Pergamon

Bioorganic & Medicinal Chemistry Letters 12 (2002) 419–421

BIOORGANIC & MEDICINAL CHEMISTRY LETTERS

A Neural Network Based Virtual Screening of Cytochrome P450 3A4 Inhibitors

László Molnár and György M. Keserű*

Computer Assisted Drug Discovery, Gedeon Richter Ltd., PO Box 27, H-1475 Budapest, Hungary

Received 29 August 2001; revised 29 October 2001; accepted 13 November 2001

Abstract—A virtual screening test to identify potential CP450 3A4 inhibitors has been developed. Molecular structures of inhibitors and non-inhibitors available in the Genetest database were represented using 2D Unity fingerprints and a feedforward neural network was trained to classify molecules regarding their inhibitory activity. Validation tests revealed that our neural net recognizes at least 89% of 3A4 inhibitors and suggest using this methodology in our virtual screening protocol.

Cytochrome P450 3A4 is one of the major polymorphic izoenzymes responsible for the metabolism of almost 50% of known drugs in humans. Inhibitors of this isoenzyme might cause drug–drug interactions because of decreasing the clearance of other drugs metabolized by 3A4. Early identification of potential 3A4 inhibitors is therefore needed to minimize the risk of clinically relevant interactions. In vitro HTS techniques can be used to screen out 3A4 inhibitors from previously synthesized compound libraries.[1] In silico virtual screening, however, has a pivotal role in the evaluation of virtual libraries. Since the high resolution X-ray structure of 3A4 is not available, structure based approaches are limited to the application of homology models. Although a high throughput docking (HTD) protocol could identify potential substrates we could not discriminate substrates and inhibitors by this technique.[2] Application of 3D QSAR models in virtual screening represents another option[3,4] that requires high throughput alignment (HTA) methodologies such as the recently launched Flex-S.[5] In addition to fitting uncertainties introduced by HTA, 3D-QSAR approaches suffer from a speed limit, that is the generation of the 3D structures for all of the compounds studied. Developing a 3D QSAR model we assume that binding modes of compounds used for the training set are similar. Considering the relatively large binding pocket of 3A4[6] it is clear that substrates should have a relatively large conformational degree of freedom within the active site (simultaneous binding of two molecules is also possible[7]). Since known 3A4 inhibitors represent a structurally diverse set of compounds it is likely that they bind

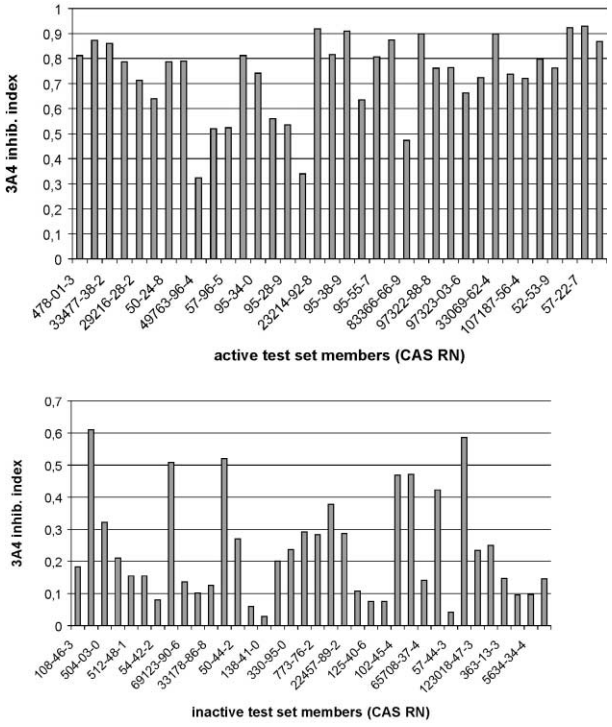


Figure 1. Distribution of 3A4 inhibitory indices calculated for inhibitors and non-inhibitors in the 3A4 test set.

*Corresponding author. Fax: +36-1-4326002; e-mail: gy.keseru@richter.hu

0960-894X/02/$ - see front matter
PII: S0960-894X(01)00771-5

in different modes, which limits the predictive power of a 3D QSAR model. Finally, a predictive 3D QSAR model requires 25–35 high quality IC_{50} data that are scarcely available even for structurally similar compounds in the literature.

