

QSAR study and molecular docking on indirubin inhibitors of Glycogen Synthase Kinase-3

Research Article

Luminita Crisan¹, Liliana Pacureanu¹, Alina Bora^{1*},
Sorin Avram¹, Ludovic Kurunczi², Zeno Simon¹

¹Department of Computational Chemistry,
Institute of Chemistry of the Romanian Academy,
300223 Timisoara, Romania

²Faculty of Pharmacy, Department of Chemical Physics,
University of Medicine and Pharmacy "Victor Babes",
300041 Timisoara, Romania

Received 10 May 2012; Accepted 23 August 2012

Abstract: The current study describes the development of *in silico* models based on a novel alternative of the MTD-PLS methodology (Partial-Least-Squares variant of Minimal Topologic Difference) developed by our group to predict the inhibition of GSK-3 β by indirubin derivatives. The new MTD-PLS methodology involves selection rules for the PLS equation coefficients based on physico-chemical considerations aimed at reducing the bias in the output information. These QSAR models have been derived using calculated fragmental descriptors relevant to binding including polarizability, hydrophobicity, hydrogen bond donor, hydrogen bond acceptor, volume and electronic effects. The MTD-PLS methodology afforded moderate but robust statistical characteristics ($R^2_v(\text{CUM}) = 0.707$, $Q^2(\text{CUM}) = 0.664$). The MTD-PLS model obtained has been validated in terms of predictive ability by joined internal-external cross-validation applying Golbraikh-Tropsha criteria and Y-randomization test. The information supplied by the MTD-PLS model has been evaluated against Fujita-Ban outcomes that afforded a statistically reliable model ($R^2 = 0.923$). Furthermore, the results originated from QSAR models were laterally validated with docking insights that suggested the substitution pattern for the design of new indirubins with improved pharmacological potential against GSK-3 β . The new restriction rules introduced in this paper are applicable and provide reliable results in accordance with physico-chemical reality.

Keywords: MTD-PLS • Fujita-Ban • Indirubin derivatives • GSK-3 β

© Versita Sp. z o.o.

1. Introduction

Protein kinases are involved in numerous physiological processes and when dysregulated may induce many diseases [1]. They belong to the largest family of enzymes characterized by a well-conserved ATP binding pocket [2]. Glycogen synthase kinase-3 (GSK-3) is a member of the serine/threonine protein kinase family. It exists pervasively in two narrowly related isoforms, GSK-3 α and GSK-3 β (encoded by distinct genes), which exhibit high reciprocal homology (90%) [3]. GSK-3 plays a critical role in glucose homeostasis, central nervous system function (*via* tau and β -catenin), and cancer (angiogenesis and tumorigenesis) [4]. Several publications have illustrated the ATP-competitive mechanism for GSK-3 inhibition by indirubins, maleimide

derivatives, paullones, staurosporines, aloisines, and the marine sponge hymenialdisine [5-7]. Hence, the design of novel potent, selective, ATP-competitive GSK-3 inhibitors is facilitated by existing knowledge concerning conspicuous structural characteristics of the ATP binding pocket [8].

The therapeutic potential of GSK-3 inhibition demonstrated by *in vitro* stimulation of glycogen synthesis and plasma glucose lowering in diabetic animals [9], has induced the pursuit of novel GSK-3 inhibitors, a challenging enterprise in both academic institutions and pharmaceutical companies.

The bis-indole (indirubin) is an active component of Danggui Longhui Wan, a traditional Chinese formula used to treat chronic disorders including leukemia and Alzheimer [10]. However, indirubin shows weak

* E-mail: alina.bora@gmail.com

pharmacokinetic and bioavailability properties including poor solubility, low absorption, and gastrointestinal toxicity [11]. Indirubin derivatives have been proven to be potent inhibitors of GSK-3 β (IC_{50} = 5–50 nM) compared to other indigoids that have limited pharmacological effects [12]. The biological assay of a series of indoles and bis-indoles against GSK-3 β , cyclin dependent kinase-1 (CDK1)/cyclin B, and cyclin dependent kinase-5 (CDK5)/p25 proved that only indirubins act upon these kinases [13].

The current literature still lacks a comprehensive investigation of indirubin derivatives displaying biological activity, although 109 active indirubins have been detected between 2001 and 2008 [12–17]. Such a broad study could lead to the rational design of indirubin inhibitors with better pharmacological properties and may lead towards a better understanding of the selective inhibition of GSK-3 β .

Since biological activity is directly related to the nature and position of substituents on a common molecular skeleton, in this paper we have applied two QSAR methodologies that assess substituent effects: a novel variant of the MTD-PLS method developed by our group, and the Fujita-Ban QSAR variant. The current investigation has been applied to 109 indirubin derivatives that display relevant structural variability and provide meaningful information in terms of SAR. Thus, the current dataset thoroughly reflects the ligand binding motif of GSK-3, offering the basis for an objective analysis that mirrors the current understanding in the field of GSK-3 inhibition. This can offer the necessary clues for rational design of new indirubin inhibitors with favorable pharmacology and the ability to forecast their activities. The PLS variant of MTD is an alignment-based three-dimensional quantitative structure activity relationships method which has been described in previous work [18,19]. The restriction rules concerning the sign of some QSAR equation coefficients (thus introducing supplementary chemical information in the equation) used previously have now been complemented with threshold values that restrict the upper and lower limit for the coefficients of independent variables that describe various interactions between ligand and receptor. The new restriction rules are aimed at reducing the bias and improving the predictive ability of the MTD-PLS equation. The sign and threshold values of the QSAR equation coefficients have been deduced on the basis of chemical and physical considerations.

The development of robust models that meaningfully suggest a reliable correlation between the experimental and predicted affinity of indirubin derivatives towards GSK-3 β is a point of interest in the current paper. The identification of consistent findings between the Fujita-

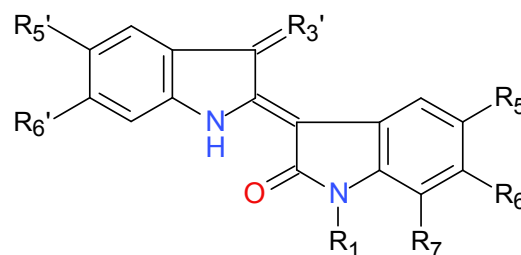


Figure 1. The template of indirubin analogues.

Ban method and the novel MTD-PLS procedure (the influence of steric, hydrophobic, donor, acceptor and electronic factors on biological activity, and binding site deformability) and some docking outcomes complement the main targets of our work.

2. Computational details

2.1. Dataset

109 Indirubin derivatives displaying biological activities have been selected from the literature [13–17] and listed in Table 1. The biological activities of indirubin derivatives are expressed as the negative logarithm of the experimental half-maximal effective concentrations, pIC₅₀ (expressed in molar units). The template (Fig. 1) of the indirubin analogues was built with the MarvinSketch v 5.4.1.1 [20] software.

2.2. MTD-PLS methodology

The PLS variant of minimal topologic difference (MTD) relates the spatially positioned properties of ligands to biological activity. The hypermolecule represents the central concept of the method: it is obtained by atom-per-atom superposition of the ligands (neglecting the H atoms) using a common skeleton, thus a topological network results with molecule fragments in its vertices [19]. Many conformational 3D descriptors which depend on the spatial distribution of the atoms in a molecule, such as the Randic molecular profile, radial distribution functions, MoRSE descriptors, and GETAWAY descriptors, are relatively insensitive in QSAR modeling to the conformational composition of the molecule [21]. Also, the application of a method such as MTD-PLS, which is based on the representation of a ligand fragment's ability to interact with the receptor (GRIND [22]), showed that the statistical parameters present moderate variation when randomly selected conformers are used. Thus, even if the three-dimensional sampling of the hypermolecule space can be provided by multiple ligand conformations, in this work only the lowest energy conformer for every molecule is retained. This is also in

Table 1. The structure of the dataset compounds, the occupied vertices (j) in hypermolecule and their biological activities, (for the MTD-PLS model).*

