Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

European Journal of Medicinal Chemistry 44 (2009) 4461–4469

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

Scoring function for DNA–drug docking of anticancer and antiparasitic compounds based on spectral moments of 2D lattice graphs for molecular dynamics trajectories

Lázaro G. Pérez-Montoto [a,b], Lourdes Santana [b], Humberto González-Díaz [a,b,*]

[a] Department of Microbiology & Parasitology, and Department of Organic Chemistry
[b] Faculty of Pharmacy, University of Santiago de Compostela, 15782, Spain

ARTICLE INFO

Article history:
Received 8 May 2009
Received in revised form
4 June 2009
Accepted 5 June 2009
Available online 17 June 2009

Keywords:
Anticancer drugs
Antiparasitic drugs
Leishmania
Trypanosoma cruzi
Antifungal compounds
Drug design
Graph theory
Markov models
Spectral moments
Complex networks
DNA–drug docking
QSAR
Molecular dynamics
Monte Carlo method

ABSTRACT

We introduce here a new class of invariants for MD trajectories based on the spectral moments $\pi_k(L)$ of the Markov matrix associated to lattice network-like (LN) graph representations of Molecular Dynamics (MD) trajectories. The procedure embeds the MD energy profiles on a 2D Cartesian coordinates system using simple heuristic rules. At the same time, we associate the LN with a Markov matrix that describes the probabilities of passing from one state to other in the new 2D space. We construct this type of LNs for 422 MD trajectories obtained in DNA–drug docking experiments of 57 furocoumarins. The combined use of psoralens + ultraviolet light (UVA) radiation is known as PUVA therapy. PUVA is effective in the treatment of skin diseases such as psoriasis and mycosis fungoides. PUVA is also useful to treat human platelet (PTL) concentrates in order to eliminate *Leishmania* spp. and *Trypanosoma cruzi*. Both are parasites that cause Leishmaniosis (a dangerous skin and visceral disease) and Chagas disease, respectively; and may circulate in blood products collected from infected donors. We included in this study both lineal (psoralens) and angular (angelicins) furocoumarins. In the study, we grouped the LNs on two sets; set1: DNA–drug complex MD trajectories for active compounds and set2: MD trajectories of non-active compounds or no-optimal MD trajectories of active compounds. We calculated the respective $\pi_k(L)$ values for all these LNs and used them as inputs to train a new classifier that discriminate set1 from set2 cases. In training series the model correctly classifies 79 out of 80 (specificity = 98.75%) set1 and 226 out of 238 (Sensitivity = 94.96%) set2 trajectories. In independent validation series the model correctly classifies 26 out of 26 (specificity = 100%) set1 and 75 out of 78 (sensitivity = 96.15%) set2 trajectories. We propose this new model as a scoring function to guide DNA-docking studies in the drug design of new coumarins for anticancer or antiparasitic PUVA therapy.

1. Introduction

Quantitative Structure–Activity Relationship (QSAR) studies unravel structural and physicochemical requirements for biological activity in a great variety of compounds [1]. The classic QSAR studies connect information of the chemical structure of the molecule, expressed by means of numbers, with the biological activity [2]. However, QSAR-like procedures are not restricted to drugs and biological activity but other systems and properties, such as allergenic character of proteins, may be predicted [3–6]. One special class of indices used in QSAR called Topological Indices (TIs) are based on the concept of molecular graph; which indicates the presence of vertices or nodes (atoms) and connections or edges between nodes (chemical bonds) [7–10]. In the same way, the field of application of TIs is, of course, not restricted to the chemistry of low-molecular-weight compounds and extends to other branches of sciences. In general, TIs of different types of graph representations or networks such as protein structure, gene polymorphisms, metabolic networks, food webs or host–parasite networks, internet, or social networks may be used. In these networks, amino acids, nucleotides, enzymes, microorganisms, cerebral cortex regions, web pages, social groups…etc, may play the role of nodes and electrostatic interactions, mutations, metabolic reactions, host–parasite relationships, brain region co-activations, links, disease propagations…etc may play the role of edges [11–17]. For

* Corresponding author. Department of Microbiology and Parasitology, Faculty of Pharmacy, University of Santiago de Compostela, 15782 Santiago de Compostela, Spain.
E-mail addresses: gonzalezdiazh@yahoo.es, humberto.gonzalez@usc.es (H. González-Díaz).

0223-5234/$ – see front matter © 2009 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2009.06.011

instance, the pseudo amino acid (PseAA) composition or PseAAC method for calculation of protein TIs was originally introduced by Chou to improve the prediction quality for protein subcellular localization and membrane protein type [18], as well as for enzyme functional class [19]. The PseAA composition can be used to represent a protein sequence with a discrete model yet without completely losing its sequence-order information. Ever since the concept of Chou's pseudo amino acid composition was introduced, various PseAAC approaches have been stimulated for enhancing the prediction quality of various protein features [20–29]. Owing to its wide usage, recently a very flexible pseudo amino acid composition generator, called "PseAAC" [30], was established at the website http://chou.med.harvard.edu/bioinf/PseAAC/, by which users can generate 63 different kinds of PseAA composition, including the dipeptide components. The publication in the area has steadily increased and, consequently, in the last years have appeared in-depth reviews that could be useful for the readers of the present manuscript [11,31–38].

Many authors prefer to use the term Quantitative Structure–Binding affinity Relationship (QSBR) when one uses QSAR-like procedures to predict drug–target binding affinity and 3D structural information [39]. Anyhow, the term QSBR has to be used carefully to avoid confusion with Quantitative Structure–Biodegradability Relationships analysis [40,41]. In this work, we use QSBR in the first sense. In any case, both approaches QSAR and QSBR diverge in some degree in the type of measure (activity or binding) and sometimes in how detailed manner we need to know the chemical structure (2D or 3D) but both use essentially the same algorithm. In addition, predicting drug activity we can use 3D drug–target QSAR/QSBR models as scoring function to guide the search of optimal drug–target interaction geometries in drug–target docking studies [42–44]. Almost all QSAR/QSBR or other types of docking scoring functions are aimed to predict protein–drug interactions. For instance, Wang et al. [45] reported a comparative study of eleven whereas Ferrara et al. [46] studied nine different docking scoring functions all for protein–drug interactions.

Conversely, DNA–drug and RNA–drug dockings are generally less investigated. In particular, we did not find a QSBR scoring function for DNA–Furocoumarin docking. The furocoumarins are a class of natural or synthetic compounds with very interesting pharmacological properties [47], commonly used in the treatment of skin diseases such as psoriasis and mycosis fungoides [48]. This treatment called PUVA consists in a therapy that combines the use of both chemicals and long-wave ultraviolet light (UVA) [49]. The molecular basis of PUVA is connected with the highly specific photo-damage in DNA of epidermal cells. This damage interferes with the DNA replication, producing an inhibition of DNA synthesis which reduces or blocks the cell duplication [50]. Although the lineal furocoumarins (psoralens) are able to form the three adduct types, the geometry of the angular ones (angelicins) only allows them to form mono-adducts with the DNA. It is well known that the side effects observed in PUVA therapy, such as skin photo-toxicity and risk of skin cancer are strictly connected with the bi-functional lesions in DNA [51]. Recently, Eastman et al. demonstrated the effectiveness of PUVA treatment with the psoralen analogue called amotosalen to inactivate the parasite *Leishmania* spp. in human platelet (PLT) concentrates intended for transfusion. *Leishmania* spp. are protozoans that cause skin and visceral diseases. *Leishmania* spp. are obligate intracellular parasites of mononuclear phagocytes and have been documented to be transmitted by blood transfusion. Both metacylic promastigotes and amastigotes were extremely susceptible to photochemical inactivation by PUVA. Promastigotes represent the infectious from the sand fly vector and amastigotes are the form that grow in mononuclear phagocytes. Thus, the PUVA of PLT concentrates inactivates both forms of *Leishmania* that would be expected to circulate in blood products collected from infected donors [52]. In addition, Gottlieb et al. reported the inactivation of the blood-borne parasite *Trypanosoma cruzi* (*T. cruzi*) by PUVA with 4′-aminomethyl-4,5′,8-trimethylpsoralen (AMT) in PLT concentrates [53]. The infectivity of the parasite is eliminated at 4.2 J/cm^2. The trypomastigote motility continues for at least 16 h-post-treatment and is inhibited only after much higher light doses. Isolation of total DNA from the parasite cells after treatment in the presence of 3H-AMT indicated that at the lethal UVA influence about 0.5 AMT adducts per kilobase pairs occurred. These results suggest that this PUVA methodology may eliminate blood-borne *T. cruzi*, the causative agent of Chagas disease. More recently, Castro and Girones demonstrated that the pathogen reduction system based on PUVA with amotosalen presents a robust efficacy for inactivation of high doses of three different strains of *T. cruzi* and offers the potential to make the PLT supply safer [54]. The biological activity of these compounds is normally studied by evaluating their capacity of forming an intercalated complex with DNA and their ability of photo-binding through mono or bi-functional addition to the same macromolecule [55]. A traditional procedure to determine the photo-biological and antiproliferative activity of furocoumarins measures ID_{50}, the UVA dose that reduces to 50% of the DNA synthesis in Ehrlich Ascites tumor cells (EATC) in presence of tested compound at certain concentration (18–20 μM). The protocols used in the activity determination are heterogeneous, however the use of the 8-MOP as reference to express the activity is very common [56–58].

