# Discriminating between Drugs and Nondrugs by Prediction of Activity Spectra for Substances (PASS)

http://www.akosgmbh.de/pass/articles/J.%20Med.%20Chem.%2044%202001.pdf

*Soheila Anzali,Gerhard Barnickel, Bertram Cezanne, Michael Krug, Dmitrii Filimonov, and Vladimir Poroikov*

*J. Med. Chem.* **2001**, *44*, 2432-2437

**Max Shneider**<br>
CS379A Case Study<br>
3/17/2006

The purpose of this paper is to describe an ADMET application that accurately predicts whether a compound is a drug or not. As we have learned in class, it is important to not only find molecules that bind to a certain target, but to also ensure that they satisfy the Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) requirements of the human body. The PASS system could potentially be used at the beginning of the drug discovery process to avoid doing unnecessary work.

PASS is a computer system that predicts more than 500 biological activities using a regression technique. Biological activities result from the interaction of chemical compounds and biological entities (for instance, the human body), and include such things as pharmacological main and side effects, mutagenicity, carcinogenicity, and embryotoxicity. PASS uses a training set to build a classifier that predicts whether the compounds in the test set are drugs or nondrugs. It has been shown to have a general prediction accuracy of around 86% in leave-one out (LOO) cross validation (which takes data from the training set to test itself with).

Both the training and test set compounds are based on their 2D structure, using descriptors called “multilevel neighborhoods of atoms” (MNA). There are several levels of these descriptors, but only the first and second are used in PASS (which describe atoms and their direct neighbors). The training set is composed of 5,000 drugs from the World Drug Index (WDI) database, and 5,000 nondrugs from the Advanced Chemicals Directory (ACD). The mean prediction accuracy using this training set in LOO cross-validation was 79.9%, which was comparable to other drug and nondrug prediction results (Sadowsky and Kubinyi).

Three different test sets were used, with each filtered to remove items that already existed in the training set, as well as compounds that had errors in their structural formulas. The first test set consisted of 864 launched and registered drug compounds from the Cipsline database (LR), of which PASS predicted 78.5% as drugs and 21.5% as nondrugs. The second test set contained 9,484 nondrug compounds with properties unfavorable for drugs, such as reactive groups or low molecular weight (NR). PASS predicted 83.8% of these as nondrugs and 16.2% as drugs. The third test set consisted of 88 drug compounds from the list of top-100 prescription pharmaceuticals (TOP100), and PASS predicted 87.5% of them as drugs and 12.5% of them as nondrugs. The improvement over the first test set results can be explained by the fact that these are well-known, established drugs, as opposed to the relatively new drugs from before.

As a further experiment, they replaced the WDI/ACD training set with a combination of the LR and ND test sets, since the original databases were relatively noisy (some compounds that were labeled as drugs were really nondrugs, and vice versa). The new LOO cross-validation was 89.9%, a vast improvement over the original 79.9%. They also re-tested TOP100, with 95.5% drug and 4.5% nondrug predictions, which were also much better than the previous 87.5% and 12.5% results.

PASS is relatively successful on new compounds that have nontraditional structures or belong to new chemical classes, since they don’t have to be the same as the items in the training set. It is also very fast (the prediction of one compound only takes 4 ms on a 3 MHz computer), so it can be run in the beginning of the drug discovery process on a large database of compounds. The experiments show that the chemical descriptors and algorithms in PASS give high predictive accuracy when discriminating between drugs and nondrugs, and that it performs favorably when compared to other methods in the field. The results were obtained without any changes to the PASS program, which opens the door to even better discrimination if more specific drug information is included.

These arguments clearly show that PASS is an effective tool for predicting drug-like properties. However, the paper was published in 2001, and new techniques have probably been developed since that time. The question remains as to whether PASS has continued to compete with these modern methods.