International Journal of Pharmaceutics 363 (2008) 78–84

Contents lists available at ScienceDirect

International Journal of Pharmaceutics

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

Application of molecular topology to the prediction of the antimalarial activity of a group of uracil-based acyclic and deoxyuridine compounds

Ramón García-Domenech*, Wanda López-Peña, Yessenia Sanchez-Perdomo, José R. Sanders, Mercedes M. Sierra-Araujo, Claudy Zapata, Jorge Gálvez

Department Química Física, Facultad de Farmacia, Universitat de Valencia, Avd. V.A. Estellés, s/n, 46100-Burjassot, Valencia, Spain

ARTICLE INFO

Article history:
Received 15 May 2008
Received in revised form 1 July 2008
Accepted 5 July 2008
Available online 15 July 2008

Keywords:
Molecular topology
Antimalarial activity
Linear discriminant analysis
Multilinear regression analysis

ABSTRACT

A topological–mathematical model has been arranged to search for new derivatives of deoxyuridine and related compounds acting as antimalarials against *Plasmodium falciparum*. By using linear discriminant and multilinear regression analysis a model with two functions was capable to predict adequately the IC_{50} for each compound of the training and test series. After carrying out a virtual screening based upon such a model, new structures potentially active against *P. falciparum* are proposed.

1. Introduction

Every year, 300–500 million people contract malaria and about 2–3 million die; most of them are children under 5 years of age. Malaria remains an important health problem in the tropics and subtropics. The emergence of chloroquine-resistant strains of *Plasmodium falciparum* and the increasing resistance against established antimalarial drugs emphasizes the demand for new effective drugs with new chemotherapeutic targets (Ridley, 2002).

Recently, it has discovered that some 5′-tritilated nucleosides are selective inhibitors of the *P. falciparum* enzyme deoxyuridine nucleotidohydrolyase (dUT-Pase) (Nguyen et al., 2005). This enzyme, is involved in nucleotide metabolism and is found in almost all organisms. In addition, dUTPase has been showed to be essential for DNA replication and cell viability and therefore, can be exploited as a drug target (Nguyen et al., 2006).

Currently, different methods are used to design new drugs; one of them being molecular topology, particularly molecular connectivity (Kier and Hall, 1976), which has demonstrated to be a useful formalism to find quantitative structure–activity relationships (QSAR). One of the most interesting advantages of molecular topology is the straightforward calculation of the topological descriptors: in this method each structure is assimilated as a hydrogen depleted graph where the atoms are represented by vertices and the bonds by edges; the connectivity between each atom to the others is included into the topological, either distance or adjacency, matrices. Mathematical manipulation of such matrices provides different sets of numbers called topological indices, TIs (Kier et al., 1976) which has widely demonstrated its ability for an easy and efficient characterization of molecular structure. When these indices are adequately selected, it is possible to obtain a very specific characterization of each chemical compound, and therefore, they can be used in QSAR models (Ivanciuc et al., 1998; Hosoya et al., 1999; de Gregorio et al., 1998; Duart et al., 2002; Garcia-Domenech et al., 2008).

In this way the TIs have demonstrated their utility for the selection and design of new drugs (Galvez et al., 1995), particularly as antimalarials (Mahmoudi et al., 2006, 2008), antivirals (de Julian-Ortiz et al., 1999), antihistaminics (Duart et al., 2005), hipoglycemics, (Calabuig et al., 2004), analgesics (Galvez et al., 1994a,b), antituberculosis (Garcia-Garcia et al., 2005), ... In some cases, the predicted structures can be regarded as new lead drugs.

In the present study, the antimalarial activities of a group of uracil-based acyclic and deoxyuridine compounds against chloroquine-resistant K1 strain of *P. falciparum* were investigated to obtain QSAR models of prediction using molecular topology, linear discriminant and the multilinear regression analysis. In addtition, a screening molecular was realized to select new compounds with theoretical higher bioactivity.

* Corresponding author. Tel.: +34 963544291.
E-mail address: ramon.garcia@uv.es (R. García-Domenech).

0378-5173/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijpharm.2008.07.003

