



Toetab TÜ ja TTÜ doktorikool
"Funktsionaalsed materjalid ja tehnoloogiad" (FMTDK)

**ESF projekt 1.2.0401.09-0079**

# CHEATING IN QSAR: HOW TO OBTAIN STATISTICALLY BRILLIANT BUT OTHERWISE USELESS MODELS

Kalev Takkis[1]

[1]*Institute of chemistry, University of Tartu, Ravila 14a, Tartu, 50411, Estonia*

e-mail: kalev.takkis@ut.ee

The question of how to choose the "correct" conformation has been a headache for QSAR community for years. Several studies have looked into the problem, describing the variance of the calculated descriptors for different conformations but none of the proposed solutions has been widely adopted. In current work, the effect of the geometry on the resulting QSAR models is being studied. For this purpose conformational search was conducted on a set of HIV protease inhibitors and activities for all conformers according to previously published model[1] were calculated. The results indicate that when choosing the conformation with the least prediction error the statistics of the new model improve significantly yet this kind of "geometry optimisation" is almost impossible to detect. This opens up a possibility to use this method to develop models that appear to be great but are in essence useless. An interesting observation about the outliers of the original model was that they remained outliers also in the new model. Whether this is a potential way to identify the model's applicability domain remains a topic of the further study.

References:

1. Takkis, K; Sild, S; QSAR Modeling of HIV-1 Protease Inhibition on Six- and Seven-membered Cyclic Ureas, *QSAR & Combinatorial Sciences,* 2009, 28, 52 - 58