Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Molecular Graphics and Modelling 24 (2006) 227–232
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

A chemometric study of phosphodiesterase 5 inhibitors

Luciana Jansen de O. Figueiredo [a,b,*], Francisco M.S. Garrido [b], Viviane Yumy M. Kunisawa [a], Carlos Eduardo Aboim [a], Otto Banho Licks [a]

[a] *Momsen, Leonardos & Cia, Rua Teófilo Otoni 63/10th floor, Rio de Janeiro 200090-080, Brazil*
[b] *Departamento de Química Inorgânica, Instituto de Química, Caixa Postal 68563, CT bloco A, Universidade Federal do Rio de Janeiro, Rio de Janeiro 21945-970, Brazil*

Received 1 February 2005; received in revised form 22 August 2005; accepted 22 August 2005
Available online 26 September 2005

Abstract

This work presents a chemometric classification for a set of phosphodiesterase 5 inhibitors, based on a pattern recognition method widely used in quantitative structure–activity (QSAR) studies, hierarchical cluster analysis (HCA) and principal component analysis (PCA), aiming to access the most relevant structural and physicochemical variables related to phosphodiesterase 5 inhibition and to quantify the similarity of the structures within the set of inhibitors. Our model is capable of classifying a test set of 26 known phosphodiesterase 5 inhibitors in terms of similarity, the results being consistent with published experimental data.

Keywords: Phosphodiesterase 5 inhibitor; Chemometric; HCA

1. Introduction

Cyclic nucleotide phosphodiesterases (PDEs) are enzymes that catalyze the hydrolysis of cyclic nucleotides cGMP and cAMP into their respective 5′ monophosphates. There are now 11 phosphodiesterase families, many of which exist as splice variants. The cAMP-specific enzymes include phosphodiesterase 4 (PDE4), -7 and -8. The cGMP-specific PDEs are PDE5, -6 and -9, whereas PDE1, -2, -3, -10 and -11 use both cyclic nucleotides [1].

Inhibition of cyclic nucleotide phosphodiesterase type 5 (PDE5) as a therapeutic target has received considerable attention in recent years, particularly for the treatment of cardiovascular diseases, e.g., angina, hypertension and congestive heart failure as well as erectile dysfunction [2–5]. The enzyme is active as a monodimer, which has a molecular mass of approximately 200 kDa. Each monomer contains a carboxy terminal catalytic domain, a highly conserved zinc binding motif, two allosteric binding pockets for cGMP and a phosphorilation site in the amino-terminal region [1].

Most known PDE5 inhibitors compete with the substrate cGMP for binding the protein at the catalytic site. An important issue in the development of PDE5 inhibitors is specificity for the other PDEs. In the case of PDE5 inhibitors, which have been described in the literature, improvements in selectivity were determined empirically and compounds were optimized on the basis of structure–activity explorations of selected chemical series [1–5].

Quantitative structure–activity relationships (QSAR) correlate biological activities of candidate compounds with their physicochemical parameters. In the present work, relationships between chemical structure and biological activity of cGMP PDE5 inhibitors (Fig. 1) were accessed using principal component analysis (PCA) and hierarchical cluster analysis (HCA), both conventional methods in pattern recognition [6–12].

Biological problems have an intrinsic multivariate nature, involving many variables at the same time and, in general, the relations between these variables and the biological response is not easily accessed, thus little useful information can be easily extracted. In order to simplify the data set in a multivariate problem and to obtain an informative picture of the data tendencies, a chemometric multivariate analysis can be used.

The easier mathematical way to represent a multivariate problem is to build a matrix relating variables and objects, as

* Corresponding author. Tel.: +55 21 25182264; fax: +55 21 25183152.
E-mail address: LJOFigueiredo@leonardos.com.br (L.J. de O. Figueiredo).