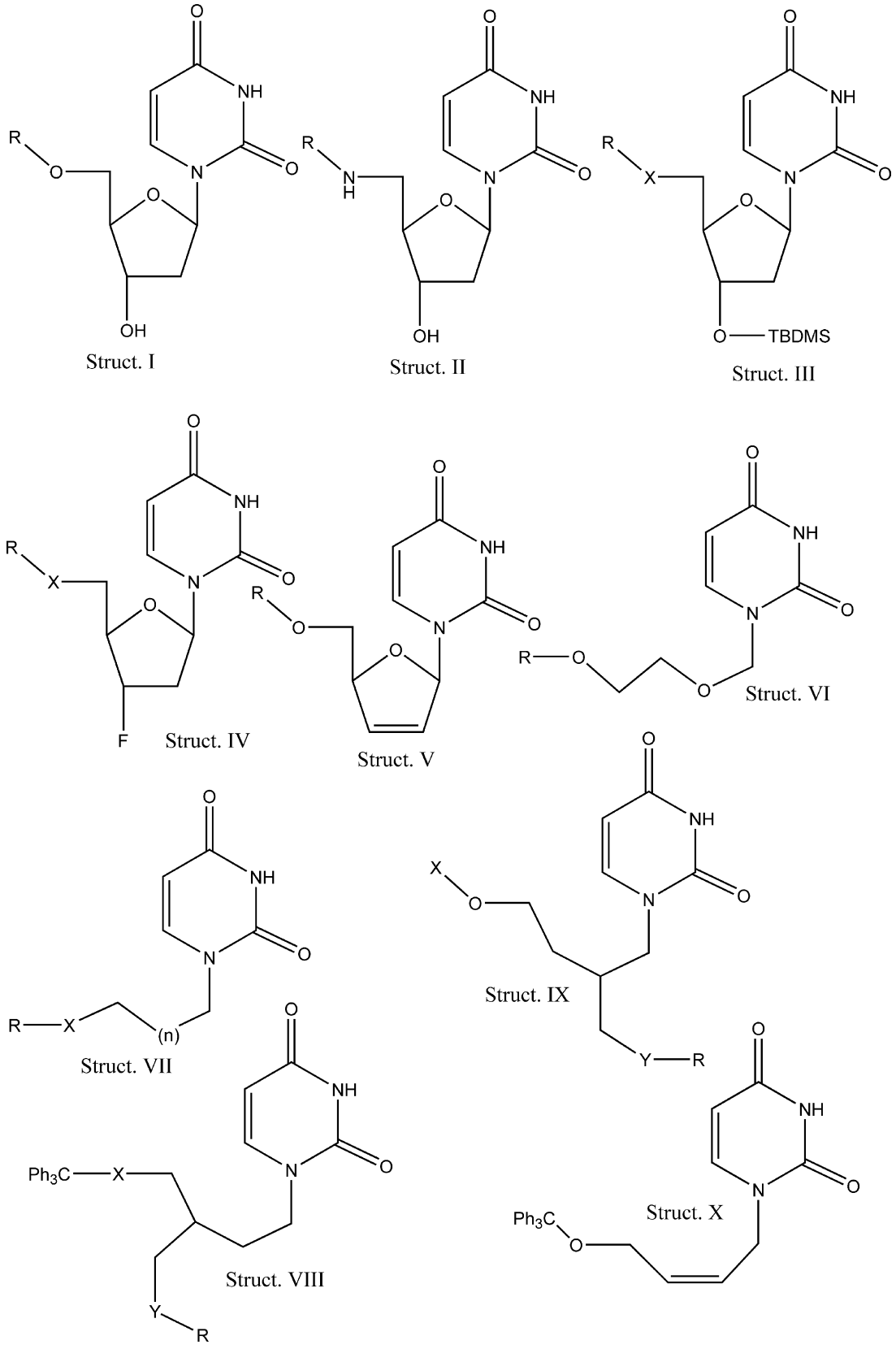


Fig. 1. Base structures of the uracil-based acyclic and deoxyuridine compounds used in this study.

Table 1
Chemical structures and values of the growth inhibition of *P. falciparum* IC_{50} (μM) for the group of deoxyuridine derivates

Compound	Structure	R	X	Y	*n*	IC_{50}exp
Acyclic-4c	VII	Ph_3Si	O	–	2	0.36
Acyclic-5c	VII	Ph_3Si	O	–	3	0.38
Acyclic-6c	VII	Ph_3Si	O	–	4	0.46
Acyclic-3c	VII	Ph_3Si	O	–	1	0.58
Acyclic-10c	IX	N_2	Ph_3C	N	–	0.64
Acyclic-2a	VI	Ph_3C	–	–	–	0.90
Acyclic-9c	VIII	$CH_3O(CO)$	O	NH	–	0.90
dURD-3g	III	Ph_3C	O	–	–	1.00
dURD-4c	IV	TPS	O	–	–	1.00
dURD-5h	V	TPS	–	–	–	1.00
Acyclic-5a	VII	Ph_3C	O	–	3	1.10
Acyclic-6b	VII	Ph_3C	NH	–	4	1.10
dURD-1m	I	TPS	–	–	–	1.10
Acyclic-10a	IX	H	Ph_3C	O	–	1.20
dURD-3h	III	Ph_3C	NH	–	–	1.80
Acyclic-10b	IX	Ph_3C	H	O	–	1.90
dURD-4e	IV	Ph_3C	O	–	–	2.00
Acyclic-5b	VII	Ph_3C	NH	–	3	2.20
Acyclic-6a	VII	Ph_3C	O	–	4	2.30
Acyclic-9a	VIII	CH_3CO	NH	O	–	2.80
dURD-5g	V	TBDPS	–	–	–	3.00
Acyclic-9b	VIII	H	NH	O	–	3.20
Acyclic-6g	VII	PhC(O)	O	–	4	3.40
Acyclic-7a	X	–	–	–	–	3.50
Acyclic-4b	VII	Ph_3C	NH	–	2	3.80
Acyclic-6d	VII	*t*-$BuPh_2Si$	O	–	4	3.90
dURD-3i	III	$(PhO)_2P(O)$	NH	–	–	4.00
Acyclic-3b	VII	Ph_3C	NH	–	1	4.40
dURD-2j	II	Ph_3C	–	–	–	4.50
Acyclic-4a	VII	Ph_3C	O	–	2	4.90
dURD-4f	IV	Ph_3C	NH	–	–	5.30
dURD-1j	I	Ph_3C	–	–	–	6.00
dURD-3a	III	$(PhO)_2P(O)$	O	–	–	6.00
dURD-1l	I	TBDPS	–	–	–	6.60
Acyclic-3a	VII	Ph_3C	O	–	1	7.50
dURD-4b	IV	TBDPS	O	–	–	8.80
Acyclic-4h	VII	$PhCH_2OC(O)$	NH	–	2	15.0
Acyclic-4i	VII	*t*-BuOC(O)	NH	–	2	17.0
dURD-3c	III	$(EtO)_2P(O)CH_2C(O)$	O	–	–	18.0
Acyclic-6e	VII	H	O	–	4	19.0
Acyclic-5e	VII	H	O	–	3	20.0
dURD-1i	I	$H_2NSO_2NHC(O)$	–	–	–	20.0
dURD-3d	III	$H_2NSO_2NHC(O)$	O	–	–	37.0
Acyclic-4g	VII	PhC(O)	O	–	2	38.0
dURD-3e	III	H_2NSO_2	O	–	–	40.0
dURD-5f	V	TBDMS	–	–	–	51.0
dURD-5a	V	$(PhO)_2P(O)$	–	–	–	54.0
Acyclic-2f	VI	*t*-$BuMe_2Si$	–	–	–	62.0
dURD-2b	II	$(EtO)_2P(O)CH_2C(O)$	–	–	–	78.0
dURD-1c	I	$(EtO)_2P(O)CH_2C(O)$	–	–	–	98.0
Acyclic-2g	VI	PhC(O)	–	–	–	127.0
dURD-5e	V	H_2NSO_2	–	–	–	163.0
dURD-5i	V	H	–	–	–	214.0
Acyclic-4e	VII	H	O	–	2	235.0
dURD-3b	III	$(EtO)_2P(O)$	O	–	–	>10
dURD-1a	I	$(PhO)_2P(O)$	–	–	–	>11
dURD-2a	II	$(PhO)_2P(O)$	–	–	–	>11
dURD-5c	V	$CH_3(CH_2)_{14}C(O)$	–	–	–	>11
dURD-5j	V	Ph_3C	–	–	–	>11
dURD-1g	I	Ts	–	–	–	>13
dURD-1k	I	TIPS	–	–	–	>13
dURD-2c	II	AdamantylC(O)	–	–	–	>13
dURD-2g	II	Ts	–	–	–	>13
dURD-5d	V	AdamantylC(O)	–	–	–	>13
dURD-1b	I	$(EtO)_2P(O)$	–	–	–	>14
dURD-4a	IV	TBDMS	O	–	–	>14
dURD-1f	I	PhC(O)	–	–	–	>15
dURD-2f	II	PhC(O)	–	–	–	>15
dURD-3f	III	H	O	–	–	>15
dURD-1e	I	$CH_3(CH_2)_3C(O)$	–	–	–	>16
dURD-2e	II	$CH_3(CH_2)_3C(O)$	–	–	–	>16
dURD-2h	II	CH_3SO_2	–	–	–	>16
dURD-2i	II	$PhCH_2$	–	–	–	>16
Acyclic-3f	VII	*t*-$BuMe_2Si$	O	–	1	>18
Acyclic-3g	VII	PhC(O)	O	–	1	>18
dURD-1d	I	$CH_3C(O)$	–	–	–	>19
dURD-2d	II	$CH_3C(O)$	–	–	–	>19
dURD-4D	IV	H	O	–	–	>21
Acyclic-3e	VII	H	O	–	1	>29
dURD-5b	V	$H_2NSO_2NHC(O)$	–	–	–	>150
dURD-1h	I	H_2NSO_2	–	–	–	>163

