Arabian Journal of Chemistry (2019) **12**, 5086–5102

King Saud University

Arabian Journal of Chemistry

www.ksu.edu.sa
www.sciencedirect.com

ORIGINAL ARTICLE

Synthesis, *in vitro* evaluation and QSAR modelling of potential antitumoral 3,4-dihydropyrimidin-2-(1*H*)-thiones

Mariana Matias [a,b], **Gonçalo Campos** [a,b], **Adriana O. Santos** [a], **Amílcar Falcão** [b,c], **Samuel Silvestre** [a,b], **Gilberto Alves** [a,b,*]

[a] *CICS-UBI – Health Sciences Research Centre, University of Beira Interior, Rua Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal*
[b] *CNC – Center for Neuroscience and Cell Biology, University of Coimbra, 3004-517 Coimbra, Portugal*
[c] *Pharmacology Department, Faculty of Pharmacy, University of Coimbra, Pólo das Ciências da Saúde, Azinhaga de Santa Comba, 3000-548 Coimbra, Portugal*

Received 2 October 2016; accepted 9 December 2016
Available online 19 December 2016

KEYWORDS

Cell cycle;
Cytotoxicity;
Dihydropyrimidinthiones;
Monastrol;
QSAR studies

Abstract A series of 3,4-dihydropyrimidin-2-(1*H*)-thiones (**1**–**22**) was synthesized through the Biginelli reaction in order to find novel anticancer drug candidates based on the structure of monastrol. The antiproliferative activity of the compounds was screened in several cell lines and the chlorinated compounds expressed considerable cytotoxicity in hepatic and/or colon and MCF-7 breast cancer cell lines. Within these, compound **11** was the most potent and showed strong antiproliferative effects on HepaRG cells (IC_{50} = 0.75 μM). Using cell proliferation data, a quantitative structure-activity relationship (QSAR) analysis was performed employing Bayesian regularized artificial neural networks to relate *in silico* calculated molecular descriptors and bioactivity of the compounds across the tested cell lines. A statistical valid QSAR model was obtained, allowing the prediction of the relative cell proliferation for hypothetical analogous compounds in the studied cell lines. Additionally, cell cycle distribution analysis showed that another potent chlorinated molecule, compound **15**, caused accumulation of cells in G_0/G_1 phase of the cell cycle in HepaRG and MCF-7 cells, which suggests a distinct mechanism of action relatively to monastrol. Overall, chlorinated

* Corresponding author at: CICS-UBI – Health Sciences Research Centre, University of Beira Interior, Av. Infante D. Henrique, 6200-506 Covilhã, Portugal. Fax: +351 275 329099.
E-mail address: gilberto@fcsaude.ubi.pt (G. Alves).
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.arabjc.2016.12.007

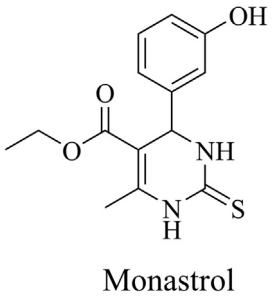
monastrol analogues showed potent cytotoxic activity in cancer cells and deserve further investigation to ascertain their potential as candidate anticancer agents.

1. Introduction

Despite the intensive research efforts to improve anticancer therapies, cancer persists as one of the major global public health concerns, having a large incidence and mortality (Siegel et al., 2015). In general, cancer treatment is based on radiotherapy, surgery, hormone therapy, immunotherapy and chemotherapy, the last pharmacological approach being the most used (Videira et al., 2014). However, due to severe side effects and lack of efficacy in advanced cancer stages, the search for novel anticancer agents with higher selectivity and lower toxicity remains a major research priority. In this context, the rational design of new drug candidates followed by biological evaluation has been a valuable strategy for the discovery and development of new drug therapies. Indeed, nowadays there is a large number of *in silico* tools available for supporting drug design/discovery, among which quantitative structure-activity relationship (QSAR) techniques play a central role (Caldwell, 2015).

In QSAR studies, a target property (bioactivity or physicochemical characteristics) is modelled for a given molecular data set, as a function of several molecular descriptors, using statistical and/or machine learning techniques. The result is the development of a predictive model that, once proven its validity, can be used to predict the target property for new related compounds. Thus, robust and well-validated QSAR models can assist the rational design of molecules with improved activity, selectivity, safety and/or physicochemical profile (Cherkasov et al., 2013; Nantasenamat et al., 2009). Among the available approaches, Bayesian regularized artificial neural networks (BRANNs) can be applied in the development, validation and use of QSAR models. To optimize the weights and regularization constant values BRANN models employ Bayesian inference to determine the posterior probability distribution of weights and related properties from a prior probability distribution. This enables the development of robust models, in which overfitting, architecture dependence and non-existent or redundant relationships are minimized (Burden and Winkler, 1999).

Although the usefulness of computer-aided drug design tools in modern drug discovery is widely recognized, the availability of robust and cost-effective synthetic routes is also essential to speed up the process. Actually, in this context, it is worthy to be noted that the Biginelli reaction has been applied to synthesize multiple bioactive molecules during drug discovery programs (Rashid et al., 2013; Suresh and Sandhu, 2012). This reaction is one of the most useful multicomponent reactions and allows the synthesis of multifunctionalized 3,4-dihydropyrimidin-2-(1*H*)-(thi)ones [DHPM(t)s] and related compounds through the simple condensation reaction of an aldehyde, an urea or thiourea, and an easily enolizable carbonyl compound (Gong et al., 2007). Over the years, DHPM(t)s and their derivatives have attracted considerable attention in organic and medicinal chemistry because they display several pharmacological and therapeutic properties (de Fátima et al., 2015) and, thus, many improved procedures with new catalysts and experimental conditions have been reported (Kolosov et al., 2009). A relevant example of a bioactive compound produced through the Biginelli reaction is monastrol (Fig. 1), which is a structurally simple DHPMt that was identified in 1999 by Mayer and collaborators as a novel bioactive cell-permeable small molecule (Mayer et al., 1999). In fact, monastrol exhibits antitumor properties by reversibly inhibiting the Kinesin-like protein KIF11 (also known as Eg5), a motor protein responsible for the formation and maintenance of the bipolar spindle in mitotic cells (Falnikar et al., 2011). However, as the antimitotic activity of monastrol is relatively weak [concentration inducing 50% inhibition of cell growth (IC_{50}) of 14 μM] (Mayer et al., 1999), this molecule is not considered a real drug candidate but rather a valuable lead molecule for the development of novel anticancer agents. Therefore, in the research for more potent compounds, several publications have described the synthesis of different entities structurally related to monastrol, such as derivatives with ester side chains and dimethylenastron analogues (Abdou et al., 2015; Abnous et al., 2013; Prokopcová et al., 2010; Svetlik et al., 2010).



Monastrol

Figure 1 Chemical structure of monastrol.

Taking into consideration the value of DHPMt scaffolds, this work aimed to find novel anticancer drug candidates structurally related to monastrol, using the Biginelli reaction. The *in vitro* evaluation of the synthesized compounds in different human cell lines included the study of their antiproliferative activity as well as cell death and cell cycle distribution. Moreover, a QSAR model for these molecules was produced, which relates *in silico* calculated molecular descriptors of the compounds to their observed cytotoxicity. In Fig. 2, it is schematically shown the workflow of the experimental and computational steps.

2. Materials and methods

2.1. Chemicals and reagents

The reagents thiourea, benzaldehyde, *p*-tolualdehyde, 4-nitrobenzaldehyde, 2,4-dichlorobenzaldehyde, 2,3-difluorobenzaldehyde and zirconium (IV) tetrachloride ($ZrCl_4$) were purchased from Acros Organics (New Jersey, USA); ethyl acetoacetate, methyl acetoacetate, acetylacetone and anisaldehyde were purchased from Merck (Hohenbrunn, Germany) and 2,3-dichlorobenzaldehyde, furaldehyde, 5-fluorouracil (5-FU) and dimethyl sulfoxide (DMSO) were purchased from Sigma-Aldrich (St. Louis, MO, USA). The ethanol 99.9% was purchased from Manuel Vieira & C^a (Torres Novas, Portugal) and deuterated DMSO (DMSO-d_6) was purchased from Armar Chemicals (Leipzig, Germany). Infrared (IR) spectra were collected on a Thermo Scientific Nicolet iS10 Smart iTR, equipped with a diamond ATR crystal. For ATR data acquisition, a sample of the solid compound was placed onto the crystal and the spectrum was recorded. An air spectrum was used as a reference in absorbance calculations. The sample spectra were collected at room temperature in the

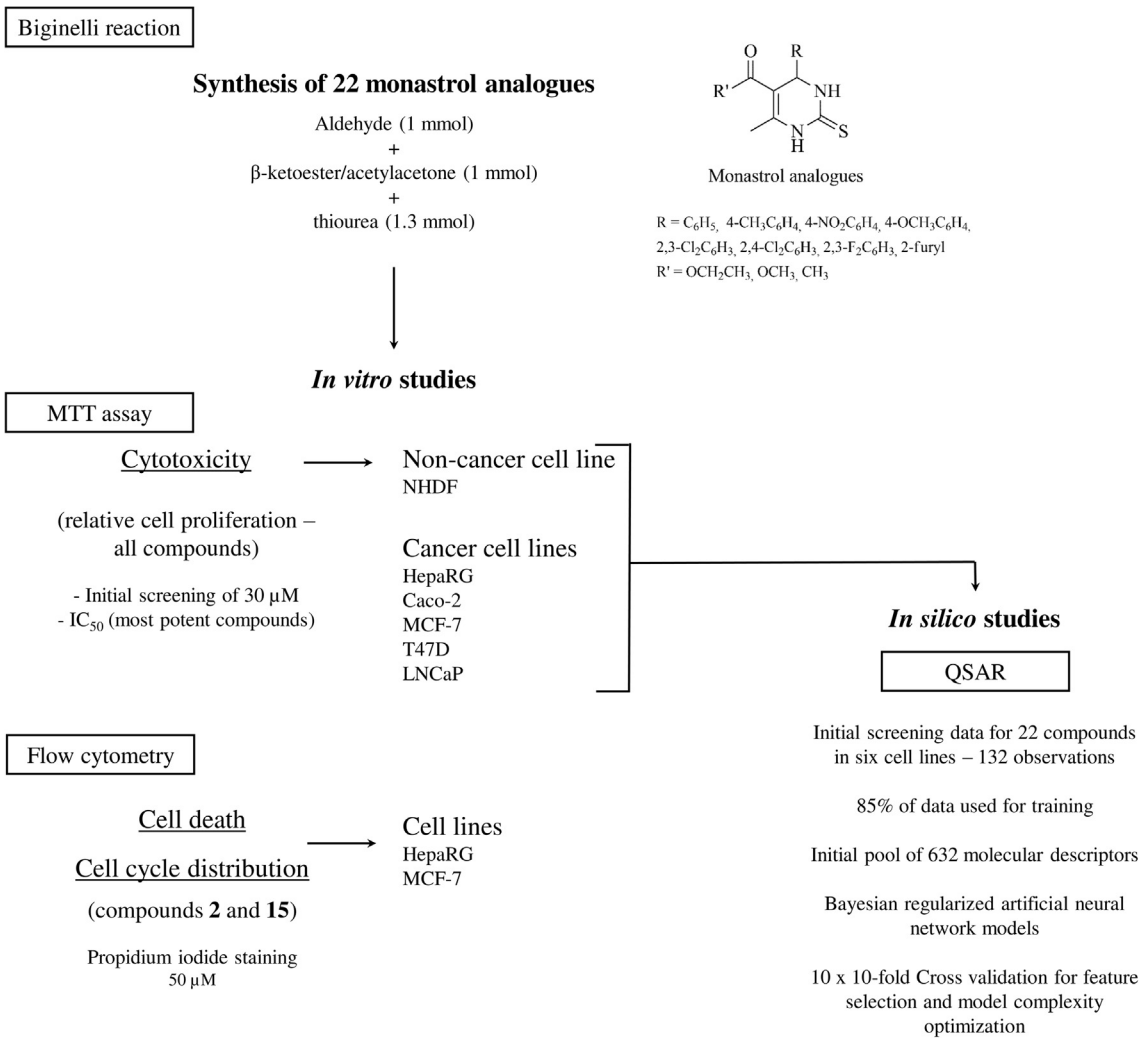


Figure 2 Schematic representation of the experimental and computational workflow.

4000–400 cm^{-1} range by averaging 16 scans at a spectral resolution of 2 cm^{-1}. Nuclear magnetic resonance (NMR) spectra (^1H NMR and ^{13}C NMR) were acquired on a Bruker Avance 400 MHz spectrometer and were processed with the software TOPSPIN 3.1 (Bruker, Fitchburg, WI, USA). DMSO-d_6 was used as solvent. Chemical shifts are reported in parts per million (ppm) relative to deuterated solvent as an internal standard. Coupling constants (J values) are reported in hertz (Hz) and splitting multiplicities are described as s = singlet; brs = broad singlet; d = doublet; dd = double doublet; t = triplet; q = quartet; dq = double quartet and m = multiplet. High resolution mass spectrometry (ESI-HRMS) was performed by the microanalysis service on a QSTAR XL instrument (Salamanca, Spain).

2.2. Synthesis and structural characterization

To a mixture of an aldehyde (1 mmol), a β-ketoester/acetylacetone (1 mmol) and thiourea (1.3 mmol), $ZrCl_4$ (5 mol%) was added, and the reaction mixture was heated with stirring at 70 °C in a preheated oil bath for the appropriate time (Table 1) until solidification. After being cooled to room temperature, cold water was added to the reaction mixture, which was then stirred for 20–30 min. The solid was separated by filtration under suction, washed with ice-cold water (40 mL), dried and then recrystallized from ethanol 99.9% to afford the pure product. All the products were characterized by IR, ^1H and ^{13}C NMR. HRMS also was included for the synthesized new products.