1093-3263/$ – see front matter
doi:10.1016/j.jmgm.2005.08.007

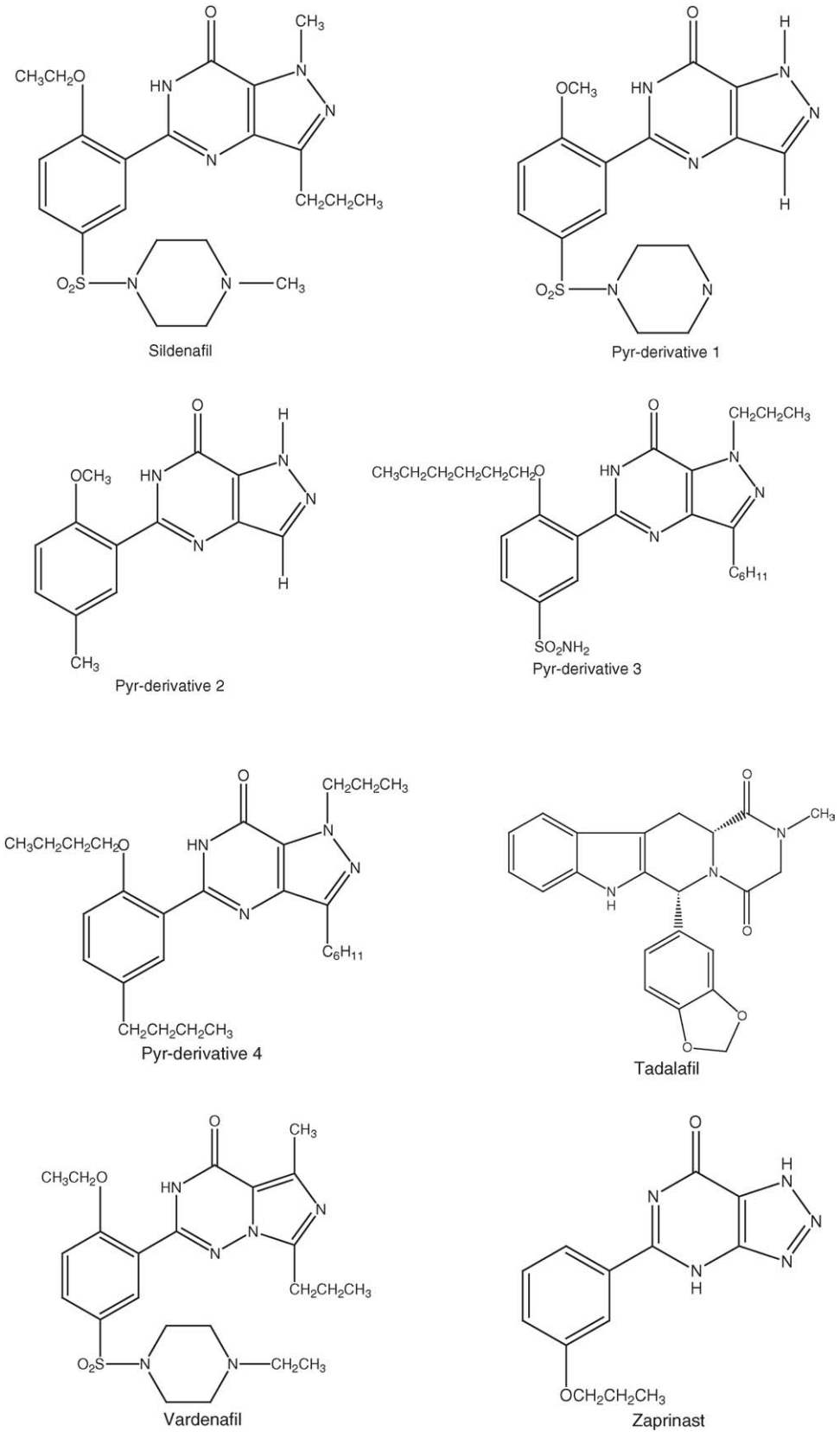

Fig. 1. Chemical structure of the selected cGMP PDE5 inhibitors. (*) For Der1-us: R^1 = pyrrolyl, $R^4 = SO_2N(Et)_2$; Der2-us: R^1 = pyrazolyl, $R^4 = SO_2N(Et)_2$; Der3-us: R^1 = piperidinyl, $R^4 = SO_2N(Et)_2$; Der4-us: R^1 = pyridinyl, $R^4 = SO_2N(Et)_2$; Der5-us: R^1 = pyridinyl, $R^4 = SO_2NH_2$; Der6-us: R^1 = pyrrolyl, $R^4 = SO_2NH_2$; Der7-us: R^1 = pyrazolyl, $R^4 = SO_2NH_2$; Der8-us: R^1 = piperidinyl, $R^4 = SO_2NH_2$. (**) For Der9-us: R^1 = pyrrolyl, $R^4 = SO_2N(Et)_2$; Der10-us: R^1 = pyrazolyl, $R^4 = SO_2N(Et)_2$; Der11-us: R^1 = piperidinyl, $R^4 = SO_2N(Et)_2$; Der12-us: R^1 = pyridinyl, $R^4 = SO_2N(Et)_2$; Der13-us: R^1 = pyridinyl, $R^4 = SO_2NH_2$; Der14-us: R^1 = pyrrolyl, $R^4 = SO_2NH_2$; Der15-us: R^1 = pyrazolyl, $R^4 = SO_2NH_2$.