TBDMS, *tert*-Butyldimethylsilyl; TBDPS, *tert*-butyldiphenylsilyl; TIPS, triisopropylsilyl; TPS, triphenylsilyl; Ts, *p*-toluenesulfonyl.

2. Materials and methods

2.1. Analysed compounds

In this study we have selected a group of 81 novel uracil-based acyclic and deoxyuridine compounds inhibitors of deoxyuridine 5′-triphosphate nucleotidohydrolase (dUTPase) with antimalarial activity. Fig. 1 and Table 1 show the structures and antimalarial activity (values of IC_{50} (μM) for the in vitro assays with erythrocytic stages of the chloroquine-resistant K1 strain of *P. falciparum*) for each compound reported through the papers (Nguyen et al., 2005, 2006).

2.2. Molecular descriptors

A set of well-known topological descriptors was used in this work: Subgraph Randić–Kier–Hall like indices up to the fourth order ($^{m}\chi_t$, $^{m}\chi_t{}^{v}$) (Kier et al., 1976; Kier and Hall, 1983), topological charge indices, TCI, up to the fifth order (J_m, G_m, J_m^v, G_m^v) (Galvez et al., 1994a,b), quotients and differences between valence and non-valence connectivity indices ($^{m}C_t = {}^{m}\chi_t/{}^{m}\chi_t^v$ and $^{m}D_t = {}^{m}\chi_t - {}^{m}\chi_t^v$). Each compound was characterized by a set of 52 descriptors. Table 2 shows the symbol, name and definition of each descriptor. All descriptors used in this work were obtained with the aid of the Desmol11 program (available by e-mail request).

2.3. QSAR algorithms

2.3.1. Linear discriminant analysis

The objective of the linear discriminant analysis (LDA), which is considered as a heuristic algorithm able to distinguish between two or more categories or objects, is to find a linear function able to discriminate between the active and inactive compounds as for their different descriptor's values. Two sets of compounds: The first with a proven antimalarial activity (in our case, all the compounds with $IC_{50} < 10$ μM) and the second comprised of inactive compounds ($IC_{50} > 10$ μM) were considered for the analysis. The discriminant ability was tested by the percentage of correct classifications into each group. LDA was performed by using the BMDP 7M package (Dixon, 1990). The selection of the descriptors was based on the *F*-Snedecor parameter, and the classification criteria was the shortest Mahalanobis distance (distance of each case to the mean of all cases used in the regression equation). 7M chooses the variables used in computing the linear classification functions in a stepwise manner. At each step the variable that adds the most to the separation of the groups is entered into (or the variable that adds the least is removed from) the discriminant function. The quality of the discriminant function is evaluated by the Wilk's lambda parameter, λ, which is a multivariate analysis of variance statistic that tests the equality of group means for the variable(s) in the discriminant function.

