a fully mature, functional virus that is capable to replicate and produce more virus. Some examples of this kind of drugs are Amprenavir, Atazanavir, Indinavir, Nelfinavir, Lopinavir, Saquinavir, Tipranavir, Ritonavir [26, 29].

CHEMOINFORMATICS METHODS FOR ANTI-HIV DRUGS DISCOVERY

In this situation, *in silico* methods allow an efficient characterization of structure-activity relationships (SARs) and ease building of diverse models to capture and codify one or several SARs, which can be used to predict activities for new molecules [30]. To increase the accuracy, artificial intelligence techniques have been applied to Quantitative structure-activity relationship (QSAR) or quantitative structure-property relationships (QSPR) analysis since the late 1980s [31-33]. Magalhães *et al.* [34] studied two- and three-dimensional QSAR (2D/3D-QSAR) on a series of HIV-1 integrase inhibitors (HIV-1 IN) contributing to the design of new and more potent derivatives. Saranya and Selvaraj [35] developed QSAR studies on HIV-1 protease inhibitors and this models have potential application in the prediction of binding affinity for the newly synthesized compounds. Debnath [36] reviewed the applications of 3D-QSAR studies in anti-HIV-1 drug design during the last decade, and highlighted that the effort of the structure-based drug design was really successful in identifying several drugs that are currently available for the treatment of HIV-1, and other applications such as understanding the drug-receptor interactions and help in the design of effective analogs. Some authors [37-39] indicate that the results of their *in silico* studies provide a useful contribution to the design of novel active molecules for the inhibition of some target protein involved in the HIV.

MULTISCALE MODELS OF ANTI-HIV DRUG PRECLINICAL ACTIVITY *VS* AIDS EPIDEMIOLOGY

A useful chemoinformatics-pharmacoepidemiology model must be multi-level to account molecular and population structure. We need to process diverse types of input data. Initially, we need the information about the anti-HIV drugs, such as chemical structure of the drug (level i) and preclinical information, like biological targets (level ii), organisms (level iii), or assay protocols (level iv). Afterwards, we need to incorporate population structure descriptors (level v) that quantify the epidemiological and socioeconomic factors affecting the population selected for the study. Last, as populations in modern society are not close systems we should quantify also the effect of interaction of the population under study with other populations that may influence the pharmacoepidemiology study (level vi). Data for levels i-iv can be obtained from public databases of biological activity of organic compounds. In addition, we can obtain data of levels v and vi from public epidemiological databases like AIDSVu (mentioned above). We can talk about three characteristic of the problem resultant from the connection of chemical, pharmacological, and epidemiological information: (1) multi-targeting, (2) multi-objective, and/or (3) multi-scaling features. Multi-targeting [40-42], refers to the existence of compounds that can interact with more than one biological target. Multi-objective optimization problem (MOOP) [43-47] refers to the necessity of prediction/optimization of results for different experimental measures obtained in different assays. Last, multi-scaling refers to the different structural levels of the organization (i-vi) of matter the input variables. It means that we need to develop models able to link the changes in the prevalence of AIDS in a given a^{th} population with the changes in the biological activity of the q^{th} drug (d_q), due to variations in chemical structure, detected in preclinical assays carry out under a set of j^{th} conditions (c_j).

REVIEW OF INFORMATION INDICES POTENTIALLY USEFUL FOR AIDS/HIV MULTISCALE STUDIES

Shannon entropy measures are universal information indices helpful in many disciplines. Shannon developed the original form of these parameters in a paper about information theory [48]. Godden *et al*. [49-51] have used Shannon entropy parameters as molecular descriptors to seek QSPR models with different applications. Other authors such as, Acharya *et al.* [52] used non-linear methods including Shannon's entropy in their study about Coronary artery disease (CAD) to evaluate the heart rate of normal and CAD-affected heart rate signals. Many other authors used Shannon's entropy parameters to encode small molecule structure [53-56]. Graham *et al.* [57-62] used entropy measures to study the information properties of organic molecules.

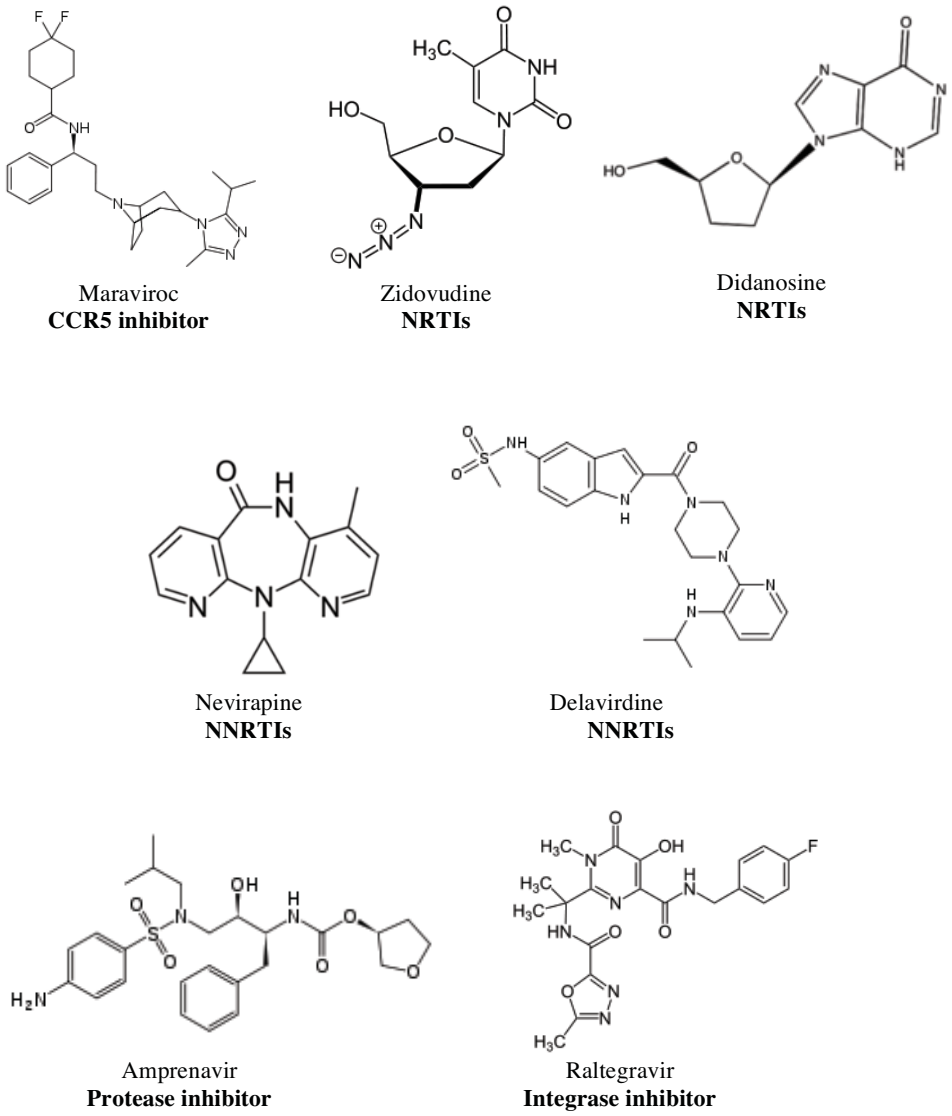


Fig. (3). Chemical structure of some anti-HIV drugs.

Symmetry Information Indices of Molecular Structure

We can use quantitative descriptors of the molecular graph of the drug. In particular, some of these parameters are useful to quantify information about the properties of biological, molecular, and/or social systems (information measures). We are going to use the information indices of neighborhood symmetry. In this work, we also call them symmetry information indices. Several authors in their publications report the use of some of this indices [63], such as Sharma *et al*. [64] used the bond and structural information content of five-order neighborhood symmetry, among others descriptors to optimize the caspase-3 inhibitory activity of isoquinoline-1,3,4-trione derivatives. Singh *et al.* [65] developed a statistical significant model considering the topological descriptors, including the indices of neighborhood symmetry in their study about the malonyl-CoA decarboxylase (MCD) inhibition activity. Prabhakar *et al.* [66] modeling the HIV-1 RT inhibitory activity of 2-(2,6-dihalophenyl)-3-(substituted pyridin-2-yl)-thiazolidin-4-ones with different topological descriptors obtained from Dragon software. Singh and Shekhawat [67] used the index of 5-order neighborhood symmetry, and highlighted the role of this descriptor in their paper of antimalarial activity of natural and synthetic prodiginines. These indices are calculated for H-included molecular graph and based on neighbor degrees and edge multiplicity [63, 68]. The symmetry information indices are calculated by partitioning graph vertices into equivalence classes; the topological equivalence of two vertices is that the corresponding neighborhoods of the k^{th} order are the same. DRAGON version 5.3 Software [69], calculates information indices of neighborhood symmetry from order 0 up to 5 (see Table **2**). The names, symbols (IC_k, TIC_k, SIC_k, BIC_k, CIC_k), and formula for the calculation of symmetry information indices of molecules appear at follow:

There are five types of neighborhood symmetry indices: the first one is the neighborhood information content (IC_k). The IC_k uses the following parameters: A_g is the cardinality of the g^{th} equivalence class and *n*AT is the total number of atoms. This index represents a measure of structural complexity per vertex. The formula of IC_k is:

$$IC_k = -\sum_{g=1}^{G} \frac{A_g}{nAT} \cdot \log_2 \frac{A_g}{nAT} \quad (1)$$