Ethyl 6-methyl-4-phenyl-2-thioxo-1,2,3,4-tetrahydropyrimidine-5-carboxylate (compound **1**) (Hingane and Shumaila, 2013). White solid (193 mg, 70%): IR (υ_{max}/cm^{-1}): 3325, 3170, 3104, 2980, 1666, 1572, 1175, 1117; ^1H NMR (400 MHz, DMSO-d_6) δ = 1.10 (t, 3H, J = 7.0 Hz, OCH_2-CH_3), 2.29 (s, 3H, CH_3), 4.01 (q, 2H, J = 7.1 Hz, OCH_2), 5.17 (d, 1H, J = 3.7 Hz, CH), 7.19–7.38 (m, 5H, ArH), 9.64 (brs, 1H, NH), 10.32 (brs, 1H, NH); ^{13}C NMR (100 MHz, DMSO-d_6) δ = 14.0, 17.2, 54.0, 59.6, 100.7, 126.4, 127.7, 128.6, 143.5, 145.0, 165.1, 174.2.

Methyl 6-methyl-4-phenyl-2-thioxo-1,2,3,4-tetrahydropyrimidine-5-carboxylate (compound **2**) (Wang et al., 2003). White solid (157 mg, 60%): IR (υ_{max}/cm^{-1}): 3312, 3179, 3116, 2996, 1661, 1579, 1178, 1112; ^1H NMR (400 MHz, DMSO-d_6) δ = 2.29 (s, 3H, CH_3), 3.56 (s, 3H, CH_3OCO), 5.18 (d, 1H,

Table 1 $ZrCl_4$-catalysed synthesis of 3,4-dihydropyrimidin-2-(1*H*)-thiones under solvent-free conditions at 70 °C.[a]

Compound	R	R′	Time	Yield (%)[b,c]
1	C_6H_5	OCH_2CH_3	14 min	70
2	C_6H_5	OCH_3	15 min	60
3	C_6H_5	CH_3	8 min	80
4	4-$(CH_3)C_6H_4$	OCH_2CH_3	21 min	44
5	4-$(CH_3)C_6H_4$	OCH_3	10 min	33
6	4-$(CH_3)C_6H_4$	CH_3	7 min	50
7	4-$(NO_2)C_6H_4$	CH_3	35 min	17
8	4-$(OCH_3)C_6H_4$	OCH_2CH_3	58 min	79
9	4-$(OCH_3)C_6H_4$	OCH_3	43 min	31
10	4-$(OCH_3)C_6H_4$	CH_3	1 h 40 min	89
11	2,3-$(Cl)_2C_6H_3$	OCH_2CH_3	45 min	21
12	2,3-$(Cl)_2C_6H_3$	OCH_3	4 h	19
13	2,3-$(Cl)_2C_6H_3$	CH_3	1 h	47
14	2,4-$(Cl)_2C_6H_3$	OCH_2CH_3	17 h	19
15	2,4-$(Cl)_2C_6H_3$	OCH_3	3 h	34
16	2,4-$(Cl)_2C_6H_3$	CH_3	6 h	42
17	2,3-$(F)_2C_6H_3$	OCH_2CH_3	7 h	16
18	2,3-$(F)_2C_6H_3$	OCH_3	1 h	19
19	2,3-$(F)_2C_6H_3$	CH_3	15 min	61
20	2-furyl	OCH_2CH_3	1 h	23
21	2-furyl	OCH_3	1 h	28
22	2-furyl	CH_3	1 h	57

[a] Reaction conditions: aldehyde (1 mmol), β-ketoester/acetylacetone (1 mmol), thiourea (1.3 mmol), $ZrCl_4$ (5 mol%) at 70 °C.
[b] Yield of isolated pure products after purification.
[c] All products were characterized by ^1H ^{13}C NMR, IR spectra and compared with available data in the literature. The new products were also characterized by high resolution mass spectrometry.

J = 3.7 Hz, CH), 7.19–7.38 (m, 5H, ArH), 9.67 (brs, 1H, NH), 10.36 (brs, 1H, NH); ^{13}C NMR (100 MHz, DMSO-d_6) δ = 17.2, 51.1, 53.9, 100.4, 126.3, 127.7, 128.6, 143.3, 145.3, 165.6, 174.3.

1-(6-Methyl-4-phenyl-2-thioxo-1,2,3,4-tetrahydropyrimidin-5-yl)ethanone (compound **3**) (Ranu and Hajra, 2000). Orange solid (197 mg, 80%): IR (υ_{max}/cm^{-1}): 3275, 3176, 2994, 1610, 1574, 1181; ^1H NMR (400 MHz, DMSO-d_6) δ = 2.16 (s, 3H, CH_3), 2.33 (s, 3H, CH_3CO), 5.30 (d, 1H, *J* = 3.9 Hz, CH), 7.20–7.38 (m, 5H, ArH), 9.75 (brs, 1H, NH), 10.27 (brs, 1H, NH); ^{13}C NMR (100 MHz, DMSO-d_6) δ = 18.3, 30.4, 53.8, 110.5, 126.6, 127.7, 128.7, 142.9, 144.6, 174.1, 194.8.

Ethyl 6-methyl-2-thioxo-4-*p*-tolyl-1,2,3,4-tetrahydropyrimidine-5-carboxylate (compound **4**) (Su et al., 2005). White solid (128 mg, 44%): IR (υ_{max}/cm^{-1}): 3322, 3172, 3105, 2982, 1670, 1574, 1464, 1175, 1118; ^1H NMR (400 MHz, DMSO-d_6) δ = 1.10 (t, 3H, *J* = 7.1 Hz, OCH_2CH_3), 2.26 (s, 3H, CH_3), 2.28 (s, 3H, CH_3), 4.00 (q, 2H, *J* = 7.1 Hz, OCH_2), 5.13 (d, 1H, *J* = 3.6 Hz, CH), 7.07–7.16 (m, 4H, ArH), 9.60 (brs, 1H, NH), 10.29 (brs, 1H, NH); ^{13}C NMR (100 MHz, DMSO-d_6) δ = 14.0, 17.1, 20.7, 53.7, 59.5, 100.8, 126.3, 129.0, 136.9, 140.6, 144.9, 165.1, 174.1.

Methyl 6-methyl-2-thioxo-4-*p*-tolyl-1,2,3,4-tetrahydropyrimidine-5-carboxylate (compound **5**) (Narahari et al., 2012). White solid (91 mg, 33%): IR (υ_{max}/cm^{-1}): 3166, 2996, 1714, 1654, 1572, 1186, 1102; ^1H NMR (400 MHz, DMSO-d_6) δ = 2.26 (s, 3H, CH_3), 2.28 (s, 3H, CH_3), 3.55 (s, 3H, CH_3-OCO), 5.13 (d, 1H, *J* = 3.7 Hz, CH), 7.07–7.159 (m, 4H, ArH), 9.62 (brs, 1H, NH), 10.32 (brs, 1H, NH); ^{13}C NMR (100 MHz, DMSO-d_6) δ = 17.2, 20.7, 51.1, 53.6, 100.5, 126.2, 129.1, 137.0, 140.4, 145.2, 165.6, 174.2.

1-(6-Methyl-2-thioxo-4-*p*-tolyl-1,2,3,4-tetrahydropyrimidin-5-yl)ethanone (compound **6**) (Fu et al., 2015). Orange solid (130 mg, 50%): IR (υ_{max}/cm^{-1}): 3282, 3174, 3002, 1617, 1582, 1451, 1180; ^1H NMR (400 MHz, DMSO-d_6) δ = 2.13 (s, 3H, CH_3), 2.26 (s, 3H, CH_3), 2.32 (s, 3H, CH_3CO), 5.25 (d, 1H, *J* = 3.9 Hz, CH), 7.09–7.17 (m, 4H, ArH), 9.71 (brs, 1H, NH), 10.24 (brs, 1H, NH); ^{13}C NMR (100 MHz, DMSO-d_6) δ = 18.2, 20.7, 30.3, 53.6, 110.4, 126.5, 129.2, 137.0, 140.0, 144.4, 174.0, 194.8.

1-(6-Methyl-4-(4-nitrophenyl)-2-thioxo-1,2,3,4-tetrahydro pyrimidin-5-yl)ethanone (compound **7**) (Fan et al., 2002). Orange solid (50 mg, 17%): IR (υ_{max}/cm^{-1}): 3266, 3171, 3011, 1574, 1519, 1344; ^1H NMR (400 MHz, DMSO-d_6) δ = 2.23 (s, 3H, CH_3), 2.36 (s, 3H, CH_3CO), 5.41 (d, 1H, *J* = 4.0 Hz, CH), 7.48 (d, 2H, *J* = 8.5 Hz, ArH), 8.22 (d, 2H, *J* = 8.5 Hz, ArH), 9.87 (brs, 1H, NH), 10.44 (brs, 1H, NH); ^{13}C NMR (100 MHz, DMSO-d_6) δ = 18.5, 30.7, 53.1, 110.2, 124.0, 127.8, 145.5, 146.9, 150.0, 174.5, 194.6.

Ethyl 4-(4-methoxyphenyl)-6-methyl-2-thioxo-1,2,3,4-tetra hydropyrimidine-5-carboxylate (compound **8**) (Hingane and Shumaila, 2013). Yellow solid (242 mg, 79%): IR (υ_{max}/cm^{-1}): 3309, 3166, 2982, 1664, 1573, 1508, 1169, 1121, 1026; ^1H NMR (400 MHz, DMSO-d_6) δ = 1.10 (t, 3H, *J* = 7.0 Hz, OCH_2CH_3), 2.28 (s, 3H, CH_3), 3.72 (s, 3H, OCH_3), 4.00 (q, 2H, *J* = 7.0 Hz, OCH_2), 5.11 (d, 1H, *J* = 3.7 Hz, CH), 6.90 (d, 2H, *J* = 8.8 Hz, ArH), 7.12 (d, 2H, *J* = 8.8 Hz, ArH), 9.59 (brs, 1H, NH), 10.28 (brs, 1H, NH); ^{13}C NMR (100 MHz, DMSO-d_6) δ = 14.0, 17.1, 53.4, 55.1, 59.5, 100.9, 113.9, 127.6, 135.7, 144.7, 158.7, 165.2, 174.0.

Methyl 4-(4-methoxyphenyl)-6-methyl-2-thioxo-1,2,3,4-tetra hydropyrimidine-5-carboxylate (compound **9**) (Liu et al., 2013). Yellow solid (91 mg, 31%): IR (υ_{max}/cm^{-1}): 3324, 3280, 2958, 1662, 1553, 1509, 1172, 1110, 1020; ^1H NMR (400 MHz, DMSO-d_6) δ = 2.29 (s, 3H, CH_3), 3.55 (s, 3H, CH_3OCO), 3.72 (s, 3H, OCH_3), 5.11 (d, 1H, *J* = 3.8 Hz,

CH), 6.90 (d, 2H, J = 8.6 Hz, ArH), 7.13 (d, 2H, J = 8.6 Hz, ArH), 9.61 (brs, 1H, NH), 10.31 (brs, 1H, NH); ^{13}C NMR (100 MHz, DMSO-d_6) δ = 17.2, 51.1, 53.3, 55.1, 100.7, 113.9, 127.6, 135.5, 145.0, 158.8, 165.7, 174.0.

1-(4-(4-Methoxyphenyl)-6-methyl-2-thioxo-1,2,3,4-tetrahydropyrimidin-5-yl)ethanone (compound **10**) (Kolosov et al., 2009). Orange solid (245 mg, 89%): IR (υ_{max}/cm^{-1}): 3308, 3230, 3002, 1618, 1564, 1509, 1179, 1024; ^{1}H NMR (400 MHz, DMSO-d_6) δ = 2.12 (s, 3H, CH_3), 2.32 (s, 3H, CH_3CO), 3.72 (s, 3H, OCH_3), 5.23 (d, 1H, J = 3.8 Hz, CH), 6.90 (d, 2H, J = 8.7 Hz, ArH), 7.14 (d, 2H, J = 8.7 Hz, ArH), 9.69 (brs, 1H, NH), 10.23 (brs, 1H, NH); ^{13}C NMR (100 MHz, DMSO-d_6) δ = 18.2, 30.3, 53.3, 55.1, 110.5, 114.0, 127.9, 135.1, 144.3, 158.8, 173.8, 194.9.

Ethyl 4-(2,3-dichlorophenyl)-6-methyl-2-thioxo-1,2,3,4-tetrahydropyrimidine-5-carboxylate (compound **11**) (Kappe, 2000). White solid (72 mg, 21%): IR (υ_{max}/cm^{-1}): 3170, 2994, 1545, 1477, 1194; ^{1}H NMR (400 MHz, DMSO-d_6) δ = 0.99 (t, 3H, J = 7.0 Hz, OCH_2CH_3), 2.33 (s, 3H, CH_3), 3.92 (q, 2H, J = 7.0 Hz, OCH_2), 5.70 (d, 1H, J = 3.0 Hz, CH), 7.28 (dd, 1H, J_1 = 7.8 Hz, J_2 = 1.43 Hz, ArH), 7.38 (t, 1H, J = 7.8 Hz, ArH), 7.58 (dd, 1H, J_1 = 7.8 Hz, J_2 = 1.42 Hz, ArH), 9.65 (brs, 1H, NH), 10.42 (brs, 1H, NH); ^{13}C NMR (100 MHz, DMSO-d_6) δ = 13.8, 17.0, 52.4, 59.5, 99.5, 127.9, 128.7, 129.9, 130.1, 131.9, 143.3, 145.9, 164.7, 173.8.

