STUDIA UBB. CHEMIA, LV, 4, 2010

STRUCTURE BASED ALGORITHM FOR CLASSIFICATION OF MALEIMIDE DERIVATIVES ATP-COMPETITIVE INHIBITORS OF GSK-3

LILIANA M. PĂCUREANU[a], ALINA BORA[a], LUMINIȚA CRIŞAN[a], LUDOVIC KURUNCZI[b]

ABSTRACT. The structure based retrospective virtual screening algorithm employed the docking engine FRED (Fast Rigid Exhaustive Docking) to dock 74 inhibitors (4-aryl-3-anilino-maleimide derivatives) and 1778 decoy molecules into glycogen synthase kinase-3 β, GSK-3β, ATP-binding site (PDB code 1Q4L).

The input database of 74 ligands was prepared following the OpenEye protocol by adding tautomers and ionization states, generating conformers, and performing charge corrections with AM1BCC option from QUACPAC software. The protein preparation has been carried out with Chimera software by deleting water molecules (except water near Thr 138), adding hydrogen and charges (AM1BCC). The energy component values of the scoring functions were subsequently submitted to PLS-DA (Projections in Latent Structures, Discriminant Analysis). The final PLS-DA result contains only the essential energy factors that describe most accurately the interactions in the ATP binding site. The results obtained are of better quality than those obtained using the total scores provided by initial scoring functions in terms of AUC (Area Under Curve) 0.938 (chemgauss2 donor + screenscore rotatable bonds) with respect to 0.887 (chemgauss3). Moreover, the early enrichment of the PLS-DA term at 1% of the database is 13.514% while for Chemgauss 3 was only 8.108%.

Keywords: *molecular docking, Projections in Latent Structures - Discriminant Analysis (PLS-DA), glycogen synthase kinase-3β (GSK-3β)*

INTRODUCTION

The identification of selective inhibitors of protein kinases by virtual screening strategies withdraw much interest in the area of drug discovery by helping in terms of time and money the high throughput screening (HTS) experiments [1]. GSK-3 is a serine/threonine protein kinase, discovered as the enzyme that inactivates the glycogen synthase (GS), the rate limiting enzyme in glycogen synthesis [2]. Besides glycogen metabolism regulation [2,3], GSK-3 controls a large number of cellular processes such as microtubule stability [4], β-catenin degradation [5], protein translation [6], etc.

[a] *Institute of Chemistry of Romanian Academy, 24 Mihai Viteazul Bvd., RO-300223, Timisoara, Romania, pacureanu@acad-icht.tm.edu.ro*

[b] *University of Medicine and Pharmacy "Victor Babes", Faculty of Pharmacy, 2 Eftimie Murgu Avenue, RO-300041, Timisoara, Romania, dick@acad-icht.tm.edu.ro*

Maleimide derivatives have been identified as ATP competitive inhibitors of GSK-3α at Smithkline Beecham pharmaceutical company by means of a high throughput screening experiment [7]. GSK-3 inhibition by maleimide derivatives caused the acceleration of glycogen synthesis in the liver suggesting the utility of maleimide inhibitors for the treatment of diabetes [3]. Moreover, additional biological investigations demonstrated that maleimide derivatives prevent neuronal death through a mechanism that involve, interactions with tau and β-catenin [8].

Structural characteristics of GSK-3 inhibitors have been investigated by various techniques including QSAR, docking and ligand based virtual screening [9,10,11,12,13,14,15].

Our investigation is directed towards a structure-based methodology due to the availability of X-ray cocrystal GSK-3β - maleimide derivative [16]. The high identity (similarity) of human GSK-3α and β 83% (89%) overall and 91% (97%) of the catalytic domain [17] permitted us to use the X-ray stucture of GSK3 β to dock the maleimide inhibitors tested in GSK 3α [7]. The docking algorithm has to check that the chemical compounds make favorable interactions with the enzyme. Therefore, the set of inhibitors were mixed with a large number of inactives (decoys) in order to reproduce the real situation.