No	R1/R5'/R6'/R7/ R5/R6	R3'	j	pIC50
1	H/H/H/H/ H/H	O	4, 80	6.000
2	H/H/H/H/ H/H	NOH	4, 5, 80	7.657
3	H/H/H/H/ H/H	NOAc	4-6, 8, 9, 80	6.698
4	H/H/H/H/ H/H	NOCH ₃	4-6, 80	6.823
5	H/H/Br/H/ H/Br	O	1, 4, 80, 82	5.346
6	H/H/Br/H/ H/Br	NOH	1, 4, 5, 80, 82	6.921
7	H/H/Br/H/ H/H	O	1, 4, 80	4.657
8	H/H/Br/H/ H/H	NOH	1, 4, 5, 80	6.468
9	H/H/H/H/ H/Br	O	4, 80, 82	7.347
10	H/H/H/H/ H/Br	NOH	4, 5, 80, 82	8.301
11	H/H/H/H/ H/Br	NOAc	4- 6, 8, 9, 80, 82	8.000
12	H/H/H/H/ H/Br	NOCH ₃	4-6, 80, 82	7.522
13	H/H/H/H/ H/Cl	O	4, 80, 82	6.854
14	H/H/H/H/ H/Cl	NOH	4, 5, 80, 82	7.699
15	H/H/H/H/ H/Cl	NOAc	4-6, 8, 9, 80, 82	7.769
16	H/H/H/H/ H/I	O	4, 80, 82	7.259
17	H/H/H/H/ H/I	NOH	4, 5, 80, 82	8.000
18	H/H/H/H/ H/I	NOAc	4- 6, 8, 9, 80, 82	7.886
19	H/H/H/H/ H/ CH=CH ₂	O	4, 80, 82, 83	6.619
20	H/H/H/H/ H/ CH=CH ₂	NOH	4, 5, 80, 82, 83	7.222
21	H/H/H/H/ H/ CH=CH ₂	NOAc	4- 6, 8, 9, 80, 82, 83	7.187
22	H/H/H/H/ H/F	O	4, 80, 82	6.187
23	H/H/H/H/ H/F	NOH	4, 5, 80, 82	6.886
24	H/H/H/H/ H/F	NOAc	4- 6, 8, 9, 80, 82	7.046
25	H/H/H/H/ CH ₃ /Br	O	4, 80, 82,85	7.602
26	H/H/H/H/ CH ₃ /Br	NOH	4, 5, 80, 82, 85	8.222
27	H/H/H/H/ CH ₃ /Br	NOAc	4- 6, 8, 9, 80, 82, 85	8.155
28	H/H/H/H/ Cl/Cl	O	4, 80, 82, 85	7.523
29	H/H/H/H/ Cl/Cl	NOH	4, 5, 80, 82, 85	8.398
30	H/H/H/H/ Cl/Cl	NOAc	4- 6, 8, 9, 80, 82, 85	8.398
31	H/H/H/H/ NO ₂ /Br	O	4, 80, 82, 85	7.000
32	H/H/H/H/ NO ₂ /Br	NOH	4, 5, 80, 82, 85	8.155
33	H/H/H/H/ NO ₂ /Br	NOAc	4- 6, 8, 9, 80, 82, 85	8.222
34	H/H/H/H/ I/H	NOH	4, 5, 80, 85	8.046
35	H/H/H/H/ SO ₂ -NH-C ₂ H ₄ -OH/H	O	4, 81, 89, 90-92	7.482
36	H/H/H/H/ SO ₂ NH ₂ /H	O	4, 81, 89, 90	7.398
37	H/H/H/H/ NO ₂ /H	O	4, 80, 85	7.377
38	H/H/H/H/ Cl/H	O	4, 80, 85	7.301
39	H/H/H/H/ Br/H	O	4, 80, 85	7.259
40	H/H/H/H/ CH ₃ /H	O	4, 80, 85	7.207
41	H/H/H/H/ I/H	O	4, 80, 85	7.167

Continued Table 1. The structure of the dataset compounds, the occupied vertices (j) in hypermolecule and their biological activities, (for the MTD-PLS model).*

No	R1/R5'/R6'/R7/ R5/R6	R3'	j	pIC50
42	H/H/H/H/ F/H	O	4, 80, 85	7.108
43	H/H/H/H/ SO ₃ H/H	NOH	4, 5, 81, 89, 90	7.097
44	H/H/H/H/ SO ₂ -NHCH ₃ /H	O	4, 81, 89-91	6.958
45	H/H/H/H/ SO ₃ -N(CH ₃) ₂ /H	O	4, 81, 89-91, 94	6.745
46	H/Br/H/H/ Br/H	O	2, 4, 80, 85	6.602
47	H/H/H/H/ SO ₃ H/H	O	4, 81, 89, 90	6.553
48	H/Br/H/H/ H/H	O	2, 4, 80	6.456
49	H/H/H/H/SO ₂ -N(C ₂ H ₄ OH) ₂ /H	O	4, 81, 89-96	6.398
50	H/Br/H/H/ SO ₃ H/H	O	2, 4, 81, 89, 90	5.398
51	Ph/ H/H/H/ H/H	O	3, 4, 80	3.699
52	H/H/H/F/ H/H	O	4, 80, 84	6.398
53	H/H/H/F/ H/H	NOH	4, 5, 80, 84	6.569
54	H/H/H/F/ H/H	NOCH ₃	4- 6, 80, 84	6.356
55	H/H/H/F/ H/H	NOCOCH ₃	4-6, 8, 9, 80, 84	6.481
56	H/H/H/Cl/ H/H	NOH	4, 5, 80, 84	4.678
57	H/H/H/Br/ H/H	NOH	4, 5, 80, 84	4.495
58	H/H/H/I/ H/H	NOH	4, 5, 80, 84	4.796
59	CH ₃ / H/H/I/ H/H	NOH	3-5, 80, 84	4.523
60	H/H/H/Br/ H/H	NOCH ₂ CH ₂ Br	4- 6, 10, 17, 80, 84	4.000
61	H/H/H/Br/ H/H	NOCH ₂ CH ₂ -N 	4-6, 10, 17, 18, 22-24, 80, 84	5.155
62	H/H/H/Br/ H/H	NOCH ₂ CH ₂ -N  *HCl	4, 5, 43-45, 80, 84	5.523
63	H/H/H/Br/ H/H	NOCH ₂ CH ₂ -N 	4-6, 10, 17, 18, 22, 34, 35, 37, 80, 84	6.244
64	H/H/H/Br/ H/H	NOCH ₂ CH ₂ -N 	4-6, 10, 17, 18, 22-24, 80, 84	5.046
65	H/H/H/Br/ H/H	NOCH ₂ CH ₂ -N  *HCl	4, 5, 43, 46, 48, 69, 70, 74, 75, 80, 84	4.959
66	H/H/H/Br/ H/H	NOCH ₂ CH ₂ -N 	4-6, 10, 17, 18, 22, 34, 35, 37, 80, 84	5.301
67	H/H/H/Br/ H/H	NOCH ₂ CH ₂ N(CH ₃) ₂	4-6, 10, 17, 18, 22, 80, 84	5.097
68	H/H/H/Br/ Br/H	NOH	4, 5, 80, 85	7.796
69	H/H/H/H/ NO ₂ /H	NOH	4, 5, 80, 85	8.678
70	CH ₃ /H/H/Br/ NO ₂ /H	NOH	3- 5, 80, 85	6.276
71	H/Br/H/H/ NO ₂ /H	NOH	2, 4, 5, 80, 85	7.259
72	CH ₃ /Br/H/H/ NO ₂ /H	NOH	2- 5, 80, 85	5.031
73	H/H/H/H/ NH ₂ /H	O	4, 80, 85	7.097

Continued Table 1. The structure of the dataset compounds, the occupied vertices (j) in hypermolecule and their biological activities, (for the MTD-PLS model).*