Methyl 4-(2,3-dichlorophenyl)-6-methyl-2-thioxo-1,2,3,4-tetrahydropyrimidine-5-carboxylate (compound **12**). White solid (63 mg, 19%): IR (υ_{max}/cm^{-1}): 3173, 3004, 1544, 1478; ^{1}H NMR (400 MHz, DMSO-d_6) δ = 2.33 (s, 3H, CH_3), 3.48 (s, 3H, CH_3OCO), 5.69 (d, 1H, J = 3.0 Hz, CH), 7.27 (dd, 1H, J_1 = 7.9 Hz, J_2 = 1.2 Hz, ArH), 7.38 (t, 1H, J = 7.8 Hz, ArH), 7.6 (dd, 1H, J_1 = 7.9 Hz, J_2 = 1.2 Hz, ArH), 9.65 (brs, 1H, NH), 10.45 (brs, 1H, NH); ^{13}C NMR (100 MHz, DMSO-d_6) δ = 17.1, 51.1, 52.4, 99.3, 127.8, 128.4, 128.8, 130.0, 132.0, 143.0, 146.1, 165.2, 138.0; HRMS (ESI-TOF): m/z [M + H]$^+$ calcd for $C_{13}H_{13}N_2O_2SCl_2$: 331.0077, found 331.0069.

1-(4-(2,3-Dichlorophenyl)-6-methyl-2-thioxo-1,2,3,4-tetrahydropyrimidin-5-yl)ethanone (compound **13**). Orange solid (148 mg, 47%): IR (υ_{max}/cm^{-1}): 3220, 2998, 1625, 1576, 1198, 1013; ^{1}H NMR (400 MHz, DMSO-d_6) δ = 2.14 (s, 3H, CH_3), 2.38 (s, 3H, CH_3CO), 5.72 (d, 1H, J = 3.6 Hz, CH), 7.21 (dd, 1H, J_1 = 8.0 Hz, J_2 = 1.4 Hz, ArH), 7.36 (t, 1H, J = 7.9 Hz, ArH), 7.58 (dd, 1H, J_1 = 8.0 Hz, J_2 = 1.42 Hz ArH), 9.68 (brs, 1H, NH), 10.39 (brs, 1H, NH); ^{13}C NMR (100 MHz, DMSO-d_6) δ = 18.3, 30.5, 52.4, 109.9, 127.6, 128.7, 130.0, 130.1, 132.1, 142.4, 145.2, 174.0, 194.4; HRMS (ESI-TOF): m/z [M + H]$^+$ calcd for $C_{13}H_{13}N_2OSCl_2$: 315.0127, found 315.0120.

Ethyl 4-(2,4-dichlorophenyl)-6-methyl-2-thioxo-1,2,3,4-tetrahydropyrimidine-5-carboxylate (compound **14**) (Wang et al., 2003). White solid (66 mg, 19%): IR (υ_{max}/cm^{-1}): 3406, 3192, 3076, 3023, 1534, 1473, 1100, 1048; ^{1}H NMR (400 MHz, DMSO-d_6) δ = 1.02 (t, 3H, J = 7.1 Hz, OCH_2-CH_3), 2.32 (s, 3H, CH_3), 3.92 (q, 2H, J = 7.1 Hz, OCH_2), 5.60 (d, 1H, J = 3.1 Hz, CH), 7.30 (d, 1H, J = 8.4 Hz, ArH), 7.44 (dd, 1H, J_1 = 8.4 Hz, J_2 = 2.1 Hz, ArH), 7.59 (d, 1H, J = 2.1 Hz, ArH), 9.62 (brs, 1H, NH), 10.40 (brs, 1H, NH); ^{13}C NMR (100 MHz, DMSO-d_6) δ = 13.9, 17.0, 51.2, 59.5, 99.3, 128.1, 128.8, 130.8, 132.8, 133.0, 139.8, 145.9, 164.6, 173.8.

Methyl 4-(2,4-dichlorophenyl)-6-methyl-2-thioxo-1,2,3,4-tetrahydropyrimidine-5-carboxylate (compound **15**) (Nasr-Esfahani et al., 2014). White solid (112 mg, 34%): IR (υ_{max}/cm^{-1}): 3406, 3178, 3090, 3025, 1545, 1470, 1101, 1047; ^{1}H NMR (400 MHz, DMSO-d_6) δ = 2.31 (s, 3H, CH_3), 3.48 (s, 3H, CH_3OCO), 5.59 (d, 1H, J = 3.1 Hz, CH), 7.29 (d, 1H, J = 8.5 Hz, ArH), 7.44 (dd, 1H, J_1 = 8.5 Hz, J_2 = 2.1 Hz, ArH), 7.59 (d, 1H, J = 2.1 Hz, ArH), 9.63 (brs, 1H, NH), 10.43 (brs, 1H, NH); ^{13}C NMR (100 MHz, DMSO-d_6) δ = 17.1, 51.0, 51.2, 99.2, 128.1, 128.9, 130.7, 132.8, 133.1, 139.7, 145.9, 165.2, 173.9.

1-(4-(2,4-Dichlorophenyl)-6-methyl-2-thioxo-1,2,3,4-tetrahydropyrimidin-5-yl)ethanone (compound **16**). Orange solid (132 mg, 42%): IR (υ_{max}/cm^{-1}): 3393, 3225, 3086, 2977, 1631, 1556, 1465 1179, 1097; ^{1}H NMR (400 MHz, DMSO-d_6) δ = 2.13 (s, 3H, CH_3), 2.36 (s, 3H, CH_3CO), 5.64 (d, 1H, J = 3.5 Hz, CH), 7.24 (d, 1H, J = 8.4 Hz, ArH), 7.42 (dd, 1H, J_1 = 8.4 Hz, J_2 = 2.1 Hz, ArH), 7.61 (d, 1H, J = 2.1 Hz, ArH), 9.66 (brs, 1H, NH), 10.37 (brs, 1H, NH); ^{13}C NMR (100 MHz, DMSO-d_6) δ = 18.3, 30.4, 51.2, 109.6, 128.1, 129.1, 130.5, 132.9, 133.2, 139.0, 145.1, 174.0, 194.4; HRMS (ESI-TOF): m/z [M + H]$^+$ calcd for $C_{13}H_{13}N_2OSCl_2$: 315.0127, found 315.0119.

Ethyl 4-(2,3-difluorophenyl)-6-methyl-2-thioxo-1,2,3,4-tetrahydropyrimidine-5-carboxylate (compound **17**). Yellow solid (50 mg, 16%): IR (υ_{max}/cm^{-1}): 3318, 3177, 3104, 2984, 1713, 1654, 1570, 1485, 1194, 1100; ^{1}H NMR (400 MHz, DMSO-d_6) δ = 1.04 (t, 3H, J = 7.1 Hz, OCH_2CH_3), 2.30 (s, 3H, CH_3), 3.95 (dq, 2H, J_1 = 7.1 Hz, J_2 = 3.1 Hz, OCH_2), 5.48 (d, 1H, J = 3.0 Hz, CH), 7.04–7.10 (m, 1H, ArH), 7.17–7.24 (m, 1H, ArH), 7.32–7.40 (m, 1H, ArH), 9.62 (brs, 1H, NH), 10.43 (brs, 1H, NH); ^{13}C NMR (100 MHz, DMSO-d_6) δ = 13.8, 17.1, 48.3, 59.6, 98.9, 116.6, 116.8, 124.2, 125.1, 133.2, 133.4, 145.7, 164.6, 174.1; HRMS (ESI-TOF): m/z [M + H]$^+$ calcd for $C_{14}H_{15}N_2O_2SF_2$: 313.0824, found 313.0817.

Methyl 4-(2,3-difluorophenyl)-6-methyl-2-thioxo-1,2,3,4-tetrahydropyrimidine-5-carboxylate (compound **18**). Yellow solid (57 mg, 19%): IR (υ_{max}/cm^{-1}): 3305, 3192, 3102, 1665, 1564, 1484, 1186, 1126; ^{1}H NMR (400 MHz, DMSO-d_6) δ = 2.30 (s, 3H, CH_3), 3.51 (s, 3H, CH_3OCO), 5.48 (d, 1H, J = 3.1 Hz, CH), 7.03–7.09 (m, 1H, ArH), 7.17–7.24 (m, 1H, ArH), 7.31–7.40 (m, 1H, ArH), 9.64 (brs, 1H, NH), 10.46 (brs, 1H, NH); ^{13}C NMR (100 MHz, DMSO-d_6) δ = 17.2, 48.4, 51.1, 98.8, 116.7, 116.9, 124.1, 125.1, 133.0, 133.1, 145.8, 165.2, 174.2; HRMS (ESI-TOF): m/z [M + H]$^+$ calcd for $C_{13}H_{13}N_2O_2SF_2$: 299.0668, found 299.0660.

1-(4-(2,3-Difluorophenyl)-6-methyl-2-thioxo-1,2,3,4-tetrahydropyrimidin-5-yl)ethanone (compound **19**). White solid (172 mg, 61%): IR (υ_{max}/cm^{-1}): 3289, 3177, 3117, 3000, 1607, 1578, 1190; ^{1}H NMR (400 MHz, DMSO-d_6) δ = 2.18 (s, 3H, CH_3), 2.35 (s, 3H, CH_3CO), 5.56 (d, 1H, J = 3.5 Hz, CH), 7.00–7.06 (m, 1H, ArH), 7.15–7.22 (m, 1H, ArH), 7.31–740 (m, 1H, ArH), 9.71 (brs, 1H, NH), 10.38 (brs, 1H, NH); ^{13}C NMR (100 MHz, DMSO-d_6) δ = 18.3, 30.5, 48.3, 109.3, 116.7, 116.9, 124.0, 125.1, 132.6, 132.7, 145.0, 174.1, 194.3; HRMS (ESI-TOF): m/z [M + H]$^+$ calcd for $C_{13}H_{13}N_2OSF_2$: 283.0718, found 283.0711.

Ethyl 4-(furan-2-yl)-6-methyl-2-thioxo-1,2,3,4-tetrahydropyrimidine-5-carboxylate (compound **20**) (Hingane and Shumaila, 2013). Brown solid (61 mg, 23%): IR (υ_{max}/cm^{-1}): 3308, 3173,

3126, 2983, 1659, 1573, 1182; ^{1}H NMR (400 MHz, DMSO-d_6) δ = 1.13 (t, 3H, J = 6.8 Hz, OCH_2CH_3), 2.28 (s, 3H, CH_3), 4.04 (q, 2H J_1 = 7.2 Hz, J_2 = 3.1 Hz, OCH_2), 5.23 (d, 1H, J = 2.9 Hz, CH), 6.14 (d, 1H, J = 3.2 Hz, ArH), 6.36–6.41 (m, 1H, ArH), 7.58 (brs, 1H, ArH), 9.64 (brs, 1H, NH), 10.398 (brs, 1H, NH); ^{13}C NMR (100 MHz, DMSO-d_6) δ = 14.1, 17.1, 47.7, 59.6, 98.2, 106.3, 110.5, 142.7, 146.0, 154.6, 174.9.

Methyl 4-(furan-2-yl)-6-methyl-2-thioxo-1,2,3,4-tetrahydro pyrimidine-5-carboxylate (compound **21**) (Liu et al., 2013). Brown solid (70 mg, 28%): IR (υ_{max}/cm^{-1}): 3309, 3183, 1655, 1569, 1183, 1114; ^{1}H NMR (400 MHz, DMSO-d_6) δ = 2.28 (s, 3H, CH_3), 3.59 (s, 3H, CH_3OCO), 5.23 (d, 1H, J = 3.7 Hz, CH), 6.14 (d, 1H, J = 3.1 Hz, ArH), 6.36–6.39 (m, 1H, ArH), 7.58 (brs, 1H, ArH), 9.66 (brs, 1H, NH), 10.42 (brs, 1H, NH); ^{13}C NMR (100 MHz, DMSO-d_6) δ = 17.2, 47.6, 51.2, 98.0, 106.3, 110.5, 142.7, 146.3, 154.5, 165.3, 174.9.

1-(4-(Furan-2-yl)-6-methyl-2-thioxo-1,2,3,4-tetrahydropyri midin-5-yl)ethanone (compound **22**) (Ramos et al., 2013). Brown solid (135 mg, 57%): IR (υ_{max}/cm^{-1}): 3285, 3192, 1606, 1573, 1179, 1012; ^{1}H NMR (400 MHz, DMSO-d_6) δ = 2.20 (s, 3H, CH_3), 2.30 (s, 3H, CH_3CO), 5.34 (d, 1H, J = 3.4 Hz, CH), 6.17 (d, 1H, J = 2.5 Hz, ArH), 6.35–6.40 (m, 1H, ArH), 7.59 (brs, 1H, ArH), 9.74 (brs, 1H, NH), 10.35 (brs, 1H, NH); ^{13}C NMR (100 MHz, DMSO-d_6) δ = 18.2, 30.1, 47.7, 106.5, 108.1, 110.5, 142.8, 145.2, 154.5, 174.8, 194.3.

2.3. In vitro studies

2.3.1. Cell culture

The MCF-7, LNCaP, NHDF, T47D and Caco-2 cell lines were obtained from American Type Culture Collection (ATCC; Manassas, VA, USA). The HepaRG cell line was obtained from Life Technologies – Invitrogen™ (through Alfagene, Portugal). They were cultured in 75 cm^2 culture flasks at 37 °C in a humidified air incubator with 5% CO_2. The high-glucose Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS; Sigma–Aldrich, St Louis, MO, USA), and 1% antibiotic/antimycotic (10,000 units/mL penicillin G, 100 mg/mL streptomycin and 25 µg/mL amphotericin B) (Ab; Sigma–Aldrich, St Louis, MO, USA) was used to culture MCF-7 cells. Caco-2 cells were cultured in DMEM supplemented with 10% FBS and 1% of the antibiotic mixture of 10,000 units/mL penicillin G and 100 mg/mL of streptomycin (sp). LNCaP and T47D cells were cultured in RPMI 1640 medium with 10% FBS and 1% sp. NHDF cells have grown in RPMI 1640 medium supplemented with 10% FBS, 2 mM L-glutamine, 10 mM HEPES, 1 mM sodium pyruvate and 1% Ab. Finally, HepaRG cells were seeded in Williams' E medium supplemented with 10% FBS, 1% sp, 5 µg/mL insulin, and 5×10^{-5} M hydrocortisone hemisuccinate (Sigma–Aldrich, St Louis, MO, USA). For all cell lines, the medium was renewed every 2–3 days until cells reach approximately 90–95% confluence. Then, they were detached by gentle trypsinization (trypsin–EDTA; Sigma–Aldrich, St Louis, MO, USA) and, before the experiments, viable cells were counted by the trypan-blue exclusion assay and suitably diluted in the adequate complete culture medium.

