

## Development of QSAR model for high-speed *in-silico* identification of potentially phototoxic organic compounds

### Introduction

The phototoxic effects of a chemical compound are of concern in numerous areas of chemistry-related industry. Pharmaceuticals, cosmetics, food additives, cleaning agents are just few examples of products that come into frequent contact with the human organism and may cause harm by means of light assisted toxicity.

The objective of this study was to create a Quantitative Structure-Activity Relationship (QSAR) computer model that could be used for rapid *in-silico* assessment of chemicals' potential to cause harmful phototoxic effects, given its structure.

### Data Collection

The structures of 114 compounds known to be phototoxic to humans were retrieved from the literature (Ref. 1-7). The same literature also yielded 36 compounds that did not exhibit noticeable phototoxic effects. Additionally, 78 compounds routinely used in cosmetic products were added to the non-phototoxic part of the training set, yielding a balanced set and increasing its chemical diversity. The structures of all studied compounds are presented in Figure 1.

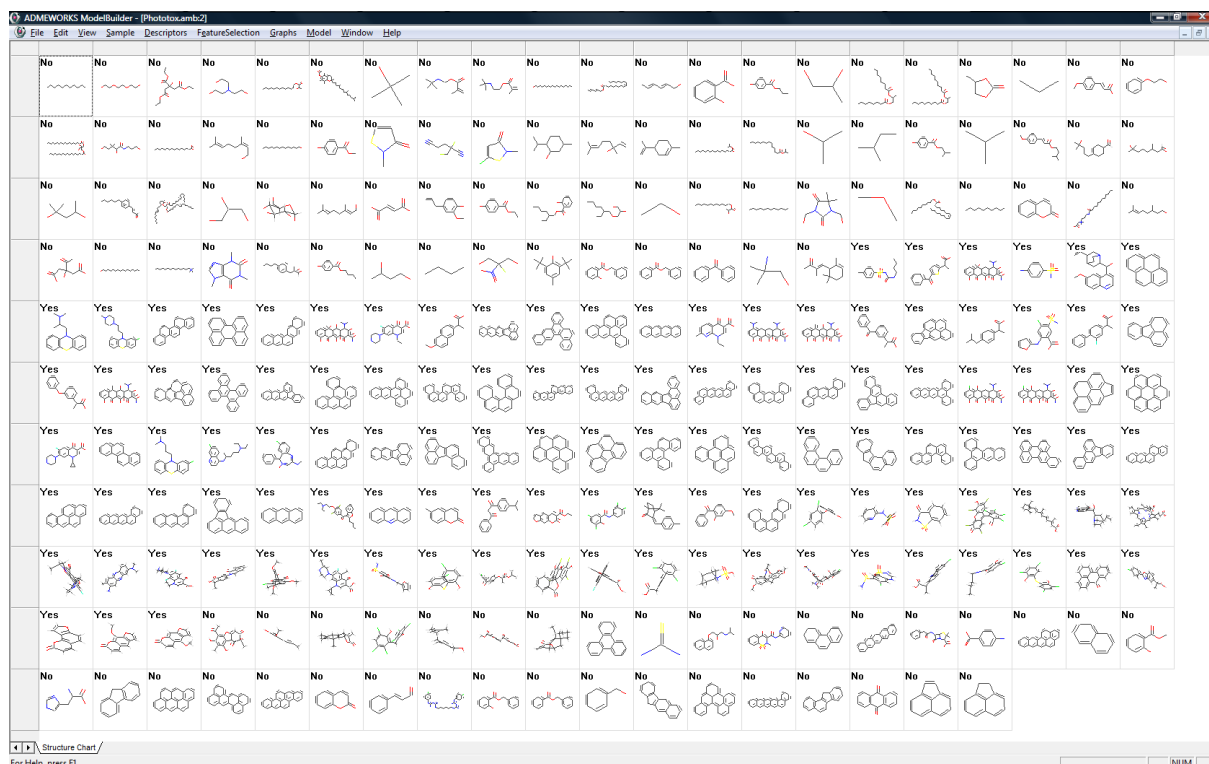


Fig. 1 Compounds used in the study and their phototoxicity.

### Model Building

The QSAR analysis aims to explain the observed (experimental) property by a mathematical expression of "descriptors" - numerical values that may be calculated for a given compound straight from its chemical structure.

The initial choice of descriptors depends upon the assumed mechanism of the reaction. As phototoxicity was the objective of this study, it was assumed that the descriptors related to the

molecules' quantum properties, shape, charge distribution and existence of specific structural parts may be of importance to the observed activity.