From the selected discriminant function, the corresponding distribution diagram of antimalarial activity, PDD, was drawn. This

Table 2
Descriptors used in this study

Symbol	Name	Definition
${}^{k}\chi_t$, k = 0–4 and t = p, c, pc	Randić-like indices of order k and type path (p), cluster (c) and path-cluster (pc)	${}^{k}\chi_t = \sum_{j=1}^{k_{n_t}} \left(\prod_{i \in S_j} \delta_i \right)^{-1/2}$, where δ_i is the number of bonds, σ or π, of the atom i to non-hydrogen atoms. S_j is the jth sub-structure of order k and type t
${}^{k}\chi_t^{v}$, k = 0–4 and t = p, c, pc	Kier–Hall indices of order k and type path (p), cluster (c) and path-cluster (pc)	${}^{k}\chi_t^{v} = \sum_{j=1}^{k_{n_t}} \left(\prod_{i \in S_j} \delta_i^{v} \right)^{-1/2}$, where δ_i^{v} is the Kier–Hall valence of the atom i. S_j is the jth sub-structure of order k and type t
G_k, k = 1–5	Topological charge indices of order k	$G_k = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left\| \mathbf{M}_{ij} - \mathbf{M}_{ji} \right\| \rightleftarrows \delta(k, \mathbf{D}_{ij})$, where $\mathbf{M} = \mathbf{AQ}$ is the product of the adjacency and inverse squared distance matrices for the hydrogen-depleted molecular graph. $\mathbf{D}$ is the distance matrix. δ is the Kronecker delta
G_k^{v}, k = 1–5	Valence topological charge indices of order k	$G_k^{v} = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left\| \mathbf{M}_{ij}^{v} - \mathbf{M}_{ji}^{v} \right\| \delta(k, \mathbf{D}_{ij})$, where $\mathbf{M}^{v} = \mathbf{A}^{v}\mathbf{Q}$ is the product of the electronegativity-modified adjacency and inverse squared distance matrices for the hydrogen-depleted molecular graph. $\mathbf{D}$ is the distance matrix. δ is the Kronecker delta
J_k, k = 1–5	Normalized topological charge indices of order k	$J_k = \frac{G_k}{N-1}$
J_k^{v}, k = 1–5	Normalized valence topological charge indices of order k	$J_k^{v} = \frac{G_k^{v}}{N-1}$
${}^{k}D_t$, k = 0–4 and t = p, c, pc	Connectivity differences of order k and type path (p), cluster (c) and path-cluster (pc)	${}^{k}D_t = {}^{k}\chi_t - {}^{k}\chi_t^{v}$
${}^{k}C_t$, k = 0–4 and t = p, c, pc	Connectivity quotients of order k and type path (p), cluster (c) and path-cluster (pc)	${}^{k}C_t = \frac{{}^{k}\chi_t}{{}^{k}\chi_t^{v}}$

Table 3
Results of prediction obtained by multilinear regression and lineal discriminant analysis with IC_{50} *P. falciparum* and the training group

Compound	IC_{50} exp[a]	log IC_{50} exp	log IC_{50} calc[b]	DF	Class
Acyclic-4c	0.36	−0.444	−0.208	15.58	A
Acyclic-5c	0.38	−0.420	−0.377	16.49	A
Acyclic-6c	0.46	−0.337	−0.546	17.35	A
Acyclic-3c	0.58	−0.237	−0.039	14.07	A
Acyclic-10c	0.64	−0.194	0.279	20.55	A
Acyclic-2a	0.9	−0.046	0.580	11.51	A
Acyclic-9c	0.9	−0.046	0.161	13.54	A
dURD-3g	1	0.000	0.302	18.32	A
dURD-4c	1	0.000	0.066	11.97	A
dURD-5h	1	0.000	−0.091	12.31	A
Acyclic-5a	1.1	0.041	0.268	18.70	A
Acyclic-6b	1.1	0.041	0.061	20.49	A
dURD-1m	1.1	0.041	0.052	12.43	A
Acyclic-10a	1.2	0.079	0.426	17.77	A
dURD-3h	1.8	0.255	0.265	18.88	A
Acyclic-10b	1.9	0.279	0.403	18.04	A
dURD-4e	2	0.301	0.711	13.98	A
Acyclic-5b	2.2	0.342	0.230	19.75	A
Acyclic-6a	2.3	0.362	0.099	19.46	A
Acyclic-9a	2.8	0.447	0.151	17.17	A
dURD-5g	3	0.477	0.613	14.92	A
Acyclic-9b	3.2	0.505	0.404	18.67	A
Acyclic-6g	3.4	0.532	1.002	−0.12	I
Acyclic-7a	3.5	0.544	0.550	13.83	A
Acyclic-4b	3.8	0.580	0.400	18.97	A
Acyclic-6d	3.9	0.591	0.158	18.97	A
dURD-3i	4	0.602	0.640	13.38	A
Acyclic-3b	4.4	0.644	0.569	17.60	A
dURD-2j	4.5	0.653	0.660	15.45	A
Acyclic-4a	4.9	0.690	0.437	17.90	A
dURD-4f	5.3	0.724	0.673	14.99	A
dURD-1j	6	0.778	0.697	14.44	A
dURD-3a	6	0.778	0.683	12.78	A
dURD-1l	6.6	0.820	0.756	15.06	A
Acyclic-3a	7.5	0.875	0.606	16.51	A
dURD-4b	8.8	0.945	0.770	14.57	A
Acyclic-4h	15	1.176	1.110	−9.14	I
Acyclic-4i	17	1.230	1.502	−21.35	I
dURD-3c	18	1.255	1.100	−8.61	I
Acyclic-6e	19	1.279	1.692	−12.75	I
Acyclic-5e	20	1.301	1.861	−16.28	I
dURD-1i	20	1.301	1.952	−26.39	I
dURD-3d	37	1.568	1.557	−10.19	I
Acyclic-4g	38	1.580	1.341	−3.89	I
dURD-3e	40	1.602	1.801	−7.10	I
dURD-5f	51	1.708	1.828	−18.35	I
dURD-5a	54	1.732	0.935	0.97	NC
Acyclic-2f	62	1.792	1.854	−28.69	I
dURD-2b	78	1.892	1.458	−19.05	I
dURD-1c	98	1.991	1.496	−19.94	I
Acyclic-2g	127	2.104	1.484	−13.79	I
dURD-5e	163	2.212	2.053	−24.21	I
dURD-5i	214	2.330	2.166	−24.00	I
Acyclic-4e	235	2.371	2.030	−20.41	I

[a] Values of IC_{50} (μM) obtained from Nguyen et al. (2005, 2006).
[b] Obtained from Eq. (1).