Scoring functions, as they have been constructed, display a series of shortcomings, especially high false positive rates. Consensus scoring has been introduced to counterweight for false positive rates of individual scoring functions. But the selection algorithm for the right, individual scoring functions represents the major challenge [18]. Jacobsson *et al.* [19] have used PLS-DA (Projections in Latent Structures - Discriminant Analysis) methodology to the total scores of individual scoring functions in order to improve the performance of individual scoring functions. In this paper we introduced the PLS-DA methodology [20] to the variables representing the components of individual scoring functions in order to get a new combination of terms that will rank more appropriately the actives with respect to inactives.

Dataset

In our study, a dataset of 74 derivatives of 3-anilino-4-arylmaleimide [7] (Figure 1) and their biological activity, measured as inhibitory activity IC_{50} (nM) evaluated against human GSK 3α, is considered. Our dataset is assembled/ mixed with a decoy set of 1778 molecules (CDK-2 decoys) downloaded from DUD (Directory of Useful Decoys) [21].

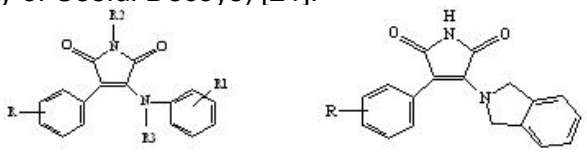

Figure 1. The template of maleimide derivatives (see ref 7):
R = H, 2-Cl, 2-OMe, 2-NO_2, 3-Cl, 3-OMe, 3-NO_2, 4-Cl, 4-OMe, 4-NO_2
R1 = H, 3-Cl, 3-OH, 4-OH, 3-Cl-4OH, 3,5-diCl-4-OH, 3-CO_2H, 4-Cl-3CO_2H, 4-SMe
R2, R3 = H, CH_3

The CDK-2 decoys were chosen on the basis of high similarity of aminoacid binding sites (85%) of GSK-3 and CDK-2 [16]. In the current study we assume the decoys are inactive, even not experimentally tested on GSK-3. Therefore, they probably can be active on this target [21]. The distribution of drug-like properties of actives and decoys are shown in Table 1.

Table 1. Drug-like properties of actives and decoys

	Molecular Weight	Rotatable Bond Number	Number of hydrogen bond donors	Number of hydrogen bond acceptors	MLOGP
Actives					
min	264.3	0	2	4	-3.569
max	575.68	12	5	12	2.773
Decoys					
min	298.37	1	0	4	-5.974
max	399.47	11	7	11	4.062

Protein preparation

The crystal structure of GSK-3 β (PDB entry: 1Q4L) in complex with inhibitor 2-chloro-5-[[4-(3-chlorophenyl)-2,5-dioxo-pyrrol-3-yl]amino]benzoic acid was downloaded from the PDB. The active site of the enzyme was prepared using Chimera package [22] deleting water molecules except water near Thr 138, that was kept as it mediates the hydrogen bonds to O^{γ} of Thr138 and $O^{\varepsilon 2}$ of Gln185, [16] adding hydrogens and AM1BCC charges.

Assignment of ionization states and generation of tautomers

Database preparation before virtual screening analysis is important for the quality of the results. Kirchmaier demonstrated that tautomerism is essential for the classification of actives in virtual screening experiments [23]. The three-dimensional structures of 74 GSK-3α inhibitors were prepared using LigPrep 2.2 module of Maestro in the Schrödinger software [24]. For the ligands, the only reasonable tautomeric forms at pH=7.4±1.5 were selected.

Conformer generation

Conformer generation for ligands and decoys was performed with Omega version 2.-2.3.2 from OpenEye package [25]. Biologically active fragment conformations are available in Omega's library. The ligand is split into fragments and next reassembled according to energetic criteria and the conformations complying with the energy window and heavy atom root mean square (RMS) distance were saved. We used an increment-based methodology for energy window of "5.0, 6.0, 7.0" kcal/mol, and RMS distance of the heavy atom coordinates for conformer detection of "0.5, 0.4, 0.3" Å. The assignments

of appropriate atomic charges were carried out with QuacPac software [25], choosing AM1BCC option (AM1 bond charge correction). The resulting conformer enriched database of actives and decoys was used as input for docking.

