

## Abstract

*J. Chromatogr. A*, **1028**, 287-295 (2004)

Garkani-Nejad Z.<sup>1</sup>, Karlovits M.<sup>2</sup>, Demuth W.<sup>2</sup>, Stimpfl T.<sup>3</sup>,  
Vycudilik W.<sup>3</sup>, Jalali-Heravi M.<sup>4</sup>, Varmuza K.<sup>2</sup>

### **Prediction of gas chromatographic retention indices of a diverse set of toxicologically relevant compounds.**

- [1] Faculty of Science, Vali-e Asr University of Rafsanjan, Iran
- [2] Institute of Chemical Engineering, Vienna University of Technology, Austria
- [3] Institute of Forensic Medicine, University of Vienna, Austria
- [4] Department of Chemistry, Sharif University of Technology, Tehran, Iran

For a set of 846 organic compounds, relevant in forensic analytical chemistry, with highly diverse chemical structures, the gas chromatographic Kovats retention indices have been quantitatively modeled by using a large set of molecular descriptors generated by software *Dragon*.

Best and very similar performances for prediction have been obtained by a partial least squares regression (PLS) model using all considered 529 descriptors, and a multiple linear regression (MLR) model using only 15 descriptors obtained by a stepwise feature selection. The standard deviations of the prediction errors (SEP), were estimated in four experiments with differently distributed training and prediction sets. For the best models SEP is about 80 retention index units, corresponding to 2.1–7.2% of the covered retention index interval of 1110–3870.

The molecular properties known to be relevant for GC retention data, such as molecular size, branching and polar functional groups are well covered by the selected 15 descriptors.

The developed models support the identification of substances in forensic analytical work by GC–MS in cases the retention data for candidate structures are not available.