No	R1/R5'/R6'/R7/ R5/R6	R3'	j	pIC50
74	H/H/H/H/ NH-Ac/H	O	4, 80, 85-88	8.125
75	H/Br/H/H/ NH ₂ /H	O	2, 4, 80, 85	6.356
76	H/Br/H/H/ NH-Ac/H	O	2, 4, 80, 85-88	7.136
77	H/H/H/H/ NH ₂ /H	NOH	4, 5, 80, 85	6.444
78	H/H/H/H/ NH-Ac/H	NOH	4, 5, 80, 85-88	6.456
79	H/Br/H/H/ NH ₂ /H	NOH	2, 4, 5, 80, 85	5.180
80	H/Br/H/H/ NH-Ac/H	NOH	2, 4, 5, 80, 85-88	4.398
81	H/H/H/H/ F/H	NOH	4, 5, 80, 85	5.887
82	H/Br/H/H/ F/H	NOH	2, 4, 5, 80, 85	4.824
83	H/Br/H/H/ Br/H	NOH	2, 4, 5, 80, 85	4.619
84	H/H/H/H/ H/Br	NOCH ₂ CH ₂ Br	4- 6, 10, 17, 80, 82	6.854
85	H/H/H/H/ H/Br	NOCH ₂ CH ₂ OH	4- 6, 10, 17, 80, 82	7.523
86	H/H/H/H/ H/Br	NOCH ₂ CH(OH)CH ₂ OH	4- 6, 10, 15- 17, 80, 82	7.468
87	H/H/H/H/ H/Br	NOCON(CH ₂ CH ₃) ₂	4-7, 10-14, 80, 82	7.523
88	H/H/H/H/ H/Br	NOCH ₂ CH ₂ N(CH ₃) ₂	4- 6, 10, 17, 18, 22, 80, 82	7.485
89	H/H/H/H/ H/Br	NOCH ₂ CH ₂ N(CH ₃) ₂ *HCl	4, 5, 43, 46, 47, 49, 63, 80, 82	7.537
90	H/H/H/H/ H/Br	NOCH ₂ CH ₂ N(CH ₂ CH ₃) ₂	4-6, 10, 17, 19, 20, 38, 39, 80, 82	7.456
91	H/H/H/H/ H/Br	NOCH ₂ CH ₂ N(CH ₂ CH ₃) ₂ *HCl	4, 5, 43, 46, 47, 76-80, 82	7.568
92	H/H/H/H/ H/Br	NOCH ₂ CH ₂ N(CH ₂ CH ₂ OH) ₂	4- 6, 10, 17, 19-21, 38, 39, 41, 80, 82	7.398
93	H/H/H/H/ H/Br	NOCH ₂ CH ₂ N(CH ₂ CH ₂ OH) ₂ *HCl	4, 5, 23, 24, 43, 46, 47, 49, 54, 63-66, 80, 82	7.387
94	H/H/H/H/ H/Br	NOCH ₂ CH ₂ N(CH ₃)CH ₂ CH(OH)CH ₂ OH	4- 6, 10, 17, 19, 38-42, 80, 82	7.174
95	H/H/H/H/ H/Br	NOCH ₂ CH ₂ N(CH ₃)CH ₂ CH(OH)CH ₂ OH*HCl	4, 5, 43, 46, 47, 49-53, 63, 80, 82	7.638
96	H/H/H/H/ H/Br	NOCH ₂ CH ₂ -N 	4- 6, 10, 17, 18, 22, 80, 82	7.585
97	H/H/H/H/ H/Br	NOCH ₂ CH ₂ -N  *HCl	4, 5, 43, 46, 47, 49, 54, 62, 63, 80, 82	7.267
98	H/H/H/H/ H/Br	NOCH ₂ CH ₂ -N 	4-6, 10, 17, 18, 22, 34, 35, 37, 80, 82	7.222
99	H/H/H/H/ H/Br	NOCH ₂ CH ₂ -N  *HCl	4, 5, 43, 46, 47, 49, 54, 63, 67, 68, 80, 82	6.958
100	H/H/H/H/ H/Br	NOCH ₂ CH ₂ -N 	4-6, 10, 17, 18, 22, 34, 35, 37, 80, 82	8.481
101	H/H/H/H/ H/Br	NOCH ₂ CH ₂ -N  *2HCl	4, 5, 43, 46, 48, 69, 71-73, 75, 80, 82	8.886
102	H/H/H/H/ H/Br	NOCH ₂ CH ₂ -N 	4- 6, 10, 17, 18, 22, 34, 35, 37, 80, 82	8.155

Continued Table 1. The structure of the dataset compounds, the occupied vertices (j) in hypermolecule and their biological activities, (for the MTD-PLS model).*

No	R1/R5'/R6'/R7/ R5/R6	R3'	j	pIC50
103	H/H/H/H/ H/Br	<chem>NOCH2CH2-N1CCN(C)CC1</chem> *2HCl	4, 5, 43, 46, 47, 49, 54-56, 62, 63, 80, 82	8.301
104	H/H/H/H/ H/Br	<chem>NOCH2CH2-N1CCN(CCO)CC1</chem>	4- 6, 10, 17, 18, 22, 25-29, 33, 80, 82	8.301
105	H/H/H/H/ H/Br	<chem>NOCH2CH2-N1CCN(CCO)CC1</chem> *2HCl	4, 5, 43, 46, 47, 49, 54-57, 58, 62, 63, 80, 82	8.376
106	H/H/H/H/ H/Br	<chem>NOCH2CH2-N1CCN(CCOCH3)CC1</chem>	4- 6, 10, 17, 18, 22, 25-30, 33, 80, 82	7.958
107	H/H/H/H/ H/Br	<chem>NOCH2CH2-N1CCN(CCOCH3)CC1</chem> *2HCl	4, 5, 43, 46, 47, 49, 54-59, 62, 63, 80, 82	7.699
108	H/H/H/H/ H/Br	<chem>NOCH2CH2-N1CCN(CCOCCO)CC1</chem>	4-6, 10, 17, 18, 22, 25-33, 80, 82	7.853
109	H/H/H/H/ H/Br	<chem>NOCH2CH2-N1CCN(CCOCCO)CC1</chem> *2HCl	4, 5, 43, 46, 47, 49, 54-63, 80, 82	7.481

agreement with the need to reduce the signal to noise ratio described below in Section 2.2.2.

The assessment of spatially extended fragment properties, as much as permitted by the current data set, mirrors the receptor binding site chart. The MTD-PLS method correlates these properties with the biological activity of $i = 1, 2, \dots, N$ compounds. The atoms (fragments) belonging to ligand i that occupy a vertex j ($j = 1, 2, \dots, M$, M being the total number of vertices) of the hypermolecule are characterized by increments of van der Waals volume (x_{jV}), polarizability (x_{jP}), hydrophobicity (x_{jH}), hydrogen bond donor and hydrogen bond acceptor properties (x_{jD} , x_{jA}), and electric charge (x_{jS}) [18,23]. The changes occurring in the physico-chemical descriptors at the vertices are related to changes of the biological activity (see Eq. 1). So far in our work [18,19,23] the steric misfit is considered as a detrimental effect (only $a_{jV} \leq 0$ accepted, see Eq. 1), while hydrophobic and polarizability interactions are considered beneficial (only a_{jH} , $a_{jP} \geq 0$ accepted, see Eq. 1), because entropic effects related to partial dehydration and van der Waals attractions between nonpolar entities are usually associated with positive contributions to the ligand binding affinity.

$$\hat{Y}_i = a_0 + \sum_{j=1}^M (a_{jH} \cdot x_{jHi} + a_{jV} \cdot x_{jVi} + a_{jP} \cdot x_{jPi} + a_{jS} \cdot x_{jSi} + a_{jA} \cdot x_{jAi} + a_{jD} \cdot x_{jDi}) \quad (1)$$

Here, $\hat{Y}_i$ represents the calculated biological activity, $a_{j\mu}$ ($\mu = H, V, P, S, A$, and D) are the PLS coefficients when the Y and x columns are unscaled and uncentered. The term a_0 denotes the intercept, and in fact represents the contribution of the hydrogen substituted compound (see Fig. 1) to the biological activity.

To carry out PLS modeling, we employed the SIMCA P9.0 [24] package. The theoretical number of descriptors is $6 \cdot M$, but a number of these are eliminated by the PLS procedure (small variance of parameter values). Other descriptors are absent at the start because of the nature of the atoms occupying the corresponding vertices (*i.e.*, lacking H-bonding characteristics).

2.2.1. New restriction rules in MTD-PLS

In order to eliminate chance correlations, new restriction rules for the $a_{j\mu}$ coefficients were introduced in the MTD-PLS methodology. These restriction rules are based

on the following rationale and relationships between the standard ligand-receptor binding free energy ΔG_i° (kcal mol⁻¹) and the experimental biological activity $Y_i = -\log K_{di}$ (see below the significance of K_{di}). The free energy variation corresponds to the equilibrium interaction of a certain ligand L_i with the biological receptor R , giving the ligand-receptor complex RL_i :



From the thermodynamics of equilibrium (Eq. 2a) at 37°C, $T = 310$ K:

$$\begin{aligned} \Delta G_i^\circ &= -2.303 \cdot R \cdot T \cdot \log K_{ai} = \\ &= 2.303 \cdot R \cdot T \cdot \log K_{di} = 1.42 \cdot \log K_{di} \text{ kcal mol}^{-1} \end{aligned} \quad (2b)$$

where K_{ai} and K_{di} are equilibrium constants (association and dissociation constants of the ligand-receptor complex RL_i respectively), R is the universal gas constant (0.00198 kcal mol⁻¹ K⁻¹). Therefore, we have:

$$-\frac{\Delta G_i^\circ}{1.42} = Y_i \quad (2c)$$

Thus, if Eq. 1 represents a good QSAR model, Eq. 2c is transformed into:

$$\begin{aligned} -\frac{\Delta G_i^\circ}{1.42} &\cong a_0 + \sum_{j=1}^M (a_{jH} \cdot x_{jHi} + a_{jV} \cdot x_{jVi} + \\ &+ a_{jP} \cdot x_{jPi} + a_{jS} \cdot x_{jSi} + a_{jA} \cdot x_{jAi} + a_{jD} \cdot x_{jDi}) \end{aligned} \quad (2d)$$

For every type of interaction in this equation, the approximate relationship (Eq. 2e) holds true:

$$-\frac{\Delta_{j\mu}(\Delta G_i^\circ)}{1.42} \cong a_{j\mu} \cdot x_{j\mu i} \quad (2e)$$

where $\Delta_{j\mu}$ refers to the contribution of parameter μ to the value of ΔG_i° .