These facts point to the stability of DNA–drug complex as a central factor in the activity of anticancer drugs in general including furocoumarins. At the light of these facts Molecular Dynamics (MD) of the DNA–drug complexes is central for drug design towards PUVA therapy. Since the advent of bioscience with the studies of Karplus et al. MD has become the by foremost well-established, computational technique to investigate structure and function of bio-molecules and their respective complexes and interactions [59–61]. In addition, after a pioneer paper entitled 'The Biological Functions of Low-Frequency Phonons' [62] published in 1977, a series of investigations into biopolymers from MD point of view have been stimulated. These studies have suggested that low-frequency (or terahertz frequency) collective motions do exist in proteins and DNA that hold a very high potential to reveal the profound dynamic mechanisms of many marvelous biological functions in biological systems (see, e.g., [63–76] and a comprehensive review [77]). This kind of inferences has been later observed by NMR [78], and been further used for medical treatments [79,80]. In view of this, to understand really the interaction mechanism of drugs with proteins or DNA, we should consider not only the static structures concerned but also the dynamical information obtained by simulating their interactions through a dynamic process.

In this sense, it is of high relevance taking into account that the previous non-covalent binding (in dark) between drug and DNA has a strong influence on the subsequent photoreaction and therefore on their biological activity [81]. Consequently, MD studies of the DNA–drug complexes in furocoumarins and anticancer drugs in general are of the major relevance. In this sense, it would be very interesting to work with invariants that encode information about the MD trajectories for the intercalation complexes of furocoumarins and anticancer drugs in general. The analysis of the MD trajectories resulting from the integration of the equations of motions in MD remains, however, the greatest challenge and requires a great deal of insight, experience, and effort. In a recent and very important work, Hamacher [82] proposed a new, theoretically sound, and versatile analysis procedure that provides scientists with a semi-quantitative invariant measure to compare various scenarios of their respective simulations.

However, using graphic approaches to study biological systems can provide useful insights, as indicated by many previous studies on a series of important biological topics. Graphs have been used to study enzyme-catalyzed reactions [83–92], protein folding kinetics [36], inhibition kinetics of processive nucleic acid polymerases and nucleases [93–99], codon usage [100–102], base frequencies in the anti-sense strands [103], and analysis of DNA sequence [104]. Moreover, graphical methods have been introduced for QSAR study [2,11,105,106] as well as utilized to deal with complicated network systems [11,38,107]. Recently, the "cellular automaton image" [108,109] has also been applied to study hepatitis B viral infections [110], HBV virus gene miss-sense mutation [111], and visual analysis of SARS-CoV [112,113], as well as representing complicated biological sequences [114] and helping to identify protein attributes [115,116]. Several authors have used pseudo-folding lattice Hydrophobicity–Polarity (HP) models to simulate protein folding making simulations to optimize the lattice structure and resemble real folding [117–124]. However, we can choose notably simpler pseudo-folding rules to avoid optimization procedures and speed up notably the construction of the lattice. In this sense, useful graph representations of DNA, RNA and/or protein sequences have been introduced by Nandy et al. [125–131], Gates [132], Leong and Morgenthaler [133], Randic et al. [134] based on 2D coordinate systems. We call these graph representations as sequence pseudo-folding lattice networks (LNs) because they look like lattice structures and in fact, we are forcing a sequence to fold in a way that not necessarily occurs in nature. In general, these LNs (as for other graphs) can be numerically characterized with TIs. These TIs describe the distribution of amino acids or nucleotides along the sequence but also encode higher order information. Thus lattice pseudo-folding TIs can be used in protein QSAR. Our group, have used different MARCH-INSIDE TIs of pseudo-folding lattice-like networks to predict diverse protein or DNA/RNA functions. For instance, we have used stochastic pseudo-folding spectral moments to predict ribonucleases [135] and dyneins [136]. In other works, we used Markov chain pseudo-folding electrostatic potentials to predict polygalacturonases [137] or human colon and breast cancer biomarkers [138]. All these MARCH-INSIDE pseudo-folding TIs can be calculated when we sum the respective indices for each node of the graph. All the above-mentioned values were used recently to predict mycobacterium promoters and compare entropies, spectral moments, and pseudo-folding electrostatic potentials [139]. The readers may see three recent reviews discussing the applications ranging from graph of small molecules to graph or network representation of protein sequences and 3D structure, DNA sequences, RNA secondary structure, or human blood proteome mass spectroscopy outcomes [11,38,140].

In any case, if we understand sequence as a type of input data we need not limit the applications of the pseudo-folding lattice network method to proteins, DNA or RNA sequences. Elaborating this line of thinking, we have proposed pseudo-folding lattice network representations of mass spectroscopy outcomes typical of blood proteome samples containing many proteins. For instance we have constructed lattice network representations for mass spectroscopy results obtained from blood proteome samples typical of drugs causing cardiotoxicity [141]. After calculation of the sum of the TIs of each sample we used them to seek a new type of classifier. The model connects TI values of the mass spectra of the blood proteome with the probability of appearance of drug cardiotoxicity. This new type of model was called Quantitative Proteome–Property Relationships (QPPR) in analogy to QSAR or QSPR [142]. We have used these lattice network TIs also to predict human prostate cancer [143].

The success of this strategy encouraged us to consider other classes of sequence data and solve different problems. For instance, the MD trajectories referred in previous paragraphs are time series obtained from simulation runs that constitute another type of sequential data. In any case, even if spectral moments of different graphs have been successfully used in QSAR before [144–147] we can see that spectral moments of an LN for MD trajectories have not been explored. Consequently, we decided to study here these indices to describe MD trajectories. In the present paper, we introduce an LN representation for the study of MD trajectories. We also obtain quantitative models able to differentiate furocoumarin derivatives according to their antiproliferative activity and the stability of the DNA–drug complex. The new model is also QSBR that has potential applications as scoring function for DNA–furocoumarin docking studies.

2. Materials and methods

2.1. Model building of DNA–drug intercalation complexes

For our study we used the decanucleotide of sequence d(CCGCTAGCGG) and the software application HyperChem [148], a fragment of DNA with double helix in B form and sugars in 2′-endo form. This decanucleotide sequence has been used in different studies concerning psoralens intercalation [149]. The structure of all the compounds selected for DNA–drug interaction studies were optimized using the interactive model building package of HyperChem [148]. The optimization of their geometries was carried out by the semi-empirical quantum mechanics calculations with method PM3 [150] using the Polak–Ribiere algorithm and the options implemented by default in the mentioned package. Thus, the minimized molecular structures were intercalated by hand approach in the DNA fragment, using the HyperChem package and taking into account the following experimentally demonstrated statements:

1. In the dark, the poly[dA-dT] poly[dA-dT] sequence in DNA is the most favorable site for intercalation since the further photoreaction takes place mainly on the 5,6 double bond of the thymine [151]. So, the optimized molecules were inserted among the thymine units in a parallel plane to the bases and, according to our decision, in a halfway position (Fig. 1. Left).
2. The furocoumarins have two reactive sites, but after photoreaction, different types of cycloadducts can be formed: mono (furan-side or pyrone-side) and di-adducts (the cross-link) [152]. Although psoralens are able to form all the cycloadduct types, angelicins form only mono-adducts owing to their angular molecular structure. Keeping this in mind, for each lineal molecule we modeled only one starting conformation, for which the cycloadduct formation by either one or other reactive site (furan or pyrone-side) is equally feasible from a geometric point of view. For each angular molecule, we decided to model two starting conformations, one for each mono-adduct formation (for the furan-side that we named as j-conformation and for the pyrone-side that we named as c-conformation).
3. The stereochemistry of the furocoumarin adducts is *cis–syn* [153,154]. Consequently, the molecules were oriented in such a way that the intercalation complex favors mainly the formation of cycloadducts with this stereochemistry. In the case of the furan-side, the stereochemistry *syn* means that the furan $O_{1'}$ and the pyrimidine N_1 are going to be on the adjacent corners of the future cyclo-butane ring. For the pyrone-side, the stereochemistry *syn* is defined as having the carbonyl-carbon of the pyrone ring and the N_1 of the pyrimidine on the adjacent corners of the future cyclo-butane ring (Fig. 1, right).