2.3.2. Preparation of test compound solutions

All compounds were dissolved in DMSO at the concentration of 10 mM and stored at 4–8 °C. From this stock solution, various working solutions of the compounds in study in different concentrations were prepared by adequate dilutions in complete culture medium before each experiment. The maximum DMSO concentration in the studies was 1% and previous experiments revealed that this solvent level has no significant effects in cell proliferation (data not shown).

2.3.3. MTT cell proliferation assay

The *in vitro* antiproliferative effects were evaluated by the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay (Sigma–Aldrich, St Louis, MO, USA). After reaching confluence, cells were trypsinized, counted using the trypan-blue exclusion assay, and diluted to 2×10^4 cells/mL, and then 100 µL of cell suspension/well was seeded in 96-well culture plates and left to adhere for 48 h. After adherence, the medium was replaced by solutions of the compounds in study (30 µM for preliminary studies and 0.01, 0.1, 1, 10, 50 and 100 µM for concentration-response studies) in the appropriate medium. Untreated cells were used as control. Each experiment was performed in quadruplicate and independently repeated. After the incubation period of 72 h, the medium was removed, 100 µL/well of phosphate buffer saline (NaCl 137 mM, KCl 2.7 mM, Na_2HPO_4 10 mM and KH_2PO_4 1.8 mM and pH adjusted to 7.4) was used to wash the cells and then 100 µL of the MTT solution (5 mg/mL), prepared in the appropriate serum-free medium, was added to each well, followed by incubation for 4–8 h at 37 °C. Then, the MTT containing medium was removed and the formazan crystals were dissolved in DMSO. The absorbance was measured at 570 nm using a microplate reader Bio-Rad Xmark spectrophotometer. Cell proliferation values were expressed as percentages from the relative absorbance measured in the treated wells *versus* control wells.

2.3.4. Cell death

The analysis of cell death was performed by flow cytometry after staining dead cells with propidium iodide (PI, solution of 1 mg/ml in 0.1% of azide and water, Sigma Aldrich, St Louis, MO, USA). Briefly, 3 mL of cells were seeded in 6-well plates (cell density of 3×10^4 cells/mL for HepaRG and MCF-7 cell lines) in complete culture medium. After 48 h they were treated with 50 µM of the compounds **2** and **15**. Untreated cells were used as negative control. At the end of 24 h of incubation, the supernatant of each well was collected, washed with PBS, and pooled with the cells harvested by trypsin treatment. The resulting cell suspension was kept on ice, pelleted by centrifugation, and resuspended in 400 µL of complete medium. Afterwards, 395 µL of the cell suspension was transferred to a FACS tube and 5 µL of PI was added. A minimum of 10,000 events was acquired using a FACSCalibur flow cytometer using the FSC, SSC and FL3 (PI) channels. Acquisition and analysis were performed with CellQuest™ Pro Software. In the FSC/FL3 contour plot, two regions were created, one corresponding to viable cells (R1) and another to dead cells (R2) in order to exclude debris, which were not considered in the analysis (data not shown). The percentage of survival is the percentage of cells in R1 as compared to the total number of events in R1 and R2.

2.3.5. Cell cycle distribution analysis

Cell cycle distribution of cells was determined through PI staining of fixed and permeabilized cells. In brief, 3 mL of cells

were seeded in 6-well plates (cell density of 2×10^4 cells/mL for HepaRG and MCF-7 cell lines) in complete culture medium. After 48 h they were treated with 50 μM of the compounds **2** and **15**. For comparison, untreated cells were used as negative control and other cells were treated with 5-FU at a concentration of 50 μM as positive control. After 48 h of incubation, the cells were trypsinized, centrifuged and resuspended in 450 μL of a cold solution of 0.5% bovine serum albumin (BSA; Amresco, USA) in PBS. The resulting cell suspension was kept on ice and then fixed by gently adding ice-cold 70% ethanol (−20 °C) with simultaneous vortexing. After at least 2 days at −20 °C, fixed cells were washed twice with PBS and resuspended in a solution of PI prepared in PBS/BSA 0.5% (50 μg/mL) and sequentially incubated with Ribonuclease A from bovine pancreas at a final concentration of 7.1 μg/mL (stock solution in 50% glycerol, 10 mM Tris–HCl, pH8, Sigma Aldrich, St Louis, MO, USA) for 15 min in the dark. The data were analysed using ModFit software (Becton Dickinson, San Jose, CA, USA). Before fitting the data, a region (R1) was created on the FL3-Width/FL3-Area contour plot to exclude cell aggregates and other region (R2) was created on the FL1-Height/FL3-Area contour plot to exclude part of the debris.

2.4. Statistics

The data are expressed as a mean ± standard deviation (SD). Comparison among multiple groups of one factor was analysed by using one-way ANOVA followed by Dunnett's post hoc tests to determine significant differences among the means. Difference between groups was considered statistically significant for a p-value lower than 0.05 ($p < 0.05$). The determination of the IC_{50} was done by sigmoidal fitting analysis considering a 95% confidence interval.

2.5. QSAR studies

2.5.1. Data handling and in silico calculation of the molecular descriptors

The antiproliferative activity (expressed as the relative cell proliferation in percentage) of the target compounds (**1**–**22**) at a concentration of 30 μM against NHDF, HepaRG, Caco-2, MCF-7, T47D and LNCaP cell lines, presented in Table 2, was initially log transformed (base 10). In order to increase the available data and the usefulness of the developed QSAR model, all activity measurements were combined using a three bit codification system to distinguish between the six tested cell lines (Table 3). Using these three binary inputs, a total of six different combinations (out of 8 possible) were selected, with no particular order, which represent each of the cell lines employed (e.g. 0, 0, 1 represents the Caco-2 cell line, while 1, 1, 0 represents MCF-7). Comparatively with the use of a dummy inputs system, an important reduction in required inputs for cell line distinction was obtained (Paixão et al., 2014). Thus, 132 cases were available for QSAR modelling, in which 112 cases (85%) were randomly selected as the training set for model development and the remaining 20 cases (15%) as the test set for external validation purposes. The random selection was performed in a way that the minimum and maximum relative cell proliferation values for each cell line were not selected for the test set, and the test set selected cases were well distributed between both cell lines and range of values in each cell line.

For the *in silico* calculation of the molecular descriptors, the titled compounds were first manually drawn in ACD/ChemSketch 2015 Pack 2 (ACD/ChemSketch, 2015), and the SMILES notation was obtained and used for tautomer and ionization state check and the calculation of the GALAS logP in ACD/Percepta 2015 (ACD/Percepta, 2015). Using E-Dragon online, 631 additional molecular descriptors were calculated (48 constitutional descriptors, 119 topological descriptors, 47 walk and path counts, 33 connectivity indices, 47 information indices, 96 2D autocorrelations, 107 edge adjacency indices, 64 BCUT descriptors, 21 topological charge indices, 44 eigenvalue-based indices and 5 molecular properties) (Tetko et al., 2005). The 3D conformation of the compounds was not taken into account.

2.5.2. BRANN modelling for QSAR development

QSAR development was performed using an in-house developed tool based on BRANNs in MATLAB R2014a (MATLAB, 2014), which allows process automation, data analysis and the use of cross-validation procedures. All calculations and modelling were performed on a 3.5 Ghz Intel i7 CPU running Windows 7 operating system. Commonly selected parameters were "trainbr" as the training function, preprocessing of input and output variables to [−1, 1] range, "tansig" transfer function in the hidden layer and "purelin" in the output layer. The remaining parameters were kept at their default value. The molecular descriptors and the three bits used for cell line identification were used as independent variables (inputs), and the log(relative cell proliferation) was used as the dependent variable (output).

Prior to BRANN modelling, the UFS algorithm (Whitley et al., 2000) was employed in the training set, in order to remove multicollinear and insufficiently discriminative molecular descriptors (selected parameters were maximum correlation value of 0.9 and minimum standard deviation value of 0.01). Following this initial input reduction, a forward selection method was performed using a 10-fold cross-validation procedure with the BRANN models, starting with the three bit inputs. Each iteration was repeated 10 times with random splits of the available data, and the average statistical evaluation was taken. For each iteration, the selected molecular descriptor was the one that returned the best average Q^2 and $RMSE_{CV}$ values. After selection of the most relevant molecular descriptors, and for simplicity sake, the same 10-fold cross-validation procedure with 10 duplicates was used to determine the optimal number of neurons in a single hidden layer between 0 (linear model) and 10. After determination of the best parameters regarding the molecular descriptors and model complexity (internal validation), a y-scrambling procedure was performed to verify the absence of chance correlations between the input and output variables (Rücker et al., 2007). The final QSAR model was then trained on all available training data. The final QSAR model was further validated using the hold-out test set (external validation), by comparison of the QSAR predicted values with those observed experimentally for the cases not used to train the model.

Finally, to elucidate the relationships between the selected molecular descriptors and relative cell proliferation, the Lek Profile method (Lek et al., 1996a, 1996b) was employed indi-

Table 2 Experimental and predicted antiproliferative activities of the target compounds (**1**–**22**) at concentration of 30 μM (as relative cell proliferation in percentage) against NHDF, HepaRG, Caco-2, MCF-7, T47D and LNCaP cell lines.

Compound	Dermal NHDF	Hepatic HepaRG	Colon Caco-2	Breast MCF-7	Breast T47D	Prostate LNCaP
1	70.67 ± 7.29*(68.43)	72.86 ± 6.03**(60.18)	105.20 ± 8.56(93.91)	75.42 ± 15.26**(71.51)	87.14 ± 3.28(78.66)	62.67 ± 12.79***(62.89)
2	83.03 ± 2.16**(69.96)	74.01 ± 9.30***(64.84)	87.28 ± 16.49(97.17)	76.87 ± 22.85(75.27)	78.97 ± 22.71(81.11)	68.23 ± 11.73***(65.13)
3	83.89 ± 1.94(83.89)	58.80 ± 9.39***(73.14)	94.14 ± 12.75(84.29)	85.05 ± 10.02*(87.52)	80.60 ± 4.26(89.02)	72.11 ± 11.27***(82.20)
4	70.12 ± 8.07(67.17)	54.84 ± 6.50***(56.01)	81.87 ± 4.89*(88.87)	78.74 ± 4.70**(69.39)	80.51 ± 8.48(77.23)	54.64 ± 5.64***(61.28)
5	76.31 ± 2.46*(68.24)	65.98 ± 7.87***(61.19)	79.80 ± 8.45(94.08)	69.63 ± 4.58(73.01)	70.44 ± 4.54**(79.61)	69.50 ± 5.19***(63.01)
6	76.31 ± 6.18(82.71)	50.84 ± 4.61***(71.88)	97.42 ± 6.86(84.71)	94.86 ± 7.74(86.80)	94.70 ± 5.62(88.56)	73.40 ± 17.33***(80.76)
7	67.37 ± 10.41(89.02)	74.04 ± 10.03*(83.55)	104.19 ± 5.56(87.63)	81.01 ± 6.69**(90.32)	96.29 ± 2.83(90.80)	62.54 ± 7.23***(88.43)
8	69.50 ± 1.44***(79.18)	45.78 ± 5.31*(48.98)	72.38 ± 3.67**(79.76)	93.39 ± 6.40(65.12)	69.02 ± 7.81***(75.31)	64.77 ± 8.81***(71.11)
9	87.68 ± 1.79(77.27)	63.69 ± 6.00***(57.16)	91.35 ± 17.36(91.73)	66.83 ± 15.26(68.22)	68.63 ± 9.60***(77.07)	73.31 ± 3.95$_{**}$(70.32)
10	96.63 ± 0.84**(82.47)	75.38 ± 10.18*(83.14)	94.08 ± 5.53(101.51)	98.07 ± 11.02(85.84)	77.24 ± 20.30**(88.00)	82.06 ± 14.73*(80.24)
11	67.26 ± 15.69***(59.45)	20.06 ± 6.28***(19.24)	51.97 ± 2.26***(29.30)	62.42 ± 15.87*(46.35)	77.30 ± 8.10**(64.01)	61.40 ± 20.47(49.28)
12	88.81 ± 8.20(63.28)	25.83 ± 3.39***(21.32)	25.36 ± 3.65***(33.12)	36.96 ± 5.48**(48.18)	59.26 ± 4.88***(64.85)	61.88 ± 5.93***(52.86)
13	69.59 ± 1.76*(77.33)	37.40 ± 2.31***(40.96)	73.83 ± 3.38*(67.12)	60.24 ± 6.64*(61.46)	56.41 ± 5.15***(72.96)	72.43 ± 4.32***(68.40)
14	65.06 ± 8.41***(62.68)	21.35 ± 1.23***(20.57)	30.21 ± 3.14***(31.73)	19.86 ± 3.55***(47.95)	59.05 ± 7.46***(65.14)	60.12 ± 6.93***(52.18)
15	56.87 ± 14.02**(65.96)	14.62 ± 2.34***(22.71)	39.27 ± 0.72***(35.63)	31.11 ± 2.78***(49.65)	52.02 ± 1.61***(65.88)	52.49 ± 5.01***(55.36)
16	85.19 ± 2.93***(79.98)	46.58 ± 4.43*(49.82)	88.11 ± 5.52(81.04)	83.05 ± 8.83*(65.50)	91.46 ± 3.28(75.60)	76.52 ± 4.97**(71.93)
17	69.15 ± 7.06**(82.87)	55.54 ± 16.43***(48.53)	88.56 ± 8.55(79.08)	56.60 ± 9.43*(65.22)	80.99 ± 4.88*(75.71)	73.61 ± 6.52**(74.31)
18	69.58 ± 5.37**(81.49)	51.66 ± 3.24***(55.37)	92.53 ± 11.28(89.32)	55.61 ± 14.22(67.63)	70.50 ± 4.79***(77.01)	78.30 ± 19.59*(73.92)
19	96.20 ± 1.55***(79.55)	76.14 ± 5.10*(85.10)	91.55 ± 6.57(112.07)	103.39 ± 4.99(82.95)	98.11 ± 11.10(86.19)	81.43 ± 18.86(76.46)
20	97.54 ± 6.80(85.00)	83.98 ± 4.00*(70.20)	90.11 ± 11.26(109.91)	69.91 ± 9.65(72.25)	78.45 ± 8.11(80.02)	83.28 ± 3.65*(78.56)
21	104.10 ± 2.43(82.15)	77.55 ± 4.50**(79.99)	103.02 ± 25.27(121.33)	84.68 ± 9.88(75.04)	95.76 ± 7.35(81.54)	82.40 ± 5.70*(76.78)
22	93.85 ± 3.81(76.08)	83.36 ± 12.76(77.17)	91.38 ± 22.26(106.74)	74.68 ± 20.52(80.51)	80.84 ± 8.98*(84.59)	82.70 ± 6.84*(72.35)

Results are expressed as means ± SD (standard deviation) after 72 h of treatment.

