ACADEMIA ROMÂNĂ
Revue Roumaine de Chimie
http://web.icf.ro/rrch/

Rev. Roum. Chim.,
2015, *60*(2-3), 183-188

Dedicated to Professor Zeno Simon
on the occasion of his 80th anniversary

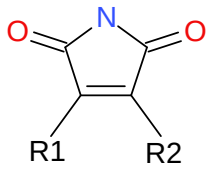
3D-QSAR STUDY OF MALEIMIDE ANALOGUES AS GLYCOGEN SYNTHASE KINASE-3 (GSK-3) INHIBITORS USING CoMSIA APPROACH**

Luminita CRISAN,[a] Sorin AVRAM,[a] Alina BORA,[a,b] Ludovic KURUNCZI[a,c] and Liliana PACUREANU[a,*]

[a]Department of Computational Chemistry, Institute of Chemistry of Roumanian Academy, Timişoara, Mihai Viteazul Avenue, 24, 300223 Timişoara, Roumania
[b] Department of Chemistry, West University of Timişoara, Faculty of Chemistry-Biology-Geography, 16 Pestalozzi Str., 300115 Timişoara, Roumania
[c] Victor Babeş University of Medicine and Pharmacy, E. Murgu 2, 300041 Timişoara, Roumania

Received August 25, 2014

Several series of substituted maleimide derivatives (3-anilino-4-aryl, indolyl, oxaindolyl), and macrocyclic maleimides were reported as potent, selective GSK-3 inhibitors. In order to gain overall insight into structural prerequisites that influence biological activity, we adopted a ligand-based methodology - three-dimensional quantitative structure–activity relationships which has been developed using the Field Based QSAR module from Schrodinger suite and a large collection of maleimide compounds (203). The CoMSIA model which was generated based on training set and validated on test set compounds (R_{train}^2=0.833, Q_{train}^2=0.673, $R_{Pearson\ test}$=0.832) is reliable and can be used to predict novel inhibitors of GSK-3 belonging to the same structural class. The molecular structural features and field contour maps portrayed by 3D-QSAR model afforded relevant insights into the structure−activity profile of maleimide derivatives. The CoMSIA model, confirm the experimental interaction pattern and is helpful for future lead compound optimization, design and prioritization of new inhibitors for chemical synthesis.



INTRODUCTION

GSK-3 emerged as a promising therapeutic target for the treatment of high clinical impact diseases including cancer, inflammatory disorders, neurodegenerative conditions, diabetes, etc. foreseeing a multitude of curative potentialities for its inhibitors. A promising class of GSK3 inhibitors is represented by maleimide derivatives (bisarylmaleimides,[1] 3-anilino-4-arylmaleimides,[2] bisindolylmaleimides,[3] azaindolylmaleimides,[4] etc.) have been reported to display a degree of selectivity towards GSK3, with respect to monoglyceride lipase,[5] PKC,[6] CDKs,[7] Cdc25B,[8] Bfl-1,[9] and DNMT-1[10] etc. Maleimide derivatives play an important role as pharmaceutical agents including antibacterial,[11] antistress,[12] cytotoxic,[13] DNA binding and apoptotic inducing activity,[13] analgesic,[14] antiangiogenic,[15] and antiprotozoal.[16] These compounds exhibit angiogenesis[17] and protein kinase inhibition[18] as well as antiproliferative[19] activity.

In the current study we propose an exhaustive analysis of the chemical and biological space of a dataset of 203 maleimide derivatives[1,2,4,20-22] aiming at establishing statistically significant Comparative Molecular Similarity Index Analysis (CoMSIA) models[23] for the inhibition of GSK-3 that can be

* Corresponding author: pacureanu@acad-icht.tm.edu.ro
** Supporting information on http: //web.icf.ro or http: //revroum/lew.ro

used for the design of further potent and selective maleimide inhibitors. This large dataset of maleimide derivatives covering many structural features and their biological responses, which were not investigated jointly in previous works, may identify significant pharmacophoric features which can give useful information in addition to lead optimization and help future rational design of novel compounds with improved pharmacology. The CoMSIA technique are widely used for rational drug design and is based on partial least square (PLS) regression[24] which have a long history of application in quantitative structure-activity relationship. In CoMSIA approach, hydrophobic, hydrogen bond donor and hydrogen bond acceptor similarity fields are computed in addition to the steric and electrostatic fields. Hence, CoMSIA establish a correlation between the 3D structures of the molecules and their activities allowing a proper and insightful interpretation of the correlations obtained.