diagram was pictured just to establish the intervals of the discriminant function in which the expectancy, *E*, of finding antifungal compounds is maximum. PDDs are histogram-like plots of connectivity functions in which expectancies appear on the ordinate axis. For an arbitrary interval of values of a given function, we define the expectancy of activity as: Ea = $a/(i+1)$; where "*a*" is the number of active compounds in the interval divided by the total number of active compounds and "*i*" is the number of inactive compounds in the interval divided by the total number of inactive compounds. The expectancy of inactivity is defined in a symmetrical way, as Ei = $i/(a+1)$. This representation provides a good visualization of the regions of minimum overlap, and allows the selection of regions

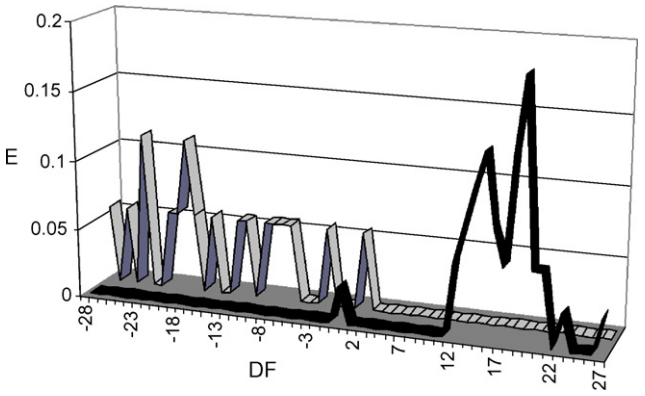


Fig. 2. Pharmacological distribution diagram for antimalarial activity by plotting expectancy (E) vs. DF (Eq. (1)) function (the black bars represent the compounds with IC_{50} < 10 μM and the grey bars, the compounds with IC_{50} > 10 μM).

in which the probability of finding active compounds reaches a maximum (Galvez et al., 1996).

2.3.2. Multilinear regression analysis

The regression equation was obtained by correlating the experimental IC_{50} values with the aforementioned TIs, by multilinear regression analysis, MLRA, through the BMDP software (Dixon, 1990). The Furnival–Wilson algorithm (Furnival, 1971), was followed to find subsets of descriptors, and minimum value of the Mallows' Cp parameter was the variable selecting criteria (Hockings, 1972). The program searched subsets with 1, 2, 3, etc., independent variables and selected the equation exhibiting the smallest Mallows Cp parameter.

Just to validate the selected prediction function, a cross-validation and an external test were carried out. For the cross-validation test, the leave-one-out algorithm was used, in which one case is eliminated from the data set and then the regression analysis with the $N-1$ remaining cases and the original descriptors (the ones selected in the first regression) is performed again. The corresponding property value for the case taken out is then predicted. The procedure is repeated as many times as there are cases in the data (Besalu, 2001).

Table 4
Results of prediction obtained by multilinear regression and lineal discriminant analysis with IC_{50} *P. falciparum* and the test group

Compound	IC_{50}exp[a]	log IC_{50}exp	log IC_{50}calc[b]	DF	Class
dURD-3b	>10	>1	1.398	−6.89	I
dURD-1a	>11	>1	1.078	0.99	I
dURD-2a	>11	>1	1.035	3.38	I
dURD-5c	>11	>1	−0.445	3.82	I
dURD-5j	>11	>1	0.554	14.42	A
dURD-1g	>13	>1	1.793	−6.35	I
dURD-1k	>13	>1	1.511	−11.39	I
dURD-2c	>13	>1	1.442	−6.15	I
dURD-2g	>13	>1	1.752	−4.84	I
dURD-5d	>13	>1	1.336	−7.04	I
dURD-1b	>14	>1	1.793	−21.25	I
dURD-4a	>14	>1	1.985	−18.93	I
dURD-1f	>15	>1	1.601	−7.47	I
dURD-2f	>15	>1	1.564	−6.47	I
dURD-3f	>15	>1	1.914	−16.68	I
dURD-1e	>16	>1	1.560	−14.72	I
dURD-2e	>16	>1	1.522	−13.62	I
dURD-2h	>16	>1	2.068	−22.47	I
dURD-2i	>16	>1	1.476	−8.80	I
Acyclic-3f	>18	>1	1.880	−21.76	I
Acyclic-3g	>18	>1	1.510	−6.61	I
dURD-1d	>19	>1	2.060	−24.17	I
dURD-2d	>19	>1	2.023	−22.90	I
dURD-4d	>21	>1	2.323	−26.71	I
Acyclic-3e	>29	>1	2.200	−25.29	I
dURD-5b	>150	>2	1.809	−26.33	I
dURD-1h	>163	>1	2.196	−24.81	I

[a] Values of IC_{50} (μM) obtained from Nguyen et al. (2005, 2006).
[b] Obtained from Eq. (1).

