Assa Reichert, George Mihalaș, Lacramioara Stoicu-Tivadar, Stefan Schulz, Rolf Engelbrecht (Eds.)

Integrating Biomedical Information: From eCell to ePatient

Proceedings of the European Federation for Medical Informatics Special Topic Conference April 6-8, 2006 Timișoara, Romania
EFMI is a nonprofit organisation concerned with the theory and practice of Information Science and Technology within Health and Health Science in a European context. Its activities focus on annual conferences, MIE (Medical Informatics Europe), STC (Special Topic conference), tutorials and workshops. EFMI publishes proceedings and special issues of its activities in scientific journals. It has national societies in 27 member countries in Europe. Personal members are welcome in working groups. The EFMI portal provides very actual international news and information about the use of computers in health care.

http://www.EFMI.org

Working Groups

CARDS - Cards in Health Care, Social Security and Welfare
Tomas TRPISOVSKY, Prague; Claudia HILDEBRAND, Munich-Neuherberg (tomas.trpisovsky@ima.cz, hildebran@gsf.de)

EDU - Education in Health Informatics
John MANTAS, Athens; Arie HASMAN, Amsterdam (jmantas@cc.uoa.gr, a.hasman@amc.uva.nl)

EHR - Electronic Health Record
Bernd BLOBEL, Erlangen (bernd.blobel@iis.fraunhofer.de)

EVAL - Assessment of Health Information Systems
Elske AMMENWERTH, Innsbruck, Pirko Nykanen, Tampere, Jytte Brender, Aalborg/elske.ammenwerth@unit.at, Pirko.Nykanen@uta.fi, jytte.brender@v-ch.dk

IPAM - Information Planning and Modelling in Health Care
Bryan MANNING, Goring-by-Sea (bryan.manning@btinternet.com)

MCMS - MDMS, Case mix and Severity of Cases
Francis ROGER-FRANCE, Brussels; Jacob HOFDIJK, Leiden (Roger@infm.ucl.ac.be, jhofdijk@wanadoo.nl)

MICIT - Medical Informatics in Countries in Transition
Lacrimioara Stoiciu-Tivadar, Timisoara; Marcelo SOSA, Madrid (stoicu@aut.ucf.ro, MSosa@EuroParl.eu.int)

MIP - Medical Image Processing
Alexander Horsch, Munich (alexander.horsch@imse.med.tu-muenchen.de)

NLU- Natural Language Understanding
Robert BAUD, Patrick RUCH (Robert.Baud@sim.hcuge.ch, patrick.ruch@liettres.unige.ch)

NURSIE - Nursing Informatics in Europe
Patrick WEBER, Lausanne (patrick.weber@nicecomputing.ch)

OMI - Organisational Impact of Medical Informatics
Jos AARTS, Rotterdam (j.aarts@bmg.eur.nl)

PCI - Primary Care Informatics
Simon De Lusignan, London (slusignan@ghms.ac.uk)

SSE – Safety, Security and Ethics
Bernd BLOBEL, Erlangen (bernd.blobel@iis.fraunhofer.de)

ISBN 3-89838-072-6 (Aka)
ISBN 973-625-303-1 (Editura Politehnica)
Automatic analysis of gait parameters in depressive disorders – a pilot study
Virgil-Radu ENĂTESCU and Virgil ENĂTESCU

Time – Frequency Representation Of Cardiac Rhythm Dynamics
Mariana MOGA, Victor Dan MOGA and Corina VERNIC

Data Mining

Mining RR and QT variability to detect risk and prevent sudden death in asymptomatic subjects
Radu M NEGOSC and Serban DINCA-PANAITESCU

An OpenSource Data Mining Application for Outcomes Measurement in Emergency Room
Miron IANCU

A PHP-based Native Bayes Classifier for Emergency Department Prediction of Car Accident-induced Trauma
Miron IANCU

Advanced Methods for Biomedical Data Processing

Molecular Descriptors Family on Structure-Activity Relationships on anti-HIV-1 potencies of HEPT and TIBO derivatives
Sorana BOLBOACĂ, Stefan ȚIGAN, and Lorentz JĂNTSCHI

Medical Predictions in Hepatitis Infection
Adriana ALBU, Andreea DRĂGULESCU and Karoly MELYHARDT

PROGNOSIS METHODS to evaluate the evolution in time of the biological parameters at HCV patients
Andreea DRĂGULESCU, Karoly MELYHARDT and Adriana ALBU

Independent Component Analysis applied in motor imagery
Mihaela UNGUREANU, Rodica STRUNGARU and Stefan COSOSCHI