MATERIALS AND METHODS

Dataset and biological activity. A number of 203 maleimide derivatives were available together with their GSK3 inhibitory activities[1,2,4,20-22] measured as IC_{50} The experimental IC_{50} values expressed in nanomolar concentrations were converted to a logarithmic scale (pIC_{50}=-logIC_{50}) before modeling (see Supplementary material, Fig. 1S and Table 1S). The distribution of dataset compounds in relation to biological activity is normal and covers a wide variation of 5.124 logarithmic units (see Supplementary material, Fig. 2S).

Training and test set selection. The data set, was randomly divided by the program into training set 163 compounds (80%) and test set of 40 compounds (20%) which were used to develop and respectively, to validate the CoMSIA models (see Supplementary material, Table 1S). This way we certify that the models are developed and tested on a sufficiently large test set necessary to build reliable 3D QSAR models. The activities of the training set ranges from 9.638 to 4.514 (pIC_{50}), with the median of 7.032, whereas the activities of the test set ranges from 9.292 to 4.905 (pIC_{50}), median 7.034 indicating that the activity range of the test set are comparable with that of the training set.

Conformational search and molecular alignment. CoMSIA methodologies require the alignment of three-dimensional structures according to a suitable conformational template, preferably a bioactive conformation which in the current case is the crystal structure of 2-chloro-5-[4-(3-chloro-phenyl)-2,5-dioxo- 2,5-dihydro-1h-pyrrol-3-ylamino]-benzoic acid co-crystallized with GSK-3β (PDBID:1Q4L)[25]. Rigid body alignment of ligands was performed with Rapid Overlay of Chemical Structures (ROCS) (www.eyesopen.com) which describe the molecular shape/volume by Gaussian functions.[26-32] The CoMSIA fields were calculated using the Field Based QSAR[33] module from Schrödinger package, accessible under Maestro[34] on a Linux cluster. The tautomers and ionization states were generated with LigPrep module[35] from Schrödinger package, in the pH range of 7.4±2.0. Multiple conformers for each compound were generated and energetically optimized under ConfGen[36] standards which employ OLPS-2005 force field.

CoMSIA methodology. CoMSIA was implemented by Klebe *et al.*[23] at BASF Ludwigshafen, Germany. This technique is most commonly used in drug discovery to find the predominant molecular characteristics that are relevant for binding to biological receptor. CoMSIA represents an extension of the CoMFA methodology which diverges only regarding the implementation of the fields. The similarity indices are calculated as follows: $A^{q}_{F,K(j)} = -\sum W_{probe,k} W_{i,k} e^{-\alpha r^{2}_{iq}}$; where A denotes the similarity index calculated at grid point q, over all atoms, i, of the molecule j; $W_{probe,k}$ stands for the probe atom with a defined radius of 1 Å, charge +1, hydrophobicity +1, hydrogen bond donor +1, hydrogen bond acceptor +1; $W_{i,k}$ represents the value of the physico-chemical parameter k, of atom i; r_{iq} is the reciprocal distance between the atom i of the molecule and the probe atom at grid point q; α, is the attenuation factor.[37] The CoMSIA models were developed using different combinations of the five descriptors (fields), steric (S), electrostatic (E), hydrophobic (H), hydrogen bond donor (D), and hydrogen bond acceptor (A). The most important statistical parameters: (i) for the training set: the correlation coefficients of regression, R^2, the scrambled correlation coefficients built using scrambled biological activities, R^2_{scr}, (low values means that the model cannot fit random data); and (ii) for the test set: Pearson correlation coefficient, R-Pearson, between predicted and observed activities for the test set, the squared Pearson correlation coefficient R^2, and predictive squared correlation coefficient Q^2. The Gaussian-type distance dependence of the physicochemical properties represents an advantage of CoMSIA since no singularities at atomic positions occurred.

The Q^2 is acknowledged as an indicator of predictive abilities. ($Q^2 > 0.6$ means a fairly good model, $Q^2 = 0.4–0.6$ means a questionable model and $Q^2 < 0.4$ a poor model).[38] To further assess the robustness of the derived models, the Y-scrambled models were generated to estimate the chance correlation.