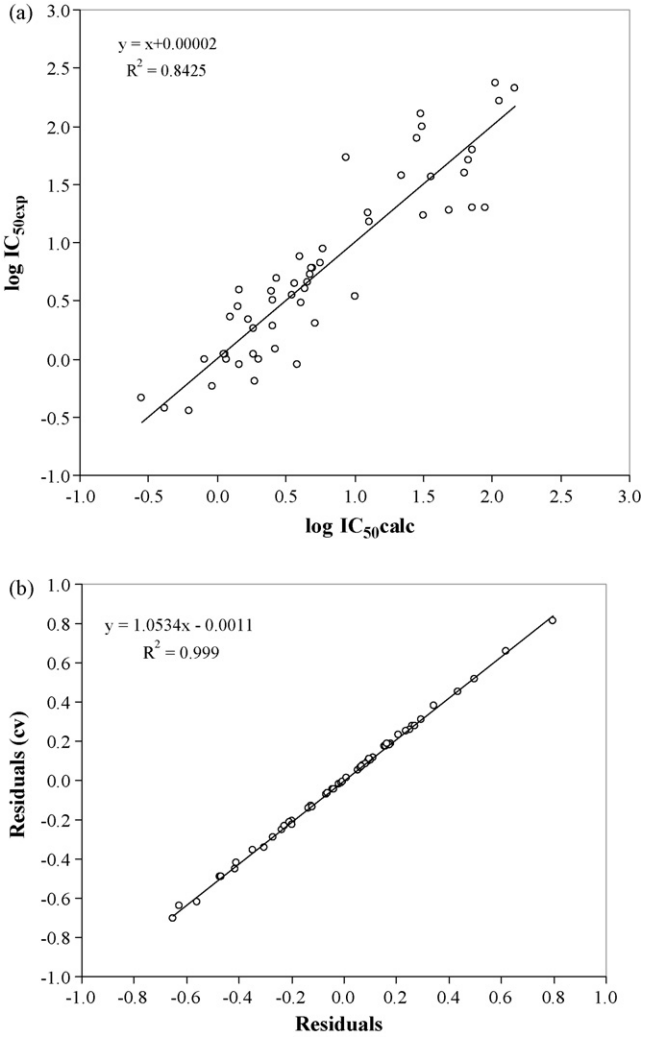


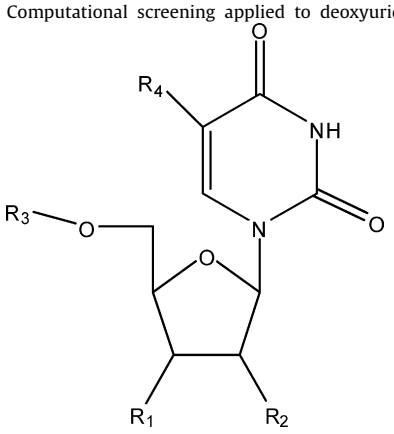
Fig. 3. Prediction of growth inhibition of *P. falciparum* IC_{50} (μM) for the group of uracil-based acyclic and deoxyuridine derivatives. (a) Graphic representation of experimental log IC_{50} against log IC_{50} calculated from Eq. (2). (b) Graphic representation of the residuals obtained in the training series against the ones obtained in the cross-validation.

3. Results and discussion

The search for an useful mathematical–topological model to predict antimalarial activity was performed in two steps. First, the selection of a discriminant function focused to distingued between the active and inactive compounds as for their antimalarial activity. Second, the obtention of a topological function capable to measure the potency of such activity in terms of IC_{50}. Both functions would comprise the framework of the mathematical model allowing the search and selection of new potent antimalarial compounds.

In order to get the discriminant function, we applied the LDA to a training set comprised of 54 compounds as well as a test set made of 27 compounds. The training set included two subsets, namely active and inactive compounds. The active compounds are referred

Table 5
Computational screening applied to deoxyuridine analogues obtained of Scifinder Scholar base and using the antimalarial model selected by molecular topology



CAS number	R_1	R_2	R_3	R_4	DF	log IC_{50}calc
137121-90-5	OH	H	CPh_3	Me	16.0	0.70
104375-88-4	OH	H	4.4′-(Phenylmethylene) bis(methoxybenzene)	I	13.9	0.27
74405-40-6	$OC(=O)CH_2CH_2COOH$	H	4.4′-(Phenylmethylene) bis(methoxybenzene)	Me	15.3	−0.03
40615-39-2	OH	H	4.4′-(Phenylmethylene) bis(methoxybenzene)	Me	14.1	0.54
23669-79-6	OH	H	4.4′-(Phenylmethylene) bis(methoxybenzene)	H	12.3	0.54
6554-10-5	OH	OH	CPh_3	H	15.3	0.92
172469-19-1	CH_2OH	H	Triphenylmethane	Me	16.3	0.54
166829-96-5	N=C(SMe)NHCN	H	((-4Methoxyphenyl) methylene)dibenzene	Me	14.2	−0.17

to as those showing $IC_{50} < 10\,\mu M$, whereas the inactive is built up from compounds showing $IC_{50} > 10\,\mu M$. The test set is comprised of all those molecules not labeled with any particular value of IC_{50}. The discriminant function selected was:

$$DF = 29.759 + 9.597G_5 - 45.479J_1^v - 1.039{}^4C_c,$$
$$N = 54, F = 121, \lambda(\text{Wilks' lambda}) = 0.121 \quad (1)$$

From here, a given compound will be selected as a potential antimalarial if DF > 0, otherwise it is classified as "inactive". The classification matrix is very significant for the training set (97.2% of correct prediction for the active group, 35 out of 36 correctly classified, and 94.4% for the inactive group, 17 correct out of 18, see Table 3, column four).

In Eq. (1) appear topological descriptors that evaluate, on the one hand, the topological aspects of each compound (^4C_c) and, on the other, the distributions of the intramolecular charge (G_5 and J_1^v).

Fig. 2 shows the antimalaric activity distribution diagram obtained with the function DF (grey and black bars represent inactive and active sets, respectively). It is apparent that the regions with minimum overlap for the compounds with theoretical antimalarial activity occur when DF > 10 and DF < 25, so the highest activity expectation occurs in these intervals. A compound will be classified as NC (non-classified) if DF > 0 and DF < 10.

A straightforward way to evaluate the quality of a given discriminant function is just applying it to an external test set. In this case, 27 compounds not present in the training set and showing IC_{50} values above 10 μM were included. Table 4 illustrates the results obtained for every compound. As may be seen, 85% of the compounds are correctly classified as inactive DF < 0 (23 out of 27). Three are non-classified (uncertain), 10 > DF > 0 (dURD-1a, dURD-2a, dURD-5c) and only one is classified as active, 25 > DF > 10 (dURD-5j).

In order to predict the theoretical value of IC_{50} for every compound, a multilinear regression analysis was carried out using the topological descriptors as independent variables and log IC_{50} as the dependent. The same training set that in LDA was used here.

