COMPARISON OF TRANSFORMATION FUNCTIONS ON THE PREDICTION OF RETENTION IN NANOFILTRATION BY MOLECULAR DESCRIPTORS

S. KOTER[1)], B. GILEWICZ-ŁUKASIK[1)], A. NOWACZYK[2)], J. NOWACZYK[2)]

[1)]Faculty of Chemistry, Nicolaus Copernicus University, ul. Gagarina 7, 87-100 Toruń, Poland; skoter@chem.uni.torun.pl
[2)]Faculty of Pharmacy, Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University, ul. Sklodowskiej-Curie 9, 85-094 Bydgoszcz, Poland

1. INTRODUCTION

The separation properties of filtration membranes (ultrafiltration, nanofiltration, reverse osmosis) are characterized by the retention coefficient:

$$R \equiv 1 - c_{permeate} / c_{feed} \tag{1}$$

where c is the concentration of solute. Usually it is correlated with the molecular weight of filtered solute (MW). This correlation is relatively high in the case of ultrafiltration, where MW of solutes to be rejected is high (2 – 100 kDa) [1, 2]. When MW and the dimensions of a solute molecules decrease, more and more important become their interactions with a solvent and membrane pore walls. Thus, in the case of nanofiltration and reverse osmosis MW as a measure of membrane retention abilities fails.

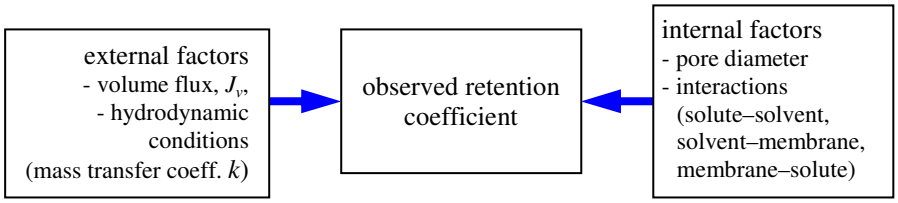

Fig.1. Factors influencing retention coefficient

In Fig.1 the factors influencing R are given. A correct estimation of R (at given J_v and k) would be possible if all the interactions were taken into account. However, the membrane producers do not reveal details on the chemical composition of membrane active layer. Thus, only an analysis of

the *R* dependence on the solute molecular structure for a given membrane-solvent system is possible. In our previous paper we discussed the correlation between *R* and molecular descriptors applying a linear relationship [3]. In this work we extend this analysis to nonlinear dependencies using two-parameter sigmoid functions (logistic, Gompertz, and arctangent). The advantage of such functions is that the range of solutes can be extended to those totally rejected or showing *R* close to zero. The finding of descriptor which well correlate with *R* has a practical significance – it would facilitate the choice of a membrane suitable for a given purpose.

2. MOLECULAR DESCRIPTORS AND SIGMOID FUNCTIONS

The descriptors were calculated using the program Dragon ver. 5.5 enabling the calculation of 1664 descriptors. Before the calculation of descriptors the geometry of each molecule were optimized at the ab initio level of theory using Gaussian®. More details on the calculation and initial selection of descriptors is given in [3].

The sigmoid functions tested in this work are as follows:

1) logistic function

$$R = \frac{1}{1+\exp(Y)} \qquad Y = \ln\left(\frac{1}{R} - 1\right) \tag{2a,b}$$

2) Gompertz function

$$R = \exp\left(-\exp(Y)\right) \qquad Y = \ln\left(-\ln(R)\right) \tag{3a,b}$$

3) arctangent

$$R = 0.5 + \arctan\left(Y\right)/\pi \quad Y = \tan\left(\pi(R-0.5)\right) \tag{4a,b}$$

where *Y* is expressed by a linear function of descriptor, *d*:

$$Y = a_0 + a_1 d \tag{5}$$

Contrary to the linear dependence *R*(*d*), the above transformations always yield R $\in$(0, 1) when Y changes from -∞ to +∞.

The applicability of the tested equations were ascertained on the base of the squared correlation coefficient, r^2, and the cross-validation coefficient, Q^2_{cv}, calculated according to the leave-one-out method [4]:

$$Q^2_{cv} = 1 - \sum_{i=1}^{n}\left(R_i - R_{i,cv}\right)^2 / \sum_{i=1}^{n}\left(R_i - \overline{R}\right)^2 \tag{6}$$

In Eq.(6) R_i is the experimental value of retention coefficient of solute *i*, $\overline{R}$ – arithmetic mean of R_i, *i*=1,..,*n*. $R_{i,cv}$ is calculated from Eqs.(2-5), where a_0 and a_d are calculated taking the data for *n*-1 solutes (excluding *i*-th solute).

The retention-descriptor relationship were tested using the experimental data on the organic solvent nanofiltration published by White [5]. These data are related to the filtration of a toluene solution containing a mixture of saturated and aromatic hydrocarbons (*n*-decane, 1-methyl-naphthalene, *n*-hexadecane, 1-phenylundecane, pristane, *n*-docosane) through a polyimide membrane.