The next index is the neighborhood total information content (TIC_k), it represents a measure of the graph complexity. The formula is as follows:

$$TIC_k = nAT \cdot IC_k \quad (2)$$

The third index is the structural information content (SIC_k), it is calculated as you can see in the formula, in a normalized form of IC_k to delete the influence of graph size:

$$SIC_k = \frac{IC_k}{\log_2 nAT} \quad (3)$$

The fourth index is the bonding information content (BIC_k), this index is calculated in a normalized form of the IC_k taking into account the number of bonds and their multiplicity. It uses the *n*BT parameter, which is the number of bonds and * is the conventional bond order (1 for single, 2 for double, 3 for triple and 1.5 for aromatic bonds). The formula is:

$$BIC_k = \frac{IC_k}{\log_2 \left(\sum_{b=1}^{nBT} \pi_b^* \right)} \quad (4)$$

Moreover, the last index is the complementary information content (CIC_k). It measures the deviation IC_k from its maximum value that corresponds to the vertex partition into equivalence classes containing one element each. The formula CIC_k is:

$$CIC_k = \log_2 nAT - IC_k \quad (5)$$

Markov-Shannon Information Indices of Income-Inequality Complex Networks

We have used Markov chains to calculate Shannon information indices of different systems including simulations of disease spreading relevant to epidemiology [70]. We can define the vector of initial absolute probabilities ${}^0\mathbf{p} \equiv [{}^0p_1, {}^0p_2, {}^0p_3..., {}^0p_a..., {}^0p_{nt}]$ for the n_t counties in the same state. We calculate the absolute probability of occurrence of the disease in a given county ${}^0p_a$ at the initial time t_0 like in other Markov chain models [71-73]:

$${}^0p_a = \frac{G_a}{\sum_{a=1}^{nt} G_c} \quad (1)$$

Here, G_a is the Gini coefficient of income inequality [74] in the a^{th} county of a given state (s) of the U.S. We should to considerate that the only epidemiological factor used as input to calculate the Shannon information indices of the county was the G_a measure of income inequality. G_a measure of income-inequality is widely used as descriptor to approach the study of the epidemiology of different diseases [75, 76]. Haidicha and Ioannidis [77] conclude in their study about Gini coefficient in multicenter clinical studies, that this measure may be routinely incorporated in the description of the characteristics of a clinical study. Using the Chapman-Kolmogorov equation, we can calculate the vector ${}^k\mathbf{p}^t \equiv [{}^kp_1, {}^kp_2, {}^kp_3..., {}^kp_a..., {}^kp_{na}]$ for the absolute probabilities ${}^kp_a$ along time t_k:

$${}^k\mathbf{p} = \left({}^1\mathbf{\Pi}\right)^k \cdot {}^0\mathbf{p} = {}^k\mathbf{\Pi} \cdot {}^0\mathbf{p} \quad (7)$$

Consequently, elements of the stochastic matrix ${}^k\mathbf{\Pi} = ({}^1\mathbf{\Pi})^k$ are the probabilities ${}^kp_{ab}$ of transmission of AIDS from one county to other in $t_k = k$ years (steps). We calculated the elements of the stochastic matrix ${}^1\mathbf{\Pi}$ as follow [71-73, 78]:

$${}^1p_{ab} = \frac{\left(G_a + G_b\right) \cdot \exp\left(G_b\right)}{\sum_{c=1}^{c=n} \left(G_a + G_c\right) \cdot \exp\left(G_c\right)} \quad (8)$$

Subsequently, we calculated the information indices ${}^aI_k(s)$ to quantify the expected income inequality/epidemiology in the counties and their neighbors using Shannon formula of entropy [70].

$${}^aI_k(s) = -\left({}^kp_a\right) \cdot \log\left({}^kp_a\right) \quad (9)$$

Moving Average (MA) Operators

The codification of the chemical structure of the compounds is the first step here. We have data about a large number of assays developed in very different conditions (c_j) for equal or different targets (molecular or not). The non-structural information here refers to different assay conditions (c_j) like concentrations, temperature, targets, organisms, *etc.* A solution may rely upon the use of the idea of Moving Average (MA) operators used in time series analysis with a similar purpose [79]. Many authors have developed Autoregressive Integrated Moving-Average (ARIMA) and other MA models based on the initial work of Box and Jenkins [80]. Langenfeld *et al.* [81] use the ARIMA models to determine the effects of the hypnotic intervention in their study about control of HIV/AIDS-related pain. Gupta *et al.* [82] use moving average analysis for predicting anti-HIV activity using a novel topological descriptor: the eccentric adjacency index, and conclude that the proposed index offers a great potential for structure-activity/property studies. Chen *et al.* [83] studied the timing and magnitude in trend for tuberculosis cases in the United States, using a combination of ARIMA and Bayesian methods and summarize the advantages of this methods for the estimation and interpretation of operational data in public health or other areas. Gupta and Madan [84] studied the development of models for the prediction of HIV integrase inhibitory activity using MA analysis. For instance, Botella-Rocamora *et al.* [85] developed SMARS: Spatial Moving Average Risk Smoothing; a model to map diseases. Two type of parameters D^q_{kj} and $\langle D^q_k \rangle_j$ are necessary to calculate a MA. The variable D^q_k is one input parameters of type (k) with average values $\langle D^q_k \rangle_j$ for all q^{th} cases measured in a set of

Table 2. Neighborhood symmetry indices calculated by Dragon v5.3 software.

k=Order	f=Family	Index $^{q}IC_{kf}$	Dragon Symbol	Family [a]
0	1	$^{q}IC_{01}$	IC0	IC index
1		$^{q}IC_{11}$	IC1	IC index
2		$^{q}IC_{21}$	IC2	IC index
3		$^{q}IC_{31}$	IC3	IC index
4		$^{q}IC_{41}$	IC4	IC index
5		$^{q}IC_{51}$	IC5	IC index
0	2	$^{q}IC_{02}$	BIC0	Bond IC
1		$^{q}IC_{12}$	BIC1	Bond IC
2		$^{q}IC_{22}$	BIC2	Bond IC
3		$^{q}IC_{32}$	BIC3	Bond IC
4		$^{q}IC_{42}$	BIC4	Bond IC
5		$^{q}IC_{52}$	BIC5	Bond IC
0	3	$^{q}IC_{03}$	CIC0	Complementary IC
1		$^{q}IC_{13}$	CIC1	Complementary IC
2		$^{q}IC_{23}$	CIC2	Complementary IC
3		$^{q}IC_{33}$	CIC3	Complementary IC
4		$^{q}IC_{43}$	CIC4	Complementary IC
5		$^{q}IC_{53}$	CIC5	Complementary IC
0	4	$^{q}IC_{04}$	SIC0	Structural IC
1		$^{q}IC_{14}$	SIC1	Structural IC
2		$^{q}IC_{24}$	SIC2	Structural IC
3		$^{q}IC_{34}$	SIC3	Structural IC
4		$^{q}IC_{44}$	SIC4	Structural IC
5		$^{q}IC_{54}$	SIC5	Structural IC
0	5	$^{q}IC_{05}$	TIC0	Total IC index
1		$^{q}IC_{15}$	TIC1	Total IC index
2		$^{q}IC_{25}$	TIC2	Total IC index
3		$^{q}IC_{35}$	TIC3	Total IC index
4		$^{q}IC_{45}$	TIC4	Total IC index
5		$^{q}IC_{55}$	TIC5	Total IC index

[a] Neighborhood Symmetry Information Content (IC).

experimental conditions (c_j) (not necessarily a molecular descriptor). The general formula of a Box-Jenkins MA operator is:

$$\Delta D^{q}{}_{kj} = D^{q}{}_{k} - \left\langle D^{q}{}_{k} \right\rangle_{j} \tag{10}$$

ALMA Models

We have developed a similar approach called ALMA (Assessing Links with Moving Averages) using also MA operators (see next section). ALMA models remember those used in ARIMA models of time series analysis. They are adaptable to all molecular descriptors and/or graphs invariants or descriptors for complex networks. In consonance with the previous section, we use a similar terminology. The inputs of one ALMA model are the descriptors $D^{q}{}_{k}$ of type k^{th} of the q^{th} system (compound or drug d_q in this case) represented by a matrix **M**. On the other hand, the outputs of one ALMA model are the links ($L_{aq} = 1$ or $L_{aq} = 0$) of a complex network with Boolean matrix **L** and formed by different pairs of input systems. Consequently, the general linear equation of the model using a generic descriptor or graph theoretical invariant $D^{q}{}_{k}$ has the following general form:

$$\begin{aligned} S_{aqj} &= \sum_{k=1}^{k=k\max} e_k \cdot D^{q}{}_{k} + \sum_{k=1}^{k=k\max} \sum_{j=1}^{j=j\max} e_{kj} \cdot \Delta D^{q}{}_{kj} + e_0 \\ &= \sum_{k=1}^{k=k\max} e_k \cdot D^{q}{}_{k} + \sum_{k=1}^{k=k\max} \sum_{j=1}^{j=j\max} e_{kj} \cdot \left(D^{q}{}_{k} - \left\langle D^{q}{}_{k} \right\rangle_{j} \right) + e_0 \end{aligned} \tag{11}$$

The output dependent variable is $S_{aqj} = S_{aq}(c_j) = S_{aqj}(c_1, c_2, c_3, \ldots c_{max})$. This variable is a numerical score of of the formation of links ($L_{aq} = 1$ or $L_{aq} = 0$) in the complex network to be predicted. In the particular case of drug-target networks is the score for the biological activity of the q^{th} drug (d_q) *vs* the a^{th} target measured in one assay carried out under the set of conditions c_j. We have published different papers with this methodology. Tenorio-Borroto *et al.* [79] studied the quantitative structure-toxicity relationships (QSTR) of drugs with entropy models for multiplex drug-target interaction endpoints. The authors Alonso *et al.* and Luan *et al.* [86, 87] introduced new QSAR models to evaluate the neurotoxicity/neuroprotective properties of drugs for the treatment of neurodegenerative diseases. Others authors have used the same methodology. For instance, Speck-Planche and Cordeiro [88-90] have reported different multi-target models using the same type of ALMA approach with molecular descriptors D^q_k of different types. They developed a multitasking QSAR model for the simultaneous prediction of anti-streptococci activity and toxic effects of drugs [88]. They developed multi-target approaches for prediction of drugs in different classes of cancer [89-92]. They also proposed ALMA models of multitarget inhibitors against different proteins associated with Alzheimer [93] and Tuberculosis [94]. Last, these authors introduced models for inhibitors for C-C chemokine receptors using sub-structural descriptors [95].