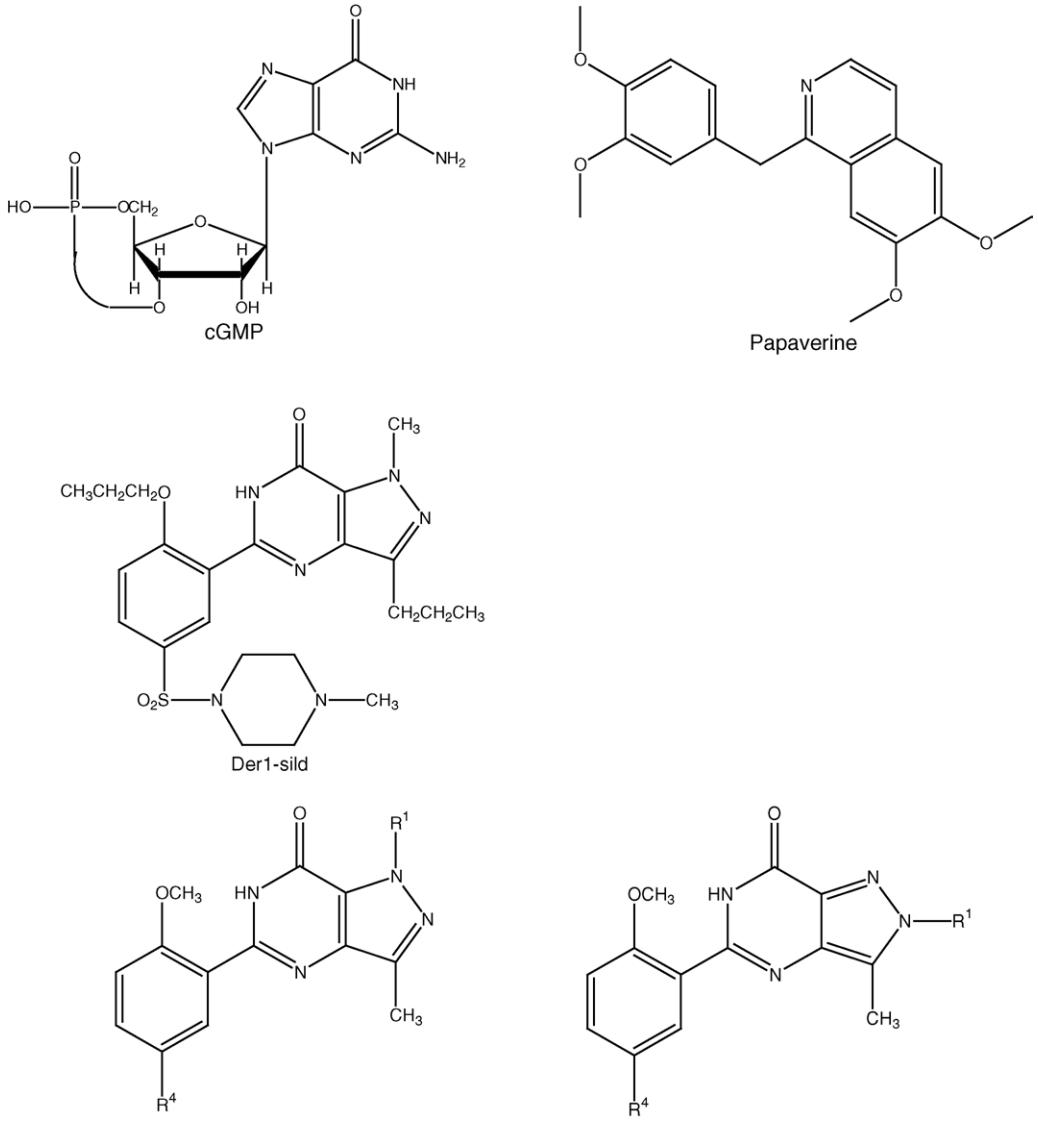

Fig. 1. (*Continued*).

can be seen in Fig. 2. In our particular case, the objects are the selected cGMP PDE5 inhibitors while the biological response is the inhibition of the enzyme (e.g., in terms of K_i) and the interaction between the inhibitors and the active site of the enzyme.

According to this methodology, each object is placed in a *n*-dimensional space (where *n* is the number of variables). However, it is more practical to work in two or three dimensions.

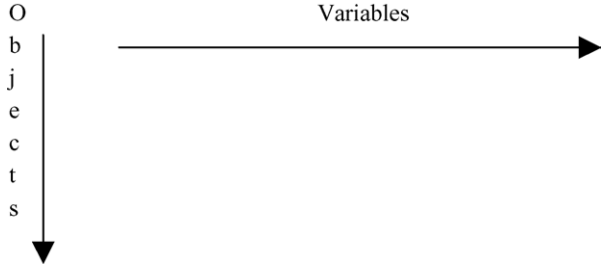

Fig. 2. Pictorial matrix correlating objects and variables.

The principal component method [10] allows the projection of higher order space in two or three dimensions with a minimal loss of statistical information. The coordinate axes of the original *n*-order space are rotated until the direction of maximum variance is coincident with one of the rotated axes (called the first principal component axis). The second principal component axis, and so on, gives the orthogonal direction of the maximum residual variance [11]. Another interesting aspect in a chemometric multivariate analysis is the possibility of separation of the objects into groups, as done with the score plot for PCA and with the HCA methods. PCA group separation is a method based on similarities perceived from approximate PCA projections of real distances whereas HCA actually utilizes exact distances in multivariate space to cluster objects based on their locations [12].

2. Material and methods