RESULTS AND DISCUSSION

3D QSAR CoMSIA analysis. CoMSIA methodology provides indirect information about binding to receptor obtained from the correlation between the biological activity of a set of compounds and their three-dimensional structures. During CoMSIA investigation forty three QSAR models were developed, but only the most significant model for each combination of fields is listed in Table 1. The best CoMSIA model obtained includes five field descriptors (**S**teric, **E**lectrostatic, **H**ydrophobic, **A**cceptor, **D**onor), with the R-Pearson coefficient for test set of 0.832, and the correlation coefficient of regression for training set R^2= 0.833. The statistical data of the models with R^2_{cv} higher than 0.5, are acceptable and indicates a good internal predictive ability. A stability value of 0.853 for the best model higher than 0.8, indicate that the models are not sensitive to the omissions from the training set, and are higher than R^2 values, which is an indication of the absence of over-fitting.

Analyzing the statistical values depicted in Table 1 we concluded that the generated models are statically acceptable and will be helpful for the detection of new compounds and design potent GSK-3 inhibitors. The fractions of field contributions for the best model are shown in Table 2.

The graphical representation of experimental versus predicted biological activity values for training and test set (see Supplementary Material, Table 1S), for the best model obtained are depicted in Fig. 1. The histogram of the residuals of the test set molecules is depicted in Fig. 3S in Supplementary material. The correlation coefficient for experimental and predicted activity value is R^2=0.833 for training set and R^2=0.692 for test set.

Graphical Interpretation of the Fields. The contour map for the most active compound (180) was analyzed in detail. The contour map provided by the CoMSIA best model including SEHAD fields, supply information about the nature and spatial arrangement of the pharmacophoric fields of maleimide derivatives which are in complementary relationship with the binding site of GSK-3 (see Fig. 2).

Table 1

The statistics regression summary of the CoMSIA models

Model	Fields	#	SD	R^2	R^2_{CV}	$R^2_{Scrambled}$	F	P	N*	S	Q^2	Pearson -R
CoMSIA	HD	5	0.505	0.762	0.485	0.377	100.300	4.79E-47	163	0.862	0.692	0.840
CoMSIA	EHD	5	0.441	0.816	0.532	0.445	139.200	7.85E-56	163	0.852	0.686	0.830
CoMSIA	EDHA	5	0.413	0.838	0.563	0.451	163.500	2.34E-60	163	0.852	0.683	0.827
CoMSIA	**ESHAD**	**5**	**0.421**	**0.839**	**0.556**	**0.446**	**156.100**	**4.97E-59**	**163**	**0.853**	**0.673**	**0.832**

# denotes the number of PLS factors; *N represents the number compounds in the training set.

Table 2

The fraction of field contributions for each PLS factor, for the best model (ESHAD)

PLS Factors	Gaussian Steric	Gaussian Electrostatic	Gaussian Hydrophobic	Gaussian Hbond Acceptor	Gaussian Hbond Donor
1	0.529	0.105	0.189	0.116	0.060
2	0.471	0.115	0.205	0.147	0.063
3	0.353	0.147	0.204	0.193	0.102
4	0.304	0.15	0.222	0.195	0.129
5	0.305	0.145	0.231	0.185	0.134