ALMA Models for AIDS/HIV Multiscale Studies

In a recent work [96], we constructed the first ALMA model for AIDS/HIV multiscale studies of U.S. at county level. We used as inputs of the model the Balaban information indices (I^q_k) of a given compound d_q and the Shannon information indices for the population (a^{th} county). We obtained the data about anti-HIV drugs activity from the database ChEMBL and we used the molecular smiles codes and the Balaban information indices to quantify information. We used Shannon information indices [49] to describe the information of the social network (income inequality characterized by Gini coefficient) [77]. The data about the AIDS prevalence and Gini coefficient at county level were obtained from the AIDSVu database. The ALMA parameters as inputs of ANNs trained/validated with 43,249 data points. The dataset included values of AIDS prevalence in 2310 U.S. counties *vs* ChEMBL results for 21,582 unique drugs, 9 viral or human protein targets, 4856 protocols, and 10 possible experimental measures. We trained different topologies of ANNs including Multilayer Perceptrons (MLPs) and Linear Neural Networks (LNNs). The LNN with Accuracy (Ac), Specificity (Sp), and Sensitivity (Sn) above 75% was the best model found. In Fig. (**4**), we show the workflow used in the present work to construct ALMA models for this problem.

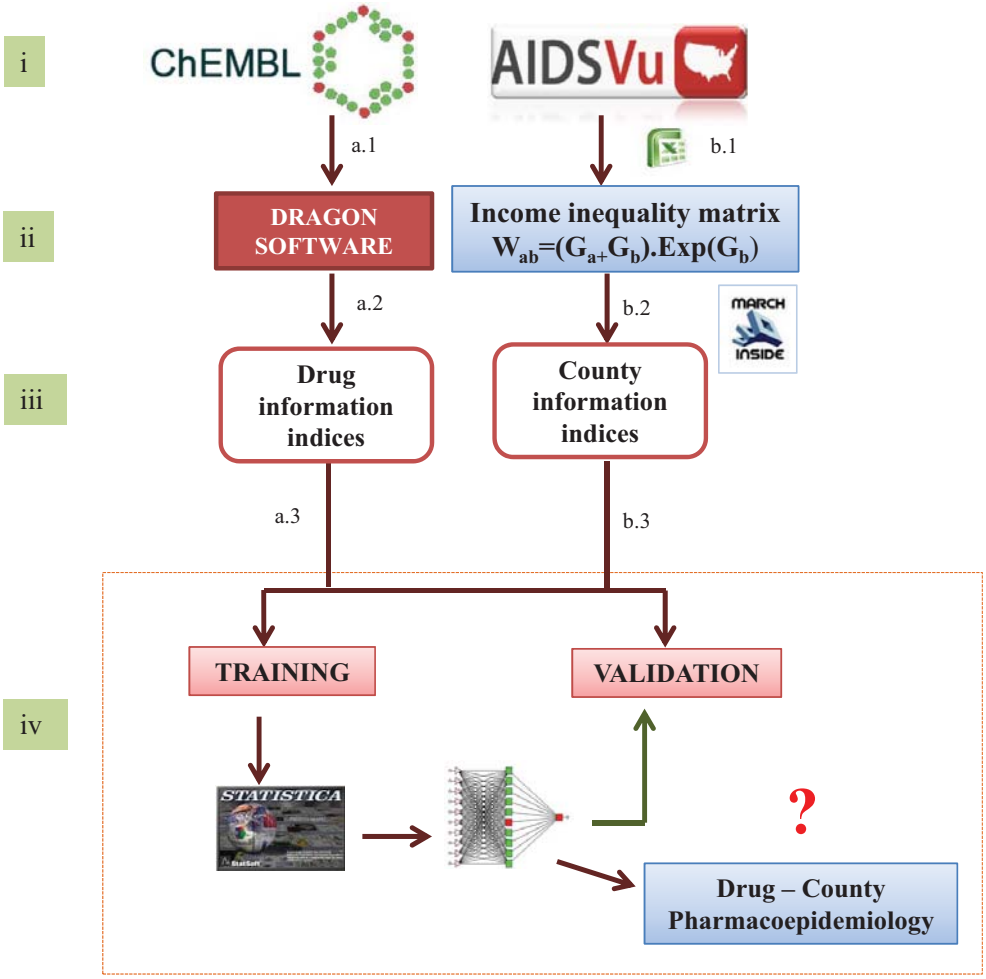


Fig. (4). Flowchart to construct the ANNs for the AIDS Pharmacoepidemiology model in U.S.

This type of models predict the formation of links (L_{aq}= 1) or not (L_{aq}= 0) in a complex network of AIDS pharmacoepidemiology in U.S. In the present context, we can use MA of properties of nodes of networks (drugs, proteins, organisms, counties, *etc.*) to predict the variable $L_{aq}(c_j)_{obs}$ in specific sub-set of conditions (c_j). This variable quantifies the formation of links between nodes. There are two different types of nodes forming this specific network. The first one represents the U.S. counties (a) and the second type of node characterizes the drugs (d_q). The value is $L_{aq}(c_j)$ = 1 when the Drug-Disease Ratio = $DDR_{aq}(c_j)$ > cutoff and $L_{aq}(c_j)_{obs}$ = 0 otherwise. In our previous work [96], we defined the ratio as follow $DDR_{aq}(c_j) = [D_q(c_j/D_a)]$. The term D_a is the AIDS prevalence rate for the a^{th} county and $D_q(c_j)$ is the biological activity of the q^{th} drug assayed in the conditions c_j. The general formula for a linear model developed using information indices was [96]:

$$\begin{aligned} S_{aqj} = & \sum_{k=1}^{k=4} e_k \cdot {I^q}_k + \sum_{k=1}^{k=4}\sum_{j=1}^{j=4} e_{kj} \cdot \Delta {I^q}_{kj} \\ & + e_{ak} \cdot {I^a}_k(t) + e_0 \\ = & \sum_{k=1}^{k=4} e_k \cdot {I^q}_k + \sum_{k=1}^{k=4}\sum_{j=1}^{j=4} e_{kj} \cdot \left({I^q}_k - \left\langle {I^q}_k \right\rangle_j \right) \\ & + e_{ak} \cdot {I^a}_k(t) + e_0 \end{aligned} \tag{12}$$

The reader should note that the predicted, output, or dependent variable S_{aqj} is not a discrete variable but a real-valued numerical score. However, the variable is directly proportional to the observed variable (L_{aq}). In general, c_j refers to different boundary conditions for the assay, *e.g.*, targets, cellular lines, organisms, experimental measures, *etc.* Therefore, c_1 = is the experimental measure of activity, c_2 = is the protein target, c_3 = is the organism that express the target, and c_4 = is the assay protocol *per se*. In order to seek the coefficients of the model we can use linear classification technique like ANNs implemented in the software package STASTITICA 6.0 [97]. The statistical parameters used to support the model were: Number of cases in training (N), and overall values of Sp, Sn, and Ac [98]. In the works of this series, we used specifically the type of descriptors called information indices (D = I). In any case, to calculate the MA we have to sum the values of I^q_k or all the n_j drugs with assay conditions c_j. Next, we divide this sum by the number of compounds n_j with this condition.

$$\Delta {I^q}_{kj} = {I^q}_k - \left\langle {I^q}_k \right\rangle_j \tag{13}$$

$$\left\langle {I^q}_k \right\rangle_j = \frac{1}{n_j} \sum_{q=1}^{q=n_j} {I^q}_k \tag{14}$$

NEW ALMA MODEL FOR AIDS/HIV MULTISCALE STUDIES

In the present paper, we changed the Balaban information indices (I^q_k) by Symmetry information content indices ($^qIC_{kf}$). However, we used the $I^a_k(s)$ indices to characterize the different populations. The new ALMA model developed using these other set of indices has the following general form:

$$\begin{aligned} S_{aqj} = & \sum_{k=0}^{k=5}\sum_{f=1}^{f=5} e_{kf} \cdot {}^qIC_{kf} \\ & + \sum_{k=0}^{k=5}\sum_{f=1}^{f=5}\sum_{j=1}^{j=4} e_{kfj} \cdot \Delta {}^qIC_{kfj} \\ & + \sum_{k=1}^{k=5} e_{ak} \cdot \Delta {I^a}_{ks} + e_0 \\ = & \sum_{k=0}^{k=5}\sum_{f=1}^{f=5} e_{kf} \cdot {}^qIC_{kf} \\ & + \sum_{k=0}^{k=5}\sum_{f=1}^{f=5}\sum_{j=1}^{j=4} e_{kfj} \cdot \left({}^qIC_{kf} - \left\langle {}^qIC_{kf} \right\rangle_j \right) \\ & + \sum_{k=1}^{k=5} e_k \cdot \left({I^a}_k - \left\langle {I^a}_k \right\rangle_s \right) + e_0 \end{aligned} \tag{15}$$

We used the software DRAGON [63] to calculate the $^qIC_{kf}$ indices for the molecules of the same dataset of anti-HIV drugs obtained from ChEMBL in the previous work [96]. In this case we calculated a total of $N_{indices} = N_k \cdot N_f = 6 \cdot 5 = 30$ values of $^qIC_{kf}$ indices with $N_k = 6$ different orders (k) that belong to $N_f = 5$ different families of descriptors (f). The families studied are the same reported in previous sections (see Table **2**). We developed different ANN models using all the set of parameters as well as simple models using different sub-sets of descriptors. At follow we discuss some of the more relevant results found.

ANN Models with All Descriptors

We obtained the ANN models using as input all descriptors. We used in total 30 $^qIC_{kf}$ indices of the molecules, 120 MA operators Δ^qIC_{kfj} for the different assay conditions for drugs (c_1,c_2,c_3,c_4), and 5 MA operators ΔI^a_{ks} for U.S. counties. It makes a total of $N_{inputs} = N_k \cdot N_f + N_k \cdot N_f \cdot N_j + {}^aN_k = 6 \cdot 5 + 6 \cdot 5 \cdot 4 + 5 = 30 + 120 + 5 = 155$ input values. In this counting formula N_k, N_f, N_j, and ^aN_k are the number of orders of molecular descriptors, families of descriptors, boundary conditions, and orders of county indices. The results obtained using the software STATISTICA show that the MLP [99] method fails to generate good prediction models. It presents values of Sp, Sn, and AUROC close to 50%. This values are typical of a random classifier and not the expected performance for a significant model [98]. Conversely, the LNN predictor based on the 155 descriptors classifies correctly above 76% of the cases in training and external validation sets (see Table **3**).