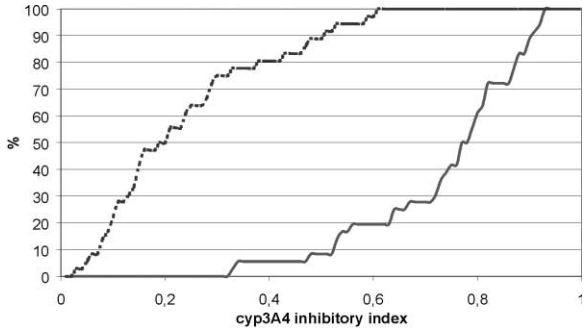
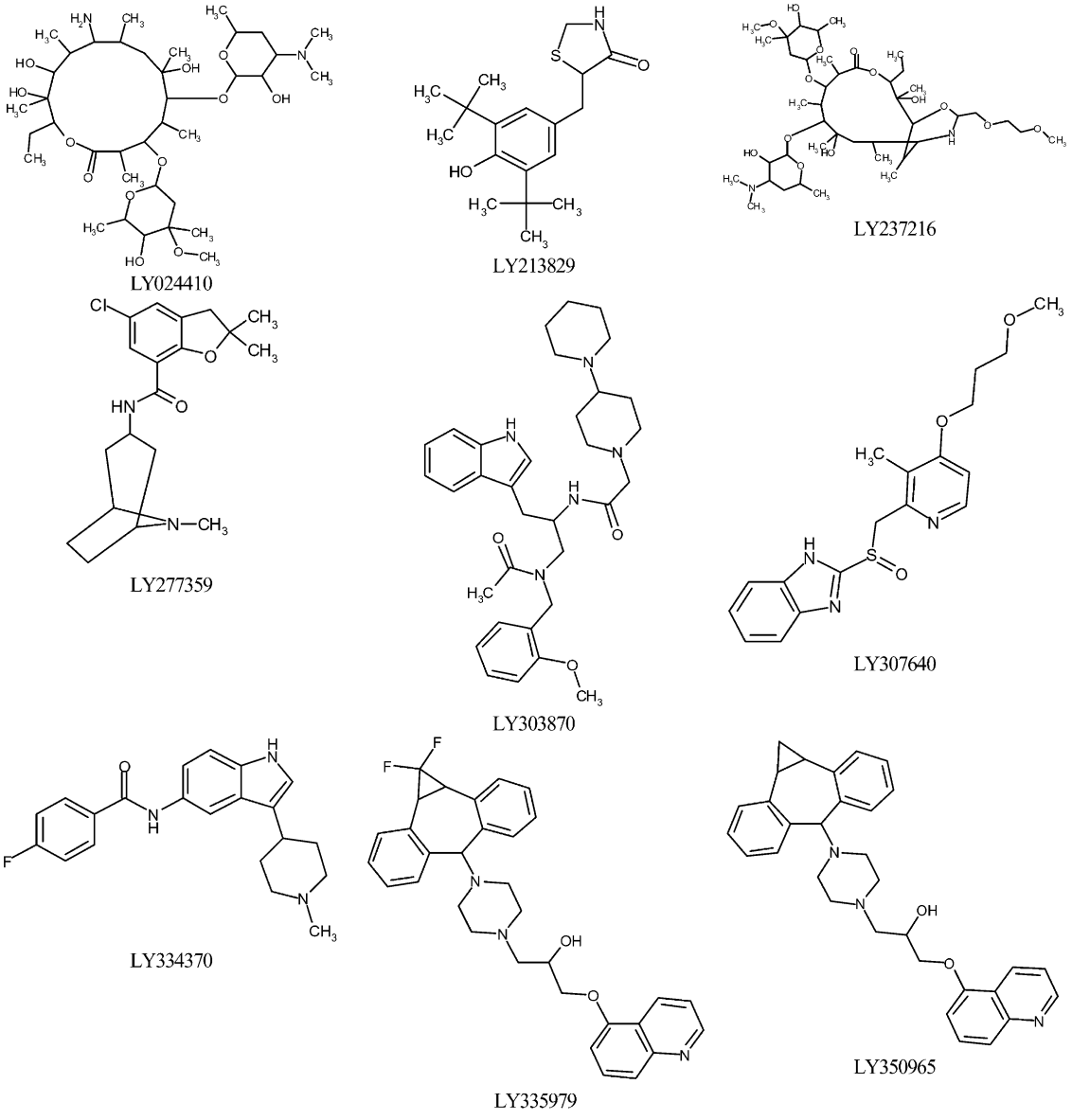


Figure 2. Cumulative distributions of inhibitory indices predicted for inhibitors and non-inhibitors in the 3A4 test set.

These limitations of 3D QSAR techniques prompted us to develop a fast neural network based method that predicts 3A4 inhibitory activity on the basis of 2D structural data. Our approach estimates the inhibitory action as an HTS-like yes/no classification and therefore does not require quantitative data.

3A4 inhibitors (INHIB) were extracted from Genetest's Human P450 Metabolism Database.[8] 3A4 substrates (SUBS) were used as non-inhibitors, since these compounds were definitely evaluated but no inhibitory action was reported. Compounds shared by both sets were removed from the substrate set. Next we assigned score values for the 3A4 inhibition of 0 to SUBS and 1 to INHIB entries. We used 145 inhibitors and 145 non-inhibitors, dividing INHIB and SUBS into training and test sets that include 109 and 36 molecules, respectively. In our previous work we found that 2D Unity fingerprint[9] can be used as a powerful 2D structural descriptor to discriminate CNS active and CNS inactive



Scheme 1. Competitive inhibitors of 3A4 synthesized at Lilly.