Docking procedure

Docking investigation was carried out with FRED (Fast Rigid Exhaustive Docking) software version 2.2.5 (www.eyesopen.com) [25]. The docking procedure occurs in two steps: shape fitting and optimization. The ligand is placed into a 0.5Å resolution grid-box incorporating all active site atoms (including hydrogen atoms) using a smooth Gaussian potential [26]. To score the ligand in the docking procedure the binding site of GSK-3β was defined using the reference ligands and an addbox of 4Å around the ligand. The best docked pose per each ligand was saved and seven classical scoring functions including Chemscore (CS), Chemgauss-2 (CG2), Chemgauss-3 (CG3), Shapegauss (SG), Screenscore (SC), OEChemscore (OECS), and PLP were used.

PLS-DA analysis

In the present work, we attempted to implement a multivariate statistical method (PLS DA), with the values of scoring function components as descriptors, in order to classify the virtual screening results in active and inactive compounds [27]. PLS is a regression method that works with two matrices, X (e.g., chemical descriptors) and Y (e.g., biological responses), and has two objectives, namely to approximate well X and Y, and to model the relationship between them [28]. For PLS DA methodology two classes are defined: the actives (1) and the inactives (2) according to ligands and decoys.

The energetic component outputs of all scoring functions (see reference [25]) were submitted to the SIMCA P 9.0 package [29] to perform initially a PCA (Principal Component Analysis) analysis [30], followed by the PLS DA analysis.

RESULTS AND DISCUSSION

In the first step of PLS DA analysis, a PCA model for the whole X matrix (N=1852 rows/compounds, and K=32 columns/energetic terms) was performed and three principal components were obtained. These three principal components explain 47.7% of the information content of the X matrix and distinguish very well the actives (*in black*) from the inactives (*in red* - Figure 2).

The PLS DA models were further constructed starting from the same X matrix. In order to improve the PLS DA models, the coefficient sign and VIP >1 (variable influence on projection) were considered as significant. Based on these criteria, six out of thirty two energetic terms were selected: CG2 Donor (Chemgauss2 contributions from donors on the ligand interacting with acceptors on the protein), CG3 Steric (Chemgauss3 steric interactions), CS HB (Chemscore hydrogen bonds), SC RB (Screenscore rotatable bond), SC Ambig (Screenscore ambiggous interactions), and SC HB (Screenscore hydrogen bonds). For these

six energetic terms, all the possible combinations were made and the first significant combination (CG2 Donor + SC RB) was selected. The sum of these terms represents the PLS-DA equal weight "mixed" scoring function.

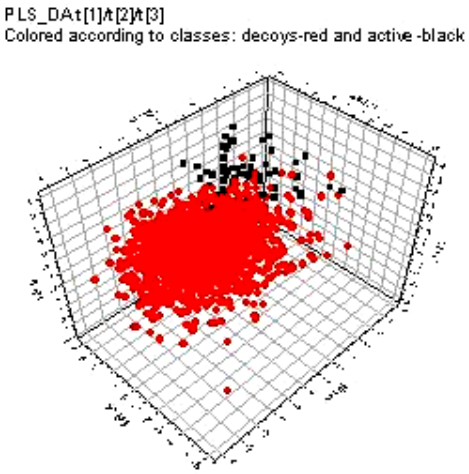

Figure 2. Classes of actives (in black) and inactives (in red)

In order to test the performance of the new "mixed" scoring function against classical scoring functions, the AUC and enrichment factors were compared. The results of ensemble AUC and enrichments are illustrated in Figure 3a and 3b.