This model presented values of Sn = 76.13 and Sp = 76.51 in training and Sn = 77.04 and Sp = 76.48 in external validation sets. The LNN network shows values of AUROC = 0.82 in training and 0.82 for external validation set. We can conclude that the linear models seem to be better than non-linear to fit the present dataset. However, the number of inputs is very high to be considered a simple model.

LNN Models for Each Family of Information Indices

We decided to train LNN classifiers with each family of indices taking into consideration that the previous LNN model presented a very large number of inputs (155 in total) with respect to the performance obtained. In this sense, we

trained ANN predictors for each family of neighborhood symmetry indices separately. It makes a total of $N_{inputs} = N_k \cdot N_f + N_k \cdot N_f \cdot N_j + {}^aN_k = 6 \cdot 1 + 6 \cdot 1 \cdot 4 + 5 = 6 + 24 + 5 = 35$ input values at least in each model for one specific family. All LNN models obtained (five in total) classified correctly above 75% of the cases and AUROC above 0.80 in train and validation series regardless of the family of indices used (see Table **4**). This represents a spectacular 5-fold reduction of the number of parameters from 155 to above 30 parameters in all models. Interestingly, all models presented a very similar performance. It could be due to the co-linearity between the different information indices of drugs. A visual inspection of the formulae used to calculate the indices shows that all use as input the IC_k index. In fact, this index is the more relevant in because it is included in the formulas for calculate the remaining indices. We studied the correlation matrix of 435 pairs of information indices for the Anti-HIV drugs studied in this work. We found that 43% of these indices, *i.e.*, 187 pairs have a significant correlation $p < 0.05$. In addition, about 15% of the pairs of indices have a strong correlation $\geq \pm 0.7$.

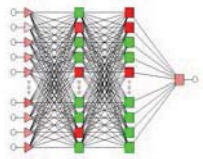
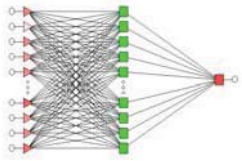
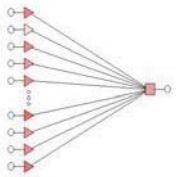
LNN Model with Information Index of Five Orders

After analysis of the previous results, we decided to test the predictive power of these indices in a simpler model. In so doing, we trained the LNN predictors using only each family of information indices of drugs ($^qIC_{5f}$) of 5- order, their MA operators (Δ^qIC_{5fj})) and the fifth MA operator of the U.S. counties (ΔI^a_{5s}). The LNN model based on $^qIC_{51}$ (LNN-IC_{51}) presented the higher values of Sn = 72.04/72.81 and Sp = 72.38/72.50 in training/ and external validation sets (see Table **5**). LNN-IC_{51} presented also the higher values for the AUROC in train and validation series (0.73 and 0.74 respectively). Analyzing all the previous results for this dataset, we found that the IC_k index appears to be the most important to predict the drug structure-activity relationships. We can conclude it by comparison to the other indices, which have lower values of classification. The equation of LNN-IC_{51} this model is the following:

$$\begin{aligned} S_{aq}(c_j) = & -25.48 \cdot {}^qIC_{51} + 1081.64 \cdot \Delta^qIC_{51}(c_1) \\ & + 29.36 \cdot \Delta^qIC_{51}(c_2) \\ & - 1084.52 \cdot \Delta^qIC_{51}(c_3) - 0.7727 \cdot \Delta^qIC_{51}(c_4) \\ & - 0.0792 \cdot \Delta I^a{}_5(s) - 0.5025 \end{aligned} \quad (16)$$

The Sensitivity analysis of this model, allowed us to quantify (rank) and order (ratio) into a sequence, the different chemoinformatics *vs* pharmacoepidemiology inputs, as you can see in the Fig. (**5**) and Table **6**. In the Fig. (**5**), we can see that the models give more relevance to the information about the molecular structure, parameters of type $^qIC_{kf}$. In second place the models rank the experimental measure used to measure the effectiveness of the drug $\Delta^qIC_{kf}(c_1)$. The third type of input is the organism $\Delta^qIC_{kf}(c_3)$, the fourth type of input in importance is the protein $\Delta^qIC_{kf}(c_2)$. The lasts ones are the information about the assay $\Delta^qIC_{kf}(c_4)$ and income inequality in the counties ΔI^a_{ks}. Whereas the LNN model with the $^qIC_{51}$ index gives a higher relevance to the information about the experimental measure $\Delta^qIC_{kf}(c_1)$, and in second place the model ranks information about the organism $\Delta^qIC_{kf}(c_3)$ used to measure the biological activity. Thus, the sensitivity analysis shows that the chosen

Table 3. ANN classifiers based on neighbourhood symmetry information indices.

ANN Models	Observed	$L_{pq} = 1$	$L_{pq} = 0$	$L_{pq} = 1$	$L_{pq} = 0$	AUROC
		Training		Validation		
MLP2 155:155-12-29-1:1 	Parameter [a]	Sn	Sp	Sn	Sp	(T / V)
	Predicted	52.44	52.68	51.11	51.95	0.53 / 0.52
	$L_{pq} = 1$	6009	9845	1948	3334	
	$L_{pq} = 0$	5449	10961	1863	3605	
MLP1 155:155-18-1:1 	Parameter [a]	Sn	Sp	Sn	Sp	(T / V)
	Predicted	52.70	52.74	51.27	51.92	0.53 / 0.52
	$L_{pq} = 1$	6039	9831	1954	3336	
	$L_{pq} = 0$	5419	10975	1857	3603	
LNN 155:155-1:1 	Parameter [a]	Sn	Sp	Sn	Sp	(T / V)
	Predicted	76.13	76.51	77.04	76.48	0.82 / 0.82
	$L_{pq} = 1$	8723	4887	2936	1632	
	$L_{pq} = 0$	2735	15919	875	5307	

[a] Parameter, Sp = Specificity, Sn = Sensitivity. Columns: Observed classifications Rows: Predicted classifications.

Table 4. LNN classifiers for the families of symmetry information indices from 0 to 5 order.

Type of Index	LNN Profile	Observed	$L_{pq} = 1$	$L_{pq} = 0$	$L_{pq} = 1$	$L_{pq} = 0$	AUROC
			Training		Validation		
$^{q}IC_{k1}$	31:31-1:1	Parameter [a]	Sn	Sp	Sn	Sp	(T / V)
		Predicted	75.88	76.17	76.59	76.27	0.81 / 0.81
		$L_{pq} = 1$	8695	4958	2919	1646	
		$L_{pq} = 0$	2763	15848	892	5293	
$^{q}IC_{k2}$	28:28-1:1	Parameter [a]	Sn	Sp	Sn	Sp	(T / V)
		Predicted	75.85	76.19	76.75	76.33	0.81 / 0.81
		$L_{pq} = 1$	8691	4952	2925	1642	
		$L_{pq} = 0$	2767	15854	886	5297	
$^{q}IC_{k3}$	29:29-1:1	Parameter [a]	Sn	Sp	Sn	Sp	(T / V)
		Predicted	75.80	76.16	76.62	76.35	0.81 / 0.81
		$L_{pq} = 1$	8686	4959	2920	1641	
		$L_{pq} = 0$	2772	15847	891	5298	
$^{q}IC_{k4}$	30:30-1:1	Parameter [a]	Sn	Sp	Sn	Sp	(T / V)
		Predicted	75.87	76.18	76.69	76.27	0.80/ 0.81
		$L_{pq} = 1$	8694	4954	2923	1646	
		$L_{pq} = 0$	2764	15852	888	5293	
$^{q}IC_{k5}$	32:32-1:1	Parameter [a]	Sn	Sp	Sn	Sp	(T / V)
		Predicted	74.84	75.31	75.67	75.25	0.78 / 0.79
		$L_{pq} = 1$	8576	5135	2884	1717	
		$L_{pq} = 0$	2882	15671	927	5222	

[a] Parameter, Sp = Specificity, Sn = Sensitivity. Columns: Observed classifications Rows: Predicted classifications.

model ranks the importance of factors in the following order (AIDS epidemiology / anti-HIV drug) ≈ organism in preclinical assay > experimental measure of activity > drug target > chemical structure of the drug > pharmacological assay> county income inequality.

In order to use this model in the future, the $^{q}IC_{51}$ information indices of the molecules, the average values of $^{q}IC_{51j}$ with the different boundary conditions, and the information of the counties in the U.S. (ΔI^{a}_{5s}) are in the Table **SM1**, Table **SM2**, and Table **SM3** respectively. In Table **7**, we illustrate some examples of values of $^{q}IC_{51}$ for drugs and ΔI^{a}_{5s} for counties of different states. In Table **8**, we show some examples of average values of $^{q}IC_{51}$ information descriptors of molecular structure for different boundary conditions.

Use of the LNN-IC_5 Model to Construct the AIDS Complex Network

Last, we used this LNN-ALMA model to generate/predict a complex network of the prevalence of AIDS in the United States at county level with respect to the preclinical activity of anti-HIV drugs. The bipartite network has two types of nodes (counties *vs* drug). Thus, this is a multiscale networksimilar to bipartite networks of drugs *vs* target proteins reported by other groups [100-104]. However, the nodes in the present network contain information about the molecules, i.e., chemical structure as well as assay conditions (target protein, organism, experimental measure, etc.). Additionally, the other set of nodes contain information about socioeconomic factors, such as the income inequality in the county.

Multiscale networks of this type have been discussed by Barabasi *et al.* [105] as one of the more important tools to perform trans-disciplinary research. The links of this complex network are the outputs $L_{aq}(c_j)_{pred} = 1$ of our model. We studied 43,249 data points to fit the model and predict the complex network. Consequently, we have to add the values of AIDS prevalence in 2310 counties in U.S. *vs* ChEMBL results for 21,582 unique drugs, 9 viral or human protein targets, 4856 protocols, and 10 possible experimental measures. In Fig. (**6**), we illustrate the sub-network of AIDS prevalence *vs* Anti-HIV drug preclinical activity for the state of Florida. For instance, the model predicts a high effectively for the drug Zidovudine [106] to treat AIDS in Nassau County. In Table **9** we include some examples of antiretroviral drugs with observed $L_{aq}(c_j)_{obs}$ and predicted $L_{aq}(c_j)_{pred}$ effects over AIDS prevalence in several counties of the same state in U.S.