Hydrophobic effects.

The terms $a_{jH} \cdot x_{jHi}$ correspond to the relocation of the ligand atom (moiety) j – described by x_{jHi} – from water to the binding site of the receptor R . This phenomenon is similar to water-octanol transfer that involves partial dehydration of the atom of ligand (L_i) at vertex j , and contact with the receptor R that occurs mainly by van der Waals interactions. As the x_{jHi} descriptor is expressed in fragmental octanol / water logP units (see below in Section 2.2.2), the numerical values are directly proportional to the standard free energy change corresponding to this transfer. The binding site of the receptor R is expected to be as lipophilic as octanol, but less hydrophilic than water. Therefore, we suggest the upper bound restriction for a_{jH} to have a value

of 1; taking into account also the presumed beneficial character of the hydrophobic interaction, we obtain Eq. 3:

$$0 \leq a_{jH} \leq 1 \quad (3)$$

Steric misfit.

The terms $a_{jV} \cdot x_{jVi}$ describes the steric misfit. It should be solved by variations of angles, especially the dihedrals (less rigid in comparison to angular vibrations or van der Waals compression). The phenomena should somehow resemble crystal melting (fusion). In the case of water, the energy change (*i.e.*, the heat of fusion) is 1.436 kcal mol⁻¹. The volume of a water molecule in cm³ is (Eq. 4a):

$$\frac{V_{\text{molar}}}{1 \text{ mol of molecules}} = \frac{18 \text{ cm}^3}{6 \cdot 10^{23}} = 3 \cdot 10^{-23} \text{ cm}^3 \quad (4a)$$

As the x_{jVi} values are expressed in Å³, the same volume in this unit is 30 Å³. Thus using Eq. 2e we obtain Eq. 4b, which in turn suggests the upper limit restriction for x_{jVi} (Eq. 4c):

$$\frac{1.436}{1.42} \cong a_{jV} \cdot 30 \quad (4b)$$

$$|a_{jV}| \leq 0.034 \quad (4c)$$

The detrimental steric misfit also imposes $a_{jV} \leq 0$, as the x_{jVi} values (volumes) are always greater than zero.

Polarizability effects.

Polarizability effects (stacking, charge transfer) are described by the $a_{jP} \cdot x_{jPi}$ terms. Supposing an energy variation of 10 ± 5 kcal mol⁻¹ and a mean number of 10 participating atoms yields an interaction energy for stacking and charge transfer of approximately 1 kcal mol⁻¹ per atom. Taking also a mean value of 5 cm³ for the polarizability of an atom (fragment), the following statement (Eq. 5a) is obvious:

$$\frac{1 \text{ kcal mol}^{-1} \text{ atom}^{-1}}{1.42 \text{ kcal mol}^{-1}} \cong a_{jP} \cdot 5 \text{ cm}^3 \text{ atom}^{-1} \quad (5a)$$

Therefore, we suggest the following restriction rule for the polarizability interactions (see also the description of the MTD-PLS methodology above, with regard to the beneficial character of this type of interaction):

$$0 \leq a_{jP} \leq 0.14 \quad (5b)$$

Hydrogen bond donor/acceptor capacity.

The terms $a_{jA} \cdot x_{jAi}$, $a_{jD} \cdot x_{jDi}$ reflect the contribution of hydrogen bond donor/acceptor interactions. The maximal value of the Gibbs free energy for a hydrogen

bond for different functional groups frequently encountered in drugs is about 6 kcal mol⁻¹ [25]. This was calculated from hydrogen bond equilibrium constants (in fact from their log values). Because the charged substituents (the most powerful category of proton donor and acceptor, respectively) were not included in the calculation, we have chosen the value of 10 kcal mol⁻¹. According to Raevsky [26], we have the following relationship (Eq. 6a):

$$\Delta G = 0.58(\text{kcal mol}^{-1}) \cdot C_a \cdot C_d \quad (6a)$$

where the experimentally determined parameters are: C_a , the H - bond acceptor free energy factor, and C_d , the H - bond donor free energy factor. The C_a and C_d values provided by Raevsky [27] range roughly between 0 and 5, but each presents an asymmetric distribution, so we have calculated a median value of 1.8 for C_a and 1.37 for C_d . The mean of these two values (~1.6) introduced in equation 6a gives Eq. 6b:

$$\frac{10 \cdot 0.58}{1.42} \cong a_{jA \text{ or } D} \cdot 1.6 \quad (6b)$$

This suggests the following upper bound restriction for the coefficients of the hydrogen bond donor/acceptor terms (Eq. 6c):

$$|a_{jA \text{ or } D}| \leq 2.55 \quad (6c)$$

Electrostatic interactions.

The contribution of electrostatic interactions is incorporated in the terms $a_{js} \cdot x_{jsi}$. The electrostatic contribution can be either attractive or repulsive. The energy of electrostatic interaction between two elementary (1e⁻) charges at a distance d of 1 Å is given by Eq. 7a:

$$U_{\text{elst}} = \frac{1}{4\pi\epsilon_0} \cdot \frac{e^2}{d} = \quad (7a)$$

$$= 9 \cdot 10^9 \cdot \frac{1.6 \cdot 10^{-19} \cdot 1.6 \cdot 10^{-19}}{10^{-10}} = 23 \cdot 10^{-19} \text{ J} = 5.5 \cdot 10^{-19} \text{ J}$$

This translates to approximately 330 kcal mol⁻¹ ($5.5 \cdot 10^{-19} - 6 \cdot 10^{-19}$) for 1 mol of electrons. If we suppose that the two interacting charges are about 0.5 e⁻, placed at a mean distance of $d = 3.5$ Å, Eq. 2e can be written as Eq. 7b:

$$\frac{330 \cdot 0.5 \cdot 0.5}{1.42 \cdot 3.5} \cong a_{js} \cdot 0.5 \quad (7b)$$

This leads to the following restriction rule (Eq. 7c):

$$|a_{js}| \leq 33.2 \quad (7c)$$

Although the new restriction rules have a strong approximate character, if their application in the MTD-PLS method leads to meaningful models, their utility is demonstrated.

2.2.2. Hypermolecule construction and structural parameter calculations

For all indirubin derivatives, the initial geometry optimizations were carried out with the molecular mechanics (MM) method using the MM+ force fields. The lowest energy conformations of the database compounds obtained with the MM method were further optimized using the semiempirical AM1 method implemented in the HyperChem 7.52 software package [28]. In the AM1 semiempirical calculations, a RMS gradient value of 0.01 kcal Å⁻¹ mol⁻¹ was employed, as a criterion for choosing an optimized conformation along with the Polak-Ribiere conjugate gradient algorithm [28]. In order to obtain a reliable spatial orientation of the indirubin substituents, a conformational search for the flexible side chains of the bis-indole rigid skeleton has been carried out with the conformational search module available in HyperChem 7.52. The lowest energy conformations were retained for each compound. Most 3D-QSAR methods require 3D alignment of molecules according to X-ray ligand coordinates, pharmacophore models or eventually docked pose. Hypermolecule (HM) construction was performed using as a template the X-ray coordinates of (Z)-1H,1'H-[2,3'] biindolyldiene-3,2'-dione-3-oxime (IXM) [5,29] (the ligand co-crystallized with GSK-3β). The lowest energy conformer for 109 indirubin derivatives [12-17] has been superimposed on the X-ray ligand structure. A RMS fit criterion defined for three atoms (C2, C10, C13) of the rigid bis-indole skeleton whose coordinates were extracted from the X-ray ligand structure was used (Fig. 2). The atoms that are not included in the common molecular skeleton are defined and numbered as distinct vertices in the hypermolecule, if they are located more than 0.7 Å apart. To further reduce the signal-to-noise ratio, we avoided introducing multiple single occupied vertices in the hypermolecule, by modifying slightly the position of several atoms. The energies of the structures obtained in this way were always situated in an energy gap of no more than 0.5 kcal mol⁻¹ above the corresponding lowest energy conformer.

Fragment descriptor calculation.