The structures of the selected cGMP PDE inhibitors were modeled using the semi-empirical MOPAC v7.0 package [13], in

Table 1
Set of cGMP PDE5 inhibitors

Inhibitor		Variables				
		ΔH	Surface Area	Volume	$E_{Hydr.}$	log P
1	Sildenafil	−6284.0679	726.17	1305.25	−2.22	−1.94
2	Pyr-derivative 1	−4662.8779	588.30	1001.58	−14.44	−4.54
3	Pyr-derivative 2	−3378.2090	461.18	741.63	−9.62	−2.10
4	Pyr-derivative 3	−7335.0923	857.72	1492.91	−5.35	1.45
5	Pyr-derivative 4	−7478.0996	836.19	1455.54	1.57	3.86
6	Tadalafil	−5470.7031	597.23	1038.62	−7.91	−0.17
7	Vardenafil	−6609.7607	763.62	1355.47	−3.53	−1.84
8	Zaprinast	−3511.5190	486.42	784.08	−15.89	1.12
9	CGMP	−3549.3386	491.40	790.95	−24.34	−0.67
10	Papaverine	−4936.5957	586.37	998.87	−7.46	−2.82
11	Der1-sild	−6081.7612	747.56	1288.81	−10.85	0.22
12	Der1-us	−6344.2817	773.81	1337.95	−5.07	−1.47
13	Der2-us	−6214.7583	768.33	1327.19	−5.78	−1.22
14	Der3-us	−6400.6631	778.76	1355.17	−6.76	−1.18
15	Der4-us	−6528.2588	781.06	1359.18	−6.09	−0.74
16	Der5-us	−5426.4219	694.75	1170.20	−11.56	−1.55
17	Der6-us	−5241.8101	680.18	1148.30	−10.54	−2.28
18	Der7-us	−5121.6821	676.55	1137.10	−11.47	−2.32
19	Der8-us	−6222.7231	772.74	1330.80	−5.61	−1.51
20	Der9-us	−6354.8550	760.19	1329.79	−8.27	−0.95
21	Der10-us	−6233.0208	754.65	1314.76	−9.96	−1.52
22	Der11-us	−6528.2227	769.76	1355.43	−6.14	−0.42
23	Der12-us	−6408.7583	767.76	1349.88	−6.74	−0.33
24	Der13-us	−5306.6548	680.68	1160.51	−12.30	−1.15
25	Der14-us	−5132.7383	668.59	1129.09	−14.64	−1.80
26	Der15-us	−5252.8442	668.67	1138.38	−13.75	−1.76