* $p < 0.05$ *versus* control;

** $p < 0.01$ *versus* control;

*** $p < 0.001$ *versus* control.

QSAR predicted values are inside parenthesis, and cases used in the external validation of the QSAR model are underlined.

Table 3 Three bit representation for the NHDF, HepaRG, Caco-2, MCF-7, T47D and LNCaP cell lines used in for QSAR modelling.

Cell line	Binary variables		
	Bit 1	Bit 2	Bit 3
NHDF	1	0	0
HepaRG	0	1	0
Caco-2	0	0	1
MCF-7	1	1	0
T47D	1	0	1
LNCaP	0	1	1

vidually for each cell line, by varying each molecular descriptor across 11 data points (10 equal intervals over the entire range of the molecular descriptor) and holding the remaining molecular descriptors at 5 different data range splits (minimum, first quartile, median, third quartile and maximum value). The average predicted responses across the five split predictions were taken as the relationship between the molecular descriptor and the response variable, and the relative importance of each molecular descriptor is taken as the maximum range between the calculated predictions (maximum predicted value – minimum predicted value).

2.5.3. Internal and external statistical evaluation

For both internal and external validation purposes, the coefficient of determination (R^2 or Q^2 in cross-validation) was used as a measurement of the goodness of fit of the model. Also, the root mean squared error (RMSE) between predicted and observed values was used as a measurement of accuracy.

3. Results and discussion

3.1. Chemical synthesis

In the present work, a Biginelli-type cyclocondensation reaction was used for the preparation of DHPMts. This multicomponent reaction allows the synthesis of several molecules with different degrees of structural diversity in a single step and therefore it is of high interest in the development and discovery of new drug candidates (Ganem, 2009; Hulme and Gore, 2003). Moreover, this method of synthesis is very simple and economical and respects environmental concerns. Thus, twenty-two DHPMts were successfully synthesized by means of the one-pot reaction among an aldehyde (benzaldehyde, *p*-tolualdehyde, 4-nitrobenzaldehyde, anisaldehyde, 2,3-dichlorobenzaldehyde, 2,4-dichlorobenzaldehyde, 2,3-difluorobenzaldehyde, or furaldehyde), a β-ketoester (ethyl acetoacetate or methyl acetoacetate)/acetylacetone and thiourea under solvent-free conditions (Table 1). Following the procedure proposed by Rodríguez-Domínguez and collaborators (Rodríguez-Domínguez et al., 2007), the catalyst used was $ZrCl_4$ (Fig. 3) and the reactions were considered concluded when the mixture solidified. The reaction time varied accordingly with the reagents from 7 min to 17 h. Although this reaction is known to be very fast (Kalita and Phukan, 2007; Nandurkar et al., 2007; Rodríguez-Domínguez et al., 2007), for the products containing halogens in their structure, with exception of compound **19**, more than 1 h was necessary to complete the reaction and lower yields were obtained. The workup was a simple wash of the product with water to remove the catalyst and then the products were recrystallized with ethanol to afford the pure products, which were characterized by IR, ^1H and ^{13}C NMR. To the best of our knowledge, several halogenated compounds are novel and, in these cases, a HRMS was also acquired to help in their characterization. Interestingly, in general, the compounds synthesized using acetylacetone as reagent were obtained with higher yields than their analogues with longer lateral chains.

3.2. Antiproliferative activity

The antiproliferative activity of the synthesized DHPMts (**1**–**22**) was evaluated in five cancer cell lines: human hepatocellular carcinoma (HepaRG), human colorectal adenocarcinoma (Caco-2), human breast adenocarcinoma (MCF-7), human breast ductal carcinoma (T47D) and human prostatic carcinoma (LNCaP), as well as in a primary non-cancerous cell line, normal human dermal fibroblasts (NHDF). The commercial anticancer drug 5-FU was also tested for comparison. Specifically, using the MTT assay, it was performed an initial screening with a single concentration of 30 μM and 72 h of exposition (Table 2). In general, the compounds did not show marked cytotoxicity in the dermal cells, in the prostatic cancer cell line and in the T47D breast cancer cell line (relative cell proliferation higher than 50% at 30 μM). Compounds **1**, **2** and **3**, which represent the most basic skeleton of this series

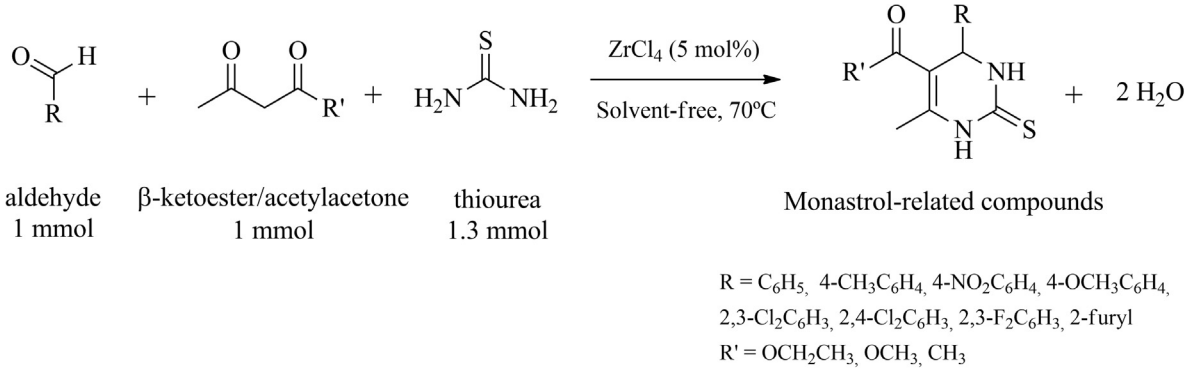


Figure 3 General scheme of one-pot synthesis of 3,4-dihydropyrimidin-2-(1*H*)-thiones catalysed by $ZrCl_4$ under solvent-free conditions.

of derivatives, without replacements in the phenyl group, did not exhibit relevant antiproliferative activity, which is consistent with the study performed by Russowky and collaborators for compound **1** against the MCF-7 cell line (Russowsky et al., 2006). Moreover, our results also suggest that methyl (**4**, **5** and **6**), nitro (**7**) and methoxy (**8**, **9** and **10**) groups introduced in the aromatic moiety at *para*-position did not strongly affect the cytotoxicity of these DHPMt derivatives. On the other hand, the molecules containing chlorine atoms attached to the aromatic ring (**11**–**16**) exhibited stronger inhibition of cell proliferation in HepaRG and/or Caco-2 and MCF-7 cancer cell lines than the molecules previously referred. Interestingly, the introduction of other halogens (fluorine atoms) into the phenyl ring (**17**, **18** and **19**) resulted in a decrease in the antiproliferative activity, when compared with their chlorine analogues. In addition, compounds **20**, **21** and **22**, which have a five-member heteroring (furan) instead of the six-member aromatic ring, also did not reveal marked antiproliferative activity.

The *in vitro* antiproliferative activity of the most cytotoxic compounds was further investigated by determining the corresponding IC_{50} against HepaRG (compounds **8**, **11**, **12**, **13**, **14**, **15** and **16**), Caco-2 (compounds **12**, **14** and **15**) and MCF-7 (compounds **12**, **14** and **15**) cells as shown in Table 4. Within these, compound **8**, containing a methoxy group attached to the aromatic ring at *para*-position, was the unique molecule evaluated that did not include a chlorine atom into its structure and was the less cytotoxic (IC_{50} = 41.48 μM) towards the HepaRG carcinoma cells. Additionally, compounds **13** and **16**, which did not express significant cytotoxicity against Caco-2 and MCF-7 cell lines, showed relatively weak cytotoxicity in the HepaRG cell line (IC_{50} = 31.86 μM and IC_{50} = 25.49 μM, respectively). These results can suggest that small lateral chains afford less cytotoxicity. On the other hand, when comparing compounds **11** and **12** (2,3-dichloro derivatives) with compounds **14** and **15** (2,4-dichloro derivatives), a higher cytotoxicity against HepaRG cells was observed with compounds **11** and **12**. In this group, compound **11** was the most potent compound, displaying stronger antiproliferative activity than 5-FU (IC_{50} = 0.75 μM *versus* IC_{50} = 2.02 μM, respectively) towards the HepaRG cell line. The structure-antiproliferative activity relationship is illustrated in Fig. 4. In fact, the reason that motivated us to evaluate the antiproliferative effect of these compounds was the existence of some reports claiming significant anticancer activity for this kind of Biginelli products. Prashantha Kumar and collaborators have assessed the antiproliferative effect of several 1,4-dihydropyrimidine derivatives in MCF-7 cells. In that work, the molecules incorporating a cinnamoyl moiety at C4 of the phenyl ring and the compounds having a furan and pyridine rings in their structure displayed promising anticancer activity (Prashantha Kumar et al., 2009). Moreover, several other DHPMts have exhibited relevant cytotoxic activity against melanoma, prostate, colon, renal, breast and ovarian cancer cell lines (Russowsky et al., 2006). In our study, the synthesized compounds evaluated in MCF-7 cells (**12**, **14** and **15**) showed stronger antiproliferative activity than described in the literature for monastrol (IC_{50} = 110.40 μM) (Guido et al., 2015). Thus, our work expanded the data on the potential antitumoral interest of this family of compounds and also allowed the identification of potent and novel antiproliferative compounds.

Table 4 Cytotoxicity (IC_{50} μM) of the most potent compounds against HepaRG, Caco-2 and MCF-7 cell lines.[a]

Compound	HepaRG	R^2	Caco-2	R^2	MCF-7	R^2
8	41.48	0.82	–	–	–	–
11	0.75	0.97	–	–	–	–
12	6.53	0.96	5.51	0.95	4.30	0.96
13	31.86	0.99	–	–	–	–
14	14.31	0.99	13.65	0.91	2.95	0.95
15	25.07	0.98	9.43	0.96	10.89	0.99
16	25.49	0.99	–	–	–	–
5-FU	2.02	0.93	1.15	0.90	1.71	0.91

[a] The cells were treated with a variety of concentrations (0.01, 0.1, 1, 10, 50 and 100 μM) during 72 h. The antiproliferative effects were determined by the MTT assay and the IC_{50} values were calculated by sigmoidal fitting. The data shown are representative of at least two independent experiments.

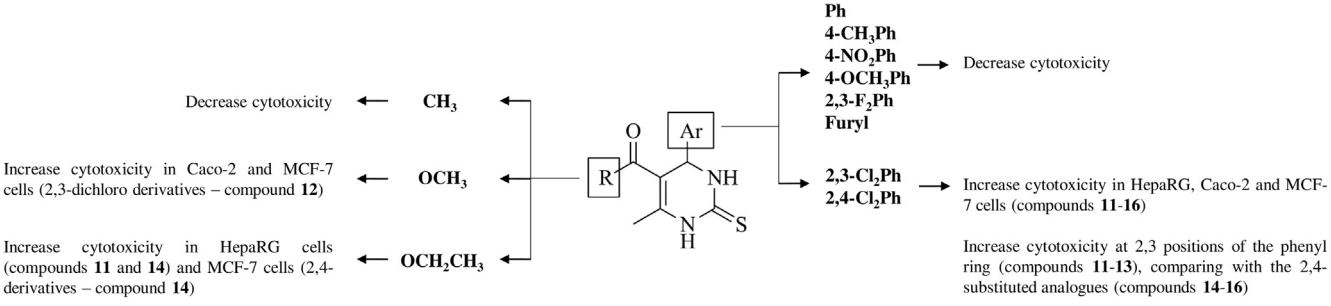


Figure 4 General structure-activity relationship of the synthesized compounds for their cytotoxicity in hepatic (HepaRG), intestinal (Caco-2) and breast (MCF-7) cancer cell lines, highlighting the most active compounds.

3.3. Cell death and cell cycle distribution

A PI flow cytometric assay was used to identify death cells with compromised cell membrane, and the results showed that compounds **2** and **15** at the concentration of 50 μM did not induce important cell death after 24 h of incubation in both MCF-7 and HepaRG cell lines (Fig. 5).