Medical Information –Philosophically Issues

Medical Error: What would Hippocrates do?
Radu LECA

The verbalization filter model from the informational level of the psychic perspective
Virgil ENĂTESCU, Virgil-Radu ENĂTESCU and Ileana ENĂTESCU

Electronic Health Record

The role of an information network in the application of secondary prevention strategies for cardiovascular diseases (Medprax - Romania)
Pal KIKELI, Mihaela HUBATSCH, Zolian PREG, Claudia IONESCU, Piroska JAKAB, Katalin Rakosi KOSA
Molecular Descriptors Family on Structure-Activity Relationships on anti-HIV-1 potencies of HEPT and TIBO derivatives

Sorana BOLBOACĂ, Stefan TÎGAN, and Lorentz JĂNTSCHI

Jiljiu Hatieganu University of Medicine and Pharmacy, Technical University of Cluj-Napoca,

Abstract A new developed methodology of Structure-Activity Relationships (SAR) was applied on a set of 57 compounds with known inhibition activity of immunodeficiency virus type 1. The methodology uses an original family of molecular structure descriptors called Molecular Descriptors Family. With a set of multiple linear regression analysis programs, the whole set of MDF members were crossed in order to find the best SAR model. The obtained model allows making important remarks on structure-activity links. The disadvantage of time consuming to analyze the entire set of descriptors is compensated by better structure-activity relationships.

Keywords: HIV-1 inhibitors, Structure Activity Relationships (SAR), Molecular Descriptors Family (MDF)

1. Introduction

Two different types of human immunodeficiency viruses (HIV-1 and HIV-2) differing in nucleotide and amino-acid sequences are responsible by the acquired immunodeficiency syndrome, but the HIV-1 type is most predominant [1].

A previous study analyzed the HEPT and TIBO derivatives potencies on HIV-1 [2] using quantitative structure-activity relationships methodology. The results obtained by Toporov & all are:

\[
\begin{aligned}
\text{n} &= 57; \text{r} = 0.9397; \text{s} = 0.520; \text{F} = 416 \text{ (all compounds)} \\
\text{n} &= 37; \text{r} = 0.9426; \text{s} = 0.513; \text{F} = 279 \text{ (training set)} \\
\text{n} &= 20; \text{r} = 0.9408; \text{s} = 0.547; \text{F} = 139 \text{ (test set)}
\end{aligned}
\]  

where \( n \) is size of the sample; \( r \) is the correlation coefficient; \( s \) is standard error and \( F \) is Fisher parameter.

Starting with the integration of complex structure information of HEPTA and TIBO derivatives, the aim of the research was to evaluate the ability of molecular

---

7 13 Emil Isac, 400023 Cluj-Napoca, Romania, sbolboaca@umfcluj.ro
8 15 Constantin Daicoviciu, 400020 Cluj - Napoca, Romania
descriptor family structure-activity relationships in modeling of the inhibition effectiveness against HIV-1.

2. Material and Method

A number of nineteen HEPT derivatives and thirty-eight TIBO derivatives with inhibition properties on HIV-1 were included into the study. The effectiveness in inhibiting HIV-1 of HEPT and TIBO derivatives (two groups of reverse transcriptase inhibitors) was taken from a previous paper [3] and is expressed as the concentration of compound required to achieve 50% protection of MT-4 cells against the virus (called log(10^6/C90)).

The use of a new original set of molecular descriptors, called Molecular Descriptors Family (MDF) into a Quantitative Structure-Activity Relationship was apply in order to study the inhibiting HIV-1 activity of 19 HEPT and 38 TIBO compounds. The steps of molecular descriptor family on structure activity relationship (MDF SAD) modeling are [4]:

- Step I: Sketch of HEPT and TIBO compounds by the use of HyperChem software [5];
- Step II: Create the file with measured inhibiting HIV-1 activity (Y_{C50}) of HEPT and TIBO compounds;
- Step III: Generate the MDF members based on topological and geometrical representations of the compounds. There were identified 296965 MDF members with real and not identical values from which only 95277 were distinct each from other. More, considering also the withdrawing of planar dependencies (one descriptor is dependent on other two) it remains only 84408.
- Step IV: Finding the SAR models for HEPT and TIBO compounds. The selected members enter into multiple linear regression analysis. Mono-varied and multi-varied models were applied. At the end of all pair’s computations the best QSAR models were selected and presented here. Note that for bi-varied model, 3562313028 pairs enters into bi-varied regression model and a multiple of enters into tri- and more varied models.
- Step V: Validation of the obtained SAR models were performed through computing the cross-validation leave-one-out correlation score [6], and the difference between this parameter and the squared correlation coefficient.
- Step VI: Analyze the selected SAR model and comparing it with previous reported model.