The semiempirical AM1 atomic partial charges were used as x_{jsi} values. In the case of vertices occupied by groups (-CH₃-NO₂, etc.), the algebraic sum of the charges was calculated. Fragmental van der Waals volumes (x_{jvi}) were evaluated using the procedure described

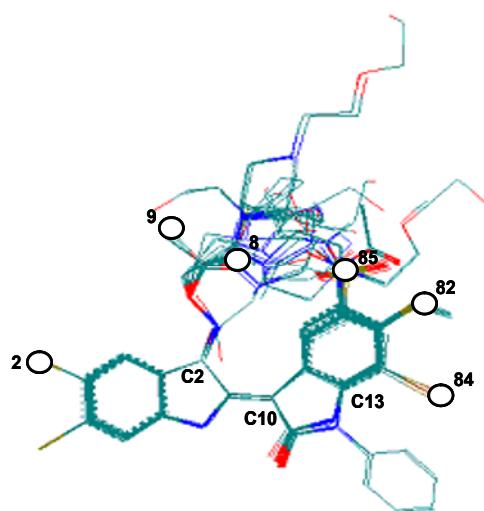


Figure 2. Hypermolecule construction and vertex numbering. The atoms belonging to the rigid skeleton which were used in the superposition step, and some significant vertices in the MTD-PLS final model are numbered.

by Olah [30]. Fragmental hydrophobicities (X_{jHi}) and fragmental polarizabilities (X_{jPi}) were calculated by the Marvin, Calculator Plugin and Chemical Terms modules of the Chemaxon software, online variant [20]. The H-bond parameters (H - bond acceptor enthalpy factor, X_{jAi} , and H - bond donor enthalpy factor, X_{jDi}) were computed with the CHED.MOLPRO demo version 2.1.0.706. (available at <http://www.timtec.net/software/demo/MolProDemo.zip>). Atomic volumes are given in Å³, hydrophobicities are provided in usual Hansch logP units, partial charges are given in fractional electronic charges, and polarizabilities P are given in cm³ mol⁻¹.

2.3. Fujita-Ban methodology

The Fujita-Ban variant of the Free-Wilson approach was applied to the full data set of 109 indirubin derivatives in order to obtain a 2D QSAR model, which directly relates structural features to biological activity in a straightforward manner. The Fujita-Ban equation expresses all the contributions relative to a hydrogen-substituted graph in the same position, as follows (Eq. 8):

$$\hat{Y}_i = Y_H + \sum_{l=1}^P b_{kl} X_{kli} \quad (8)$$

where k is a substitution position, l depicts a substituent different from hydrogen, Y_H is the biological activity calculated for the “parent” compound (all the k positions substituted with H); X_{kli} is equal to 1 when the substituent l is present in position k, and 0 otherwise; b_{kl} are the regression coefficients, or the activity enhancement factors. The positive sign of the b_{kl} coefficients indicates a beneficial effect of the independent variable while

the negative sign of these coefficients represents a detrimental influence.

2.4. Validation of QSAR models

The MTD-PLS model developed in the current work was validated by a Y-randomization test and four fold cross-validation alongside with the fulfillment of Golbraikh-Tropsha validation criteria, while the Fujita-Ban QSAR equation was tested with respect to Y-randomization. The Y-randomization test [31] is applied to demonstrate the statistical reliability of a QSAR model. The dependent variables (biological activities) are randomly permuted and a novel QSAR model is built based on the same independent variable matrix. If the QSAR models obtained in this manner tend to exhibit minimal R^2 and Q^2 values, the chance correlation bias is excluded. The exclusion of chance correlation of a QSAR model can be quantitatively assessed by the penalty parameter R_p^2 [32], which amends the squared correlation coefficient by subtracting the average squared correlation coefficient (R_r^2) of the randomized models from the squared correlation coefficient of the non-randomized model (R^2), as follows (Eq. 9):

$$R_p^2 = R^2 \cdot \sqrt{R^2 - R_r^2} \quad (9)$$

Roy *et al.* [32] considered that R_p^2 should take values greater than 0.5 for a statistically adequate model that excludes chance correlation. In order to apply the cross-validation procedure, a methodology for constructing training and test sets was set up as follows [33,34]:

- (i) All data set compounds were ordered in ascending order of the experimental activity values;
- (ii) Four subsets (denoted as 1, 2, 3 and 4) were constructed by selection: for every consecutive quartet of the ordered set, the first molecule was introduced in subset 1, the second in subset 2, etc...;
- (iii) The construction of the four training sets was carried out by combining any three subsets, while the fourth was designed as a test set (for which the activities were predicted using the model resulting from the corresponding training set): 1 + 2 + 3, 1 + 3 + 4, 1 + 2 + 4, and 2 + 3 + 4.

To provide confidence in the validation methodology, the Golbraikh-Tropsha criteria have been applied to all four assembled test sets, as described before. Moreover, to certify the predictive ability of the MTD-PLS model, the following conditions for the test set were considered:

- (i) cross validation correlation coefficient $Q^2 > 0.5$;
- (ii) the squared correlation coefficient R^2 between predicted and observed activities $R^2 > 0.6$;
- (iii) the coefficients of determination for predicted versus observed activities R_0^2 , and observed versus

predicted activities R_0^2 , both equations forced through the origin, have to comply with the following conditions:

$$\frac{R^2 - R_0^2}{R^2} < 0.1 \text{ or } \frac{R^2 - R_0^2}{R^2} < 0.1 \text{ and } |R_0^2 - R_0^2| < 0.3;$$

(iv) the slopes k and k' of the regression lines through the origin ($0.85 \leq k \leq 1.15$ or $0.85 \leq k' \leq 1.15$) [35-37].

2.5. Ligand and protein preparation

The conformational sampling of indirubin derivatives was carried out with the Omega 2.2 module from the OpenEye suite [38]. The crystal structure of GSK-3 β (PDB entry: 1Q41) in complex with (Z)-1H,1'H-[2,3']biindolylidene-3,2'-dione-3-oxime were downloaded from the RCSB Protein Data Bank (RCSB PDB) [29]. The enzyme was prepared for docking using the FRED RECEPTOR facility from the OpenEye package. FRED RECEPTOR generated an active site box of 7222 Å³, inner and outer contours of 63Å³ and 2009Å³, respectively. We preserved two crystallographic water molecules adjacent to THR138 that were demonstrated to be mediators for binding to the kinase. Docking constraints were introduced that force indirubin to make hydrogen bonds to ASP133 and VAL135 in order to comply with pharmacophoric conditions that were observed in the X-ray co-crystal. Ten alternant poses for each molecule were retained. The pose that displayed the best overlay to the X-ray ligand was designated as the correct pose.

3. Results and discussion

3.1. Fujita-Ban results

Statistically significant Fujita Ban models (M1) were obtained for all 109 compounds, where N represents the number of compounds, R^2 is the squared correlation coefficient, SEE denotes the standard error of estimates and F designates the Fisher's test (Eq. 10). The corresponding individual contributions of the substituents to the binding affinity are shown in Eq. 10.

$$\begin{aligned} \text{pIC}_{50} = & 6.956(\pm 0.551) - 0.210(\pm 0.055)R1_CH_3 - \\ & 0.250(\pm 0.047)R1_Phenyl + 0.115(\pm 0.056)R5_Cl \\ & + 0.185(\pm 0.054)R5_Br + 0.180(\pm 0.050)R5_NHAc - \\ & 0.184(\pm 0.052)R5_F - 0.225(\pm 0.055)R5'_Br \\ & + 0.224(\pm 0.092)R6_Br - 0.234(\pm 0.053)R6'_Br \\ & - 0.404(\pm 0.103)R7_Cl - 0.418(\pm 0.082)R7_Br - \\ & 0.237(\pm 0.055)R7_I \end{aligned} \quad (10)$$

N = 105; $R^2=0.923$; SEE = 0.506; F = 7.505

The substituents that were detected significant by the Fujita-Ban model (Eq. 10) are placed at positions 1, 5, 5', 6, 6', and 7. The physico-chemical properties

of the substituents that are essential for modulating the activity of indirubins are hydrogen bond acceptors (fluorine, chlorine, bromine and iodine at positions 5, 5', 6, 6', and 7), hydrogen bond donor-acceptor NH-Ac (position 5), and hydrophobic methyl or phenyl at position 1. A favorable effect is provided by high molecular weight halogens and the NH-Ac group at positions 5 and 6, while a small volume substituent (fluorine) is detrimental at position 5. Bromine at positions 5', 6', 7 is detrimental as well as chlorine and iodine at position 7, and methyl and phenyl groups at position 1. Four outliers have been detected in this model (the compounds 42, 78, 80, 83 listed in Table 1), for which standard residuals are higher than $\pm 2\sigma$ (standard deviation). Detailed interpretations of the substituent effects are provided in comparison with complementary methods in section 3.5. The validity of this model was confirmed by the Y randomization test (see Table 5).