As monastrol affects the cell cycle, mainly as a specific inhibitor of the human motor protein Eg5 (Asraf et al., 2015), this encouraged us to evaluate the cell cycle distribution induced by the compound **15** in MCF-7 and HepaRG cell lines. This compound was chosen because it previously showed cytotoxicity in both cell lines and the compound **2**, without substituents in the aromatic ring, was also evaluated to understand whether chlorine atoms present on compound **15** have influence in the cell cycle distribution. Untreated cells and 5-FU-treated cells were used as negative and positive controls, respectively. Interestingly, compound **15** at 50 μM for 48 h arrested MCF-7 cells in G_0/G_1 phase (the phase before DNA replication), increasing the proportion of cells in this cell cycle phase from 65.96 ± 1.01% (control) to 89.63 ± 0.92% (Fig. 6, Panel B). This phenomenon was accompanied by a 4.4-fold decrease in the populations of cells in S phase (the DNA synthesis phase where DNA replication occurs) and 2.1-fold decrease of cells in G_2/M phase (phase where cell division occurs). Although less pronounced, compound **2** also had effect on the percentage of cells in G_0/G_1 phase (77.06 ± 2.61%), which suggests that the basic skeleton of DHPMts could itself have some effects in the distribution of the cell cycle. However, chlorine atoms are probably important to increase this effect. This is in contrast to monastrol effect, which is described to induce mitotic arrest in several cell lines (Asraf et al., 2015), and did not seem to significantly affect the distribution of the cell cycle of MCF-7 cells even at 1 mM (Guido et al., 2015). As expected, in HepaRG cells, compounds **2** and **15** did not share the effect of 5-FU either, which strongly arrested the cells in S phase (68.97 ± 2.06% *versus* the control 17.01 ± 2.508%) (Fig. 7, Panel B). Similarly, both compounds significantly increased G_0/G_1 stage (86.16 ± 1.66% for compound **2** and 84.99 ± 1.42% for compound **15** *versus* the control 72.31 ± 3.92%) after treatment with 50 μM for 48 h (Fig. 7, Panel B). Interestingly, the tested DHPMt derivatives did not lead to the typical cell cycle phenotype of monastrol, related to impaired mitosis due to Eg5 inhibition. For example, it is described that at 130 μM, monastrol increased the proportion of cells in G_2/M phase (around 62%) in retinal pigmented epithelial cells stably transfected with hTERT after 24 h of incubation (Bartoli et al., 2011). In contrast, our data suggest that cell cycle arrest in G_0/G_1 stage could contribute to the antiproliferative effects of compound **15** in MCF-7 and HepaRG cells, but this hypothesis does not exclude that other mechanisms may be involved in the cytotoxicity of this compound.

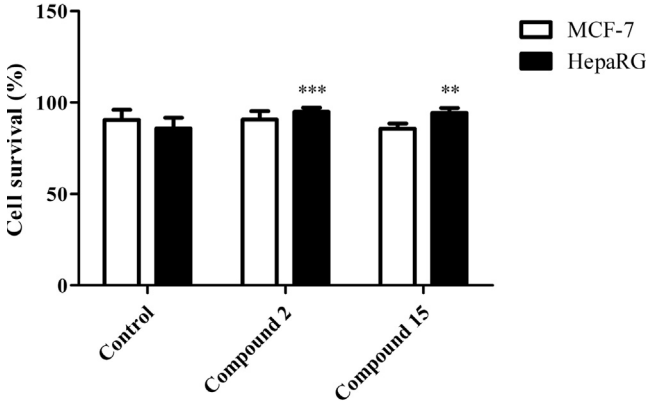


Figure 5 Percentage of cell survival after 24 h treatment with 50 μM of compounds 2 and 15 in MCF-7 and HepaRG cell lines through PI flow cytometric assay. The control corresponds to untreated cells. The percentage of survival is the percentage of live cells as compared to the total number of events of both live and dead cells. Each bar represents the mean ± SD. $^{**}p < 0.01$ *versus* control; $^{***}p < 0.001$ *versus* control.

3.4. QSAR studies

In order to develop a QSAR model to relate the *in silico* calculated molecular descriptors of the titled compounds to their experimentally obtained antiproliferative activities, and predict the given response of hypothetical related compounds, we used BRANN models coupled with an optimization process for the selection of the most relevant molecular descriptors and required model complexity.

Since a QSAR model that can reliably predict the bioactivity of compounds in multiple cell lines is of greater utility, and that the use of larger data sets for training can generate better predictive models, we choose to incorporate all available data in Table 2 in a single output QSAR model (Tropsha, 2010). Thus, a three bit codification was assigned to each cell line to distinguish between them (Table 3). Given the importance of an external test set to ultimately validate a QSAR model (Tropsha et al., 2003), the available data were split as described in the experimental section to ensure that the test set was representative across both cell lines and experimental value ranges.

Following the optimization procedure described in the experimental section, the initial 632 calculated descriptors were reduced to 14 by removal of multicollinear and insufficiently discriminative molecular descriptors using the Unsupervised Forward Selection (UFS) algorithm (Whitley et al., 2000). Further selection of the most relevant descriptors using a forward selection method, by maximizing the 10-fold cross-validated coefficient of determination (Q^2), returned three molecular descriptors as the most relevant: BLI (Kier benzene-likeliness index), GATS1m (Geary autocorrelation of lag 1 weighted by mass) and GATS5v (Geary autocorrelation of lag 5 weighted by van der Waals volume). The calculated descriptors values for the titled compounds are given in Table 5. Finally, in what concerns the tested model complexities (0 to 10 neurons in one hidden layer), two neurons returned the best average Q^2 and $RMSE_{CV}$. Thus, the final model 6-2-1, trained on all training cases, contains six inputs (three bits for cell line differentiation and the three selected molecular descriptors), two neurons in one hidden layer and one output neuron that returns the logarithm of the relative cell proliferation. The Pearson linear correlation matrix for the molecular descriptors and relative cell proliferation values for the tested cell lines is presented in Table 6. One requirement for a valid QSAR model is the use of non-collinear molecular

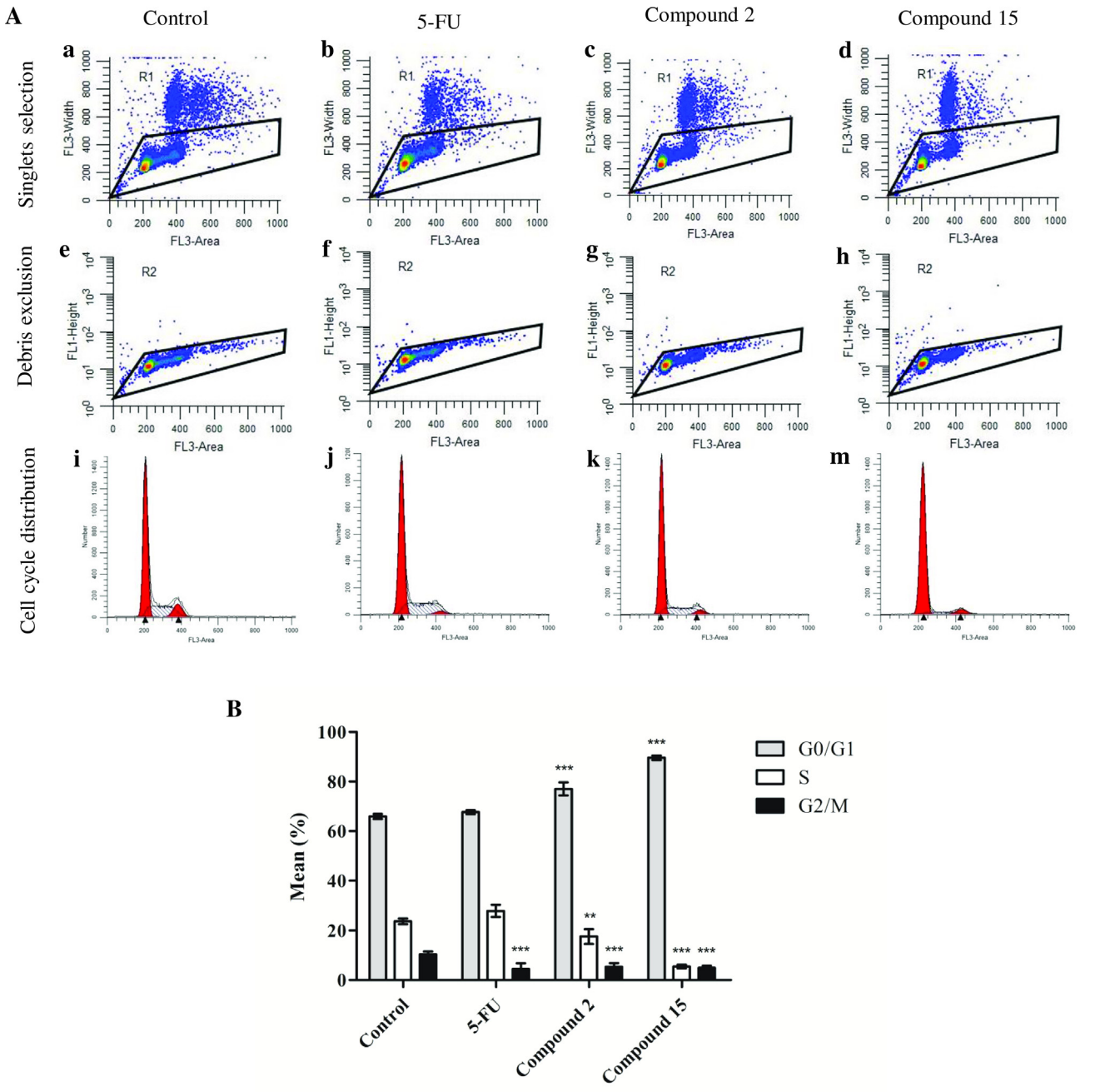


Figure 6 Cell cycle distribution analysis of MCF-7 breast cancer cells after treatment with compounds 2 and 15 (50 μM) for 48 h. A negative control (untreated cells) and a positive control (5-FU, 50 μM) were included. The analysis of the cell cycle distribution was performed using the PI staining and by flow cytometry. A – Representative cell cycle distribution analysis showing in a, b, c, and d, gating of singlets by region R1 created on the FL3-Width/FL3-Area contour plot; in e, f, g, and h, debris exclusion by region (R2) created on the FL1-Height/FL3-Area contour plot; and in i, j, k, and m, cell cycle distribution fit, respectively for negative control, 5-FU, compound 2, and compound 15. B – Quantification of the proportion of cells in G_0/G_1, S, and G_2/M phases of the cell cycle. Each bar represents the mean ± SD of four samples (originating from two independent experiments). $^{**}p < 0.01$ *versus* control; $^{***}p < 0.001$ *versus* control.

descriptors (Dearden et al., 2009). In Table 6, it can be observed that the chosen descriptors are poorly correlated among themselves and, in general, they are negatively correlated with the relative cell proliferation experimental values.

The internal validation statistics (coefficient of determination and RMSE) of the developed QSAR model, for both cross-validation and training, are presented in Table 7. In this table are also shown the external validation statistics obtained when comparing the predicted output of the cases left out of the training process, for which the experimental result is known. Analysing the obtained cross-validated statistics (Q^2 of 0.686 and $RMSE_{CV}$ of 0.086), and external test set statistics (R^2_{pred} of 0.699 and RMSE of 0.087), it can be concluded that the model presents great predictive ability, generating reliable