ΔH is expressed in kcal mol^{-1}, surface areas are in square angstroms while molecular volumes are in cubic angstroms.

order to obtain the physicochemical variables for multivariate analysis, starting with geometry optimization and then collecting the structural and physicochemical parameters. The starting geometries were obtained in a conformational search initial step, wherein the molecules were edited and optimized with molecular mechanics and molecular dynamics calculations, as implemented by TINKER [14] package. The force field used was AMBER. These results are summarized in Table 1.

The data set showed in Table 1 was converted in a matrix of 26 inhibitors (called objects) and 5 variables (working matrix with 26 lines and 5 columns).

The chemometric analysis was conducted using the ARTHUR/UNICAMP, package [11]. The programs used were ENTER for the edition of the training set, SCALE for the autoscaling, KARLOV for principal component analysis, CORREL for the calculation of the correlation matrix (correlation between variables), DISTAN for the calculation of the distance matrix and HIER for the cluster analysis. These calculations were performed on a personal computer.

Our selection of variables was made considering chemical parameters of interest. The variables surface area and molecular volume were chosen to classify inhibitors in terms of their ability to achieve different multiple stable interactions, that increases with the size of the inhibitor, in addition, surface area can be directly related to molecular topology. Hydration energy ($E_{Hydr.}$) is an energetic variable that classifies the set of inhibitors in terms of their water affinity, this being related to the electron flow expected in the inhibition mechanism involved herein. Another energetic variable is (ΔH) that classifies the set of inhibitors in terms of relative thermodynamic stability, while the partition coefficient (log P) is a variable strictly related to hydrophobic/hydrophilic profile of the inhibitors and widely used in QSAR studies [15].

All inhibitors were minimized with MOPAC v7.0 [13] package with the AM1 Hamiltonian; a method that is widely used in molecular modeling and gives good results for organic molecules like the inhibitors of our training set. Some key-words were included in order to obtain a suitable calculation: MMOK, is a key-word that incorporates molecular mechanics parameters to the calculation and is desirable for systems that contain N–H and/or O–H conjugated to carbonyls; in addition, the key-words PRECISE and GNORM = 0.05 were employed to ensure the choice of the most stable conformers.

The chemometric multivariate analysis (PCA and HCA) was applied to *autoscaled* data. Since we are working with variables with different orders of magnitudes and different units autoscaling is specifically recommended since the transformed data are adimensional and all variables have the same variance.

Docking simulations were also applied to simulate the interactions of the set of inhibitors with the crystallographic structure of the cGMP PDE5 enzyme (protein data bank deposited coordinate 1RKP.ent). This methodology fills a binding site cavity with spheres locally complementary to the macromolecular surface. The ''sphere'' program creates a space-filling image that complements the geometric characteristics of the receptor site, and these spheres are used to find matches with potential ligands. Other implementations address different scoring functions, such as molecular force fields, chemical or contact scores.

The DOCK v4.0 [16] program was used to perform the docking simulations, using GRID and DOCK programs, default parameter values in energy scoring and in intramolecular scoring were chosen and selected van der Waals parameters were set.