On the other hand, some of the studied angular molecules present ramifications in the C3 carbon that hindered us to model

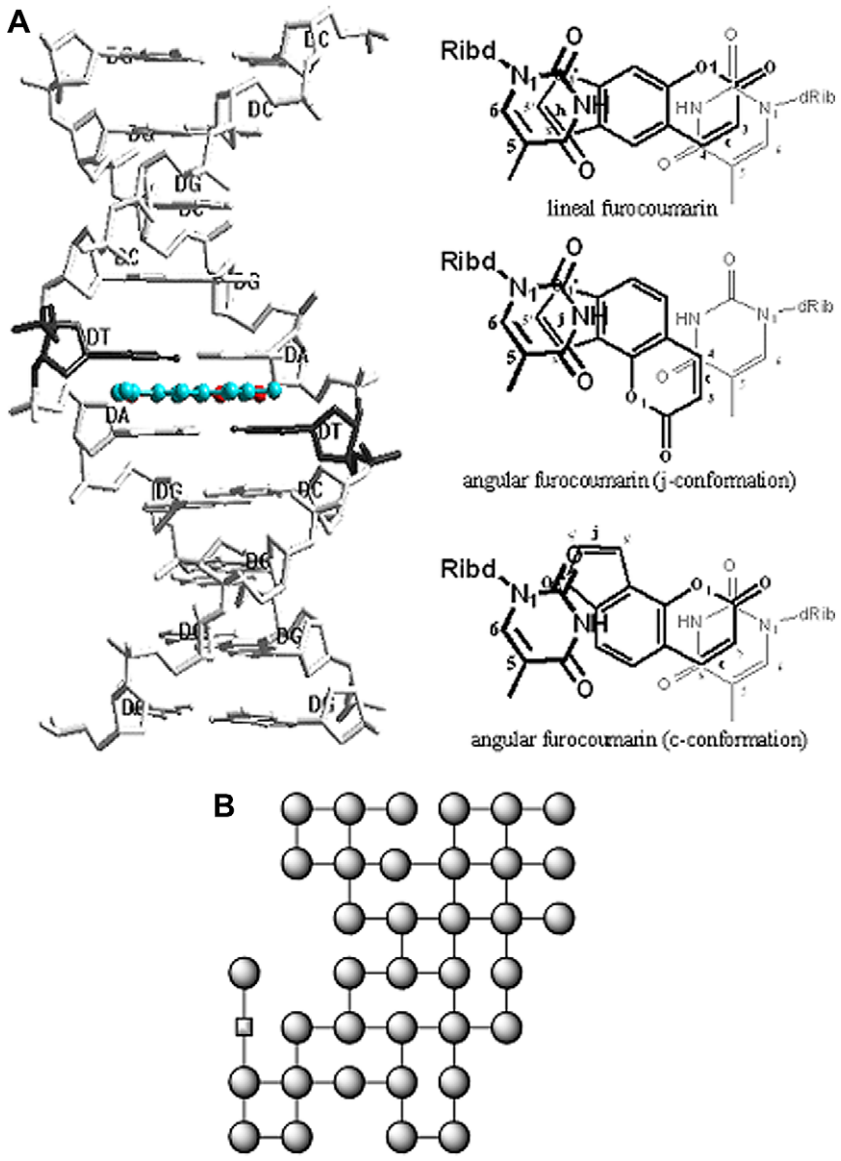


Fig. 1. DNA–drug complex (A) and MD lattice network (B) used to calculate the π_k(L) values.

appropriately their j-conformation, due to steric problems with the thymine ring. We also found steric impediments in the backbone of the DNA when these ramifications are much bigger. In all these cases, we decided to model several alternative starting conformations for which the steric effects were eliminated. For the majority of the cases we just varied the insertion degree of molecule in the DNA; in the most critical cases we also had to rotate the molecule clockwise, see Figs. 1 and 2 and Table 1 for details.

Both, the displacement outwards DNA and the molecule rotation were carried out in the halfway and parallel plane to the nitrogen bases. In this sense, the geometric criterion used was the relative distance (in the plane projection) between the geometric centers of the double bonds (j or c bond for furocoumarins and 5,6 bond for the thymine) that will take part in the photo-addition and the relative angle between them. In both Table 1 and Fig. 2, the variations of these geometric parameters used to model the j-conformations are represented in a simplified way. Taking these aspects into consideration the notation of an MD trajectory is given here as: *m*-[Bond/Dist./Ang.], where, *m* is the number of the compound in Table 2 or Table 3, Bond = j, c, or j and c are the chemical bonds susceptible to photo-addition in this position; whereas Dist. and Ang. are the distance and angular intercalation parameters, respectively (see Table 1).

2.2. Obtaining Monte Carlo MD energetic profiles

The DNA–drug docking molecular dynamics trajectories or energetic profiles of all the starting intercalation complexes were obtained by means of the Monte Carlo [155] method, using the HyperChem package. In this sense, the force field AMBER94 of molecular mechanics was used with distant-dependent dielectric constant (scale factor 1), electrostatic and Van der Waals values by default and cutoffs shifted with outer radius of 14 Å. All the components of the force field were included and the atom type was recalculated keeping their current charges. Finally, the simulation was executed in vacuo at 300 K and 100 optimization steps obtaining MD trajectories with 100 potential energy ${}^{d}E_j$ ($j = 1, 2, 3, \ldots 100$) values each. We obtained 21 MD trajectories for psoralens and 154 MD trajectories for the 36 different angelicins. We also analyzed 36 averaged MD trajectories for each angelicin taking the

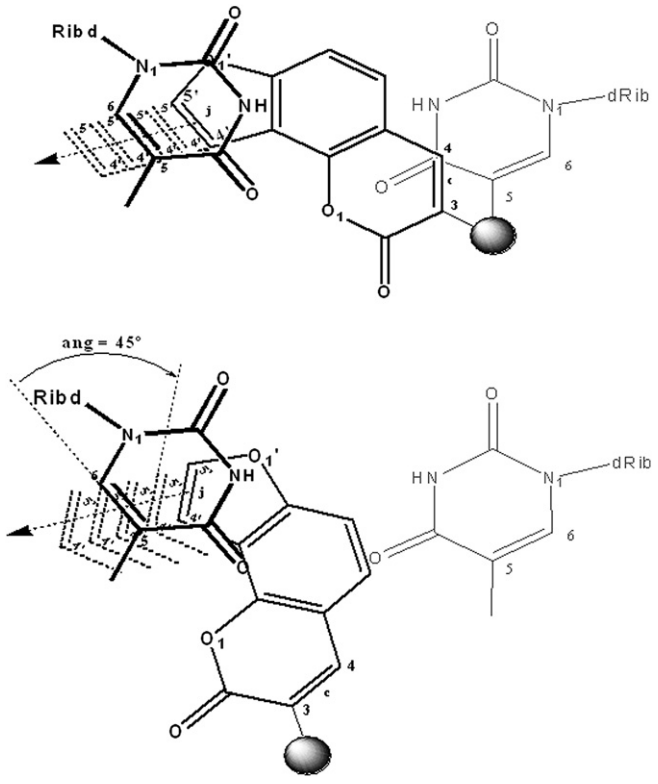


Fig. 2. Modifications made to the j-conformations to avoid the steric impediments. (Top) Decrease of the insertion degree of the molecule in the DNA. The relative distance between the geometric centers of the double bonds takes discrete values of −1, −0.5, 0, 0.5 and 1 times the distance of a C–C bond. (Bottom) Decrease of the insertion degree accompanied by a rotation clockwise of the molecule to a magnitude of 45°.

average energy ${}^{d}E_j(\text{avg})$ for all the initial positions of one compound at each one of the 100 steps. All these MD trajectories form a total of $21 + 154 + 36 = 211$ MD trajectories. In addition, we analyzed other 211 MD trajectories (decoy trajectories) obtained as a random deviation from each one of the previous 211 MD trajectories.

$$ {}^{d}E_j(\text{rnd}) = {}^{d}E_j + \text{random}(j, \max({}^{d}E_j), \min({}^{d}E_j)) \quad (1) $$

These random MD trajectories contain 100 energy values ${}^{d}E_j(\text{rnd})$ obtained with the random generator of Excel by adding a random deviation term to each ${}^{d}E_j$ within the max–min limits of ${}^{d}E_j$ for all the previous MD trajectories. The utility of these decoy trajectories is to test the robustness of the method to deviations of the MD trajectories selected. In total, we studied 422 MD trajectories. The information about all these 422 MD trajectories including $\pi_k(\text{d})$ and $\pi_k(\text{L})$ values relevant to this work was recorded in the Supplementary material.