The structures of all compounds were put in an industry standard Structure Data File and all further analyses were performed in the ADMEWORKS ModelBuilder software (Ref. 8). 152 descriptors were calculated for the whole training set. The quantum and charge descriptors were calculated using a fast and robust AM1 semiempirical method.

A subsequent data set analysis was performed using the Particle Swarm Optimization algorithm for the feature selection. Next, the 19 descriptors (3 topological, 4 substructure-count and 12 quantum/charge) with the highest potential for explaining the experimentally determined phototoxicity were selected. A Linear Discriminant Function (LDA) model was built using the Stochastic Gradient Perceptron algorithm.



Fig. 2 Populations of descriptors for the training set phototoxic compounds - red, non-phototoxic - green.

## Results and Conclusion

The Figures 2-4 illustrate the relationships between descriptors' values and the observed phototoxicity. The population analysis yields no conclusive results as to the significance of any single descriptor to the phototoxicity. However, both the clustering and principal component analysis show clearly noticeable tendencies in the training set - the samples with the same phototoxic properties tend to form continuous regions, which is an indication of an existing order in the training set, making it suitable for the creation of a QSAR model.

The statistical parameters of the final LDA model created are as follows:

- Overall classification rate: 96.05 %
- Phototoxic compound classification rate: 100%
- Non-Phototoxic compound classification rate: 92.11%
- Leave-1-out internal validation rate: 92.54%

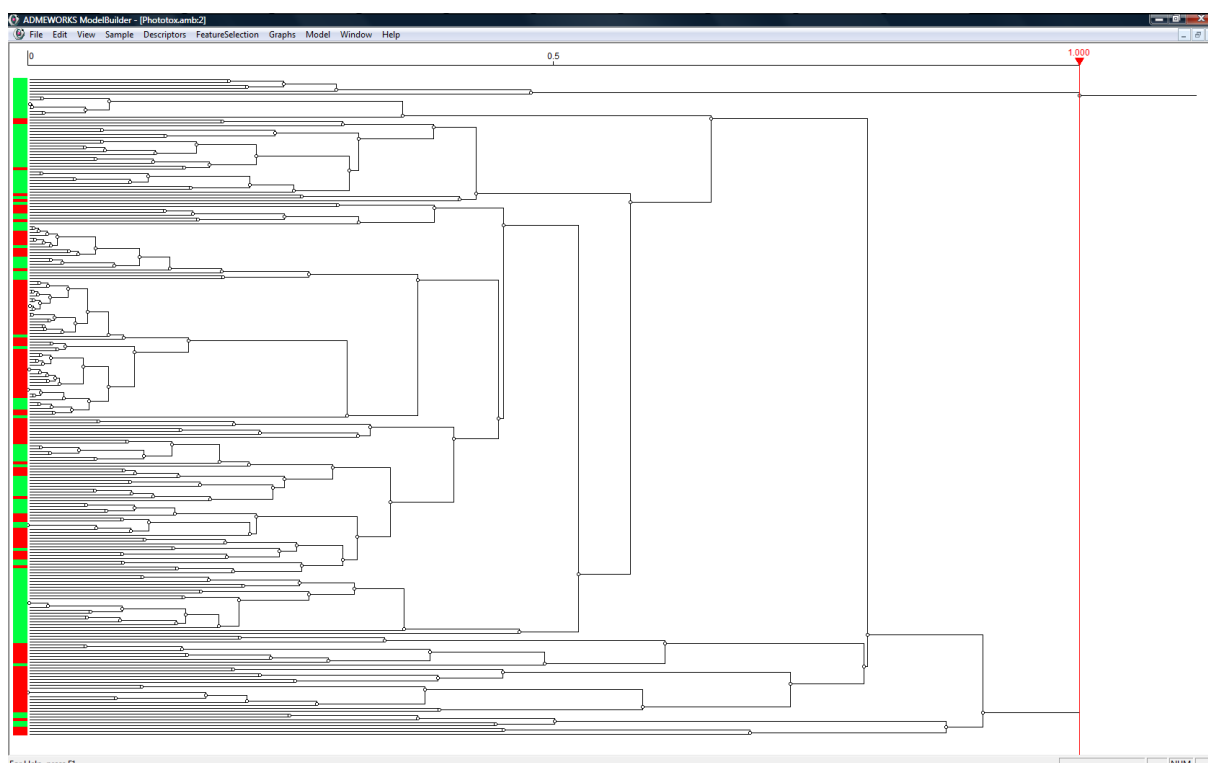


Fig. 3 Clustering of compounds in multidimensional descriptor space.