The selected function was:

$$\log IC_{50} = 3.154 - 0.338{}^1\chi^v + 0.381{}^4\chi_{pc}, \quad N = 54, R^2 = 0.842,$$
$$Q^2 = 0.825, \text{SEE} = 0.308, F = 136.4, p < 0.001 \quad (2)$$

The predictive equation, Eq. (2), exhibits a R^2 value above 0.80 (0.842) what explain over 84% of the variance. The descriptors selected are χ_i indices, which encode topological information on molecular assembling. Particularly $^1\chi^v$, would take into account the degree of branching as well as molecular volume and $^4\chi_{pc}$, the existence of quaternary ramifications. Both descriptors, namely $^1\chi^v$, and $^4\chi_{pc}$, have decisive influence on the activity of those compounds showing big groups (Ph_3C or Ph_3Si over the position R in the scaffolds) (Fig. 1).

Table 3 and Fig. 3a, summarize the predictions achieved through Eq. (2) for each compound in the training set. Altogether, we can realize that there is an acceptable performance as far as 75% of compounds show residuals shorter than ±1SEE. The largest residual compound, dURD-5a, exhibits a value of 0.797. LDA classifies that compound as uncertain (NC) what suggests that either the experimental IC_{50} is not correct or the topological model used here lacks a correct classification as for its antimalarial activity.

The validation of Eq. (2) was done by a leave-one-out cross validation and using an external test. Prediction coefficient, $Q^2 = 0.825$ suggests a high predictive capability of Eq. (2) (see Fig. 3b). The results of the test are outlied in Table 4, where we can observe that the predicted values for log IC_{50} are in all cases above 1.0, what implies an estimated IC_{50} above 10 μM. There are only two exceptions corresponding to the compounds dURD-5c and dURD-5j.

Upon the results from LDA and multilinear regression (Eqs. (1) and (2)) a topological model for the search of novel antimalarial agents can be setup. The search is on acyclic compounds derivatives of uracil as well as deoxiuridine derivatives according to the following requirements.

If DF > 10 and DF < 25 and log $IC_{50} < 1$ then the compound is labeled as *potential antimalarial*. Otherwise the compound is classified as inactive.

Table 3, column 6, illustrates the classification for every compound analyzed according to the above rules. For the training set,

the rate of correctly classified compounds is 98% (53 out of 54 compounds) whereas for the test set, the overall degree of success is over 96% (26 out 27 compounds).

Based on the topological model achieved, a virtual screening on the Scifinder Scholar database was carried out searching for new deoxyuridene derivatives showing a better activity. Table 5 shows the number and position of the substitutes on R_1, R_2, R_3 and R_4, from the scaffold used together with some of the compounds selected.

Some of the compounds, namely 104375-88-4, 40615-39-2, 6554-10-5, 172469-19-1 and 166829-96-5, have showed some antiviral activity (Wengel et al., 1995; Hernandez et al., 2002; Andronova et al., 2003); Others, such as 23669-79-6 show anti-leishmania activity (actually *Leichmania major*) (Hidalgo-Zarco et al., 2001). The discovery of a new application, i.e. the antimalarial activity, for these compounds would be very interesting since it may spread the field of their therapeutical landscape.

These suggestive results need to be corroborated with the corresponding antimalarial activity assays, which should allow the validation or evaluation of the model proposed and serve as useful tool for the search of novel compounds with a higher activity against *P. falciparum*.

4. Conclusions

Molecular topology has been used with success for finding a QSAR model to predict the antimalarial activity of a group of uracil-based acyclic and deoxyuridine compounds. All the molecular descriptors used are graph-theoretical ones. The mathematical model employed in this work retains the main structural features that involve the correlated property, IC_{50}, and, therefore, can be applicable to the search of new active compounds by virtual screening throughout databases. We have built up a virtual library with several hundreds of deoxyuridine derivatives for virtually seeking and optimizing the antimalarial activity against *P. falciparum*. Interesting improvements in the activity have been obtained.

Acknowledgements

We thank the Fondo de Investigación Sanitaria, Ministerio de Sanidad, Spain (project SAF2005-PI052128) and Master Internacional en Enfermedades Parasitarias Tropicales, Universitat de Valencia, for support of this work.

References

Andronova, V.L., Skorobogatyi, M.V., Manasova, E.V., Berlin Iu, A., Korshun, V.A., Galegov, G.A., 2003. Antiviral activity of some 5-arylethynyl 2′-deoxyuridine derivatives. Bioorg. Khim. 29, 290–295.

Besalu, E., 2001. Fast computation of cross-validated properties in full linear leave-many-out procedures. J. Math. Chem. 29, 191–204.

Calabuig, C., Anton-Fos, G.M., Galvez, J., Garcia-Domenech, R., 2004. New hypoglycemic agents selected by molecular topology. Int. J. Pharmaceut. 278, 111–118.

de Gregorio, C., Kier, L.B., Hall, L.H., 1998. QSAR modeling with the electrotopological state indices: corticosteroids. J. Comput. Aided Mol. Des. 12, 557–561.

de Julian-Ortiz, J.V., Galvez, J., Munoz-Collado, C., Garcia-Domenech, R., Gimeno-Cardona, C., 1999. Virtual combinatorial syntheses and computational screening of new potential anti-herpes compounds. J. Med. Chem. 42, 3308–3314.

Dixon, W.J., 1990. BMDP Statistical Software. University of California, Berkeley.

Duart, M.J., Anton-Fos, G.M., de Julian-Ortiz, J.V., Gozalbes, R., Galvez, J., Garcia-Domenech, R., 2002. Use of molecular topology for the prediction of physico-chemical, pharmacokinetic and toxicological properties of a group of antihistaminic drugs. Int. J. Pharmaceut. 246, 111–119.

Duart, M.J., Anton-Fos, G.M., Aleman, P.A., Gay-Roig, J.B., Gonzalez-Rosende, M.E., Galvez, J., Garcia-Domenech, R., 2005. New potential antihistaminic compounds. Virtual combinatorial chemistry, computational screening, real synthesis, and pharmacological evaluation. J. Med. Chem. 48, 1260–1264.