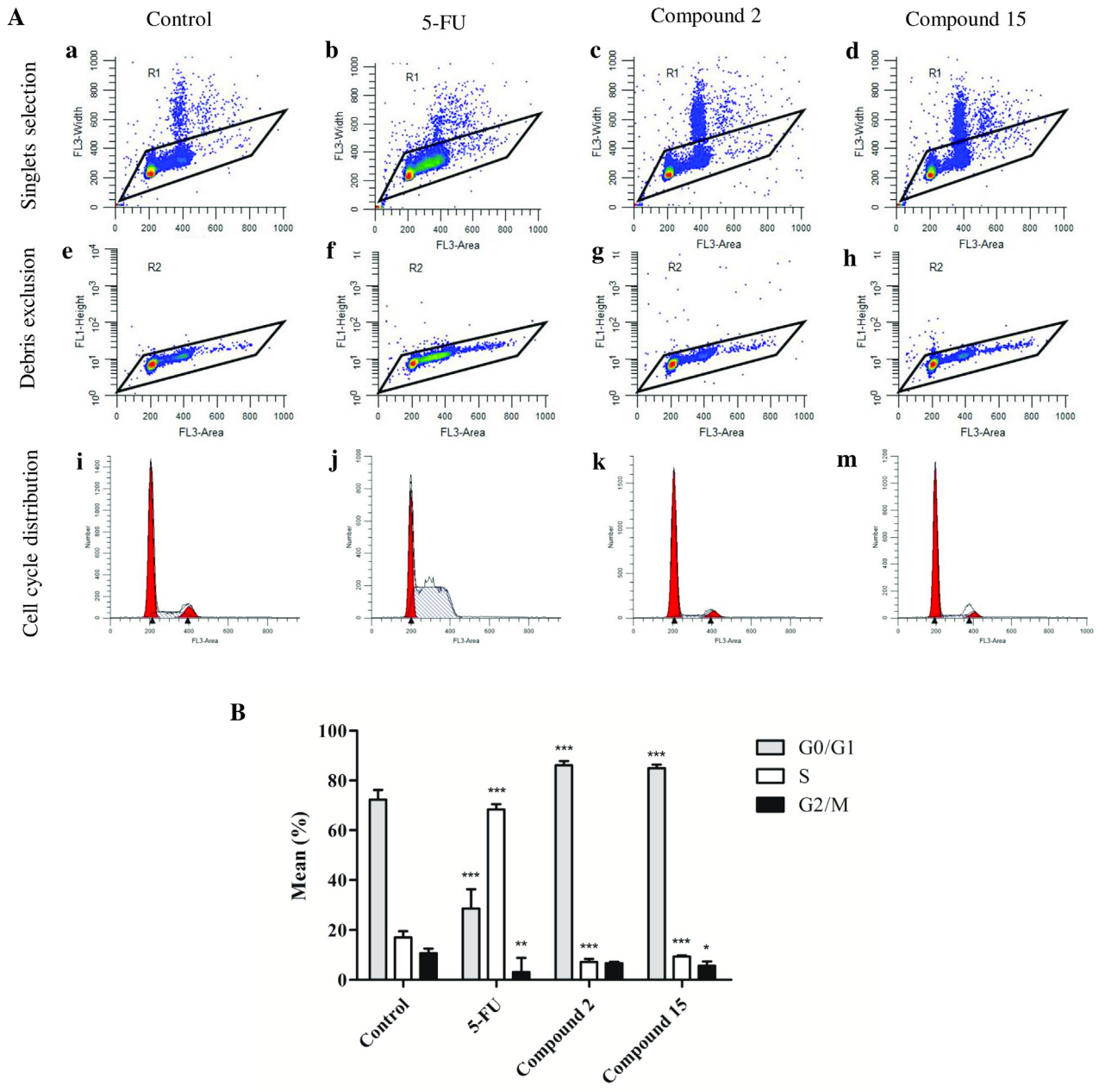


Figure 7 Cell cycle distribution analysis of HepaRG hepatic cancer cells after treatment with compounds 2 and 15 (50 μM) for 48 h. The negative control consists of untreated cells and 5-FU (50 μM) was used as positive control. The analysis of the cell cycle distribution was performed by using PI staining and by flow cytometry. A – Representative cell cycle distribution analysis showing in a, b, c, and d, gating of singlets by the region R1 created on the FL3-Width/FL3-Area contour plot; in e, f, g, and h, debris exclusion by gating in the region R2 created on the FL1-Height/FL3-Area contour plot, and in i, j, k, and m, cell cycle distribution fit, respectively for negative control, 5-FU, compound 2, and compound 15. B – Quantification of the accumulation of cells in G_0/G_1, S, and G^2/M phases of the cell cycle. Each bar represents the mean ± SD of four samples (originating from two independent experiments). *$p < 0.05$ *versus* control; **$p < 0.01$ *versus* control; ***$p < 0.001$ *versus* control.

predictions. This can be seen in Fig. 8, in which the predicted log(relative cell proliferation) of both training and test set cases is similar to the respective experimental results for the majority of the cases. Also, for the y-scrambling procedure, the highest value of Q^2 achieved in 10 trials was 0.006, which confirms the absence of chance correlations between the molecular descriptors and the response. The developed model can be effectively used to guide future efforts in improving activity of new hypothetical related compounds, by predicting the relative cell proliferation in any of the used cell lines, given that the required molecular descriptors for the new compounds are calculated. Since only *in silico* molecular descriptors are used, there is no prior requirement of synthesis of new compounds for model prediction, and future synthesized and assayed compounds can be used to validate even further the developed model.

Table 5 Selected *in silico* calculated molecular descriptors values for the titled compounds (**1**–**22**) in the QSAR model.

Compound	BLI[a]	GATS1m[b]	GATS5v[c]
1	0.981	0.540	1.158
2	0.940	0.541	1.195
3	0.974	0.508	1.041
4	0.993	0.539	1.166
5	0.955	0.540	1.202
6	0.987	0.508	1.047
7	0.906	0.518	0.850
8	0.964	0.571	1.032
9	0.926	0.573	1.067
10	0.955	0.540	0.890
11	1.024	0.572	1.115
12	0.989	0.576	1.141
13	1.022	0.561	1.016
14	1.023	0.572	1.098
15	0.988	0.576	1.126
16	1.021	0.561	0.968
17	0.921	0.584	1.034
18	0.881	0.588	1.066
19	0.908	0.555	0.984
20	0.952	0.577	0.913
21	0.907	0.580	0.952
22	0.940	0.544	1.063

[a] BLI: Kier benzene-likeliness index.
[b] GATS1m: Geary autocorrelation of lag 1 weighted by mass.
[c] GATS5v: Geary autocorrelation of lag 5 weighted by van der Waals volume.

Table 6 Pearson's linear correlation matrix for the *in silico* calculated molecular descriptors and tested NHDF, HepaRG, Caco-2, MCF-7, T47D and LNCaP cell lines relative cell proliferation.

Correlation	BLI[a]	GATS1m[b]	GATS5v[c]
BLI	1.000		
GATS1 m	−0.086	1.000	
GATS5 v	0.310	0.036	1.000
NHDF	−0.350	−0.019	−0.391
HepaRG	−0.651	−0.341	−0.369
Caco-2	−0.565	−0.427	−0.422
MCF-7	−0.307	−0.529	−0.414
T47D	−0.377	−0.446	−0.426
LNCaP	−0.487	0.050	−0.551

[a] BLI: Kier benzene-likeliness index.
[b] GATS1m: Geary autocorrelation of lag 1 weighted by mass.
[c] GATS5v: Geary autocorrelation of lag 5 weighted by van der Waals volume.

Although artificial neural networks are commonly known as "black-boxes" due to the challenging interpretation of the inputs used (Olden and Jackson, 2002), several approaches have been developed to overcome this limitation (Olden et al., 2004). In this work, we employed the Lek profile method as depicted in the experimental section. Fig. 9 shows the obtained trends between the molecular descriptors and the response variable for each cell line. Although the trends obtained differ between the cell lines, as would be expected,

Table 7 Statistical evaluation of the developed QSAR model for the cross-validation, training and test data.

Parameter	Value
Train cases	112
Q^2 (10-fold cross-validation)	0.686
$RMSE_{CV}$ (10-fold cross-validation)	0.086
R^2 (non cross-validated)	0.764
RMSE (non cross-validated)	0.074
Test cases	20
R^2_{pred}	0.699
RMSE	0.087

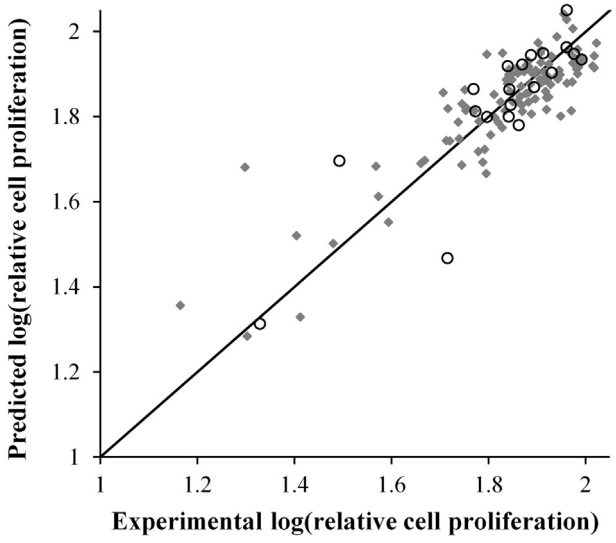


Figure 8 Plot of the experimental and predicted log(relative cell proliferation) for the developed 6-2-1 BRANN QSAR model. Solid line represents the line of unity, grey marks indicate cases used for training and open circles represent cases used for external testing.

even so they are fairly similar. In general, as the value of the molecular descriptors BLI, GATS1m and GATS5v increases, the relative cell proliferation decreases, and this relationship is similar to the Pearson's linear correlation coefficient values found in Table 6. BLI is a measure of aromaticity calculated from molecular topology, obtained by dividing the first-order valence connectivity index $^1\chi^v$ by the number of bonds in the molecule (excluding hydrogen bonds). The value is normalized on the benzene molecule, for which BLI takes a value of 1. Thus, as suggested by the Lek trends obtained, an increase in molecular aromaticity favours the antiproliferative activity (Kier and Hall, 1986; Todeschini and Consonni, 2000). GATS1m and GATS5v belong to the 2D autocorrelation group of molecular descriptors, which employ the Geary algorithm. These molecular descriptors describe the distribution of a specific atomic property in a molecule. Higher values are obtained when pairs of atoms in a molecule at a specified topological distance (lag value) present differences in the selected atomic property. Thus, as suggested by the Lek trends

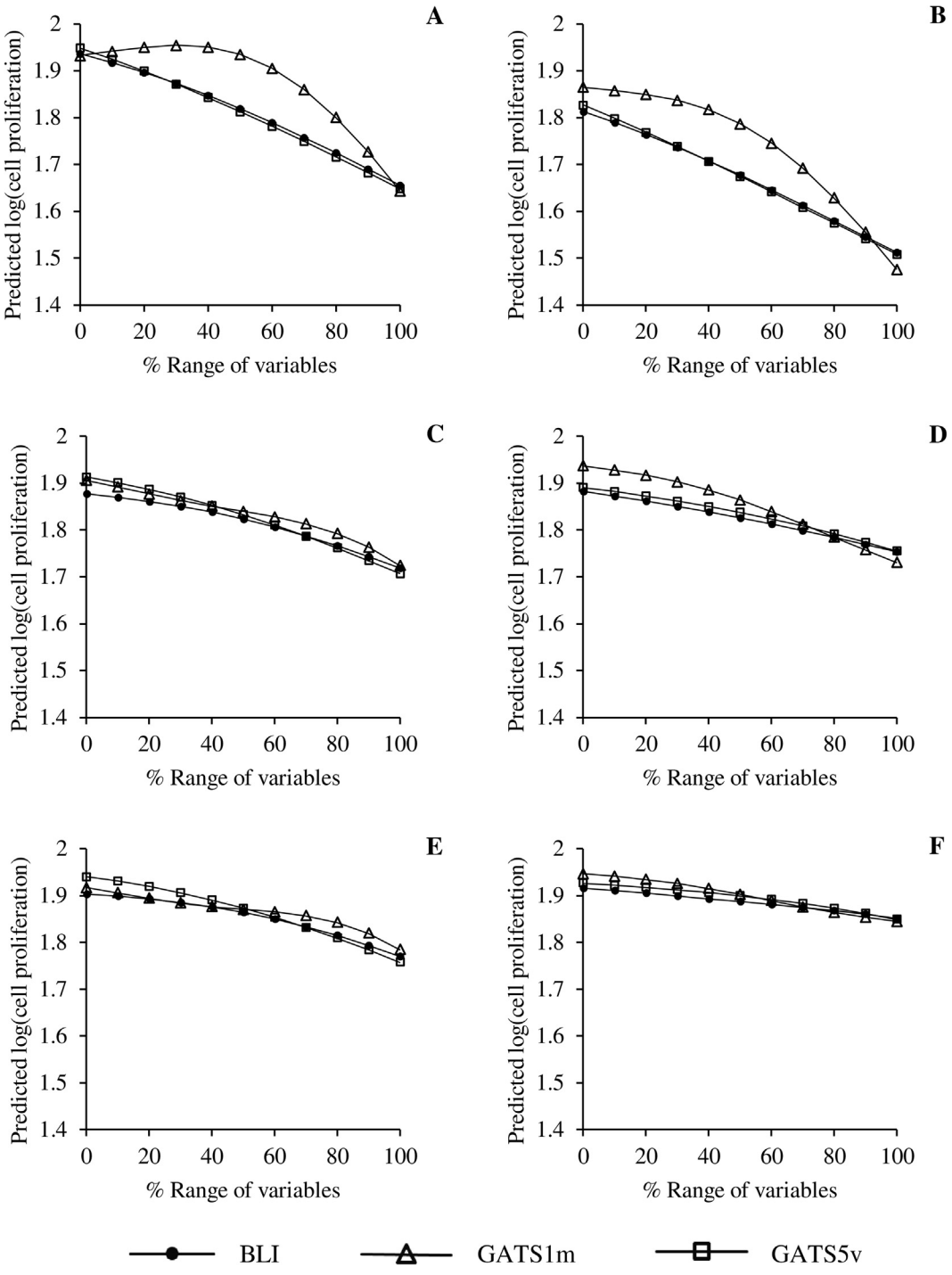


Figure 9 Contribution profile of the molecular descriptors BLI, GATS1m and GATS5v to the prediction of the log(relative cell proliferation) by the BRANN QSAR model for the (A) Caco-2, (B) HepaRG, (C) LNCaP, (D) MCF-7, (E) NHDF and (F) T47D cell lines. Each data point is obtained as the average predicted output when each variable is varied across its minimum and maximum value and the remaining variables are fixed at their minimum, first quartile, median, third quartile and maximum value.

obtained, larger differences in atomic weight and van der Waals volume of atom pairs at a topological distance of 1 and 5, respectively, tend to favour the antiproliferative activity (Geary, 1954; Todeschini and Consonni, 2000). In terms of relative importance of each molecular descriptor to the prediction of the antiproliferative activity, it appears that GATS1m is the most relevant input, followed by GATS5v, and finally BLI.

4. Conclusion

In summary, a series of DHPMt was successfully synthesized via the Biginelli three-component condensation reaction between an aldehyde, a β-ketoester/acetylacetone and thiourea, and was evaluated for their potential anticancer activity. From the *in vitro* antiproliferative screening, it was clear that chlorine-incorporating compounds had a signifi-

cant effect on the proliferation of hepatic (HepaRG), colon (Caco-2) and breast (MCF-7) cancer cell lines, without significant cytotoxicity for the normal dermal cell line (NHDF). Our data also showed that, when comparing with the anticancer drug 5-FU, the derivative **11** had higher potency towards the HepaRG cell line. Considering the antiproliferative data, a QSAR model was developed based only on three *in silico* calculated molecular descriptors (BLI, GATS1m, and GATS5v), which is able to reliably predict the relative cell proliferation in multiple cell lines (NHDF, HepaRG, Caco-2, MCF-7, T47D and LNCaP). This model can be a valuable tool to guide future efforts in improving antiproliferative activity of new structurally related compounds, while maintaining selectivity. Furthermore, flow cytometric analysis revealed that compound **15** caused cell cycle arrest and led to accumulation of cells in G_0/G_1 phase. Although our results afford some interesting information, additional studies are required for a better understanding of the mechanisms of cytotoxicity.