Table 5. LNN classifiers for each family of symmetry information indices of 5-order.

Type of Index	Observed	$L_{pq} = 1$	$L_{pq} = 0$	$L_{pq} = 1$	$L_{pq} = 0$	AUROC
		Training		Validation		
$^{q}IC_{51}$	Parameter [a]	Sn	Sp	Sn	Sp	(T / V)
	Predicted	72.04	72.38	72.81	72.50	0.73 / 0.73
	$L_{pq} = 1$	8255	5746	2775	1908	
	$L_{pq} = 0$	3203	15060	1036	5031	
$^{q}IC_{52}$	Parameter [a]	Sn	Sp	Sn	Sp	(T / V)
	Predicted	69.36	69.81	70.21	69.95	0.73 / 0.74
	$L_{pq} = 1$	7948	6281	2676	2085	
	$L_{pq} = 0$	3510	14525	1135	4854	
$^{q}IC_{53}$	Parameter [a]	Sn	Sp	Sn	Sp	(T / V)
	Predicted	66.51	66.94	67.04	66.73	0.70/ 0.70
	$L_{pq} = 1$	7621	6878	2555	2308	
	$L_{pq} = 0$	3837	13928	1256	4631	
$^{q}IC_{54}$	Parameter [a]	Sn	Sp	Sn	Sp	(T / V)
	Predicted	67.69	68.03	68.56	67.99	0.73 / 0.73
	$L_{pq} = 1$	7757	6651	2613	2221	
	$L_{pq} = 0$	3701	14155	1198	4718	
$^{q}IC_{55}$	Parameter [a]	Sn	Sp	Sn	Sp	(T / V)
	Predicted	54.90	54.80	55.81	55.16	0.52 / 0.51
	$L_{pq} = 1$	5167	11403	1684	3828	
	$L_{pq} = 0$	6291	9403	2127	3111	

[a] Parameter, Sp = Specificity, Sn = Sensitivity. Columns: Observed classifications Rows: Predicted classifications.

Table 6. Parameters used for preprocessing data, carry out model predictions, and sensitivity analysis.

Data Operation	[a] Parameter / Effect	Drug $^{q}IC_{51}$	Measure $\Delta\,^{q}IC_{51}(c_1)$	Protein $\Delta\,^{q}IC_{51}(c_2)$	Organism $\Delta\,^{q}IC_{51}(c_3)$	Assay $\Delta\,^{q}IC_{51}(c_4)$	County $\Delta I^{a}_{5}(s)$
Preprocessing	Missing data value	0.9976678	0.9975952	0.9965166	0.9976395	0.9512190	0.5069762
	Linear shift parameter	0.0009935	0.0009935	0.0009937	0.0009935	0.0009945	0.0060280
	Scale coefficient	0.9925405	0.9975564	0.9964782	0.9976007	0.9512009	0.5070032
Model	LNN coefficient	-25.48	1081.64	29.36	-1084.52	-0.7727	0.0792
Sensitivity	Train ratio	1.047460	13.26668	1.062159	13.30156	1.000043	1.000075
	Train rank	4	2	3	1	6	5
	Validation ratio	1.139811	22.88760	1.172580	22.96505	1.000254	0.999974
	Validation rank	4	2	3	1	5	6

[a]Parameters used to transform data; LNN coefficients are the coefficients of the variables in the model, the independent term of the model is $e_0 = -0.5026$ (see equation in the text).

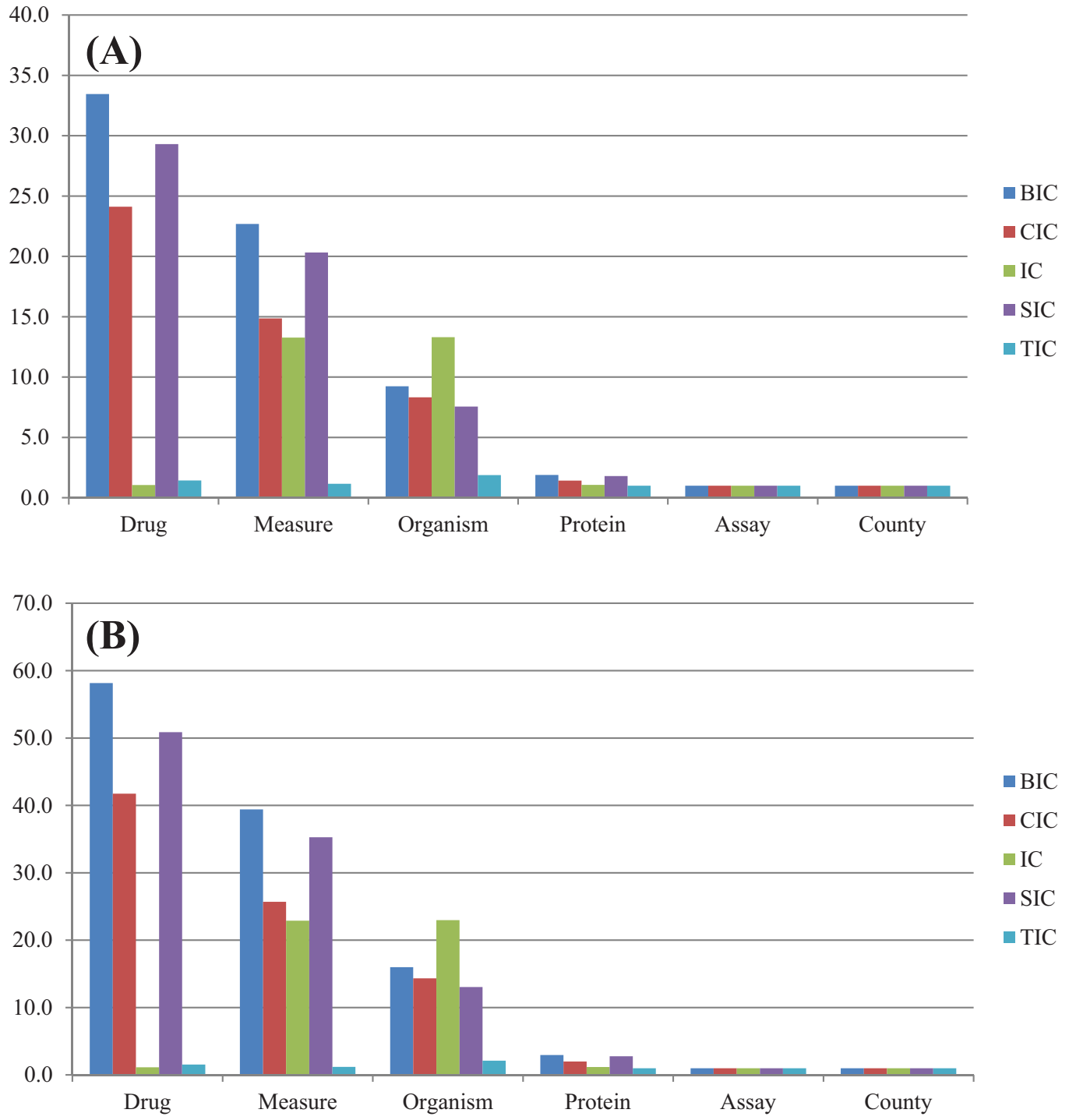


Fig. (5). Sensitivity analysis of LNN 6:6-1:1 classifiers. (A) training and (B) validation.

CONCLUSION

This work presents a review of several aspects of the disease, including the epidemiology, pathophysiology, treatments, etc. We also developed a model called LNN-ALMA to generate complex networks of the prevalence of AIDS in the counties of the U.S. with respect to the preclinical activity of anti-HIV drugs. The best classifier found was the LNN-$IC_{51;}$ this classifier has only six inputs based on neighborhood information content indices, compared to the other models, the IC_k index seems to be the most important to predict the drug structure-activity relationships. The new model has similar performance but is notably simpler than a previous model based on Balaban's information indices with >20 inputs. In future work, we will continue to improve the models and we will include other information indices, socioeconomic factors, machine-learning techniques, etc.

Table 7. Some values of IC^q_5 for drugs and $\Delta I^a_5(s)$ for counties of different states.

Name	ID	Target	Organism	$^qIC_{51}$
Delavirdine	593	HIV-1	HIV-1	5.169
Zidovudine	129	HIV-1	HIV-1	4.726
Entecavir	713	HIV	HIV	4.901
Lopinavir	729	HIV-1	HIV-1	5.629
Lamivudine	1230	HIV-1	HIV-1	5.114
Ciprofloxacin	8	HIV-1	HIV-1	4.678
Apigenin	28	HIV-1	HIV-1	4.64
Stavudine	991	HIV-1 RT	HIV-1	4.566
Hept	31871	HIV-1 RT	HIV-1	4.703
Quercetin	50	HIV-1 IN	HIV-1	5
Loviride	37624	HIV-1 PR	HIV-1	4.785
Vicriviroc	82301	CC-CKR-5	hsa	5.241
Dipyridyl	39879	CC-CKR-5	hsa	3.322
Plerixafor	18442	CXCR-4	hsa	4.647
Disulfiram	964	CC-CKR-2	hsa	2.642
County Name	**State (s)**	**D_a [i]**	**G_a [ii]**	**$\Delta I^a_5(s)$**
Autauga County	AL	181	0.405	9.1376
Arkansas County	AR	165	0.467	-1.4848
Apache County	AZ	124	0.488	-5.8417
Alameda County	CA	396	0.456	13.1466
Adams County	CO	179	0.403	2.1596
Fairfield County	CT	375	0.537	0.0
Kent County	DE	240	0.406	0.0
Baker County	FL	380	0.429	0.1173
Atkinson County	GA	256	0.447	-4.0347
Honolulu County	HI	201	0.422	0.0
Boone County	IA	58	0.407	-16.9929
Bannock County	ID	100	0.429	5.6071
Adams County	IL	65	0.453	-2.1575
Allen County	IN	136	0.428	2.738
Allen County	KS	44	0.394	0
Allen County	KY	71	0.42	-3.4678
Anderson County	KY	76	0.376	14.6872
Acadia Parish	LA	174	0.452	-4.7955
Ascension Parish	LA	178	0.409	9.3625
Berkshire County	MA	102	0.462	18.2985
Allegany County	MD	180	0.446	-38.5628
Calvert County	MD	124	0.369	12.8543
Hancock County	ME	73	0.437	0.2091

[i] D_a is the AIDS prevalence rate in the county p in 2010.
[ii] G_a is the Gini income-inequality measure of US county in 2010.