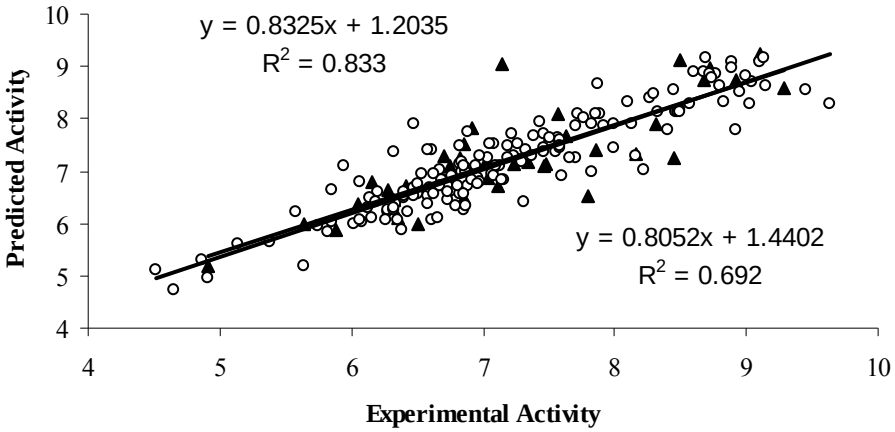
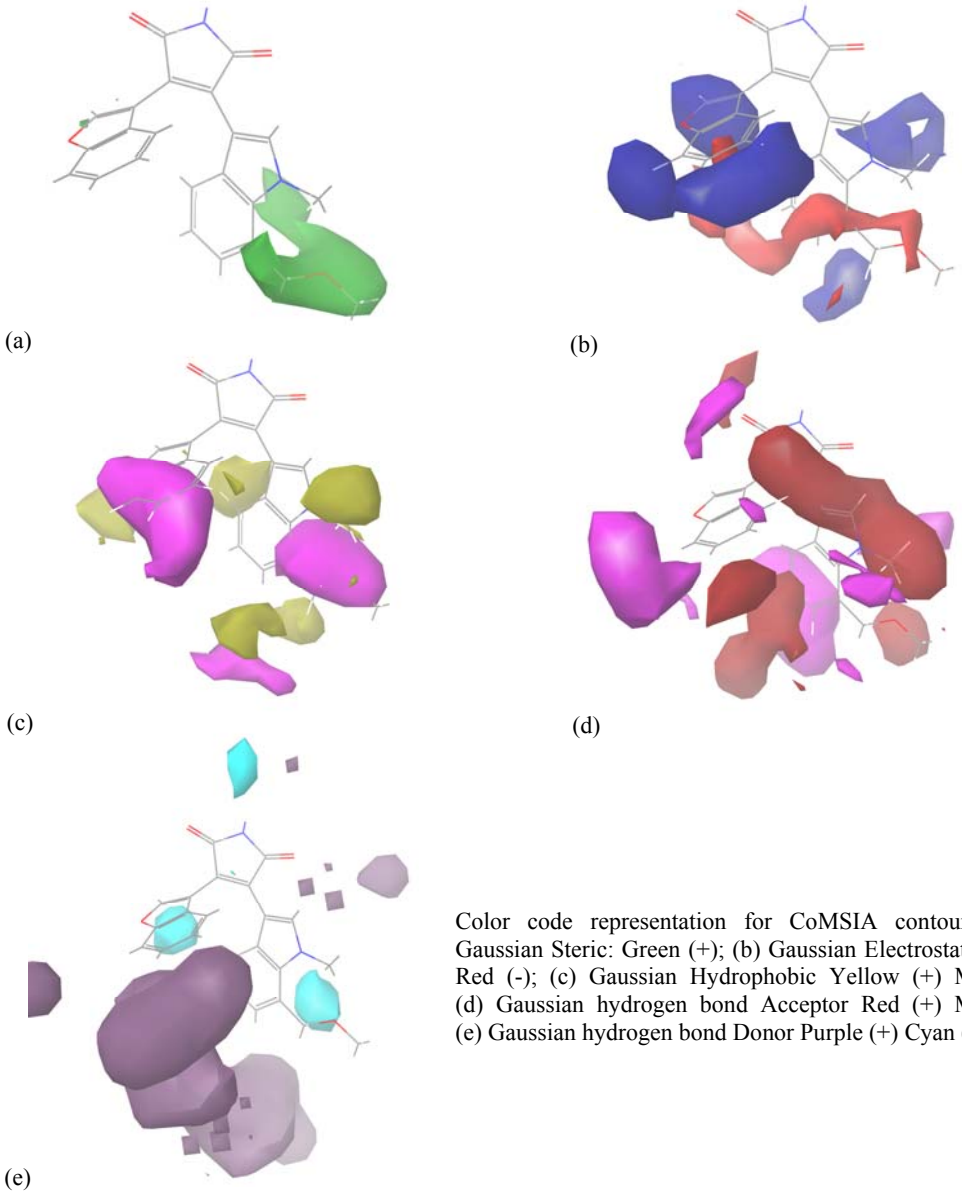


Fig. 1 – Experimental vs predicted biological activities for the training (circles) and the test sets (triangles).