Acknowledgements

The authors are grateful to Fundação para a Ciência e a Tecnologia (Lisbon, Portugal) for the PhD fellowships of Mariana Matias (SFRH/BD/85279/2012) and Gonçalo Campos (SFRH/BD/95505/2013). This work was also supported by FEDER funds through the POCI - COMPETE 2020 - Operational Programme Competitiveness and Internationalisation in Axis I - Strengthening research, technological development and innovation (Project POCI-01-0145-FEDER-007491) and National Funds by FCT - Foundation for Science and Technology (Project UID/Multi /00709/2013).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.arabjc.2016.12.007.

References

Abdou, A.M., Botros, S., Hassan, R.A., Kamel, M.M., Taber, D.F., Taher, A.T., 2015. Useful four-carbon synthons en route to monastrol analogs. Tetrahedron 71, 139–146.

Abnous, K., Barati, B., Mehri, S., Farimani, M.R.M., Alibolandi, M., Mohammadpour, F., Ghandadi, M., Hadizadeh, F., 2013. Synthesis and molecular modeling of six novel monastrol analogues: evaluation of cytotoxicity and kinesin inhibitory activity against HeLa cell line. DARU 21, 1–8.

ACD/ChemSketch, 2015. Advanced Chemistry Development, Inc., Toronto, ON, Canada. <http://www.acdlabs.com>.

ACD/Percepta, 2015. Advanced Chemistry Development, Inc., Toronto, ON, Canada, <http://www.acdlabs.com>.

Asraf, H., Avunie-Masala, R., Hershfinkel, M., Gheber, L., 2015. Mitotic slippage and expression of survivin are linked to differential sensitivity of human cancer cell-lines to the Kinesin-5 inhibitor monastrol. PLoS ONE 10, e0129255.

Bartoli, K.M., Jakovljevic, J., Woolford, J.L., Saunders, W.S., 2011. Kinesin molecular motor Eg5 functions during polypeptide synthesis. Mol. Biol. Cell 22, 3420–3430.

Burden, F.R., Winkler, D.A., 1999. Robust QSAR models using Bayesian regularised artificial neural networks. J. Med. Chem. 42, 3183–3187.

Caldwell, G.W., 2015. In silico tools used for compound selection during target-based drug discovery and development. Expert Opin. Drug Discov. 10, 901–923.

Cherkasov, A., Muratov, E.N., Fourches, D., Varnek, A., Baskin, I.I., Cronin, M., Dearden, J.C., Gramatica, P., Martin, Y.C., Todeschini, R., Consonni, V., Kuz, V.E., Cramer, R.D., Benigni, R., Yang, C., Rathman, J.F., Terfloth, L., Gasteiger, J., Richard, A. M., Tropsha, A., 2013. Perspective QSAR modeling: where have you been ? Where are you going to? J. Med. Chem. 57, 4977–5010.

de Fátima, Â., Braga, T.C., da S Neto, L., Terra, B.S., Oliveira, B.G.F., da Silva, D.L., Modolo, L.V., 2015. A mini-review on Biginelli adducts with notable pharmacological properties. J. Adv. Res. 6, 363–373.

Dearden, J.C., Cronin, M.T.D., Kaiser, K.L.E., 2009. How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSPR). SAR QSAR Env. Res. 20, 241–266.

Falnikar, A., Tole, S., Baas, P.W., 2011. Kinesin-5, a mitotic microtubule-associated motor protein, modulates neuronal migration. Mol. Biol. Cell 22, 1561–1574.

Fan, X., Zhang, X., Zheng, Y., 2002. Samarium chloride catalysed Biginelli reaction: one-pot synthesis of 3,4-dihydropyrimidin-2(1*H*)-ones. J. Chem. Res. 2002. 438–438.

Fu, R., Yang, Y., Lai, W., Ma, Y., Chen, Z., Zhou, J., Chai, W., Wang, Q., Yuan, R., 2015. Efficient and green microwave-assisted multicomponent Biginelli reaction for the synthesis of dihydropyrimidinones catalyzed by heteropolyanion-based ionic liquids under solvent-free conditions. Synth. Commun. 45, 467–477.

Ganem, B., 2009. Strategies for innovation in multicomponent reaction design. Acc. Chem. Res. 42, 463–472.

Geary, R.C., 1954. The contiguity ratio and statistical mapping. Incorp. Stat. 5, 115–127.

Gong, L.-Z., Chen, X.-H., Xu, X.-Y., 2007. Asymmetric organocatalytic Biginelli reactions: a new approach to quickly access optically active 3,4-dihydropyrimidin-2-(1*H*)-ones. Chem. Eur. J. 13, 8920–8926.

Guido, B.C., Ramos, L.M., Nolasco, D.O., Nobrega, C.C., Andrade, B.Y., Pic-Taylor, A., Neto, B.A., Corrêa, J.R., 2015. Impact of kinesin Eg5 inhibition by 3,4-dihydropyrimidin-2(1*H*)-one derivatives on various breast cancer cell features. BMC Cancer 15, 283.

Hingane, D.G., Shumaila, A.M.A., 2013. Note Silica gel supported bismuth nitrate pentahydrate: a highly active catalyst under solvent free conditions towards the synthesis of dihydropyrimidin-2(1*H*)-ones and their sulphur analogues. Ind. J. Chem. 52, 1161–1165.

Hulme, C., Gore, V., 2003. Multi-component reactions: emerging chemistry in drug discovery" "from xylocain to crixivan. Curr. Med. Chem. 10, 51–80.

Kalita, H.R., Phukan, P., 2007. CuI as reusable catalyst for the Biginelli reaction. Catal. Commun. 8, 179–182.

Kappe, C.O., 2000. Highly versatile solid phase synthesis of biofunctional 4-aryl-3,4-dihydropyrimidines using resin-bound isothiourea building blocks and multidirectional resin cleavage. Bioorg. Med. Chem. Lett. 10, 49–51.

Kier, L.B., Hall, L.H., 1986. Molecular Connectivity in Structure-Activity Analysis. Wiley-VCH, United Kingdom.

Kolosov, M.A., Orlov, V.D., Beloborodov, D.A., Dotsenko, V.V., 2009. A chemical placebo: NaCl as an effective, cheapest, non-acidic and greener catalyst for Biginelli-type 3,4-dihydropyrimidin-2(1*H*)-ones (-thiones) synthesis. Mol. Divers. 13, 5–25.

Lek, S., Belaud, A., Baran, P., Dimopoulos, I., Delacoste, M., 1996a. Role of some environmental variables in trout abundance models using neural networks. Aquat. Living Resour. 9, 23–29.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., Aulagnier, S., 1996b. Application of neural networks to modelling nonlinear relationships in ecology. Ecol. Model. 90, 39–52.

Liu, Q., Pan, N., Xu, J., Zhang, W., Kong, F., 2013. Microwave-assisted and iodine-catalyzed synthesis of dihydropyrimidin-2-thiones via Biginelli reaction under solvent-free conditions. Synth. Commun. 43, 139–146.

MATLAB and Neural Network Toolbox Release 2014a, The MathWorks, Inc., Natick, Massachusetts, United States, 2014.

Mayer, T.U., Kapoor, T.M., Haggarty, S.J., King, R.W., Schreiber, S. L., Mitchison, T.J., 1999. Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen. Science 286, 971–974.

Nandurkar, N.S., Bhanushali, M.J., Bhor, M.D., Bhanage, B.M., 2007. $Y(NO_3)_3{\cdot}6H_2O$: a novel and reusable catalyst for one pot synthesis of 3,4-dihydropyrimidin-2(1*H*)-ones under solvent-free conditions. J. Mol. Catal. A Chem. 271, 14–17.

Nantasenamat, C., Isarankura-Na-Ayudhya, C., Naenna, T., Prachayasittikul, V., 2009. A practical overview of quantitative structure-activity relationship. EXCLI J. 8, 74–88.

Narahari, S.R., Reguri, B.R., Gudaparthi, O., Mukkanti, K., 2012. Synthesis of dihydropyrimidinones via Biginelli multi-component reaction. Tetrahedron Lett. 53, 1543–1545.

Nasr-Esfahani, M., Montazerozohori, M., Aghel-Mirrezaee, M., Kashi, H., 2014. Efficient and green catalytic synthesis of dihydropyrimidinone (thione) derivatives using cobalt nitrate in solvent-free conditions. J. Chil. Chem. Soc. 59, 2311–2314.

Olden, J.D., Jackson, D.A., 2002. Illuminating the "black box": a ramdomization approach for understanding variable contributions in artificial neuronal networks. Ecol. Model. 154, 135–150.

Olden, J.D., Joy, M.K., Death, R.G., 2004. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. Ecol. Model. 178, 389–397.

Paixão, P., Aniceto, N., Gouveia, L.F., Morais, J.A.G., 2014. Prediction of drug distribution in rat and humans using an artificial neural networks ensemble and a PBPK model. Pharm. Res. 31, 3313–3322.

Prashantha Kumar, B.R., Sankar, G., Nasir Baig, R.B., Chandrashekaran, S., 2009. Novel Biginelli dihydropyrimidines with potential anticancer activity: a parallel synthesis and CoMSIA study. Eur. J. Med. Chem. 44, 4192–4198.

Prokopcová, H., Dallinger, D., Uray, G., Kaan, H.Y.K., Ulaganathan, V., Kozielski, F., Laggner, C., Kappe, C.O., 2010. Structure-activity relationships and molecular docking of novel dihydropyrimidine-based mitotic Eg5 inhibitors. ChemMedChem 5, 1760–1769.

Ramos, L.M., Guido, B.C., Nobrega, C.C., Corrêa, J.R., Silva, R.G., De Oliveira, H.C.B., Gomes, A.F., Gozzo, F.C., Neto, B.A.D., 2013. The Biginelli reaction with an imidazolium-tagged recyclable iron catalyst: kinetics, mechanism, and antitumoral activity. Chem. Eur. J. 19, 4156–4168.

Ranu, B.C., Hajra, A., 2000. Indium (III) chloride-catalyzed one-pot synthesis of dihydropyrimidinones by a three-component coupling of 1,3-dicarbonyl compounds, aldehydes, and urea: an improved procedure for the Biginelli reaction. J. Org. Chem. 65, 6270–6272.

Rashid, U., Batool, I., Wadood, A., Khan, A., ul-Haq, Z., Chaudhary, M.I., Ansari, F.L., 2013. Structure based virtual screening-driven identification of monastrol as a potent urease inhibitor. J. Mol. Graph. Model. 43, 47–57.

Rodríguez-Domínguez, J.C., Bernardi, D., Kirsch, G., 2007. $ZrCl_4$ or $ZrOCl_2$ under neat conditions: optimized green alternatives for the Biginelli reaction. Tetrahedron Lett. 48, 5777–5780.

Rücker, C., Rücker, G., Meringer, M., 2007. y-Randomization and its variants in QSPR/QSAR. J. Chem. Inf. Model. 47, 2345–2357.

Russowsky, D., Canto, R.F.S., Sanches, S.A.A., D'Oca, M.G.M., de Fátima, Â., Pilli, R.A., Kohn, L.K., Antônio, M.A., de Carvalho, J.E., 2006. Synthesis and differential antiproliferative activity of Biginelli compounds against cancer cell lines: monastrol, oxomonastrol and oxygenated analogues. Bioorg. Chem. 34, 173–182.

Siegel, R.L., Miller, K.D., Jemal, A., 2015. Cancer statistics, 2015. CA Cancer J. Clin. 65, 5–29.

Su, W., Li, J., Zheng, Z., Shen, Y., 2005. One-pot synthesis of dihydropyrimidiones catalyzed by strontium (II) triflate under solvent-free conditions. Tetrahedron Lett. 46, 6037–6040.

Suresh, Sandhu, J.S., 2012. Past, present and future of the Biginelli reaction: a critical perspective. ARKIVOC 2012, 69–73.

Svetlik, J., Veizerová, L., Mayer, T.U., Catarinella, M., 2010. Monastrol analogs: a synthesis of pyrazolopyridine, benzopyranopyrazolopyridine, and oxygen-bridged azolopyrimidine derivatives and their biological screening. Bioorg. Med. Chem. Lett. 20, 4073–4076.

Tetko, I.V., Gasteiger, J., Todeschini, R., Mauri, A., Livingstone, D., Ertl, P., Palyulin, V.A., Radchenko, E.V., Zefirov, N.S., Makarenko, A.S., Tanchuk, V.Y., Prokopenko, V.V., 2005. Virtual computational chemistry laboratory - design and description. J. Comput. Aided Mol. Des. 19, 453–463.

Todeschini, R., Consonni, V., 2000. Handbook of Molecular Descriptors. Wiley-VCH, United Kingdom.

Tropsha, A., 2010. Best practices for QSAR model development, validation, and exploitation. Mol. Inf. 29, 476–488.

Tropsha, A., Gramatica, P., Gombar, V.K., 2003. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. QSAR Comb. Sci. 22, 69–77.

Videira, M., Reis, R.L., Brito, M.A., 2014. Deconstructing breast cancer cell biology and the mechanisms of multidrug resistance. Biochim. Biophys. Acta 1846, 312–325.

Wang, L., Qian, C., Tian, H., Ma, Y., 2003. Lanthanide triflate catalyzed one-pot synthesis of dihydropyrimidin-2(1*H*)-thiones by a three-component of 1,3-dicarbonyl compounds, aldehydes, and thiourea using a solvent-free Biginelli condensation. Synth. Commun. 33, 1459–1468.

Whitley, D.C., Ford, M.G., Livingstone, D.J., 2000. Unsupervised forward selection: a method for eliminating redundant variables. J. Chem. Inf. Comput. Sci. 40, 1160–1168.