3.2. MTD-PLS results

PLS calculations have been applied to MTD fragmental descriptors (444 independent variables in the X-matrix) and biological activities (Y-vector) for 109 compounds. The SIMCA P 09 package provided the best correlation equation that describes the relationship between the biological activities and the independent variables. In the PLS model, the variables that display a_{ju} coefficients that do not obey the restriction rules regarding the sign and upper limit were eliminated during the successive phases of the model construction. The model M2_1 (Table 2) where all restriction rules are fulfilled was obtained after gradually eliminating the outliers (1, 5, 7, 22, 51, 56 - 59, 60, 69, 71, 74, 76, 81, 82, 83, 100) in the intermediary models, whose standard residuals exceeded $\pm 2\sigma$ units. In order to eliminate the chance correlation from the PLS model, a new, final model was constructed (M2_2, and Table 2), retaining only the a_{ju} coefficients that are statistically different from zero. This model, although it possesses apparently weaker statistical characteristics, displays a degree of stability and also internal predictive power ($R^2_Y(\text{CUM})$ - $Q^2(\text{CUM})$ = 0.043) for a description of these terms, see Table 2).

In the analysis of the PLS models, we used the Variable Importance in the Projection (VIP) method and the sign of the coefficients (b_j) to assess the importance and the effect of the variable in the model. The VIP values are computed by the SIMCA software and reflect the importance of the descriptor in the PLS model with respect to Y, i.e., its correlation with the responses. The variables with higher VIP scores are more relevant for explaining the activity [39], and are associated with molecular fragments having higher significance. Among 23 significant descriptors, 16 exhibit $VIP > 0.5$, and their

Table 2. The statistical parameters of the MTD-PLS models obtained.*

Model	R ² X(CUM)	R ² Y(CUM)	Q ² (CUM)	A	K-1	N
M2_1 (MTD-PLS model)	0.121	0.805	0.606	2	245	91
M2_2(MTD-PLS model)	0.210	0.707	0.664	1	23	91

*R²_X(CUM) and R²_Y(CUM) are the cumulative sum of squares for X and Y values, explained by the extracted principal components; Q²(CUM) is the fraction of the total variation of Y that can be predicted by means of the same principal components; A – number of PLS significant components; K – the number of x variables; N – the number of compounds used in model construction.

Table 3. The statistical parameters obtained for the training sets used in the cross-validation of the MTD-PLS model.*

For M2_2	R ² Y(CUM)	Q ² (CUM)	A	N	K
M2_2_1 (subset 2+3+4)	0.763	0.699	1	68	24
M2_2_2 (subset 1+3+4)	0.685	0.618	1	68	24
M2_2_3 (subset 1+2+4)	0.764	0.705	2	68	24
M2_2_4 (subset 1+2+3)	0.708	0.662	1	69	24

* for notations see Table 2.

influence is as follows [j (fragmental property)]: beneficial 8(A, S), 9(S, P), 82(H, A, P, S), 85(H), and detrimental 2(S, A, V), 84(S, A, V), 85(D).

3.3. Model validation

Four training set models were constructed according to the selection described in section 2.4. Although the number of compounds (N) from these training set models differs from the whole set model (see Table 2), the statistical parameters, the number of principal components (A), and the number of significant x variables (K) are very similar (see Table 2 and Table 3). Also, the most important variables according to VIP are mainly the same as in the whole set model.

The Golbraikh-Tropsha criteria have been verified for each test set corresponding to the four training sets (see section 2.4). The calculated parameters respect the imposed criteria, demonstrating statistical robustness (see Table 4).

The Y-randomization test was performed for the MTD-PLS and Fujita-Ban QSAR models. The models obtained for each randomization round exhibit significantly lower R²_Y(CUM)_r, Q²(CUM)_r, and R²_r values than the non-randomized models (see Table 5). Also, the R²_p values are above the 0.5 limit. This means that the chance correlation bias is not present in our models. Therefore, our models are considered statistically reliable and validate the methodologies used and the dataset constitution.

3.4. Molecular docking results

Molecular docking of indirubins into the ATP binding site of GSK-3β establishes the ligand positions and

the GSK-3β - indirubin interactions responsible for biological activity. The best pose was designated on the basis of consensus scoring, and these orientations were considered to evaluate the binding position and interactions with the GSK-3β binding site. Besides the three hydrogen bonds with ASP133 and VAL135 (two hydrogen bonds), the current docking investigation highlights several significant interactions of indirubin derivatives with ILE62, VAL70, LYS85, VAL110, LEU132, ARG141, CYS199, ASP200. All these amino acids have been confirmed by X-ray diffraction investigation to interact with GSK-3β ATP-mimetic inhibitors [5]. The comparison of MTD-PLS, Fujita-Ban and docking outcomes is discussed in terms of substitution patterns and interactions with the protein in section 3.5.

3.5. Comparison of MTD-PLS outcomes with X-ray crystallography, molecular docking and previous investigations

The information provided by the MTD-PLS model was interpreted by taking into account the signs and the values of the a_{jm} coefficients. A discussion of the current MTD-PLS model outcomes in comparison with experimental X-ray co-crystal structure of indirubin-GSK-3β, docking results and preceding CoMFA and CoMSIA [40] results is presented here. First, we have analyzed the contacts of binding site atoms and indirubin in the X-ray structure. We have subsequently correlated the interactions observed in the X-ray co-crystal with the information provided by MTD-PLS analysis regarding the vertices of the hypermolecule with relevant statistical significance. To provide a direct comparison between MTD-PLS outcomes and the three dimensional interaction pattern, the compounds whose atoms occupy the vertices which provided significant information in the MTD-PLS model have been docked into the ATP-binding site of GSK-3β.

Vertex 2 (R5') is occupied only by bromine in eleven molecules. To explain the detrimental effect of bromine concerning the volume, negative charge and hydrogen bonding acceptor capacity as shown by the MTD-PLS analysis, we have docked all the molecules that possess bromine in vertex 2. This molecule forms hydrogen bonds with VAL135 and ASP133. It was observed also that the Br atom is involved in relatively close sterical contacts

Table 4. The Golbraikh-Tropsha validation of the MTD-PLS model.*

For M2_2	R^2	R_o^2	$R_o'^2$	k	k'	$(R^2 - R_o^2) / R^2$	$ R^2 - R_o'^2 $
M2_2_1 (subset 2+3+4)	0.681	0.661	0.576	1.030	0.965	0.029	0.105
M2_2_2 (subset 1+3+4)	0.751	0.732	0.500	1.010	0.988	0.025	0.251
M2_2_3 (subset 1+2+4)	0.684	0.682	0.573	0.999	0.995	0.003	0.111
M2_2_4 (subset 1+2+3)	0.729	0.728	0.575	1.020	0.981	0.001	0.154

* R^2 - the correlation coefficient of predicted versus observed activities for the test set; R_o^2 -the coefficient of determination for predicted versus observed activities; $R_o'^2$ -the coefficient of determination for observed versus predicted activities for the regression line passing through the origin; k, k'- the slopes of the regression lines passing through the origin.

Table 5. The statistical parameters for the Y-randomized models: $R_y^2(CUM)_r$ and $Q^2(CUM)_r$ for the MTD-PLS model; and the p significance level for the Fujita-Ban model.

Random Model	MTD-PLS model			Fujita-Ban model	
	A	$R_y^2(CUM)_r$	$Q^2(CUM)_r$	R_r^2	p
1	1	0.056	-0.100	0.504	0.947
2	1	0.055	-0.100	0.451	0.991
3	1	0.087	-0.045	0.561	0.790
4	1	0.041	-0.098	0.449	0.992
5	1	0.043	-0.100	0.496	0.959
6	1	0.048	-0.100	0.553	0.823
7	1	0.174	-0.018	0.555	0.813
8	1	0.105	-0.026	0.575	0.727
9	1	0.133	0.045	0.401	0.999
10	1	0.111	-0.068	0.571	0.746
R_p^2		0.558		0.594	

with ILE62 and ARG141. Moreover, the negative partial charge of bromine interacts unfavorably (repulsion) with the ILE62 oxygen (O=). This is evidenced by the highest VIP value (1.236) of the charge descriptor coefficients at this vertex. Most probably this sterical constraint and repulsive contact determine a limited possibility of movement for the ligands possessing bromine at this position. The same detrimental effect of bromine at this position is suggested also by the Fujita-Ban QSAR model.