3. Results

The best performing SAR (five-varied model) was selected and is presented here. The selection of the best performing five-varied model was made first after the greatest squared correlation coefficient and then after the greatest values of cross-validation leave-one-out (loo) score (r^2_{cv(loo)}).

The models have the following equation:

\[ Y = 17.7 - 7.11 \cdot InMdTHg - 1.23 \cdot fDMeET + 8.36 \cdot AlMrKQr + 6.59 \cdot 10^3 \cdot ImDMtQr + 5.98 \cdot ImdEMg \]
where $\hat{Y}$ is predictor of measured inhibition activity ($Y_{C50}$) and InMTHg, IFDM$\text{wEr}$, AiMrKQt, ImDMtQt, and ILMdEMg are molecular descriptors.

The characteristics associated with the above-described models are in table 1 and is graphically represented in figure 1.

![Figure 1. Measured inhibition activity (Measured) vs. estimated (Estimated) with five-varied SAR model](image)

Assessment of the MDF SAR model was performed by applying a correlated correlation analysis, which took into consideration the five-varied SAR models and compared it with the best performing (model with four variables, $r = 0.9397$, $n = 57$ – equation 1) previous reported model [2] by the use of Steiger's Z test. The results of comparison are in table 2.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r(Y_{C50}, \hat{Y}_{SAR})$</td>
<td>0.9579</td>
</tr>
<tr>
<td>$r(Y_{C50}, \hat{Y}_{\text{previous}})$</td>
<td>0.9399</td>
</tr>
<tr>
<td>$r(\hat{Y}<em>{SAR}, \hat{Y}</em>{\text{previous}})$</td>
<td>0.9252</td>
</tr>
<tr>
<td>Steiger's Z test parameter</td>
<td>1.3462</td>
</tr>
<tr>
<td>$p_{\text{Steiger's Z}}$ (%)</td>
<td>0.0891</td>
</tr>
</tbody>
</table>

4. Discussions

The selected best found five-varied SAR model of HEPT and TIBO QSAR HIV-1 inhibiting activity shows that atoms mass and attached hydrogen's has significance on activity behavior using geometrical model of the molecule (InMTHg and ILMdEMg). Partial charge has significance on activity behavior using strictly topological model. More, two descriptors use the Qt association in the selected best found five-varied model (AiMrKQt and ImDMtQt).

The atomic and group electronegativity, as a composed property tends with increasing of number of descriptors to be replaced by more accurate properties:
attached hydrogen's, partial charge and mass. Thus, if bi-varied model has only atomic and group electronegativity as atomic descriptors, in tri-varied model disappear one electronegativity based descriptor and appear one attached hydrogen's based and one partial charge based, and for five-varied are two descriptors based on partial charge, one based on attached hydrogen’s and one based on atomic mass.

Looking at the five-varied model, we can say that the inhibitory activity it is of molecular topology as well as molecular geometry and depend on partial change of molecule, molecular mass and number of bounded hydrogen's. The values of squared correlation coefficient ($r^2 = 0.9175$), the student parameter, associated p-values and 95% confidence intervals (see table 1) demonstrate the goodness of fit of the five-varied MDF SAR model. The power of the five-varied model in prediction of the inhibitory activity of compounds is demonstrate by the cross-validation leave-one-out correlation score ($r^2_{cv(loo)} = 0.8997$). The stability of the best performing five-varied MDF SAR model is give by the difference between the squared correlation coefficient and the cross-validation leave-one-out correlation score ($r^2 - r^2_{cv(loo)} = 0.0178$).

Comparing with previous reported model (equations (1)) [2], our model (equations (2)) is better ($r^2 = 0.918$ – see table 1 and equation 1). At modeling level, our approach is more software independent than previous reported. We use software dependent procedures only for constructing a basic geometrical model of the molecules and compute the partial charge distribution inside the molecules. We do not “optimize” the geometrical shape according to an arbitrary choosed model and/or algorithm.

Starting with the knowledge learned from the studied set, inhibition property of new compound from the same class can be predicted by the use of an original software, which is available at the following address:

http://vl.academicdirect.org/molecular_topology/mdf_findings/sar/

Thus, the software is able to predict the inhibitory activity of new compounds from the same class with low costs. It can be concluded that the use of MDF for SARs finding on HIV-1 potent compounds offers accurate models and allow making of important remarks about structure-activity links.

5. Acknowledgement

The research was supported in part through project ET36/2005 by UEFISCSU Romania.

References