Color code representation for CoMSIA contour maps (a) Gaussian Steric: Green (+); (b) Gaussian Electrostatic: Blue (+) Red (-); (c) Gaussian Hydrophobic Yellow (+) Magenta (-); (d) Gaussian hydrogen bond Acceptor Red (+) Magenta (-); (e) Gaussian hydrogen bond Donor Purple (+) Cyan (-).

Fig. 2 – The CoMSIA steric (a), electrostatic (b), hydrophobic (c), hydrogen bond acceptor (d) and hydrogen bond donor (e) contour map based on model SEHAD.

The CoMSIA steric field (green) (Fig. 2a) maps the sterically favorable contours corresponding to the regions in space (R1 (the substituents on the indole ring) and R2 (denotes substituents on benzofuran) where bulky substituents are favorable to biological activity. The effect of positive electrostatic field (Fig. 2b), designated by the blue contours observed in the vicinity of R1 and R2 substituents of maleimide ring, especially nearby negatively charged atoms (N,O) point out the areas where the positive groups can facilitate the increase in activity, whereas the red contours designate the areas where the partial negative charge is favorable for biological activity. The evaluation of CoMSIA hydrophobic contour maps (Fig. 2c) where the yellow contour indicate the regions where hydrophobic substituents are favorable and magenta contours indicate the unfavorable regions. These areas correspond to the region in space in the vicinity of both R1 and R2, where hydrophobic interaction with CYS199 and ILE62 may occur. The significant influence of the hydrophobic areas of the compound 180 can be seen at methyl substituent of nitrogen at position 1 of indole group. This fact is correlated with the trend of experimental biological activity of meleimides where the most active compounds display hydrophobic substituents at the N-alkyl group. According to the previous observations the hydrophilic groups are favorable at positions 6 and 7 of indolyl substituent, the most active compound 180 bears a methoxymethyl moiety at position 7.[39] Hydrogen bond acceptor map includes the red contours which indicate the regions where hydrogen bond acceptor substituents are favorable whereas magenta contours represent the sites where hydrogen bond acceptor groups are unfavorable (Fig. 2d). The hydrogen bond donor map (Fig. 2e) display also positive and negative areas depicted in cyan and respectively purple contours. The large purple contours reveal the necessity of the hydrogen bond donor groups on these areas to the inhibitory activities. The NH group belonging to maleimide ring is mandatory for inhibitory properties because the substitution of NH lead to steep decrease of biological activity[2] and the interaction of NH group with CO group of ASP133 was observed in all co-crystal structures of GSK-3β with maleimide derivatives (PDBID 1Q4L, 2R0E, 2OW3, 2SD0). Based on these results the CoMSIA contour maps offer valuable information necessary to understand the relationship between the physico-chemical structure and inhibitory activities. Since the inhibitory activity of the most active maleimides are situated in the picomolar area and are selective to the GSK-3 target, the chemical and biological information provided are valuable and can be exploited to design novel, potent and selective compounds. The current model represents an improvement over the CoMSIA model obtained by Kim *et al.*[39] which is based on 51 compounds, steric and electrostatic fields, and display lower statistical characteristics (R^2=0.746).

CONCLUSIONS

In this study, the CoMSIA approach has been successfully applied to a large set of structurally diverse maleimide derivatives whose activity gap range equal five orders of magnitude in logarithmic units. The constructed CoMSIA model, generated by means of steric, electrostatic, hydrophobic, H-bond acceptor and H-bond donor fields showed reliable statistical parameters and acceptable predictive abilities. This model can be extrapolated, under the limitations of applicability domain, to predict novel and more potent maleimide derivatives. The CoMSIA contour maps are in accordance with experimental observations and provided useful information about the relationship between the chemical structure and inhibitory activity. This model can guide further plausible modification of molecular structure to optimize biological activity.

Acknowledgments: The work of Alina Bora was supported by the strategic grant POSDRU/159/1.5/S/137750, Project "Doctoral and Postdoctoral programs support for increased competitiveness in Exact Science research" co-financed by the European Social Fund within the Sectoral Operational Programme Human Resources Development 2007-2013 and performed at West University of Timisoara. Authors thanks to Rita Podzuna from Schrödinger Inc for providing an evaluation license for Field based QSAR module. We thank OpenEye for providing academic licenses. The work of L. Crisan, S. Avram, and L. Pacureanu was supported by the Project No. 1.2/2014 of the Institute of Chemistry of Roumanian Academy, Timişoara.