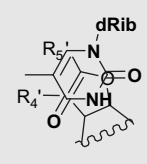
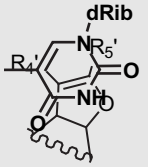
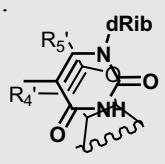
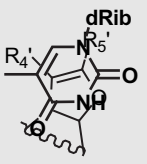
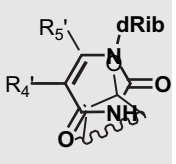
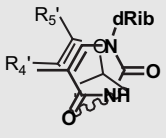
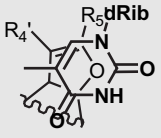
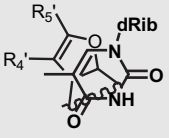
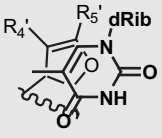
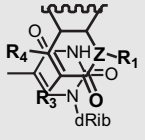
2.3. Markov spectral moments for 2D lattice representation of DNA–drug MD trajectories

The MARCH-INSIDE approach is extended here to the study of LN representations for MD trajectories obtained in DNA–drug docking studies. The key of the method we propose is the regrouping into four groups of the energy values ${}^{d}E_s$ obtained for different steps (s) of one MD trajectory after docking one drug with DNA. These four groups characterize the deviation of the energy value ${}^{d}E_s$ from the average energy of the same DNA–drug complex at other steps (MD-average); or the deviation from average energy values of the same step for other drugs (step-average). First, the values of energy for the MD profile of one DNA–drug complex is placed in a Cartesian 2D space starting with the first energy value at the coordinates (0, 0). We calculated coordinates of the successive energy values using simple heuristic rules, in a similar manner than it can be used for a DNA or protein sequences [136,138]:

a) Increases in +1 the x axis; if ${}^{d}E_s >$ MD-average and ${}^{d}E_s >$ step-average (rightwards-step) or:
b) Decreases in −1 the x axis; if ${}^{d}E_s >$ MD-average and ${}^{d}E_s <$ step-average (leftwards-step) or:
c) Increases in +1 the y axis; if ${}^{d}E_s <$ MD-average and ${}^{d}E_s >$ step-average (upwards-step) or:
d) Decreases in −1 the y axis; if ${}^{d}E_s <$ MD-average and ${}^{d}E_s <$ step-average (downwards-step).

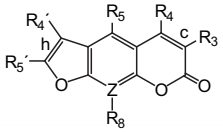
Secondly, the method uses the matrix ${}^{1}\boldsymbol{\Pi}$, which is a squared matrix to characterize the MD profile embedded into the lattice-like graph. Please, note that the number of nodes (n) in the graph may be equal or even smaller than the number of steps given to

Table 1
Representations for starting conformations used.

					
Dist.[a]	1	1	0.5	0.5	0
Ang.[a]	0°	45°	0°	45°	0°
					
Dist.[a]	−0.5	−0.5	−1	−1	1
Ang.[a]	0°	45°	0°	45°	0°

[a] **Dist.:** discrete distance between the geometric centers of the two double bonds (in the plane projection). The possible modular values are 0; 0.5 and 1. Positive value if the compound was moved inwards DNA pocket and negative if it was moved outwards DNA. **Ang.:** magnitude of clockwise rotation of compound (0° or 45°).

Table 2
Lineal furocoumarins (psoralens) and their aza-analogues used.



Drug	Z	R_3	R_4	R_5	$R_{4'}$	$R_{5'}$	R_8	ID_{50}[a]	Ref.[b]
1	C	Me	Me	H	Me	H	H	0.34	[35]
2	C	H	H	OMe	H	H	H	0.66	[30]
3	C	H	CH_2OH	H	Me	H	OMe	0.84	[32]
4	C	Me	H	H	Me	H	OMe	0.89	[76]
5	C	H	H	H	H	H	OMe	1.00	[34]
6	C	Me	H	H	Me	H	H	1.01	[77]
7	C	H	H	H	Me	Me	H	1.26	[35]
8	C	Me	H	H	Me	H	Me	1.34	[77]
9	C	H	H	H	H	H	H	1.52	[78]
10	C	Me	H	H	Me	Me	H	1.79	[35]
11	C	H	CH_2OH	H	Me	H	H	2.32	[32]
12	C	H	Me	H	H	Me	Me	27.6	[30]
13	N	H	H	H	Me	H	–	0.13	[34]
14	N	H	Me	H	H	H	–	0.14	[34]
15	N	H	H	H	Me	Me	–	0.18	[79]
16	N	H	H	Me	Me	Me	–	0.25	[34]
17	N	Me	Me	H	Me	H	–	0.67	[34]
18	N	H	Me	H	Me	H	–	0.68	[79]
19	N	H	H	Me	Me	H	–	0.97	[34]
20	N	Me	Me	H	Me	Me	–	1.83	[79]
21	N	H	Me	H	Me	Me	–	3.66	[79]

[a] The experimental antiproliferative activity in Ehrlich Ascites tumor cells expressed as ID_{50} relative to 8-MOP.
[b] Ref.: References in which the activity of compounds was reported.

obtain the MD profile. The same happens for amino acids or DNA bases in the polymeric chain. Accordingly, the matrix $^1\Pi$ contains the probabilities $^1p_{ij}$ to reach a node n_i moving throughout a walk of length $k = 1$ from other node n_j [139,141]:

$$p_{ij} = \frac{\left(\frac{1}{D_{0j}}\right)\cdot\left(\sum_{s\in j}{}^dE_s\right)}{\sum_{m=s}^{n}\alpha_{il}\cdot\left(\frac{1}{D_{0s}}\right)\cdot\left(\sum_{s\in j}{}^dE_s\right)} = \frac{\left(\frac{{}^dE_j}{D_{0j}}\right)}{\sum_{m=l}^{n}\alpha_{il}\cdot\left(\frac{{}^dE_l}{D_{0s}}\right)} \quad (2)$$

where, ${}^dE_j$ is the sum of all energy values of the steps ${}^dE_s$ that overlap on the same node j. The parameter α_{ij} equals to 1 if the nodes n_i and n_j are adjacent in the graph and equals to 0 otherwise. The value D_{0j} gives the geometric location of the node and represents the Euclidean distance between the node and the center of coordinates. In Fig. 1 (bottom part) we depicted an example of LN for an MD trajectory. Afterwards, it is straightforward to realize the calculation of π_k(L) values:

$$\pi_k(\mathrm{L}) = \sum_{i=j}^{n} {}^kp_{ij}({}^dE_s) = \mathrm{Tr}({}^k\Pi) = \mathrm{Tr}\left[({}^1\Pi)^k\right] \quad (3)$$

2.4. Markov spectral moments of classic molecular graph representations

Using MCM we can calculate these π_k(d) values associated to the electronic distribution in molecule of the drug (d). The theoretic foundations of the method have been given in previous works, so we do not detail it here but refer the reader to these works [156–158]. We can use π_k(d) values in addition to the π_k(L) values to describe only the drug (not the MD trajectory). The π_k(d) values are spectral moments of the classic electronic Markov matrix ($^1\Pi$). These values have been used in QSAR before and depend only on connections (chemical bonds) between node (atoms) in the molecular graph and the electronegativity (χ_j) of these atoms. The values π_k(d) are referred to atoms (nodes) in molecular graphs. These vectors are elements of a Markov chain based on the stochastic matrix $^1\Pi$, which contains elements that describe the probabilities of transition of electrons $p_1(i,j)$ from node (atom) i-th to, j-th.

$$\pi_k(\mathrm{d}) = \sum_{i=j}^{n} {}^kp_{ij}(\chi) = \mathrm{Tr}({}^k\Pi) = \mathrm{Tr}\left[({}^1\Pi)^k\right] \quad (4)$$

In order to ensure that the $p_1(i,j)$ values describe the probabilities of transition of electrons from node (atom) i-th to, j-th we use atmoic electronegativity values. At following, we give the formula for both the transition probabilities (elements of the matrix) and the atoms set entropy centrality measures.

$${}^1p_{ij}(\chi) = \frac{\delta_{ij}\cdot\chi_j}{\sum_{k=1}^{n}\delta_{ik}\cdot\chi_k} \quad (5)$$

2.5. The dataset statistical analysis

In this study, we selected different furocoumarins and some of their aza-analogues, whose antiproliferative activities in Ehrlich Ascites tumor cells have been determined (Tables 2 and 3). We obtained in total 422 MD trajectories for these compounds. We constructed 422 LNs (one for each MD trajectory) transformed them in a vector of 11 π_k(d) values for the compound and 11 π_k(L) values for the MD trajectory (see previous sections). We grouped all these 422 MD trajectories into two sets, one composed of MD trajectories of complexes between DNA and active compounds and the other composed of trajectories of active compounds with no-optimal MD trajectories and/or trajectories of non-active compounds. In general, compounds such as 4′-MAP and the 4-MBAP, with activities (ID_{50} relative to 8-MOP) of 0.13 and 0.14 are considered as poorly active [57,58]. Keeping in mind all the above-mentioned aspects, we classified the 57 compounds, compiled for our dataset into two observed activity groups: 0 for the inactive compounds ($LD_{50} \leq 0.1$) and 1 for the active ones ($LD_{50} > 0.1$).