Very high overall classification rate (by Leave-1-out cross-validation) as well as 100% classification rate of the phototoxic compounds show the potential of the model for practical use in filtering compounds with unwanted phototoxic effects.

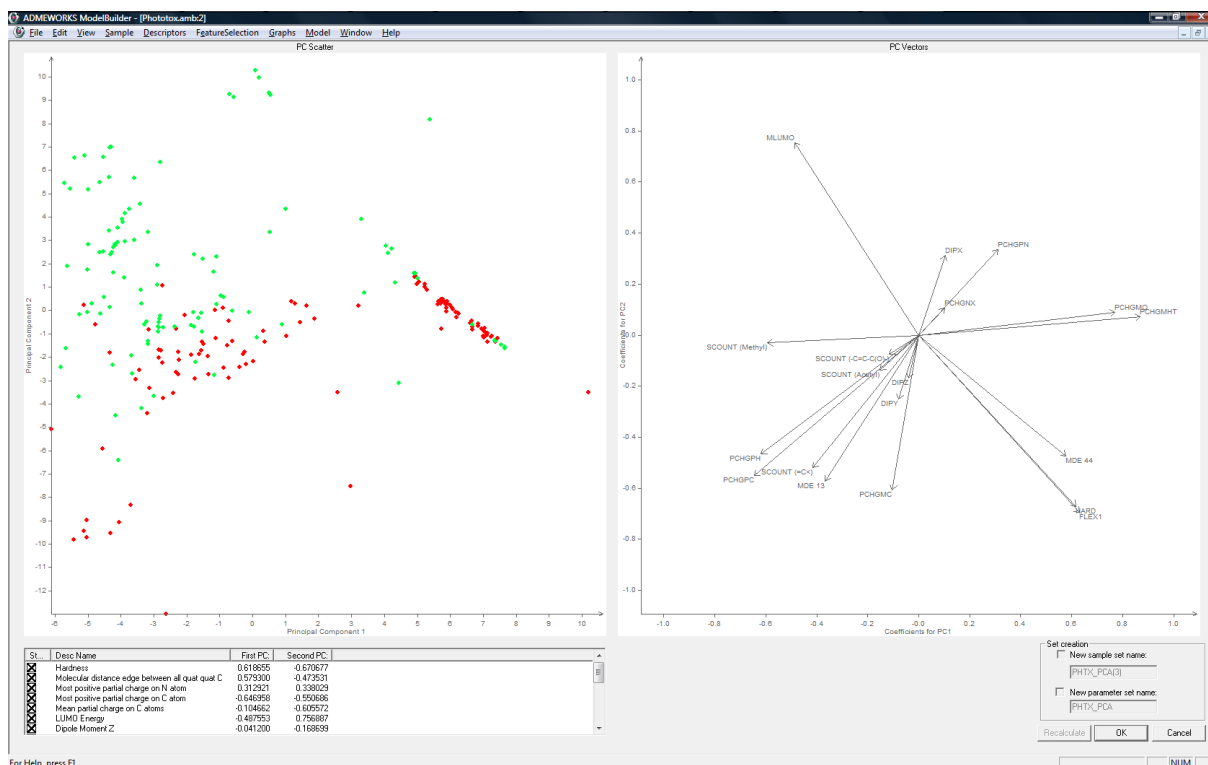


Fig. 4 Principal component analysis of descriptor values.

## References

1. F.A. de Lima Ribeiro, M.M.C. Ferreira, J. Mol. Struct.: THEOCHEM, 719, 2005, s. 191–200.
2. Dr. D. Forbes, The Toxicology Forum - Winter Meeting 2000.
3. H. Spielmann et. al., ATLA, 22, 1994, s. 314-348.
4. H. Spielmann et. al., Toxicology in Vitro, 12, 1998, s. 305-327.
5. J. Cotovio, et. al., AATEX, 14, Special Issue, s. 389-396.
6. K. H. Kaidbey, A. M. Klingman, J. Invest. Dermatol., 70, 1978, s. 272-274.
7. J. Ferguson, R. Dawe, J. Antimicrob. Chemother., 40, 1997, Suppl. A, s. 93–98.
8. ADMEWORKS ModelBuilder, Fujitsu Kyushu Systems Ltd. 2010