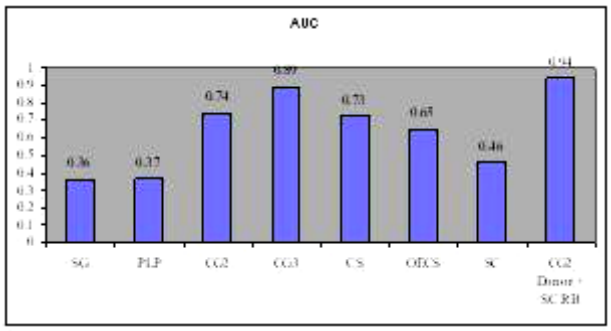

Figure 3.a) Bar chart showing AUC values obtained with seven classical and the new „mixed" scoring functions

The AUC of 0.887 and enrichment factor of 8.108% at 1% of database show good performances of the classical CG3 at the beginning, but these results were surpassed by the corresponding values of the "mixed" components (CG2 donor + SC RB) scoring functions AUC (0.938 and enrichment factor 13.513 % at 1% of database for this combination).

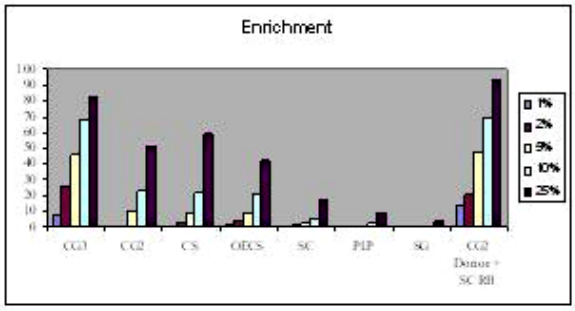

Figure 3.b) Enrichment performances at 1%, 2%, 5%, 10% and 25% of the database

Analyzing the classical CG2 and SC scoring functions, AUC is 0.735 and respectively 0.459 while the enrichment factor is 0.011% / 0.011% by the top 1% database and show low performances, but the donor + rotatable bond components (CG2 Donor + SC RB) seems to be significant in this combination.

The CG2 Donor energetic term into the "mixed" components scoring function measures the H-bond interaction energy between ligand and protein. The SC RB component is a penalty term proportional to the number of rotatable bonds in the ligand. SC RB is an important term in our situation since a number of compounds display a considerable number of flexible bonds in the decoys (up to 11) and ligands (up to 12).

In the top 2% - 25% of the database, the number of detected actives increases and the largest percentage (93.243%) was retrieved at 25% in the case of new "mixed" scoring functions.

CONCLUSIONS

Here we reported a promising workflow for structure-based virtual screening using rigid docking (FRED software) followed by PLS DA analysis. A new "mixed" scoring function was built. It collects the energy factors from different scoring functions that illustrate the particular interactions in the GSK3β site. In this way, the results here reported, are of better quality than those obtained by using every single scoring function available in the OpenEye package. The present study enabled us to indentify the optimal protocol for the highest enrichment of actives in the top 1% to 25% of the database for seven classical and one "mixed' scoring function. Therefore, in the following studies the algorithm for docking scoring aiming at ranking the actives versus decoys will be based on all possible combinations.

ACKNOWLEDGMENTS

We thank to Dr. Maria Mracec for kindly providing the Schrödinger software, to Dr. Erik Johansson (Umetrics, Sweden) for the SIMCA P v. 9.0 package and to PhD. student Sorin Avram for the computational support.

