5\_p7\_p

Quantitative prediction of K values

- Introduction
- Fragment models
- sp-LFERs
- pp-LFERs
- Comparison of the various methods
- Predictive models based on molecular structure

• Critical remarks on approaches from chemical engineering

- Selftest
- ▼ Problems
- ↓ Problem 1
- ↓ Problem 2
- ↓ Problem 3
- ↓ Answer
- ↓ Problem 4
- ↓ Answer
- ↓ Problem 5
- ↓ Answer
- ↓ Problem 6
- ↓ Answer
- ↓ 🌻 Problem 7
- ↓ 🌔 Problem 8
- ↓ Problem 9

## **Problem 7**

In Tenax\_results.xls, sheet 2 (reference: Schneider, M. and K.-U. Goss *Anal. Chem.* **2009**, *81*, 3017-3021. Journal link Download pdf ) you will find the compound descriptors that you will need in order to set up a pp-LFER. Go ahead and try it!

As you may have noticed, there are no compounds in the data set that do have an A-value > 0 which means that there is no H-bond donating compound. Hence, this data set is certainly not sufficient to establish a complete pp-LFER equation for Tenax, because we will not be able to draw any conclusions on the H-bond accepting properties of Tenax. For a more complete picture, it would therefore be desirable to add some H-bond donating compounds such as alcohols. Such data are shown in sheet 3. With these we can try to establish now a pp-LFER by calculating a multiple regression (sheet 3). If we use the resulting parameters we can backcalculate the experimental values in order to see how good our equation fits the data (sheet 3). It appears that the result is very satisfying.

According to our theoretical consideration, this model should also be suitable to predict the sorption of other data that had not been included in the model calibration. On sheet 4 the result of such an evaluation can be seen. It appears that the resulting predictions are ok for most compounds with the exception of the highly fluorinated compounds in line 44 - 48. The problem here is that the calibration done on sheet 3 was completely insufficient because the used data set was not diverse enough. Diverse means that the whole chemical space should be covered, (i.e. a large part of the range of the single descriptors should be covered and –even more important– the descriptors of the compounds in the calibration data set must not be cross correlated). In practice, it is often quite challenging to measure experimental data for such a diverse data set. The specific problem here was the strong cross correlation between the V<sub>i</sub> and L<sub>i</sub> descriptor in the calibration set and the completely different properties of the highly

fluorinated compounds in this respect. On sheet 5 we have recalibrated the pp-LFER model including the experimental data of the highly fluorinated molecules. The resulting model is able to describe all data well, much better than the the calibrated model on

sheet 3. Note that the r<sup>2</sup> and se in the prediction are similar to those on sheet 3. This is a nice example for the very limited information that one can gain from these two statistical parameters. What could have made you suspicious on sheet 3 are the large standard errors (se) in the coefficients. These errors are much smaller on sheet 5. The problem shown here is a general one for highly fluorinatd compounds: any kind of calibrated model should only be applied to highly fluorinated compounds if such compounds had been part of the calibration data set. For more information see Goss, K.-U. and G. Bronner *J. Phys. Chem. A* **2006**, *110*, 9518–9522. Journal link Download pdf . The general problem of a cross correlation of descriptors in the calibration data set, however, is not at all limited to highly fluorinated compounds or to the descriptors L<sub>i</sub> and V<sub>i</sub>.