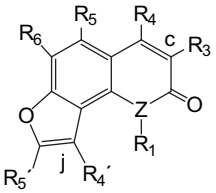
QSBR studies were carried out to obtain models that allow us to classify the furocoumarin derivatives in one of these two activity groups. We selected Linear Discriminant Analysis (LDA) [159,160] to fit the discriminant function as implemented in the LDA module of the STATISTICA 6.0 software package [161]. We used forward-stepwise algorithm for variable selection [162–164]. The strength of the correlation was determined by the Canonical Regression Coefficient (Rc) and the statistical significance of the LDA model was determined with U-statistics (U) and the respective p-level (p). We standardized all the variables included in the model in order to bring it into the same scale. Subsequently, a standardized linear discriminant equation that allows to compare their coefficients is obtained [165]. We also inspected the percentage of good classification, cases/variables' ratios (ρ parameter), and number of variables to be explored to avoid over-fitting or chance correlation [162,163]. The general form of this model is:

$$\text{MD-score} = \sum_{k=0}^{10} a_k\cdot\pi_k(\mathrm{L}) + \sum_{k=0}^{10} c_k\cdot\pi_k(\mathrm{d}) + b_0 \quad (6)$$

3. Results and discussion

Computational approaches, such as structural bioinformatics [76,77], molecular docking [66,78,79], Monte Carlo simulated annealing approach [80] and QSAR [50,81,82] can timely provide very

Table 3
Angular furocoumarins (angelicins) and their aza-analogues used.



Compd.	Z	R_1	R_3	R_4	R_5	R_6	$R_{4'}$	$R_{5'}$	ID_{50}[a]	Ref.[b]
22	O	–	COMe	H	H	H	H	H	<0.01	[80]
23	O	–	COPh	H	H	H	H	H	<0.01	[80]
24	O	–	$CON(Et)_2$	H	H	H	H	H	<0.01	[80]
25	O	–	$CONH(CH_2)_2OH$	H	H	H	H	H	<0.01	[80]
26	O	–	$CONH(CH_2)_2OEt$	H	H	H	H	H	<0.01	[80]
27	O	–	$CONH(CH_2)_2NMe_2$	H	H	H	H	H	<0.01	[80]
28	O	–	$CON[(CH_2)_2OH]_2$	H	H	H	H	H	<0.01	[80]
29	O	–	$CON(CH_2)_2NMe$	H	H	H	H	H	<0.01	[80]
30	O	–	$CONH_2$	H	H	H	H	H	0.05	[80]
31	O	–	$CON(CH_2)_2O$	H	H	H	H	H	0.06	[80]
32	O	–	CO_2H	H	H	H	H	H	0.07	[80]
33	O	–	$CON(Me)_2$	H	H	H	H	H	0.07	[80]
34	O	–	CO_2Me	H	H	H	H	H	0.20	[80]
35	O	–	Me	H	H	H	H	H	0.20	[81]
36	O	–	Me	Me	H	H	Me	H	0.03	[35]
37	O	–	Me	Me	H	H	H	H	0.35	[81]
38	O	–	CO_2Et	H	H	H	H	H	0.40	[81]
39	O	–	H	H	H	H	H	H	0.55	[82]
40	O	–	H	Me	H	H	H	H	0.55	[35]
41	O	–	H	Me	H	H	CH_2OMe	Me	0.60	[81]
42	O	–	H	H	H	H	H	Me	0.80	[82]
43	O	–	H	H	H	H	Me	H	0.81	[82]
44	O	–	H	Me	H	H	H	Me	1.27	[81]
45	O	–	H	H	H	H	Me	Me	1.47	[81]
46	O	–	H	H	Me	H	Me	H	5.30	[82]
47	O	–	H	Me	H	H	Me	H	5.75	[82]
48	O	–	H	Me	Me	H	Me	H	5.78	[81]
49	N	H	H	Me	H	H	Me	H	0.48	[83]
50	N	H	H	CH_2OH	H	Me	H	Me	0.66	[84]
51	N	H	H	Me	H	Me	Me	CH_2OH	1.07	[83]
52	N	H	H	Me	H	Me	H	Me	1.36	[83]
53	N	Me	H	CH_2OMe	H	Me	H	Me	2.09	[84]
54	N	H	H	Me	H	H	Me	Me	2.59	[83]
55	N	H	H	Me	H	Me	Me	H	4.62	[83]
56	N	Me	H	CH_2OH	H	Me	H	Me	5.60	[84]
57	N	H	H	Me	H	Me	Me	Me	9.25	[83]

[a] The experimental antiproliferative activity in Ehrlich Ascites tumor cells expressed as ID_{50} relative to 8-MOP.
[b] Ref.: References in which the activity of compounds was reported.

useful information to stimulate basic research and drug development. In view of this, the present study was attempted to propose a new scoring function for studying docking of drugs to DNA in hopes that our method may become a useful tool for drug development. Using both the classic $\pi_k(d)$ and the new $\pi_k(L)$ values as inputs we can obtain a classifier to discriminate DNA–drug complexes of the two classes defined in Section 2. The best model we found was:

$$\begin{aligned}\text{MD-score} &= 3.17\cdot\pi_2(L) - 15.47\cdot\pi_4(d) - 17.66\cdot\pi_0(d) \\ &\quad +111.19\cdot\pi_{10}(d) - 114.87\cdot\pi_9(d) - 2.96 \\ n &= 318 \quad Rc = 0.77 \quad F = 88.5 \quad p < 0.01 \end{aligned} \tag{7}$$

The output of the model, MD-score, is a real value variable that scores the predicted goodness of fit for one MD trajectory. In statistical prediction, the following three cross-validation methods are often used to examine a predictor for its effectiveness in practical application: independent dataset test, sub-sampling test, and jackknife test [166]. However, as elucidated by [167] and demonstrated in [168], among the three cross-validation methods, the jackknife test is deemed the most objective that can always yield a unique result for a given benchmark dataset, and hence has been increasingly and widely used by investigators to examine the accuracy of various predictors (see, e.g., [20–25,27,169–172]). In the current study, because the jackknife test would take a lot of computational time, we choose to use the independent dataset test to examine the prediction accuracy. The model was trained with a training series and later validated with an external validation series. In training series the model correctly classifies 79 out of 80 (specificity = 98.75%) optimal and 226 out of 238 (sensitivity = 94.96%) no-optimal MD trajectories. In external validation series the model correctly classifies 26 out of 26 (specificity = 100%) optimal and 75 out of 78 (sensitivity = 96.15%) no-optimal MD trajectories. These results represent total accuracy = 95.91% and 97.12% in training and validation respectively. Previous QSAR works that use LDA as the classification technique accept this level of sensitivity, specificity, and accuracy as indicative of high quality of the model [106,173–178+].

4. Conclusions

We can obtain new types of 2D graph theoretical representation for Molecular Dynamics (MD) trajectories that resemble LNs used for DNA and protein sequences. At the same time, it is possible to

calculate new classes of invariants for MD trajectories based on the spectral moments π_k(L) of the Markov matrix associated to these LNs. The π_k(L) values can be used as inputs to train new classifiers in order to discriminate between optimal and no-optimal intercalation modes relevant to the biological activity. The new models can be used as scoring functions to guide DNA-docking studies in the drug design of new coumarins for PUVA therapy.