3. Results and discussion

The results of PCA were quite interesting; we decided to work with a model based on the two first principal components, responsible for 92% of the total variance of the set of inhibitors. For the first two principal components (PC1 and PC2) all the five chemical variables have importance, as shown in the following Eqs. (1) and (2):

$$\mathrm{PC1} = 0.498(\mathrm{Surf.\ area}) + 0.502(\mathrm{Volume}) + 0.510(\Delta H) + 0.431(E_{\mathrm{Hydr.}}) + 0.235(\log P) \quad (1)$$

$$\mathrm{PC2} = 0.970(\log P) + \mathrm{smaller\ terms} \quad (2)$$

The first principal component (PC1) clearly shows a separation in terms of the spatial orientation of the inhibitor's cyclic backbone, with four groups as follows:

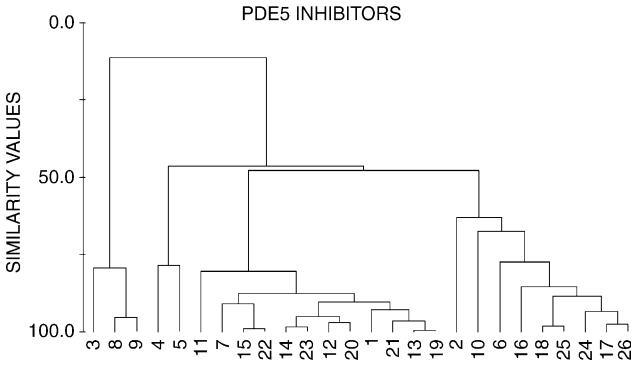

Fig. 3. Hierarchical dendrogram plot for the set of 26 inhibitors.

- *Group 1:* cGMP, zaprinast and the pyrazolopyrimidinone derivative 2 (Pyr-derivative 2);
- *Group 2:* Tadalafil and papaverine, the pyrazolopyrimidinone derivative 1 (Pyr-derivative 1) and derivatives revealed in the US patent US6,916,927 bearing the NH_2 substituent at the sulphonyl group;
- *Group 3:* Sildenafil, the sildenafil derivative (Der1-sild), vardenafil and the remaining set of derivatives revealed in the US patent US6,916,927;
- *Group 4:* The pyrazolopyrimidinone derivatives 3 and 4 (Pyr-derivative 3 and 4).

The results of hierarchical cluster analysis do corroborate the group separation observed with the PCA analysis (Fig. 3, hierarchical dendrogram plot).

Hierarchical cluster analysis is performed using 100% of the variance of the five dimensional space and complements the PCA study. HCA clusters the inhibitors together, generating groups with a degree of similarity. There were found four major groups and two inhibitors with low similarity regarding the set of inhibitors. The first group observed has 63% of similarity and comprises the inhibitors Pyr-derivative 1, papaverine, tadalafil and derivatives revealed in the US patent US6,916,927 bearing the NH_2 substituent at the sulphonyl group. The second group observed has 80% of similarity and comprises the inhibitors sildenafil, sildenafil derivative, vardenafil and the remaining set of derivatives revealed in the US patent US6,916,927. The third group shows 78% of similarity among the pyrazolopyrimidinone derivatives inhibitors the Pyr-derivative 3 and Pyr-derivative 4. Finally a separated group comprises of Pyr-derivative 2, zaprinast and cGMP were found in the cluster separation, with 11% similarity in relation to the rest of the groups.

X-ray diffraction studies [4] clearly show that the binding mode of the inhibitors tadalafil and sildenafil with the active site of the cGMP PDE5 enzyme is quite different. Both PCA and HCA analyses were able to differentiate the set of inhibitors in terms of spatial structure. Initial docking simulations were performed with the set of inhibitors to verify if the group separation observed with our chemometric analysis could be inferred as differences in binding modes at the active site of the cGMP PDE5 enzyme (Table 2).

The obtained results were quite interesting, suggesting different interactions between the different groups of inhibitors, as shown in Fig. 4 for zaprinast, tadalafil and sildenafil (inhibitors classified in different groups).