Vertex 8 is occupied by an oxygen atom in ten molecules. Concerning the negative charge and the hydrogen bond acceptor capacity, MTD-PLS indicates an augmenting effect of the activity in the presence of a negatively charged atom. The rigid docking procedure shows that the presence of the oxygen atom in close proximity to LYS85 is beneficial from the electrostatic point of view (lysine bears a positive charge at the nitrogen atom in the NH_3^+ moiety). Additionally, the hydrogen bond acceptor character is also favorable, with an optimal distance between the H-bond donor and acceptor (C=O).

Vertex 9 is occupied by a CH_3 group in ten molecules. The MTD-PLS model suggests favorable positive charge and dispersion interactions for this vertex. Indeed the docking results suggest that around the analyzed ligand methyl group, there are some adjacent receptor atoms situated at van der Waals contact distance which are favorable for these types of interactions. These interactions are: the non-polar VAL70 methyl group, and an advantageous electrostatic interaction with ASP200. The above mentioned findings are in agreement with a previous CoMFA study which underlines the beneficial effect of carboxyl group in this region [40].

Vertex 82 in the MTD-PLS corresponds to substitution position R6. There are 38 compounds with bromine at this position, six compounds with chlorine, three compounds with iodine, three with fluorine and three with a CH=group. This position is situated in a hydrophobic environment (VAL110 and LEU132), thus hydrophobicity is favored in this area, as shown also by a previous CoMFA and CoMSIA study [40]. This fact is substantiated by the MTD-PLS analysis. We have found for this vertex beneficial polarizability and hydrogen

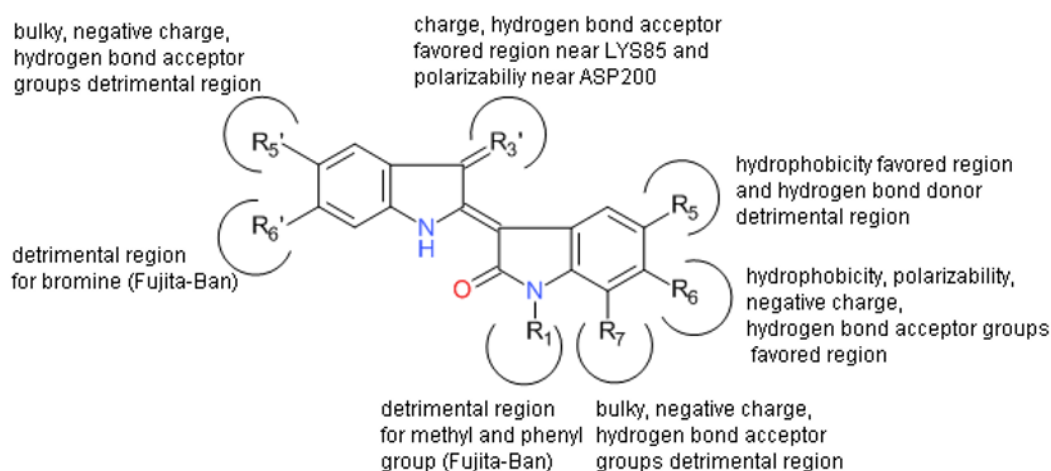


Figure 3. Structure–activity relationships revealed by QSAR and molecular docking.

bond acceptor interactions. The beneficial effect of hydrophobic interactions for this vertex is supported by the high VIP value (1.744) of the corresponding coefficient. Our docking results confirm the MTD-PLS results. The potential hydrophobic and van der Waals interactions with LYS85, LEU132, and VAL110 validate hydrophobicity and polarizability as favorable interactions at this vertex. The beneficial hydrogen bond acceptor ability resulting from the MTD-PLS model might be explained by the presence of ASP200 nearby. These results are in good agreement with the positive values for the coefficients A82 and P82, and the negative value for S82. The favorable effect of large volume substituents at this position is also highlighted by the Fujita-Ban analysis.

Vertex 84 (R7). In our series, this vertex is occupied by halogen atoms (three molecules display fluorine, one molecule chlorine, nine molecules bromine, and two molecules iodine). The coefficients of the negative charge, hydrogen bonding acceptor capacity and volume resulting from the MTD-PLS analysis suggest that the presence of the halogens at this vertex is detrimental. Distances smaller than the sum of van der Waals radii with respect to LEU132 seem to induce the steric hindrance that is reflected by the volume coefficient. Furthermore, some compounds containing bulky substituents such as chlorine (compound 56), bromine (compounds 57 and 60) and iodine (compounds 58 and 59) were eliminated as outliers in the MTD PLS model and show low biological activity (Table 1). Hydrogen bonding acceptor capacity is not influential because the donor groups are missing in this area. These two parameters display high VIP values ($VIP_V=1.547$, $VIP_A=1.549$), indicating the same importance of both variables with respect to binding to the receptor. The low experimental activities (pIC_{50} below 5.13) of compounds containing halogens

at vertex 84 also suggest corroborate the statements above. The same conclusion results from the Fujita-Ban model.

Vertex 85 corresponds to substitution position R5 on the indirubin skeleton. There are thirty two molecules that have substituents at this position: fluorine, chlorine, bromine, iodine, $-NO_2$ groups, $-NH_2$, $-NH$, $-CH_3$ are among them. Previous CoMFA and CoMSIA investigations [40] suggest that this position is situated in close proximity to a hydrophobic zone near VAL70 and CYS199. The positive coefficient of H85 ($VIP=0.810$) provided by the MTD-PLS equation points out favorable hydrophobic interactions in agreement with the mentioned study. The arguments are obtained from docking experiments: optimal distances to VAL70 were displayed by the majority of the substituents.

Vertices 1 and 3 correspond to positions R6'/R1 of the indirubin ring and do not offer any information in MTD-PLS because in this region our series does not possess substitutional variance (bromine, methyl, phenyl). Some information about vertex 1 is provided by CoMFA that suggests a beneficial effect of hydrophilic groups [40]. On the contrary, our Fujita-Ban analysis suggests that bromine atoms at position R6' and a methyl or phenyl substituent at R1 is detrimental.

We have demonstrated that the key interactions predicted by MTD-PLS are in agreement with those inferred from the X-ray co-crystal. The information extracted during PLS analysis (Fig. 3) can be at best as accurate as the information introduced into the model via experimental activities and structural parameters related to ligand geometry. To summarize, our MTD-PLS model provides additional information with respect to previous CoMFA and COMSIA investigations as follows: at the substitution position R6, the acceptor capacity (putative H-bond between the $-NH$ group (ASP200)

and the halogens) and charge has a beneficial effect; at position R7, the acceptor capacity, negative charge and volume are detrimental; at position R5' the negative charge, acceptor capacity and volume are detrimental; at position R5 the donor capacity is detrimental.

The new restriction rules introduced in this paper display applicability and provide reliable results in accordance with physico-chemical reality. The weaker statistical performances of the MTD-PLS model were surmounted by "lateral" validation using docking experiments. Taking into account the information provided by the MTD-PLS method we can refine the picture that represents the substitution pattern of indirubins, and explain the biological activity variance within this series.

4. Conclusions

The comparison of the MTD-PLS analysis with crystallographic X-ray data yields valuable information regarding the rigidity and deformability of certain binding site regions, which are not available directly from crystallographic investigation. There are situations when the interpretations of the QSAR coefficients from the resulting models are not in agreement with experimental data. In these cases a detailed investigation of the phenomena is required, but in some cases it is possible to offer a rational explanation. The information obtained through MTD-PLS analysis suggests that principally the MTD-PLS works well, but the final result depends on the quality of the input data, *e.g.* the number of data points, experimental errors, *etc.* If the input data are not sufficient or contradictory, the PLS model cannot provide stable or correct outcomes. It is possible to obtain concordant results between QSAR models and experimental data by investigation of statistically robust models, or on the contrary using poor quality models, displaying lower correlation coefficients but containing a large variety of substituents at certain positions. Valuable information concerning ligand-receptor interaction can be obtained by applying the restrictive conditions of the

classical MTD-PLS method regarding the detrimental effect of the volume parameter, the favorable effect of hydrophobic and polarizability descriptors [18], as well as the new restrictive conditions originating from physico-chemical considerations, introduced herein. The CoMFA results on a series of 42 indirubins [40] of which 39 are included in our dataset and investigated in the current study are confirmed by MTD-PLS analysis. The MTD-PLS outcomes bring consistent additional information about the following substitution positions: R6 where the acceptor capacity and negative charge had beneficial effects, R7 wherein the acceptor capacity and volume has detrimental effects, R5' at which the negative charge, acceptor capacity and volume are detrimental, and R5 where the donor capacity is detrimental.

The new restriction rules introduced in this paper are justified by the current investigation, where the information introduced is based on physico-chemical reality, and not only on statistical calculation. The QSAR models have been validated laterally by docking experiments that confirmed the correct prediction.