Acknowledgments

The authors sincerely thank kind attention and useful comments from Editor Prof. Dr. Antonio Monge-Vega and two reviewers. González-Díaz H acknowledges financial support of Program Isidro Parga Pondal funded by Xunta de Galicia and European Social Fund (E.S.F.).

Appendix. Supplementary material

Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.ejmech.2009.06.011.

References

[1] F. Giordanetto, P. Fossa, G. Menozzi, L. Mosti, J. Comput.-Aided Mol. Des. 17 (2003) 53–64.
[2] F.J. Prado-Prado, H. González-Díaz, O.M. de la Vega, F.M. Ubeira, K.C. Chou, Bioorg. Med. Chem. 16 (2008) 5871–5880.
[3] C.H. Schein, O. Ivanciuc, W. Braun, Immunol. Allergy Clin. North Am. 27 (2007) 1–27.
[4] C.H. Schein, O. Ivanciuc, W. Braun, J. Agric. Food Chem. 53 (2005) 8752–8759.
[5] O. Ivanciuc, C.H. Schein, W. Braun, Nucleic Acids Res. 31 (2003) 359–362.
[6] O. Ivanciuc, V. Mathura, T. Midoro-Horiuti, W. Braun, R.M. Goldblum, C.H. Schein, J. Agric. Food Chem. 51 (2003) 4830–4837.
[7] A.T. Balaban, A. Beteringhe, T. Constantinescu, P.A. Filip, O. Ivanciuc, J. Chem. Inf. Model. 47 (2007) 716–731.
[8] O. Ivanciuc, T. Ivanciuc, D.J. Klein, W.A. Seitz, A.T. Balaban, J. Chem. Inf. Comput. Sci. 41 (2001) 536–549.
[9] O. Ivanciuc, J. Chem. Inf. Comput. Sci. 40 (2000) 1412–1422.
[10] D. Bonchev, J. Chem. Inf. Comput. Sci. 40 (2000) 934–941.
[11] H. González-Díaz, Y. González-Díaz, L. Santana, F.M. Ubeira, E. Uriarte, Proteomics 8 (2008) 750–778.
[12] O. Ivanciuc, T. Ivanciuc, D.J. Klein, SAR QSAR Environ. Res. 12 (2001) 1–16.
[13] D. Bonchev, G.A. Buck, J. Chem. Inf. Model. 47 (2007) 909–917.
[14] D. Bonchev, SAR QSAR Environ. Res. 14 (2003) 199–214.
[15] D. Bonchev, Chem. Biodivers. 1 (2004) 312–326.
[16] L.B. Kier, D. Bonchev, G.A. Buck, Chem. Biodivers. 2 (2005) 233–243.
[17] S. Bornholdt, H.G. Schuster, Handbook of Graphs and Complex Networks: From the Genome to the Internet, Wiley-VCH GmbH & CO. KGa, Weinheim, 2003.
[18] K.C. Chou, Proteins 43 (2001) 246–255;
Erratum, Proteins 44 (2001) 60.
[19] K.C. Chou, Bioinformatics 21 (2005) 10–19.
[20] G.Y. Zhang, H.C. Li, J.Q. Gao, B.S. Fang, Protein Pept. Lett. 15 (2008) 1132–1137.
[21] F.M. Li, Q.Z. Li, Protein Pept. Lett. 15 (2008) 612–616.
[22] H. Lin, H. Ding, F.B. Guo, A.Y. Zhang, J. Huang, Protein Pept. Lett. 15 (2008) 739–744.
[23] H. Lin, J. Theor. Biol. 252 (2008) 350–356.
[24] G.Y. Zhang, B.S. Fang, J. Theor. Biol. 253 (2008) 310–315.
[25] X.B. Zhou, C. Chen, Z.C. Li, X.Y. Zou, J. Theor. Biol. 248 (2007) 546–551.
[26] D.N. Georgiou, T.E. Karakasidis, J.J. Nieto, A. Torres, J. Theor. Biol. 257 (2009) 17–26.
[27] X. Jiang, R. Wei, T. Zhang, Q. Gu, Protein Pept. Lett. 15 (2008) 392–396.
[28] C. Chen, L. Chen, X. Zou, P. Cai, Protein Pept. Lett. 16 (2009) 27–31.
[29] K.C. Chou, H.B. Shen, J. Proteome Res. 6 (2007) 1728–1734.
[30] H.B. Shen, K.C. Chou, Anal. Biochem. 373 (2008) 386–388.
[31] O. Mason, M. Verwoerd, IET Syst. Biol. 1 (2007) 89–119.
[32] A. Krishnan, J.P. Zbilut, M. Tomita, A. Giuliani, Curr. Protein Pept. Sci. 9 (2008) 28–38.
[33] K. Kayser, H.J. Gabius, Prog. Histochem. Cytochem. 32 (1997) 1–106.
[34] R.B. Glassman, Brain Res. Bull. 60 (2003) 25–42.
[35] R. Garcia-Domenech, J. Galvez, J.V. de Julian-Ortiz, L. Pogliani, Chem. Rev. 108 (2008) 1127–1169.
[36] K.C. Chou, Biophys. Chem. 35 (1990) 1–24.
[37] E. Estrada, E. Uriarte, Curr. Med. Chem. 8 (2001) 1573–1588.
[38] H. González-Díaz, S. Vilar, L. Santana, E. Uriarte, Curr. Top. Med. Chem. 7 (2007) 1025–1039.
[39] S. Zhang, A. Golbraikh, A. Tropsha, J. Med. Chem. 49 (2006) 2713–2724.
[40] B. Cuissart, F. Touffet, B. Cremilleux, R. Bureau, S. Rault, J. Chem. Inf. Comput. Sci. 42 (2002) 1043–1052.
[41] C.W. Andrews, L. Bennett, L.X. Yu, Pharm. Res. 17 (2000) 639–644.
[42] C. Hetenyi, G. Paragi, U. Maran, Z. Timar, M. Karelson, B. Penke, J. Am. Chem. Soc. 128 (2006) 1233–1239.
[43] M.A. Lill, A. Vedani, M. Dobler, J. Med. Chem. 47 (2004) 6174–6186.
[44] R. Smith, R.E. Hubbard, D.A. Gschwend, A.R. Leach, A.C. Good, J. Mol. Graph. Model. 22 (2003) 41–53.
[45] R. Wang, Y. Lu, S. Wang, J. Med. Chem. 46 (2003) 2287–2303.
[46] P. Ferrara, H. Gohlke, D.J. Price, G. Klebe, C.L. Brooks 3rd, J. Med. Chem. 47 (2004) 3032–3047.
[47] L. Santana, E. Uriarte, F. Roleira, N. Milhazes, F. Borges, Curr. Med. Chem. 11 (2004) 3239–3261.