Table 8. Average values of information descriptors of molecular structure for different boundary conditions.

c_1	**Experimental Measure**	$N(c_1)$	$<^qIC_{51}>$
IC_{50}(nM)	Inhibitory concentration 50%	20332	5.049
EC_{50}(nM)	Effective concentration 50%	14981	5.139
K_i(nM)	Inhibitory constant	3736	5.422
IC_{95}(nM)	Inhibitory concentration 95%	1290	5.368
IC_{90}(nM)	Inhibitory concentration 90%	1118	5.022
ED_{50}(nM)	Effective dose 50%	860	5.036
EC_{50}(ug.mL^{-1})	Effective concentration	526	5.109
IC_{50}(ug.mL^{-1})	Inhibitory concentration	335	5.111
EC_{90}(nM)	Effective concentration	67	4.763
c_2	**Target Protein**	$N(c_2)$	
CC-CKR-5	C-C chemokine receptor type 5	2304	5.383
CC-CKR-2	C-C chemokine receptor type 2	2009	4.867
CC-CKR-3	C-C chemokine receptor type 3	1206	5.397
CC-CKR-4	C-C chemokine receptor type 4	345	5.256
CXCR-4	C-X-C chemokine receptor type 4	332	5.789
HIV-1 RT	HIV-1 reverse transcriptase	4029	4.922
HIV-1 IN	HIV-1 integrase	1702	3.769
HIV-1 PR	HIV-1 protease	5946	5.495
GP160	Envelope polyprotein GP160	34	4.952
c_3	**Organism**	$N(c_3)$	
HIV-1	HIV-1	34544	5.093
mmu	Mus musculus	68	5.536
hsa	Homo sapiens	6128	5.230
HIV-2	HIV-2	1030	5.330
HIV	HIV	1479	5.169
c_4	**Assay**	$N(c_4)$	
1033994	Antiviral activity against HIV1	282	4.874
708445	Effective conc. required for the inhibition of HIV-1 IIIB in MT-4 cells	176	5.835
859312	Inhibitory activity was determined against HIV type 1 protease	175	5.942
659084	Inhibitory conc. for displacement of [125I]-MIP-1 alpha from recombinant human CCR5 expressed in CHO cell	141	5.789
974332	Displacement of [125I]MIP1alpha from human CCR5 expressed in CHO cells	109	5.090

Table 9. Examples of antiretroviral drugs with observed $L_{aq}(c_j)_{obs}$ and predicted $L_{aq}(c_j)_{pred}$ effects.

Compound ID	$L_{aq}(c_j)_{obs}$	$L_{aq}(c_j)_{pred}$	c-level	Compound Name	Target	Organism	Assay ID	Measure	State, County
57	1	1	0.60	Nevirapine	HIV-1 RT	HIV-1	804261	EC_{50}(nM)	VA, Covington city
57	1	1	0.60	Nevirapine	HIV-1 RT	HIV-1	804184	EC_{50}(nM)	VA, Culpeper
114	1	1	0.57	Saquinavir	HIV-1	HIV-1	833275	EC_{50}(nM)	KY, Jackson
114	1	1	0.57	Saquinavir	HIV-1	HIV-1	833117	EC_{50}(nM)	KY, Montgomery
115	1	1	0.57	Indinavir	HIV-1	HIV-1	935845	EC_{50}(nM)	KY, Pulaski
115	1	1	0.58	Indinavir	HIV-1	HIV-1	952983	EC_{50}(nM)	KY, Rowan
116	1	1	0.58	Amprenavir	HIV-1	HIV-1	935841	EC_{50}(nM)	MA, Berkshire
116	1	1	0.58	Amprenavir	HIV-1	HIV-1	935840	EC_{50}(nM)	MA, Hampshire
116	1	1	0.28	Amprenavir	HIV-1 PR	HIV-1	828911	K_i(nM)	CA, Lake
116	1	1	0.28	Amprenavir	HIV-1 PR	HIV-1	830292	K_i(nM)	GA, Union
116	1	1	0.29	Amprenavir	HIV-1 PR	HIV-1	909837	K_i(nM)	IN, Clinton
116	1	1	0.28	Amprenavir	HIV-1 PR	HIV-1	909838	K_i(nM)	IN, Hancock
129	1	1	0.61	Zidovudine	HIV-1	HIV-1	1640111	EC_{50}(nM)	TN, Sevier
129	1	1	0.60	Zidovudine	HIV-1	HIV-1	1640126	EC_{50}(nM)	TN, Smith
57	1	0	0.70	Nevirapine	HIV-1	HIV-1	695511	IC_{50}(nM)	MS, Walthall
57	1	0	0.70	Nevirapine	HIV-1	HIV-1	695519	IC_{50}(nM)	MS, Wayne
38380	0	0	0.85	Fasudil	CC-CKR-2	hsa	915971	IC_{50}(nM)	GA, Baldwin
1185005	0	0	0.85	Cenicriviroc	CC-CKR-2	hsa	2184889	IC_{50}(nM)	MO, Laclede
39879	0	0	0.84	Dipyridyl	CC-CKR-5	hsa	2215079	IC_{50}(nM)	MS, Alcorn
82301	0	0	0.84	Vicriviroc	CC-CKR-5	hsa	1697613	IC_{50}(nM)	NJ, Essex
82301	0	0	0.84	Vicriviroc	CC-CKR-5	hsa	1697612	IC_{50}(nM)	NJ, Middlesex
1172035	0	0	0.84	Nifeviroc	CC-CKR-5	hsa	1174016	IC_{50}(nM)	MO, Laclede
1172035	0	0	0.84	Nifeviroc	CC-CKR-5	hsa	1174015	IC_{50}(nM)	MO, Macon

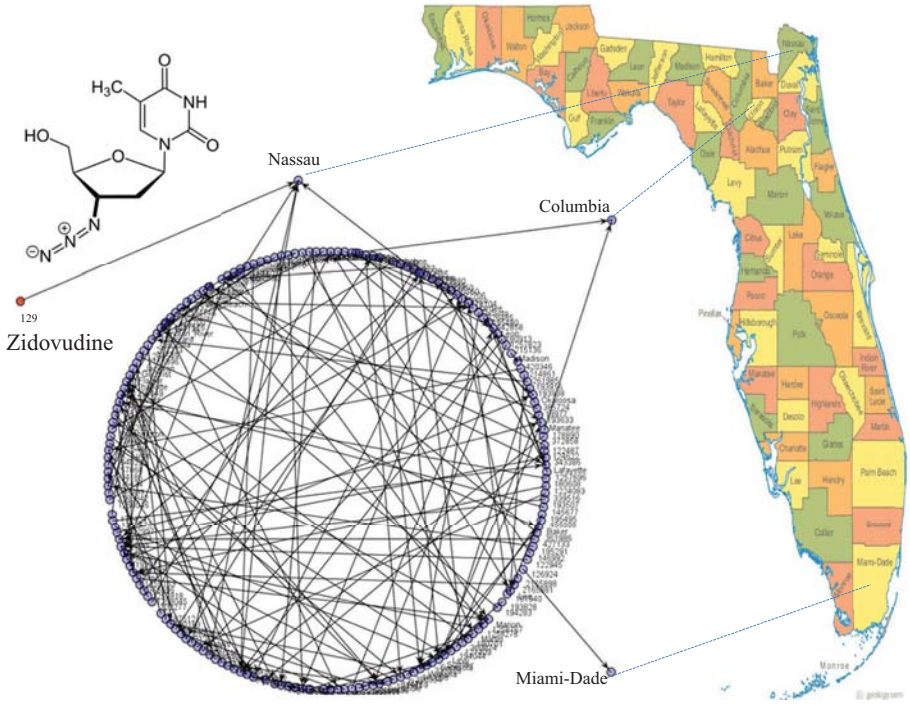


Fig. (6). Sub-network of AIDS prevalence *vs* Anti-HIV drug activity for U.S. state of Florida (FL).

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

R.O.M acknowledges financial support of FPI fellowship associated to research project (AGL2011-30563-C03-01) funded by MECD (Spanish Ministry of Education, Culture and Sport).

REFERENCES

[1] Lindemann D, Steffen I, Pohlmann S. Cellular entry of retroviruses. Adv Exp Med Biol 2013; 790: 128-49.

[2] Rinaldo CR. HIV-1 Infection of CD4 T Cells by Professional Antigen Presenting Cells. Scientifica (Cairo) 2013; 2013: 164203.

[3] Noorizadeh H, Sajjadifar S, Farmany A. A quantitative structure-activity relationship study of anti-HIV activity of substituted HEPT using nonlinear models. Med Chem Res 2013; 22: 5442-52.

[4] HIV surveillance--United States, 1981-2008. MMWR Morb Mortal Wkly Rep 2011; 60: 689-93.

[5] Moss JA. HIV/AIDS Review. Radiol Technol 2013; 84: 247-67; quiz p. 68-70.

[6] Piot P, Quinn TC. Response to the AIDS pandemic--a global health model. N Engl J Med 2013; 368: 2210-8.

[7] Heikamp K, Bajorath J. Large-scale similarity search profiling of ChEMBL compound data sets. J Chem Inf Model 2011; 51: 1831-9.

[8] Gaulton A, Bellis LJ, Bento AP, *et al*. ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res 2012; 40: D1100-7.