The MTD-PLS method provided a handful of information that helped to complete the picture of the indirubin substitution pattern, to explain biological activity and to design indirubins with favorable pharmacology. The results of the MTD-PLS method have been largely confirmed by the interactions between indirubin derivatives and the active site of the protein observed by docking and by Fujita-Ban analysis.

Acknowledgements

We thank to Dr. Erik Johansson (Umetrics, Sweden) for kindly providing the SIMCA program package, to Dr. Mircea Mracec for access to the HYPERCHEM package, to Dr. Simona Funar-Timofei for access to the STATISTICA software, and OpenEye Scientific Software, Inc., for providing us an academic license (Simona Funar-Timofei laboratory). This project was financially supported by Project No. 1.2 of the Institute of Chemistry of the Romanian Academy, Timisoara.

References

- [1] N. Embi, D.B. Rylatt, P. Cohen, *Eur. J. Biochem.* 107, 519 (1980)
- [2] A. Vulpetti, P. Crivori, A. Cameron, J. Bertrand, M.G. Brasca, R. D'Alessio, P. Pevarello, *J. Chem. Inf. Model.* 45, 1282 (2005)
- [3] A. Ali, K.P. Hoeflich, J.R. Woodgett, *Chem. Rev.* 101, 2527 (2001)
- [4] H.C. Zhang, H. Ye, B.R. Conway, C.K. Derian, M.F. Addo, G.H. Kuo, L.R. Hecker, D.R. Croll, J. Li, L. Westover, J.Z. Xu, R. Look, K.T. Demarest, P. Andrade-Gordon, B.P. Damiano, B.E. Maryanoff, *Bioor. Med. Chem. Lett.* 14, 3245 (2004)
- [5] J.A. Bertrand, S. Thieffine, A. Vulpetti, C. Cristiani, B. Valsasina, S. Knapp, H.M. Kalisz, M. Flocco, *J. Mol. Biol.* 333, 393 (2003)
- [6] Y. Mettey, M. Gompel, V. Thomas, M. Garnier,

- M. Leost, I. Ceballos-Picot, M. Noble, J. Endicott, J.M. Vierfond, L. Meijer, *J. Med. Chem.* 46, 222 (2003)
- [7] Y. Wan, W. Hur, C.Y. Cho, Y. Liu, F.J. Adrian, O. Lozach, S. Bach, T. Mayer, D. Fabbro, L. Meijer, N.S. Gray, *Chem. Biol.* 11, 247 (2004)
- [8] E. Lescot, R. Bureau, J. Sopkova-de Oliveira Santos, C. Rochais, V. Lisowski, J.C. Lancelot, S. Rault, *J. Chem. Inf. Model.* 45, 708 (2005)
- [9] E.J. Henriksen, T.R. Kinnick, M.K. Teachey, M.P. O'Keefe, D. Ring, K.W. Johnson, S.D. Harrison, *Am. J. Physiol. Endocrinol. Metab.* 284, E892 (2003)
- [10] Z. Xiao, Y. Hao, B. Liu, L. Qian, *Leuk. Lymphoma* 43, 1763 (2002)
- [11] Z.H. Wang, Y. Dong, T. Wang, M.H. Shang, W.Y. Hua, Q.Z. Yao, *Chin. Chem. Lett.* 21, 297 (2010)
- [12] S. Leclerc, M. Garnier, R. Hoessel, D. Marko, J.A. Bibb, G.L. Snyder, P. Greengard, J. Biernat, Y.Z. Wu, E.M. Mandelkow, G. Eisenbrand, L. Meijer, *J. Biol. Chem.* 276, 251 (2001)
- [13] J.A. Martinez, A. Castro, I. Dorronsoro, M. Alonso, *Med. Res. Rev.* 22, 373 (2002)
- [14] P. Polychronopoulos, P. Magiatis, A.L. Skaltsounis, V. Myrianthopoulos, E. Mikros, A. Tarricone, A. Musacchio, S.M. Roe, L. Pearl, M. Leost, P. Greengard, L. Meijer, *Med. Chem.* 47, 935 (2004)
- [15] A. Beauchard, Y. Ferandin, S. Fre'ere, O. Lozach, M. Blairvacq, L. Meijer, V. Thie'ry, T. Besson, *Bioorg. Med. Chem.* 14, 6434 (2006)
- [16] Y. Ferandin, K. Bettayeb, M. Kritsanida, O. Lozach, P. Polychronopoulos, P. Magiatis, A.L. Skaltsounis, L. Meijer, *J. Med. Chem.* 49, 4638 (2006)
- [17] K. Vougiogiannopoulou, Y. Ferandin, K. Bettayeb, V. Myrianthopoulos, O. Lozach, Y. Fan, C.H. Johnson, P. Magiatis, A.L. Skaltsounis, E. Mikros, L. Meijer, *J. Med. Chem.* 51, 6421 (2008)
- [18] L. Kurunczi, M. Olah, T.I. Oprea, C. Bologa, Z. Simon, *J. Chem. Inf. Comput. Sci.* 42, 841 (2002)
- [19] L. Kurunczi, E. Seclaman, T.I. Oprea, L. Crisan, Z. Simon, *J. Chem. Inf. Model.* 45, 1275 (2005)
- [20] Marvin version 5.4.1.1. (ChemAxon, Budapest, Hungary, 2011)
- [21] S.K. Dogra, A. Das, K. Subramanian, In: G. Cruciani (Ed.), *The 233th ACS National Meeting*, Chicago, IL, March 25-29, 2007, General Papers, Division of Chemical Information, http://www.qsarworld.com/files/ACS_3D_Descriptors_Conformers.pdf
- [22] M. Pastor, *Alignment-independent Descriptors from Molecular Interaction Fields*, in *Molecular Interaction Fields. Applications in Drug Discovery and ADME Prediction* (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2006) 117-142
- [23] T.I. Oprea, L. Kurunczi, M. Olah, Z. Simon, *SAR QSAR Environ. Res.* 12, 75 (2001)
- [24] SIMCAP, version 9.0 (Umetrics AB, Umea, Sweden, 2001)
- [25] M.H. Abraham, P.P. Duce, D.V. Prior, D.G. Barratt, G.G. Morris, P.J. Taylor, *J. Chem. Soc. Perkin Trans. II* 1355 (1989)
- [26] O.A. Raevsky, In: R. Mannhold (Ed.), *H - bonding Parameterization in Quantitative Structure - Activity Relationships and Drug Design*, in *Molecular Drug Properties. Measurement and Prediction* (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2008) 127-154
- [27] O.A. Raevsky, V.Ju. Grigor'ev, D. Kireev, N.S. Zefirov, *Quant. Struct. Act. Relat.* 11, 49 (1992)
- [28] Hyperchem 7.52 release for Windows (HyperCube, Inc., Gainesville, Florida, USA, 2005) <http://www.hyper.com>
- [29] F.C. Bernstein, T.F. Koetzle, G.J. Williams, E.E. Meyer Jr., M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, *J. Mol. Biol.* 112, 535 (1977)
- [30] M. Olah, *Rev Roum. Chim.* 45, 1123 (2000)
- [31] F. Lindgren, B. Hansen, W. Karcher, M. Sjöström, L. Eriksson, *J. Chemom.* 10, 521 (1996)
- [32] P.P. Roy, S. Paul, I. Mitra, K. Roy, *Molecules* 14, 1660 (2009)
- [33] A.R. Katritzky, S. Slavov, D. Dobchev, M. Karelson, *Comp. Chem. Eng.* 31, 1123 (2007)
- [34] A.R. Katritzky, L.M. Pacureanu, D. Dobchev, D.C. Fara, P.R. Duchowicz, M. Karelson, *Bioorg. Med. Chem.* 14, 4987 (2006)
- [35] A. Golbraikh, A. Tropsha, *J. Comp. Aid. Molecul. Des.* 16, 357 (2002)
- [36] A. Golbraikh, A. Tropsha, *J. Comp. Aid. Mol. Des.* 120, 269 (2002)
- [37] A. Golbraikh, M. Shen, Z. Xiao, K.H. Lee, A. Tropsha, *J. Comp. Aid. Mol. Des.* 17, 241 (2003)
- [38] OpenEye Scientific Software (OpenEye Inc., Santa Fe, USA, 2009)
- [39] L. Eriksson, E. Johansson, N. Kettaneh-Wold, S. Wold, *Multi- and MegaVariate Data Analysis. Principles and Applications* (Umetrics AB: Umeå, Sweden, 2001) 94-97
- [40] N. Zhang, Y. Jiang, J. Zou, B. Zhang, H. Jin, Y. Wang, Q. Yu, *Eur. J. Med. Chem.* 41, 373 (2006)