[48] M.A. Pathak, T.B. Fitzpatrick, J. Photochem. Photobiol. B 14 (1992) 3–22.
[49] J.A. Parrish, R.S. Stern, M.A. Pathak, T.B. Fitzpatrick, in: J.A. Parrish, J.D. Regan, J.A. Parrish (Eds.), The Science of Photomedicine, Plenum Press, New York, 1982, p. 595.
[50] F. Dall'Acqua, D. Vedaldi, F. Baccichetti, F. Bordin, D. Averbeck, Farmaco [Sci.] 36 (1981) 519–535.
[51] M.A. Pathak, J.A. Parrish, T.B. Fitzpatrick, Farmaco [Sci.] 36 (1981) 479–491.
[52] R.T. Eastman, L.K. Barrett, K. Dupuis, F.S. Buckner, W.C. Van Voorhis, Transfusion (Paris) 45 (2005) 1459–1463.
[53] P. Gottlieb, H. Margolis-Nunno, R. Robinson, L.G. Shen, E. Chimezie, B. Horowitz, et al., Photochem. Photobiol. 63 (1996) 562–565.
[54] E. Castro, N. Girones, J.L. Bueno, J. Carrion, L. Lin, M. Fresno, Transfusion (Paris) 47 (2007) 434–441.
[55] G. Zagotto, O. Gia, F. Baccichetti, E. Uriarte, M. Palumbo, Photochem. Photobiol. 58 (1993) 486–491.
[56] L. Musajo, P. Visentini, F. Baccichetti, M.A. Razzi, Experientia 23 (1967) 335–336.
[57] F. Baccichetti, F. Bordin, M. Simonato, L. Toniolo, C. Marzano, P. Rodighiero, et al., Il Farmaco 47 (1992) 1529–1541.
[58] C. Antonello, G. Zagotto, S. Mobilio, C. Marzano, O. Gia, E. Uriarte, Il Farmaco 49 (1994) 277–280.
[59] J.A. McCammon, B.R. Gelin, M. Karplus, Nature 267 (1977) 585–590.
[60] M. Karplus, J.A. McCammon, Nat. Struct. Biol. 9 (2002) 646–652.
[61] J.A. McCammon, M. Karplus, Nature 268 (1977) 765–766.
[62] K.C. Chou, N.Y. Chen, Sci. Sin. 20 (1977) 447–457.
[63] K.C. Chou, Biopolymers 26 (1987) 285–295.
[64] K.C. Chou, Biophys. Chem. 25 (1986) 105–116.
[65] K.C. Chou, Y.S. Kiang, Biophys. Chem. 22 (1985) 219–235.
[66] K.C. Chou, Biophys. J. 48 (1985) 289–297.
[67] K.C. Chou, Biophys. Chem. 20 (1984) 61–71.
[68] K.C. Chou, Biochem. J. 221 (1984) 27–31.
[69] K.C. Chou, Biophys. J. 45 (1984) 881–889.
[70] K.C. Chou, Biochem. J. 215 (1983) 465–469.
[71] K.C. Chou, Biochem. J. 209 (1983) 573–580.
[72] P. Martel, Prog. Biophys. Mol. Biol. 57 (1992) 129–179.
[73] Z. Sinkala, J. Theor. Biol. 241 (2006) 919–927.
[74] K.C. Chou, N.Y. Chen, S. Forsen, Chem. Scr. 18 (1981) 126–132.
[75] K.C. Chou, C.T. Zhang, G.M. Maggiora, Biopolymers 34 (1994) 143–153.
[76] K.C. Chou, B. Mao, Biopolymers (Biospectrosc.) 27 (1988) 1795–1815.
[77] K.C. Chou, Biophys. Chem. 30 (1988) 3–48.
[78] J.J. Chou, S. Li, C.B. Klee, A. Bax, Nat. Struct. Biol. 8 (2001) 990–997.
[79] G. Gordon, J. Cell. Physiol. 212 (2007) 579–582.
[80] G. Gordon, J. Biomed. Sci. Eng. 1 (2008) 152–156.
[81] O. Gia, S. Marciani Magno, H. Gonzalez-Diaz, E. Quezada, L. Santana, E. Uriarte, et al., Bioorg. Med. Chem. 13 (2005) 809–817.
[82] K. Hamacher, J. Comput. Chem. 28 (2007) 2576–2580.
[83] E.L. King, C. Altman, J. Phys. Chem. 60 (1956) 1375–1378.
[84] K.C. Chou, S.P. Jiang, W.M. Liu, C.H. Fee, Sci. Sin. 22 (1979) 341–358.
[85] K.C. Chou, S. Forsen, Biochem. J. 187 (1980) 829–835.
[86] K.C. Chou, S. Forsen, Can. J. Chem. 59 (1981) 737–755.
[87] K.C. Chou, W.M. Liu, J. Theor. Biol. 91 (1981) 637–654.
[88] D. Myers, G. Palmer, Comput. Appl. Biosci. 1 (1985) 105–110.
[89] K.C. Chou, J. Biol. Chem. 264 (1989) 12074–12079.
[90] P. Kuzmic, K.Y. Ng, T.D. Heath, Anal. Biochem. 200 (1992) 68–73.
[91] G.P. Zhou, M.H. Deng, Biochem. J. 222 (1984) 169–176.
[92] J. Andraos, Can. J. Chem. 86 (2008) 342–357.
[93] I.W. Althaus, J.J. Chou, A.J. Gonzales, M.R. Deibel, K.C. Chou, F.J. Kezdy, et al., J. Biol. Chem. 268 (1993) 6119–6124.
[94] I.W. Althaus, A.J. Gonzales, J.J. Chou, D.L. Romero, M.R. Deibel, K.C. Chou, et al., J. Biol. Chem. 268 (1993) 14875–14880.
[95] I.W. Althaus, J.J. Chou, A.J. Gonzales, M.R. Deibel, K.C. Chou, F.J. Kezdy, et al., Biochemistry 32 (1993) 6548–6554.
[96] I.W. Althaus, J.J. Chou, A.J. Gonzales, R.J. LeMay, M.R. Deibel, K.C. Chou, et al., Experientia 50 (1994) 23–28.
[97] I.W. Althaus, J.J. Chou, A.J. Gonzales, M.R. Deibel, K.C. Chou, F.J. Kezdy, et al., Biochem. Pharmacol. 47 (1994) 2017–2028.
[98] I.W. Althaus, K.C. Chou, R.J. Lemay, K.M. Franks, M.R. Deibel, F.J. Kezdy, et al., Biochem. Pharmacol. 51 (1996) 743–750.
[99] K.C. Chou, F.J. Kezdy, F. Reusser, Anal. Biochem. 221 (1994) 217–230.
[100] K.C. Chou, C.T. Zhang, AIDS Res. Hum. Retroviruses 8 (1992) 1967–1976.
[101] C.T. Zhang, K.C. Chou, J. Protein Chem. 12 (1993) 329–335.
[102] C.T. Zhang, K.C. Chou, J. Mol. Biol. 238 (1994) 1–8.