REFERENCES

1. T.A. Engler, S. Malhotra, T.P. Burkholder, J.R. Henry, D. Mendel, W.J. Porter, K. Furness, C. Diefenbacher, A. Marquart, J.K. Reel, Y. Li, J. Clayton, B. Cunningham, J. McLean, J.C. O'Toole, J. Brozinick, E. Hawkins, E. Misener, D. Briere, R.A. Brier, J.R. Wagner, R.M.

Campbell, B.D. Anderson, R. Vaughn, D.B. Bennet, T.I. Meier and J.A. Cook, *Bioorg. Med. Chem. Lett.*, **2005**, *15*, 899-903.
2. D.G. Smith, M. Buffet, A.E. Fenwick, D. Haigh, R.J. Ife, M. Saunders, P.S. Brian, R. Stacey and R.W. Ward, *Bioorg. Med. Chem. Lett.*, **2001**, *11*, 635-639.
3. Z. Hang-Cheng, B.W. Kimberly, Y. Hong, F.M. David, K.D. Claudia, F.A. Michael, A.G. Patricia, J.E. Annette, R.C. Bruce, W. Lori, Z.X. Jun, A.L. Richard, T.D. Keith, E. Stuart and E.M. Bruce, *Bioorg. Med. Chem. Lett.*, **2003**, *13*, 3049-3053.
4. G. H. Kuo, C. Prouty, A. DeAngelis, L. Shen, D. J. O'Neill, C. Shah, P. J. Connolly, W. V. Murray, B. R. Conway, P. Cheung, L. Westover, J. Z. Xu, R. A. Look, K. T. Demarest, S. Emanuel, S. A. Middleton, L. Jolliffe, M. P. Beavers and X. Chen, *J. Med. Chem.*, **2003**, *46*, 4021-4031.
5. N. Matuszak, G. Muccioli, G. Laber and D. Lambert, *J. Med. Chem.*, **2009**, *52*, 7410-7420.
6. P. D. Davis, C. H. Hill, G. Lawton, J. S. Nixon, S. E. Wilkinson, S. A. Hurst, E. Keech and S. E. Turner, *J. Med. Chem.*, **1992**, *35*, 177-184.
7. A. Castro and A. Martinez, *Expert Opinion on Therapeutic Patents*, Oct **2000**, Vol. 10, No. 10: 1519–1527.
8. H. Chen, Y. Liu, L. Wang, Q. Shen, J. Li and F. Nan, *Bioorg. Med. Chem.*, **2010**, *20*, 2876-2879.
9. J.R. Cashman, M. MacDonald, S. Ghirmai, K.J. Okolotowicz, E. Sergienko, B. Brown, X. Garcia, D. Zhai, R. Dahl and J.C. Reed, *Bioorg. Med. Chem.*, **2010**, *20*, 6560-6564.
10. T. Suzuki, R. Tanaka, S. Hamada, H. Nakagawa and N. Miyata, *Bioorg. Med. Chem.*, **2010**, *20*, 1124-1127.
11. S.N. Lopez, M. Sortino, A. Escalante, F. Campos, R. Correa, V. Cechinel Filho, R.J. Nunes and S.A. Zacchino, *Arzneim-Forsch. Drug. Res.*, **2003**, *53*, 280-288.
12. R. Badru, P. Anand and B. Singh, *Eur. J. Med. Chem.*, **2012**, *48*, 81-91.
13. Alaa A.-M. Abdel-Aziz, *Eur. J. Med. Chem.*, **2007**, *42*, 614-626.
14. F. Mahle, T. Guimaraes, A. Meira, R. Correa, R. Cruz, R. Nunes, V. Cechinel-Filho and F. Campos-Buzzi, *Eur. J. Med. Chem.*, **2010**, *45*, 4761-4768.
15. N. Acero, M.F. Brana, L. Anorbe, G. Dominguez, D. Munoz-Mingarro, F. Mitjans and J. Piulats, *Eur. J. Med. Chem.*, **2012**, *48*, 108-113.
16. Y. Durust, H. Karakus, M. Kaiser and D. Tasdemir, *Eur. J. Med. Chem.*, **2012**, *48*, 296-304.