Furnival, G.M., 1971. All possible regression with less computation. Technometrics 13, 03.

Galvez, J., Garcia-Domenech, R., De Julian-Ortiz, V., Soler, R., 1994a. Topological approach to analgesia. J. Chem. Inform. Comput. Sci. 34, 1198–1203.

Galvez, J., Garcia, R., Salabert, M.T., Soler, R., 1994b. Charge indexes. New topological descriptors. J. Chem. Inform. Comput. Sci. 34, 520–525.

Galvez, J., Garcia-Domenech, R., de Julian-Ortiz, J.V., Soler, R., 1995. Topological approach to drug design. J. Chem. Inform. Comput. Sci. 35, 938 (Erratum to document cited in CA122:177672).

Galvez, J., Garcia-Domenech, R., de Gregorio Alapont, C., de Julian-Ortiz, J.V., Popa, L., 1996. Pharmacological distribution diagrams: a tool for de novo drug design. J. Mol. Graph. 14, 272–276.

Garcia-Domenech, R., Galvez, J., de Julian-Ortiz, J.V., Pogliani, L., 2008. Some new trends in chemical graph theory. Chem. Rev. 108, 1127–1169.

Garcia-Garcia, A., Galvez, J., Vicente de Julian-Ortiz, J., Garcia-Domenech, R., Munoz, C., Guna, R., Borras, R., 2005. Search of chemical scaffolds for novel antituberculosis agents. J. Biomol. Screening 10, 206–214.

Hernandez, A.I., Balzarini, J., Karlsson, A., Camarasa, M.J., Perez-Perez, M.J., 2002. Acyclic nucleoside analogues as novel inhibitors of human mitochondrial thymidine kinase. J. Med. Chem. 45, 4254–4263.

Hidalgo-Zarco, F., Camacho, A.G., Bernier-Villamor, V., Nord, J., Ruiz-Perez, L.M., Gonzalez-Pacanowska, D., 2001. Kinetic properties and inhibition of the dimeric dUTPase-dUDPase from *Leishmania major*. Protein Sci. 10, 1426–1433.

Hockings, R.R., 1972. Criteria for selection of a subset regression: which one should be used? Technometrics 14, 967–970.

Hosoya, H., Gotoh, M., Murakami, M., Ikeda, S., 1999. Topological index and thermodynamic properties. 5. How can we explain the topological dependency of thermodynamic properties of alkanes with the topology of graphs? J. Chem. Inform. Comput. Sci. 39, 192–196.

Ivanciuc, O., Ivanciuc, T., Balaban, A.T., 1998. Quantitative structure–property relationship study of normal boiling points for halogen-/oxygen-/sulfur-containing organic compounds using the CODESSA program. Tetrahedron 54, 9129–9142.

Kier, L.B., Hall, L.H., 1976. Molecular connectivity. VII. Specific treatment of heteroatoms. J. Pharm. Sci. 65, 1806–1809.

Kier, L.B., Hall, L.H., 1983. General definition of valence delta-values for molecular connectivity. J. Pharm. Sci. 72, 1170–1173.

Kier, L.B., Murray, W.J., Randic, M., Hall, L.H., 1976. Molecular connectivity. V. Connectivity series concept applied to density. J. Pharm. Sci. 65, 1226–1230.

Mahmoudi, N., de Julian-Ortiz, J.-V., Ciceron, L., Galvez, J., Mazier, D., Danis, M., Derouin, F., Garcia-Domenech, R., 2006. Identification of new antimalarial drugs by linear discriminant analysis and topological virtual screening. J. Antimicrob. Chemother. 57, 489–497.

Mahmoudi, N., Garcia-Domenech, R., Galvez, J., Farhati, K., Franetich, J.-F., Sauerwein, R., Hannoun, L., Derouin, F., Danis, M., Mazier, D., 2008. New active drugs against liver stages of *Plasmodium* predicted by molecular topology. Antimicrob. Agents Chemother. 52, 1215–1220.

Nguyen, C., Kasinathan, G., Leal-Cortijo, I., Musso-Buendia, A., Kaiser, M., Brun, R., Ruiz-Perez, L.M., Johansson, N.G., Gonzalez-Pacanowska, D., Gilbert, I.H., Nguyen, C., Kasinathan, G., Leal-Cortijo, I., Musso-Buendia, A., Kaiser, M., Brun, R., Ruiz-Perez, L.M., Johansson, N.G., Gonzalez-Pacanowska, D., Gilbert, I.H., 2005. Deoxyuridine triphosphate nucleotidohydrolase as a potential antiparasitic drug target. J. Med. Chem. 48, 5942–5954.

Nguyen, C., Ruda, G.F., Schipani, A., Kasinathan, G., Leal, I., Musso-Buendia, A., Kaiser, M., Brun, R., Ruiz-Perez, L.M., Sahlberg, B.L., Johansson, N.G., Gonzalez-Pacanowska, D., Gilbert, I.H., Nguyen, C., Ruda, G.F., Schipani, A., Kasinathan, G., Leal, I., Musso-Buendia, A., Kaiser, M., Brun, R., Ruiz-Perez, L.M., Sahlberg, B.-L., Johansson, N.G., Gonzalez-Pacanowska, D., Gilbert, I.H., 2006. Acyclic nucleoside analogues as inhibitors of *Plasmodium falciparum* dUTPase. J. Med. Chem. 49, 4183–4195.

Ridley, R.G., 2002. Medical need, scientific opportunity and the drive for antimalarial drugs. Nature 415, 686–693.

Wengel, J., Schinazi, R.F., Caruthers, M.H., 1995. Stereoselective synthesis, chemistry and antiviral evaluation of 1-(2,-dideoxy-3-C-hydroxymethyl-beta-D-threo-pentofuranosyl)thymine derivatives. Bioorg. Med. Chem. 3, 1223–1229.