Crystallographic data shows that the active site of PDE5 is located at the center of the C-terminal helical bundle domain and that it is composed of four subsites (M_{site} = metal binding site, Q_{pocket} = core pocket, H_{pocket} = hydrophobic pocket and L_{region} = lid region). The pyrazolopyrimidinone group of sildenafil binds at the Q_{pocket} region of the enzyme while the ethoxyphenyl group fits into the hydrophobic H_{pocket}, the L_{region} of the enzyme surrounds the methylpirazine group of sildenafil. Tadalafil, on the other hand, makes no interaction with the L_{region} of the enzyme while the hydrophobic H_{pocket} is filled with the methylenedioxyphenyl group [4]. Such differences are clearly visualized in Fig. 4a, which shows the superimposed structures of "docked" sildenafil (blue) and tadalafil (cyan). Zaprinast, classified neither in the tadalafil nor in the sildenafil groups, shows potential interactions with the hydrophobic H_{pocket} and Q_{pocket} region of

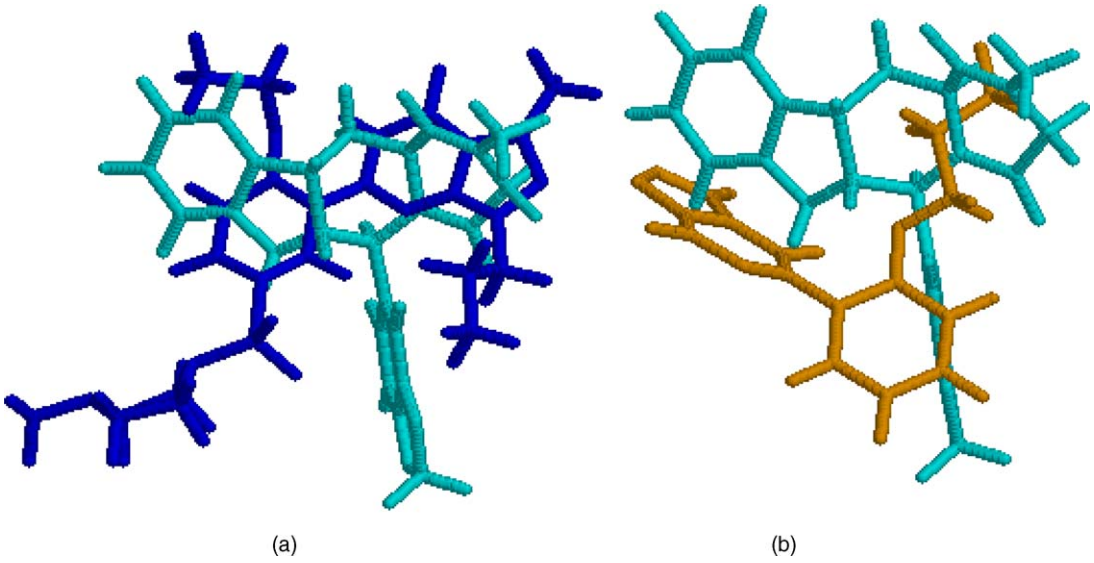

Fig. 4. (a) Superimposed docked structures of sildenafil (blue) and tadalafil (cyan) and (b) superimposed docked structures of tadalafil (cyan) and zaprinast (orange).

Table 2
Docking simulation parameters

General_parameters	
Flexible_ligand	yes
Orient_ligand	yes
Score_ligand	yes
Minimize_ligand	yes
Multiple_ligands	no
Random_seed	
Variable flexible_ligand_parameters	
Multiple_anchors	yes
Anchor_search	yes
Anchor_size	20
Write_partial_structures	no
Torsion_drive	yes
Clash_overlap	0.5
Configuration_per_cycle	25
Torsion_minimize	yes
Reminimize_layer_number	2
Minimize_anchor	yes
Reminimize_anchor	yes
Reminimize_ligand	yes
Minimization_parameters	
Contact_minimize	yes
Chemical_minimize	yes
Energy_minimize	yes
Initial_translation	1
Initial_rotation	0.1
Initial_torsion	10
Maximum_iterations	100
Contact_convergence	0.01
Chemical_convergence	0.01
Energy_convergence	0.01
Maximum_cycles	30
Cycle_convergence	1
Contact_termination	1
Chemical_termination	1
Energy_termination	1

the enzyme suggesting different binding sites than the ones experimentally determined for tadalafil.