REFERENCES

1. P. Cohen, *Nature Reviews Drug Discovery*, **2002**, *1*, 309.
2. P. Cohen, "Muscle Glycogen Synthase", Enzymes Academic Press, New York, **1986**, vol. XVII, 461-497.
3. M.P. Coghlan, A.A. Culbert, D.A.E. Cross, S.L. Corcoran, J.W. Yates, N.J. Pearce, O.L. Rausch, G.J. Murphy, P.S. Carter, R. Roxbee Cox, D. Mills, M.J. Brown, D. Haigh, R.W. Ward, D.G. Smith, K.J. Murray, A.D. Reith, J.C. Holder, *Chem. Biol.*, **2000**, *7*, 793.
4. M. Hong, D.C. Chen, P.S. Klein, V.M. Lee, *J. Biol. Chem.*, **1997**, *272*, 25326.
5. M.J. Hart, R. de los Santos, I.N. Albert, B. Rubinfeld, P. Polakis, *Curr. Biol.*, **1998**, *8*, 573.
6. G.I. Welsh, C.M. Miller, A.J. Loughlin, N.T. Price, C. G. Proud, *FEBS Lett.*, **1998**, *421*, 125.
7. D.G. Smitth, M. Buffet, A.E. Fenwick, D. Haigh, R. Ife, M. Saunders, B.P. Slingsby, R. Stacey, R.W. Ward, *Bioorg. Med. Chem. Lett.*, **2001**, *11*, 635.
8. D.A.E. Cross, A.A. Cubert, K.A. Chalmers, L. Facci, S.D. Skaper, A.D. Reith, *J. Neurochem.*, **2001**, *77*, 94.
9. 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, *J. Med. Chem.*, **2004**, *47,* 935.
10. F.X. Tavares, J.A. Boucheron, S.H. Dickerson, R.J. Griffin, F. Preugschat, S.A. Thomson, T.Y. Wang, H.Q. Zhou, *J. Med. Chem.*, **2004**, *47*, 4716.
11. A. Peat, D. Garrido, J.A. Boucheron, S.L. Schweiker, S.H. Dickerson, J.R. Wilson, T.Y. Wang, S.A. Thomson, *Bioorg. Med. Chem. Lett.*, **2004**, *14*, 2127.
12. M. Zeng, Y. Jiang, B. Zhang, Z. Kewen, N. Zhang, Q. Yu, *Bioorg. Med. Chem. Lett.*, **2005**, *15*, 395.
13. A.R. Katritzky, L.M. Pacureanu, D.A. Dobchev, D. Fara, P. Duchowitz, M. Karelson, *Bioorg. Med. Chem.,* **2006**, *14*, 4987.
14. N. Dessalev, P.V. Bharatam, *Eur. J. Med. Chem.*, **2007**, *42*, 1014.
15. N. Dessalev, P.V. Bharatam, *Biophys. Chem.*, **2007**, *128*, 165.
16. J.A. Bertrand, S. Thieffine, A. Vulpetti, C. Cristiani, B. Valsasina, S. Knapp, H.M. Kalisz, M. Flocco, *J. Mol. Biol.*, **2003**, *333*, 393.
17. A. Adnan, K.P. Hoeflich, J.R. Woodgett, *Chem. Rev.*, **2001**, *101,* 2527.
18. M. Stahl, M. Rarey, *J. Med. Chem.*, **2001**, 5, 375.
19. M. Jacobsson, P. Lidn, E. Stjernschantz, H. Bostrm, U. Norinder, *J. Med. Chem.*, **2003**, *46*, 5781.
20. S. Wold, E. Johansson, M. Cocchi, *ESCOM: Leiden*, **1993**, 523.
21. N. Huang, B.K. Shoichet, J.J. Irwin, *J. Med. Chem.*, **2006**, *49*, 6789. http://dud.docking.org./r2/
22. UCSF Chimera v 1.3 http://www.cgl.ucsf.edu/chimera
23. J. Kirchmaier, P. Markt, S. Distinto, G. Wolber, T. Langer, *J. Comput. Aided Mol. Des.*, **2008**, *22*, 213.

24. LigPrep, version 2.2, Schrödinger, LLC, New York, NY, **2005** [http://www.schrodinger.com/].
25. OpenEye Scientific Software, Inc., 9 Bisbee Ct, Suite D Santa Fe, NM 87508 http://www.eyesopen.com/.
26. M.R. McGann, H.R. Almond, A. Nicholls, J.A. Grant, F.K. Brown, *Biopolymers*, **2003**, *68*, 76.
27. M. Jacobsson, P. Lidn, E. Stjernschantz, H. Bostrm, U. Norinder, *J. Med. Chem., **2003**, 46,* 5781.
28. L. Eriksson, J. Gottfries, E. Johansson, S. Wold, *Chemometrics and Intelligent Laboratory Systems*, **2004**, *73*, 73.
29. SIMCA P, version 9.0; Umetrics AB: Umea, Sweden. http://www.umetrics.com.
30. M. Daszykowski, K. Kaczmarek, Y. Vander Heyden, B. Walczak, *Chemom. Intel. Lab. Syst.*, **2007**, *85*, 203.