[9] Bento AP, Gaulton A, Hersey A, *et al*. The ChEMBL bioactivity database: an update. Nucleic Acids Res 2013.

[10] Alkhatib G. The biology of CCR5 and CXCR4. Curr Opin HIV AIDS 2009; 4: 96-103.

[11] Wilkin TJ, Gulick RM. CCR5 antagonism in HIV infection: current concepts and future opportunities. Annu Rev Med 2012; 63: 81-93.

[12] Blanpain C, Libert F, Vassart G, Parmentier M. CCR5 and HIV infection. Receptors Channels 2002; 8: 19-31.

[13] Tan JH, Ludeman JP, Wedderburn J, *et al*. Tyrosine sulfation of chemokine receptor CCR2 enhances interactions with both monomeric and dimeric forms of the chemokine monocyte chemoattractant protein-1 (MCP-1). J Biol Chem 2013; 288: 10024-34.

[14] Hu WS, Hughes SH. HIV-1 reverse transcription. Cold Spring Harb Perspect Med 2012; 2: pii: a006882.

[15] Karmon SL, Markowitz M. Next-generation integrase inhibitors : where to after raltegravir? Drugs 2013; 73: 213-28.

[16] Qiu X, Liu ZP. Recent developments of peptidomimetic HIV-1 protease inhibitors. Curr Med Chem 2011; 18: 4513-37.

[17] Castro HC, Abreu PA, Geraldo RB, *et al*. Looking at the proteases from a simple perspective. J Mol Recognit 2011; 24: 165-81.

[18] Wu B, Chien EY, Mol CD, *et al*. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. Science 2010; 330: 1066-71.

[19] Tan Q, Zhu Y, Li J, *et al*. Structure of the CCR5 chemokine receptor-HIV entry inhibitor maraviroc complex. Science 2013; 341: 1387-90.

[20] Lee WG, Gallardo-Macias R, Frey KM, *et al*. Picomolar Inhibitors of HIV Reverse Transcriptase Featuring Bicyclic Replacement of a Cyanovinylphenyl Group. J Am Chem Soc 2013; 135: 16705-13.

[21] Wielens J, Headey SJ, Rhodes DI, *et al*. Parallel screening of low molecular weight fragment libraries: do differences in methodology affect hit identification? J Biomol Screen 2013; 18: 147-59.

[22] Zhang H, Wang YF, Shen CH, *et al*. Novel P2 tris-tetrahydrofuran group in antiviral compound 1 (GRL-0519) fills the S2 binding pocket of selected mutants of HIV-1 protease. J Med Chem 2013; 56: 1074-83.

[23] Qian K, Morris-Natschke SL, Lee KH. HIV entry inhibitors and their potential in HIV therapy. Med Res Rev 2009; 29: 369-93.

[24] Perno CF. The discovery and development of HIV therapy: the new challenges. Ann Ist Super Sanita 2011; 47: 41-3.

[25] Saag MS. New and investigational antiretroviral drugs for HIV infection: mechanisms of action and early research findings. Top Antivir Med 2012; 20: 162-7.

[26] Arts EJ, Hazuda DJ. HIV-1 antiretroviral drug therapy. Cold Spring Harb Perspect Med 2012; 2: a007161.

[27] Powderly WG. Integrase inhibitors in the treatment of HIV-1 infection. J Antimicrob Chemother 2010; 65: 2485-8.

[28] Adams JL, Greener BN, Kashuba AD. Pharmacology of HIV integrase inhibitors. Curr Opin HIV AIDS 2012; 7: 390-400.

[29] Chougrani I, Luton D, Matheron S, Mandelbrot L, Azria E. In: HIV AIDS (Auckl). V: 2013; 5: pp. 253-62.

[30] Guha R. On exploring structure-activity relationships. Methods Mol Biol 2013; 993: 81-94.

[31] Burbidge R, Trotter M, Buxton B, Holden S. Drug design by machine learning: support vector machines for pharmaceutical data analysis. Comput Chem 2001; 26: 5-14.

[32] Manallack DT, Livingstone DJ. Neural networks in drug discovery: have they lived up to their promise? Eur J Med Chem 1999; 34: 195-208.

[33] Si H, Yuan S, Zhang K, Fu A, Duan Y-B, Zhide H. Quantitative structure activity relationship study on EC50 of anti-HIV drugs. Chemometr Intell Lab 2008; 90: 15-24.

[34] Magalhaes Ude O, Souza AM, Albuquerque MG, *et al*. Hologram quantitative structure-activity relationship and comparative molecular field analysis studies within a series of tricyclic phthalimide HIV-1 integrase inhibitors. Drug Des Devel Ther 2013; 7: 953-61.

[35] Saranya N, Selvaraj S. QSAR studies on HIV-1 protease inhibitors using non-linearly transformed descriptors. Curr Comput Aided Drug Des 2012; 8: 10-49.

[36] Debnath AK. Application of 3D-QSAR techniques in anti-HIV-1 drug design--an overview. Curr Pharm Des 2005; 11: 3091-110.

[37] Debnath U, Verma S, Jain S, Katti SB, Prabhakar YS. Pyridones as NNRTIs against HIV-1 mutants: 3D-QSAR and protein informatics. J Comput Aided Mol Des 2013; 27: 637-54.

[38] Sun XH, Guan JQ, Tan JJ, Liu C, Wang CX. 3D-QSAR studies of quinoline ring derivatives as HIV-1 integrase inhibitors. SAR QSAR Environ Res 2012; 23: 683-703.

[39] Swiderek K, Marti S, Moliner V. Theoretical studies of HIV-1 reverse transcriptase inhibition. Phys Chem Chem Phys 2012; 14: 12614-24.

[40] Hu Y, Bajorath J. Molecular scaffolds with high propensity to form multi-target activity cliffs. J Chem Inf Model 2010; 50: 500-10.

[41] Erhan D, L'Heureux P J, Yue SY, Bengio Y. Collaborative filtering on a family of biological targets. J Chem Inf Model 2006; 46: 626-35.

[42] Namasivayam V, Hu Y, Balfer J, Bajorath J. Classification of compounds with distinct or overlapping multi-target activities and diverse molecular mechanisms using emerging chemical patterns. J Chem Inf Model 2013; 53: 1272-81.

[43] Cruz-Monteagudo M, Cordeiro MN, Tejera E, Dominguez ER, Borges F. Desirability-based multi-objective QSAR in drug discovery. Mini Rev Med Chem 2012; 12: 920-35.

[44] Machado A, Tejera E, Cruz-Monteagudo M, Rebelo I. Application of desirability-based multi(bi)-objective optimization in the design of selective arylpiperazine derivates for the 5-HT1A serotonin receptor. Eur J Med Chem 2009; 44: 5045-54.

[45] Saiz-Urra L, Bustillo Perez AJ, Cruz-Monteagudo M, *et al*. Global antifungal profile optimization of chlorophenyl derivatives against Botrytis cinerea and Colletotrichum gloeosporioides. J Agric Food Chem 2009; 57: 4838-43.

[46] Cruz-Monteagudo M, Borges F, Cordeiro MN, *et al*. Desirability-based methods of multiobjective optimization and ranking for global QSAR studies. Filtering safe and potent drug candidates from combinatorial libraries. J Comb Chem 2008; 10: 897-913.

[47] Nicolaou CA, Brown N, Pattichis CS. Molecular optimization using computational multi-objective methods. Curr Opin Drug Discov Devel 2007; 10: 316-24.

[48] Shannon CE. A Mathematical Theory of Communication. Bell Syst Tech J 1948; 27: 379-423.

[49] Godden JW, Stahura FL, Bajorath J. Variability of molecular descriptors in compound databases revealed by Shannon entropy calculations. J Chem Inf Comput Sci 2000; 40: 796-800.

[50] Stahura FL, Godden JW, Bajorath J. Differential Shannon entropy analysis identifies molecular property descriptors that predict aqueous solubility of synthetic compounds with high accuracy in

binary QSAR calculations. J Chem Inf Comput Sci 2002; 42: 550-8.

[51] Stahura FL, Godden JW, Xue L, Bajorath J. Distinguishing between natural products and synthetic molecules by descriptor Shannon entropy analysis and binary QSAR calculations. J Chem Inf Comput Sci 2000; 40: 1245-52.

[52] Acharya UR, Faust O, Sree V, *et al*. Linear and nonlinear analysis of normal and CAD-affected heart rate signals. Comput Methods Programs Biomed 2014; 113: 55-68.

[53] Roy K, Saha A. Comparative QSPR studies with molecular connectivity, molecular negentropy and TAU indicesPart I: Molecular thermochemical properties of diverse functional acyclic compounds. J Mol Model (Online) 2003; 9: 259-70.

[54] Agrawal VK, Khadikar PV. Modelling of carbonic anhydrase inhibitory activity of sulfonamides using molecular negentropy. Bioorg Med Chem Lett 2003; 13: 447-53.

[55] Katritzky AR, Lomaka A, Petrukhin R, Jain R, Karelson M, Visser AE, *et al*. QSPR correlation of the melting point for pyridinium bromides, potential ionic liquids. J Chem Inf Comput Sci 2002; 42: 71-4.

[56] Katritzky AR, Perumal S, Petrukhin R, Kleinpeter E. Codessa-based theoretical QSPR model for hydantoin HPLC-RT lipophilicities. J Chem Inf Comput Sci 2001; 41: 569-74.

[57] Graham DJ, Schacht D. Base Information Content in Organic Molecular Formulae. J Chem Inf Comput Sci 2000; 40: 942.

[58] Graham DJ. Information Content in Organic Molecules: Structure Considerations Based on Integer Statistics. J Chem Inf Comput Sci 2002; 42: 215.

[59] Graham DJ, Malarkey C, Schulmerich MV. Information Content in Organic Molecules: Quantification and Statistical Structure *via* Brownian Processing. J Chem Inf Comput Sci 2004; 44.

[60] Graham DJ, Schulmerich MV. Information Content in Organic Molecules: Reaction Pathway Analysis *via* Brownian Processing. J Chem Inf Comput Sci 2004; 44.

[61] Graham DJ. Information Content and Organic Molecules: Aggregation States and Solvent Effects. J Chem Inf Model 2005; 45.