4. Conclusion

The chemometric analysis performed on the set of 26 known phosphodiesterase 5 inhibitors, more specifically PCA and HCA analyses, was capable of classifying the test set in terms of similarity, the results being consistent with published experimental data. Docking simulations were used to correlate the grouped inhibitors and the expected binding mode with the active site of the cGMP PDE5 enzyme.

References

[1] D.P. Rotella, Phosphodiesterase 5 inhibitors: current status and potential applications, Nature 1 (2002) 674–684.
[2] A. Daugan, P. Grondin, C. Rualt, A.-C.M. de Gouville, H. Coste, J. Kirilovsky, F. Hyafil, R. Labaudinière, The discovery of tadalafil: a novel and highly selective PDE5 inhibitor. 1: 5,6,11,11a-Tetrahydro-1H-imidazo [1′,5′:1,6] pyrido [3,4-*b*]-indole-1,3-(2H)-dione analogues, J. Med. Chem. 46 (2003) 4522–4532.
[3] A. Daugan, P. Grondin, C. Rualt, A.-C.M. de Gouville, H. Coste, J.M. Linget, J. Kirilovsky, F. Hyafil, R. Labaudinière, The discovery of tadalafil: a novel and highly selective PDE5 inhibitor. 2. 2,3,6,7,12,12a-Hexahydropyrazino [1′,2′:1,6] pyrido [3,4-*b*]-indole-1,4-dione analogues, J. Med. Chem. 46 (2003) 4533–4542.
[4] B.-J. Sung, K.Y. Hwang, Y.H. Jeon, J.I. Lee, Y.-S. Heo, J.H. Kim, J. Moon, J.M. Yoon, Y.-L. Hyun, J.H. Kim, S.J. Eum, S.-Y. Park, J.-O. Lee, T.G. Lee, S. Ro, J.M. Cho, Structure of the catalytic domain of human phosphodiesterase 5 with bound drug molecules, Nature 425 (2003) 98–102.
[5] J.C. Lanter, Z. Sui, M.J. Macielag, J.J. Fiordeliso, W. Jiang, Y. Qiu, S. Bhattacharjee, P. Kraft, T. Mathew John, D. Haynes-Johnson, E. Craig, J. Clancy, Structure–activity relationships of *N*-acyl pyrroloquinolone PDE-5 inhibitors, J. Med. Chem. 47 (2004) 656–662.
[6] G.J. Niemiin, in: W. Karcher, J. Devilers (Eds.), Pratical Application of Quantitative Structure–Activity Relationships (QSAR) in Enviromental Chemistry and Toxicology, Kluwer Academic, Dordrecht, 1990.
[7] L.J.O. Figueiredo, F.M.S. Garrido, Mol. Struct. (TEOCHEM) 539 (2001) 75–81.
[8] S. So, M. Karplus, J. Med. Chem. 39 (1996) 1521–1530.
[9] G. Cruciani, M. Pastor, J. Med. Chem. 38 (1995) 4637–4647.
[10] K.V. Mardia, J.T. Kent, J.M. Bibby, Multivariate Analysis, Academic Press, London, 1979.
[11] I.S. Scarminio, R.E. Bruns, Trends Anal. Chem. 8 (1989) 326.
[12] B.R. Kowalski, C.F. Bender, J. Am. Chem. Soc. 9 (1972) 5632.
[13] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, J. Am. Chem. Soc. 107 (1985) 3902.
[14] TINKER molecular modelling package, public domain, PC Version, downloaded from the TINKER Website (http://www.tinker.com).
[15] C. Hansch, A. Leo, Exploring QSAR: Fundamentals and Applications in Chemistry and Biology, ACS, Washington, 1995.
[16] M.L. Connolly, Science 221 (1983) 709–713.