Table 1. Mean Tanimoto coefficients calculated between substrates (SU) and inhibitors (IN) in the training (TR) and test (TE) sets

	SUTR	SUTE	INTR	INTE
SUTR	1.0000			
SUTE	0.1838	1.0000		
INTR	0.2080	0.1847	1.0000	
INTE	0.2220	0.1941	0.2658	1.0000

Table 2. Calculated inhibitory index for competitive Lilly inhibitors of 3A4

Compd	3A4 inhibitory index
LY024410	0.86
LY213829	0.39
LY237216	0.87
LY277359	0.80
LY303870	0.86
LY307640	0.82
LY334370	0.80
LY335979	0.88
LY350965	0.88

compounds.[10] Therefore in this study we used Unity fingerprints again and generated them for the elements of the training and test sets using Sybyl.[11] Fingerprint bits were then transformed to 992 input neurons by an in-house software. All neural network operations were performed using the Stuttgart Neural Network Simulator (SNNS) program.[12] Using 992 input (bits of the Unity fingerprint), 31 hidden and one output (score) neurons we constructed feedforward nets. All input and output values are scaled between 0.1 and 0.9. The net was trained by the standard backpropagation algorithm as implemented in SNNS. Training set elements were mixed in each cycle to present them to the net in different orders. Training involved 2000 cycles with a learning rate of 0.2, which assured sufficient convergence in test runs. Overtraining (memory effect) was avoided by stopping the training process at the minimum of sum of square errors (SSE) calculated on the test set.

Our $992\times31\times1$ feedforward neural network was first used to re-evaluate the training set. Considering the borderline scoring value of 0.5 we found that the net classified 97% of inhibitors and 95% of non-inhibitors correctly. The predictive power of our net has also been demonstrated by the jack-knife approach. Retraining using randomly selected multiple training and test sets caused no difference in predictivity.

Next we evaluated the network on the preselected test set containing 36 inhibitors and 36 non-inhibitors never presented to the neural network before. Figure 1 shows the distribution of scores for the test set. This study was performed to check the robustness and the predictive power of the trained net.

Cumulative curves (Fig. 2) demonstrate that, using the borderline scoring value of 0.5, 91.7% of the inhibitors and 88.9% of the non-inhibitor compounds were classified correctly.

To show that the high predictive power achieved is not a consequence of structural similarity between the training and test sets we performed a comparative similarity analysis. Histogram of all Tanimoto coefficients calculated between training and test set entries of both inhibitors and non-inhibitors revealed that these compounds were not similar to each other (Table 1).

Our result proves that the trained net has a high predictive power in distinguishing between previously not considered compounds. Compounds available in the Genetest database are mainly known drugs. In a final test we calculated the 3A4 inhibitory index for drug candidates synthesized at Lilly (Scheme 1, Table 2).[3]

All these compounds were found to be competitive inhibitors of 3A4. Our trained net successfully classified eight out of the nine compounds to be an inhibitor.

Since tests above simulated the situation of virtual screening, the particular ability to select inhibitors from a never used dataset demonstrates the effectiveness of this approach for the virtual high throughput screening of 3A4 inhibitors.

References and Notes

1. Masimirembwa, C. M.; Thompson, R.; Andersson, T. B. *Comb. Chem. HTS* **2001**, *4*, 245.
2. Keserű, G. M. *J. Comp. Aided Mol. Des.* **2001**, *15*, 649.
3. Ekins, S.; Bravi, G.; Binkley, S.; Gillespie, T. A.; Ring, B. J.; Wikel, J. H.; Wrighton, S. A. *J. Pharmacol. Exp. Ther.* **1999**, *290*, 429.
4. Ekins, S.; Bravi, G.; Wikel, J. H.; Wrighton, S. A. *J. Pharmacol. Exp. Ther.* **1999**, *291*, 424.
5. Lemmen, C.; Lengauer, T.; Klebe, G. *J. Med. Chem.* **1998**, *41*, 4502.
6. Szklarz, G. D.; Halpert, J. R. *J. Comp. Aided Mol. Des.* **1997**, *11*, 265.
7. Korzekwa, K. R.; Krishnamachary, N.; Shou, M.; Ogai, A.; Parise, R. A.; Rettie, A. E.; Gonzalez, F. J.; Tracy, T. S. *Biochemistry* **1998**, *37*, 4137.
8. Genetest's Human P450 Metabolism Database (http://www.gentest.com/human/p450_database/index.html).
9. Unity 4.0.3. Tripos Inc., 1699 South Hanley Road, St. Louis, MO 63144, USA.
10. Keserű, G. M.; Molnár, L.; Greiner, I. *Comb. Chem. HTS* **2000**, *3*, 535.
11. Sybyl 6.5.3. Tripos Inc., 1699 South Hanley Road, St. Louis, MO 63144, USA.
12. SNNS. Stuttgart Neural Network Simulator, University of Stuttgart, 1995.