17. C. Peifer, T. Stoiber, E. Unger, F. Totzke, C. Schachtele, D. Marme, R. Brenk, G. Klebe, D. Schollmeyer, and G. Dannhardt, *J. Med. Chem.*, **2006**, *49*, 1271-1281.
18. J. Trapp, A. Jochum, R. Meier, L. Saunders, B. Marshall, C. Kunick, E. Verdin, P. Goekjian, W. Sippl and M. Jung, *J. Med. Chem.*, **2006**, *49*, 7307-7316.
19. S. Routier, J. Merour, N. Dias, A. Lansiaux, C. Bailly, O. Lozach and L. Meijer, *J. Med. Chem.*, **2006**, *49*, 789-799.
20. 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 and B. E. Maryanoff, *Bioorg. Med. Chem. Lett.*, **2004**, *14*, 3245-3250.
21. I. N. Gaisina, F. Gallier , K. H. Kim, A. V. Ougolkov, S. Guo, D. Holzle, D. N. Luchini, T. Kurome, S. Y. Blond, D. Billadeau and A. P. Kozokowski, *J. Med. Chem.*, **2009**, *52*,1853-1863.
22. Ye Q, Guiqing Xu, Lv Dan, Cheng Zhe, Li Jia and Yongzhou Hu, *Bioorg. Med. Chem.*, **2009**, *17*, 4302-4312.
23. G. Klebe, U. Abraham and T. Mietzner, *J. Med. Chem.*, **1994**, *37*, 4130-4146.
24. S. Wold, C. Albano, W. Dunn, U. Edlund, K. Esbensen, P. Geladi, S. Hellberg, E. Johansson, W. Lindberg and M. Sjostrom, "Chemometrics: mathematics and statics in chemistry" The Netherlands: Reidel, Dordrecht, 1987, p.17.
25. RCSB Protein Data Bank. http://www.rcsb.org/pdb/home/home.do. Accessed on 15 Feb 2014.
26. ROCS version 3.1.0 OpenEye Scientific Software, Inc. Santa Fe, NM, USA, 2010, www.eyesopen.com.
27. J. Venhorst, S. Nunez, J. W. Terpstra and C.G. Kruse, *J. Med. Chem.*, **2008**, *51*, 3222-3229**.**
28. R. P. Sheridan, G. B. McGaughey and W. D. Cornell, *J. Comput. Aided Mol. Des.*, **2008**, *22*, 257-265.
29. T. S. Rush, J. A. Grant, L. Mosyak and A.Nicholls, *J. Med. Chem.*, **2005,** *48*, 1489-1495.
30. J. A. Grant, M. A. Gallardo and B. Pickup, *J. Comp. Chem.*, **1996**, *17*, 1653-1666.
31. T. Tuccinardi, G. Ortore, M. Amelia Santos, S. M. Marques, E. Nuti, A. Rosello and A. Martinelli, J. *Chem. Inf. Model*, **2009**, *49*, 1715-1724.
32. J. J. Sutherland, R. K. Nandigam, J. A. Erickson and M. Vieth, *J. Chem. Inf. Model*, **2007**, *47*, 2293-2302.
33. Field-Based QSAR, version 2014-1, Schrödinger, LLC, New York, NY, 2014.
34. Maestro, version 9.7, Schrödinger, LLC, New York, NY, 2014.
35. LigPrep, version 2.9, Schrödinger, LLC, New York, NY, 2014.
36. K. S. Watts, P. Dalal, R. B.Murphy, W. Sherman, R. A. Friesner and J. C. Shelley, *J. Chem. Inf. Model.*, **2010**, *50*, 534-546.
37. M. Bohm, J. St Rzebecher and G. Klebe, *J. Med. Chem.*, **1999**, *42*, 458-477.
38. M. Clark and R. D. Cramer, *Quant. Struct. Act. Relat.*, **1993**, *12*, 137-145.
39. K. H. Kim, I. Gaisina, F. Gallier, D. Holzle, S. Y. Blond, A. Mesecar and A. P. Kozikowski, *J. Mol. Model.*, **2009**, *15*, 1463-1479.