[103] K.C. Chou, C.T. Zhang, D.W. Elrod, J. Protein Chem. 15 (1996) 59–61.
[104] X.Q. Qi, J. Wen, Z.H. Qi, J. Theor. Biol. 249 (2007) 681–690.
[105] H. Gonzalez-Diaz, A. Sanchez-Gonzalez, Y. Gonzalez-Diaz, J. Inorg. Biochem. 100 (2006) 1290–1297.
[106] H. Gonzalez-Diaz, I. Bonet, C. Teran, E. De Clercq, R. Bello, M.M. Garcia, et al., Eur. J. Med. Chem. 42 (2007) 580–585.
[107] Y. Diao, M. Li, Z. Feng, J. Yin, Y. Pan, J. Theor. Biol. 247 (2007) 608–615.
[108] S. Wolfram, Nat. Protoc. 311 (1984) 419–424.
[109] S. Wolfram, A New Kind of Science (2002) Champaign, IL.
[110] X. Xiao, S.H. Shao, K.C. Chou, Biochem. Biophys. Res. Commun. 342 (2006) 605–610.
[111] X. Xiao, S. Shao, Y. Ding, Z. Huang, X. Chen, K.C. Chou, J. Theor. Biol. 235 (2005) 555–565.
[112] M. Wang, J.S. Yao, Z.D. Huang, Z.J. Xu, G.P. Liu, H.Y. Zhao, et al., Med. Chem. 1 (2005) 39–47.
[113] L. Gao, Y.S. Ding, H. Dai, S.H. Shao, Z.D. Huang, K.C. Chou, J. Pharm. Biomed. Anal. 41 (2006) 246–250.
[114] X. Xiao, S. Shao, Y. Ding, Z. Huang, X. Chen, K.C. Chou, Amino Acids 28 (2005) 29–35.
[115] X. Xiao, S.H. Shao, Y.S. Ding, Z.D. Huang, K.C. Chou, Amino Acids 30 (2006) 49–54.
[116] X. Xiao, K.C. Chou, Protein Pept. Lett. 14 (2007) 871–875.
[117] M. Chen, W.Q. Huang, Genomics Proteomics Bioinformatics 3 (2005) 225–230.
[118] C. Thachuk, A. Shmygelska, H.H. Hoos, BMC Bioinformatics 8 (2007) 342.
[119] X.S. Zhang, Y. Wang, Z.W. Zhan, L.Y. Wu, L. Chen, J. Bioinform. Comput. Biol. 3 (2005) 385–400.
[120] M. Jiang, B. Zhu, J. Bioinform. Comput. Biol. 3 (2005) 19–34.
[121] A. Gupta, J. Manuch, L. Stacho, J. Comput. Biol. 12 (2005) 1328–1345.
[122] A. Gupta, J. Manuch, L. Stacho, Proc. IEEE Comput. Syst. Bioinform. Conf. (2004) 311–318.
[123] B. Berger, T. Leighton, J. Comput. Biol. 5 (1998) 27–40.
[124] R. Agarwala, S. Batzoglou, V. Dancik, S.E. Decatur, S. Hannenhalli, M. Farach, et al., J. Comput. Biol. 4 (1997) 275–296.
[125] A. Nandy, Comput. Appl. Biosci. 12 (1996) 55–62.
[126] A. Nandy, M. Harle, S.C. Basak, ARKIVOC 9 (2006) 211–238.
[127] A. Nandy, Indian J. Biochem. Biophys. 31 (1994) 149–155.
[128] A. Nandy, S.C. Basak, J. Chem. Inf. Comput. Sci. 40 (2000) 915–919.
[129] C. Raychaudhury, A. Nandy, J. Chem. Inf. Comput. Sci. 39 (1999) 243–247.
[130] M. Randic, M. Vracko, A. Nandy, S.C. Basak, J. Chem. Inf. Comput. Sci. 40 (2000) 1235–1244.
[131] A. Nandy, A. Ghosh, P. Nandy, In Silico Biol. 9 (2009).
[132] M.A. Gates, J. Theor. Biol. 119 (1986) 319–328.
[133] P.M. Leong, S. Morgenthaler, Comput. Appl. Biosci. 11 (1995) 503–507.
[134] M. Randic, X. Guo, S.C. Basak, J. Chem. Inf. Comput. Sci. 41 (2001) 619–626.
[135] G. Aguero-Chapin, H. Gonzalez-Diaz, G. de la Riva, E. Rodriguez, A. Sanchez-Rodriguez, G. Podda, et al., J. Chem. Inf. Model. 48 (2008) 434–448.
[136] M.A. Dea-Ayuela, Y. Perez-Castillo, A. Meneses-Marcel, F.M. Ubeira, F. Bolas-Fernandez, K.C. Chou, et al., Bioorg. Med. Chem. 16 (2008) 7770–7776.
[137] G. Agüero-Chapin, H. Gonzalez-Diaz, R. Molina, J. Varona-Santos, E. Uriarte, Y. Gonzalez-Diaz, FEBS Lett. 580 (2006) 723–730.
[138] S. Vilar, H. González-Díaz, L. Santana, E. Uriarte, J. Comput. Chem. 29 (2008) 2613–2622.
[139] A. Perez-Bello, C.R. Munteanu, F.M. Ubeira, A. Lopes De Magalhaes, E. Uriarte, H. Gonzalez-Diaz, J. Theor. Biol. (2008).
[140] H. Gonzalez-Diaz, F. Prado-Prado, F.M. Ubeira, Curr. Top. Med. Chem. 8 (2008) 1676–1690.
[141] M. Cruz-Monteagudo, H. González-Díaz, F. Borges, E.R. Dominguez, M.N. Cordeiro, Chem. Res. Toxicol. (2008) 619–632.
[142] M. Cruz-Monteagudo, C.R. Munteanu, F. Borges, M.N. Cordeiro, E. Uriarte, H. Gonzalez-Diaz, Bioorg. Med. Chem. 16 (2008) 9684–9693.
[143] G. Ferino, H. Gonzalez-Diaz, G. Delogu, G. Podda, E. Uriarte, Biochem. Biophys. Res. Commun. 372 (2008) 320–325.
[144] E. Estrada, E. Molina, J. Mol. Graph. Model. 20 (2001) 54–64.
[145] E. Estrada, E. Molina, E. Uriarte, SAR QSAR Environ. Res. 12 (2001) 445–459.
[146] E. Molina, H.G. Diaz, M.P. Gonzalez, E. Rodriguez, E. Uriarte, J. Chem. Inf. Comput. Sci. 44 (2004) 515–521.
[147] E. Estrada, E. Molina, I. Perdomo-Lopez, J. Chem. Inf. Comput. Sci. 41 (2001) 1015–1021.
[148] Hypercube Inc., Hyperchem Software. Release 7.5 for Windows, Molecular Modeling System, Hypercube Inc, Gainesville, FL, USA, 2002.
[149] B.F. Eichman, B.H. Mooers, M. Alberti, J.E. Hearst, P.S. Ho, J. Mol. Biol. 308 (2001) 15–26.
[150] T. Clark, A Handbook of Computational Chemistry, John Wiley & Sons, New York, 1985.
[151] N. Kitamura, S. Kohtani, R. Nakagaki, J. Photochem. Photobiol. C: Photochem. Rev. 6 (2005) 168–185.
[152] J.W. Tessman, S.T. Isaacs, J.E. Hearst, Biochemistry (Mosc) 24 (1985) 1669–1676.
[153] G.D. Cimino, H.B. Gamper, S.T. Isaacs, J.E. Hearst, Ann. Rev. Biochem. 54 (1985) 1151–1193.
[154] S. Caffieri, G. Miolo, F. Dall'Acqua, F. Benetollo, G. Bombieri, Photochem. Photobiol. 72 (2000) 23–27.
[155] Y. Tominaga, W.L. Jorgensen, J. Med. Chem. 47 (2004) 2534–2549.
[156] H. González-Díaz, L.A. Torres-Gomez, Y. Guevara, M.S. Almeida, R. Molina, N. Castanedo, et al., J. Mol. Model. 11 (2005) 116–123.
[157] H. González-Díaz, O. Gia, E. Uriarte, I. Hernadez, R. Ramos, M. Chaviano, et al., J. Mol. Model. 9 (2003) 395–407.
[158] H. González-Díaz, E. Olazabal, N. Castanedo, I.H. Sanchez, A. Morales, H.S. Serrano, et al., J. Mol. Model. 8 (2002) 237–245.
[159] H. Van Waterbeemd, Discriminant analysis for activity prediction, in: H. Van Waterbeemd (Ed.), Chemometric Methods in Molecular Design, Wiley-VCH, New York, 1995, pp. 265–282.
[160] E. Estrada, E. Molina, J. Chem. Inf. Comput. Sci. 41 (2001) 791–797.
[161] STATISTICA. 6.0 for Windows, Statsoft Inc., 2001.
[162] R.B. Kowalski, S. Wold, Pattern recognition in chemistry, in: P.R. Krishnaiah, L.N. Kanal (Eds.), Handbook of Statistics, North Holland Publishing Company, Amsterdam, 1982, pp. 673–697.
[163] H. Van Waterbeemd, Chemometric Methods in Molecular Design, Wiley-VCH, New York, 1995.
[164] M. Cruz-Monteagudo, H. Gonzalez-Diaz, G. Aguero-Chapin, L. Santana, F. Borges, E.R. Dominguez, et al., J. Comput. Chem. 28 (2007) 1909–1923.
[165] M.H. Kutner, C.J. Nachtsheim, J. Neter, W. Li, Standardized multiple regression model, Applied Linear Statistical Models, fifth ed. McGraw Hill, New York, 2005, pp. 271–277.
[166] K.C. Chou, C.T. Zhang, Crit. Rev. Biochem. Mol. Biol. 30 (1995) 275–349.
[167] K.C. Chou, H.B. Shen, Nat. Protoc. 3 (2008) 153–162.
[168] K.C. Chou, H.B. Shen, Anal. Biochem. 370 (2007) 1–16.
[169] G.P. Zhou, J. Protein Chem. 17 (1998) 729–738.
[170] G.P. Zhou, N. Assa-Munt, PROTEINS: Struct. Funct. Genet. 44 (2001) 57–59.
[171] Y.S. Ding, T.L. Zhang, Pattern Recognit. Lett. 29 (2008) 1887–1892.
[172] G.P. Zhou, K. Doctor, PROTEINS: Struct. Funct. Genet. 50 (2003) 44–48.
[173] A. Meneses-Marcel, O.M. Rivera-Borroto, Y. Marrero-Ponce, A. Montero, Y. Machado Tugores, J.A. Escario, et al., J. Biomol. Screen. 13 (2008) 785–794.
[174] Y. Marrero-Ponce, A. Meneses-Marcel, O.M. Rivera-Borroto, R. Garcia-Domenech, J.V. De Julian-Ortiz, A. Montero, et al., J. Comput. Aided Mol. Des. 22 (2008) 523–540.
[175] G.M. Casanola-Martin, Y. Marrero-Ponce, M. Tareq Hassan Khan, F. Torrens, F. Perez-Gimenez, A. Rescigno, J. Biomol. Screen. 13 (2008) 1014–1024.
[176] Y.M. Alvarez-Ginarte, Y. Marrero-Ponce, J.A. Ruiz-Garcia, L.A. Montero-Cabrera, J.M. Garcia de la Vega, P. Noheda Marin, et al., J. Comput. Chem. 29 (2008) 317–333.
[177] G.M. Casanola-Martin, Y. Marrero-Ponce, M.T. Khan, A. Ather, K.M. Khan, F. Torrens, et al., Eur. J. Med. Chem. 42 (2007) 1370–1381.
[178] Y.M. Alvarez-Ginarte, Y. Marrero-Ponce, J.A. Ruiz-Garcia, L.A. Montero-Cabrera, J.M. Vega, P. Noheda Marin, et al., J. Comput. Chem. (2007).