[62] Graham DJ. Information Content in Organic Molecules: Brownian Processing at Low Levels. J Chem Inf Model 2007; 47: 376-89.

[63] Todeschini R, Consonni V. Handbook of Molecular Descriptors: Weinheim, Germany 2000.

[64] Sharma BK, Pilania P, Singh P, Prabhakar YS. Combinatorial protocol in multiple linear regression/partial least-squares directed rationale for the caspase-3 inhibition activity of isoquinoline-1,3,4-trione derivatives. SAR QSAR Environ Res 2010; 21: 169-85.

[65] Singh P, Kumar R, Sharma BK, Prabhakar YS. Topological descriptors in modeling malonyl coenzyme A decarboxylase inhibitory activity: N-Alkyl-N-(1,1,1,3,3,3-hexafluoro-2-hydroxypropylphenyl)amide derivatives. J Enzyme Inhib Med Chem 2009; 24: 77-85.

[66] Prabhakar YS, Rawal RK, Gupta MK, Solomon VR, Katti SB. Topological descriptors in modeling the HIV inhibitory activity of 2-aryl-3-pyridyl-thiazolidin-4-ones. Comb Chem High Throughput Screen 2005; 8: 431-7.

[67] Singh P, Shekhawat N. Chemometric descriptors in the rationale of antimalarial activity of natural and synthetic prodiginines. J Curr Chem Pharm Sc 2012; 2: 244-60.

[68] Magnuson VR, Harriss DK, Basak SC. In: Studies in Physical and Theoretical Chemistry; King, RB. Elsevier. V: Amsterdam (The Netherlands) 1983; pp. 178-91.

[69] Todeschini R, Consonni V, Mauri A, Pavan M. DRAGON [computer program]. Milano, Italy: Talete srl; 2005.

[70] Riera-Fernandez P, Munteanu CR, Escobar M, *et al*. New Markov-Shannon Entropy models to assess connectivity quality in complex networks: from molecular to cellular pathway, Parasite-Host, Neural, Industry, and Legal-Social networks. J Theor Biol 2012; 293: 174-88.

[71] Gonzalez-Diaz H, Aguero G, Cabrera MA, *et al*. Unified Markov thermodynamics based on stochastic forms to classify drugs considering molecular structure, partition system, and biological species: distribution of the antimicrobial G1 on rat tissues. Bioorg Med Chem Lett 2005; 15: 551-7.

[72] Gonzalez-Diaz H, Cruz-Monteagudo M, Molina R, Tenorio E, Uriarte E. Predicting multiple drugs side effects with a general drug-target interaction thermodynamic Markov model. Bioorg Med Chem 2005; 13: 1119-29.

[73] Van Waterbeemd H. In: Chemometric methods in molecular design. VCH. V: New York, NY 1995; 2: pp. 265-82.

[74] Pabayo R, Kawachi I, Gilman SE. Income inequality among American states and the incidence of major depression. J Epidemiol Community Health 2013.

[75] Burns JK, Tomita A, Kapadia AS. Income inequality and schizophrenia: Increased schizophrenia incidence in countries with high levels of income inequality. Int J Soc Psychiatry 2013.

[76] Green C, Yu BN, Marrie RA. Exploring the implications of small-area variation in the incidence of multiple sclerosis. Am J Epidemiol 2013; 178: 1059-66.

[77] Haidich AB, Ioannidis JP. The Gini coefficient as a measure for understanding accrual inequalities in multicenter clinical studies. J Clin Epidemiol 2004; 57: 341-8.

[78] González-Díaz H, Prado-Prado FJ, Santana L, Uriarte E. Unify QSAR approach to antimicrobials. Part 1: Predicting antifungal activity against different species. Bioorg Med Chem 2006; 14 5973-80.

[79] Tenorio-Borroto E, Garcia-Mera X, Penuelas-Rivas CG, *et al*. Entropy model for multiplex drug-target interaction endpoints of drug immunotoxicity. Curr Top Med Chem 2013; 13: 1636-49.

[80] Box GEP, Jenkins GM. Time series analysis: Forecasting and control. Holden-Day: San Francisco, California 1970.

[81] Langenfeld MC, Cipani E, Borckardt JJ. Hypnosis for the control of HIV/AIDS-related pain. Int J Clin Exp Hypn 2002; 50: 170-88.

[82] Gupta S, Singh M, Madan AK. Predicting anti-HIV activity: computational approach using a novel topological descriptor. J Comput Aided Mol Des 2001; 15: 671-8.

[83] Chen MP, Shang N, Winston CA, Becerra JE. A Bayesian analysis of the 2009 decline in tuberculosis morbidity in the United States. Stat Med 2012; 31: 3278-84.

[84] Gupta M, Madan AK. Diverse models for the prediction of HIV integrase inhibitory activity of substituted quinolone carboxylic acids. Arch Pharm (Weinheim) 2012; 345: 989-1000.

[85] Botella-Rocamora P, Lopez-Quilez A, Martinez-Beneito MA. Spatial moving average risk smoothing. Stat Med 2013; 32: 2595-612.

[86] Alonso N, Caamano O, Romero-Duran FJ, *et al*. Model for High-Throughput Screening of Multi-Target Drugs in Chemical Neurosciences; Synthesis, Assay and Theoretic Study of Rasagiline Carbamates. ACS Chem Neurosci 2013.

[87] Luan F, Cordeiro MN, Alonso N, *et al*. TOPS-MODE model of multiplexing neuroprotective effects of drugs and experimental-theoretic study of new 1,3-rasagiline derivatives potentially useful in neurodegenerative diseases. Bioorg Med Chem 2013; 21: 1870-9.

[88] Speck-Planche A, Kleandrova VV, Cordeiro MN. Chemoinformatics for rational discovery of safe antibacterial drugs: simultaneous predictions of biological activity against streptococci and toxicological profiles in laboratory animals. Bioorg Med Chem 2013; 21: 2727-32.

[89] Speck-Planche A, Kleandrova VV, Luan F, Cordeiro MN. Chemoinformatics in multi-target drug discovery for anti-cancer therapy: *in silico* design of potent and versatile anti-brain tumor agents. Anticancer Agents Med Chem 2012; 12: 678-85.

[90] Speck-Planche A, Kleandrova VV, Luan F, Cordeiro MN. Chemoinformatics in anti-cancer chemotherapy: multi-target QSAR model for the *in silico* discovery of anti-breast cancer agents. Eur J Pharm Sci 2012; 47: 273-9.

[91] Speck-Planche A, Kleandrova VV, Luan F, Cordeiro MN. Rational drug design for anti-cancer chemotherapy: multi-target QSAR models for the *in silico* discovery of anti-colorectal cancer agents. Bioorg Med Chem 2012; 20: 4848-55.

[92] Speck-Planche A, Kleandrova VV, Luan F, Cordeiro MN. Multi-target drug discovery in anti-cancer therapy: fragment-based approach toward the design of potent and versatile anti-prostate cancer agents. Bioorg Med Chem 2011; 19: 6239-44.

[93] Speck-Planche A, Kleandrova VV, Luan F, Cordeiro MN. Multi-target inhibitors for proteins associated with Alzheimer: *in silico* discovery using fragment-based descriptors. Curr Alzheimer Res 2013; 10: 117-24.

[94] Speck-Planche A, Kleandrova VV, Luan F, Cordeiro MN. *In silico* discovery and virtual screening of multi-target inhibitors for proteins in Mycobacterium tuberculosis. Comb Chem High Throughput Screen 2012; 15: 666-73.

[95] Speck-Planche A, Kleandrova VV. *In silico* design of multi-target inhibitors for C-C chemokine receptors using substructural descriptors. Mol Divers 2012; 16: 183-91.

[96] González-Díaz H, Herrera-Ibatá DM, Duardo-Sanchez A, Munteanu CR, Orbegozo-Medina RA, Pazos A. Model of the Multiscale Complex Network of AIDS prevalence in US at county level *vs* Preclinical activity of anti-HIV drugs based on information indices of molecular graphs and social networks. J Chem Inf Model 2014; Submitted.

[97] StatSoft.Inc. STATISTICA (data analysis software system), version 6.0, www.statsoft.com.Statsoft, Inc. 6.0 ed2002.

[98] Hill T, Lewicki P. STATISTICS Methods and Applications. A Comprehensive Reference for Science, Industry and Data Mining: Tulsa 2006

[99] Rosenblatt F. Principles of neurodynamics; perceptrons and the theory of brain mechanisms: Washington D.C. 1962.

[100] Hecker N, Ahmed J, von Eichborn J, Dunkel M, Macha K, Eckert A, *et al*. SuperTarget goes quantitative: update on drug-target interactions. Nucleic Acids Res 2012; 40: D1113-7.

[101] Prado-Prado F, Garcia-Mera X, Escobar M, *et al*. 3D MI-DRAGON: new model for the reconstruction of US FDA drug-target network and theoretical-experimental studies of inhibitors of rasagiline derivatives for AChE. Curr Top Med Chem 2012; 12: 1843-65.

[102] Prado-Prado F, Garcia-Mera X, Abeijon P, *et al.* Using entropy of drug and protein graphs to predict FDA drug-target network: theoretic-experimental study of MAO inhibitors and hemoglobin peptides from Fasciola hepatica. Eur J Med Chem 2011; 46: 1074-94.

[103] Vina D, Uriarte E, Orallo F, Gonzalez-Diaz H. Alignment-free prediction of a drug-target complex network based on parameters of drug connectivity and protein sequence of receptors. Mol Pharm 2009; 6: 825-35.

[104] Araujo RP, Liotta LA, Petricoin EF. Proteins, drug targets and the mechanisms they control: the simple truth about complex networks. Nat Rev Drug Discov 2007; 6: 871-80.

[105] Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet 2011; 12: 56-68.

[106] Shey MS, Kongnyuy EJ, Alobwede SM, Wiysonge CS. Co-formulated abacavir-lamivudine-zidovudine for initial treatment of HIV infection and AIDS. Cochrane Database Syst Rev 2013; 3: CD005481.

Received: February 20, 2015 Revised: April 23, 2015 Accepted: June 26, 2015