3. RESULTS AND DISCUSSION

In Fig.2a it is seen that the correlation between the retention coefficient, *R*, and the molecular weight of solute, *MW*, is too poor to be used in the prediction of *R*. Applying a linear approximation *R(d)* only two descriptors (H4p and RTu from the GETAWAY descriptors) were found fulfilling the conditions: the determination coefficient r^2 exceeds 0.99 and the cross-validation coefficient Q^2_{cv} is not lower than 0.98.

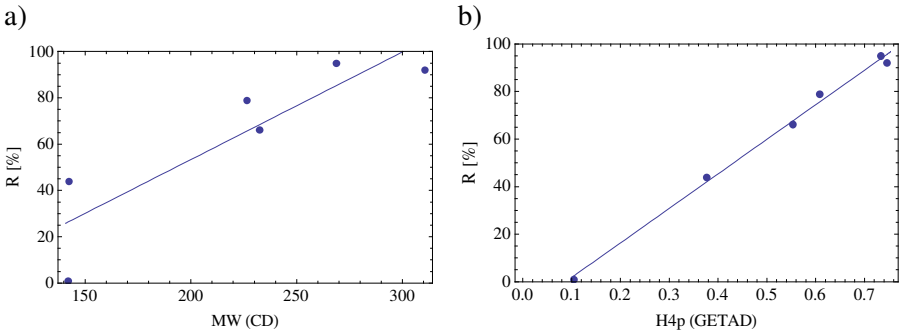

Fig.2. Dependence of *R* on a molecular descriptor of solute, a) molecular weight, MW, b) H4p (GETAWAY descriptor)

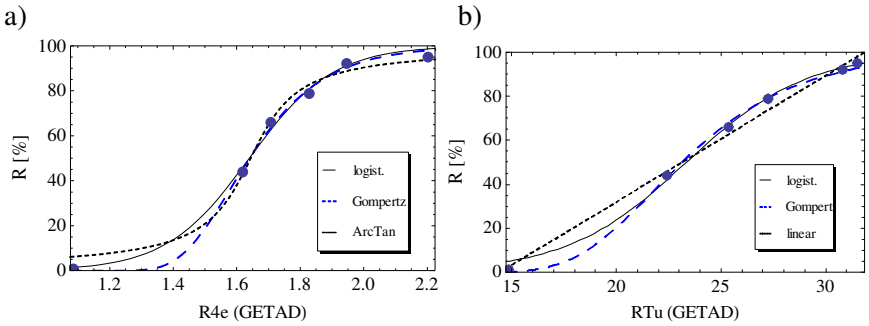

Fig.3. a) R4e – the best descriptor common for all the tested sigmoid functions, b) RTu – the best common descriptor for linear, Gompertz and logistic functions

Applying the sigmoidal functions the number of descriptors satisfying the above mentioned criteria is higher. For Gompertz function – 8, for logistic function – 4, whereas for arctangent only 2 such descriptors were found. For all these descriptors r^2 is higher than 0.998. The best descriptors are: HATS5e (GETAWAY), MAXDP (topological descriptor), R4u (GETAWAY) for Gompertz, logistic, arctangent functions, respectively. It should be noted that the descriptor R4e (GETAWAY) fulfills the condition $Q^2_{cv} \geq 0.98$ for all the sigmoid functions simultaneously (Fig.3a), whereas RTu (GETAWAY) fulfills that condition for the linear, Gompertz and logistic functions (Fig.3b).

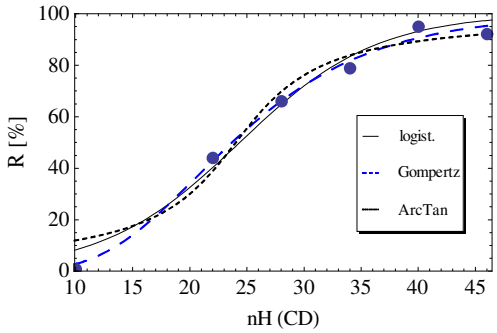

Fig.4. nH – for Gompertz function $r^2 = 0.9986$, $Q^2_{cv}=0.987$

The above mentioned descriptors are of rather complicated nature and there is no place here to explain their meaning in detail. However, among the selected descriptors there is one of a simple meaning – nH from the constitutional descriptors which denotes the number of hydrogen atoms in molecule. For Gompertz function it yields $Q^2_{cv} = 0.987$, for logistic and arctangent Q^2_{cv} is substantially lower. The high correlation of nH with R can be explained by the fact that all the solutes consist of C and H atoms only.

4. CONCLUSIONS

Three sigmoid functions (logistic, Gompertz, arctangent) have been tested in the prediction of retention coefficients, R, of the saturated and aromatic hydrocarbons using their molecular descriptors, d. Comparing to the linear relationship $R(d)$ more descriptors fulfill the condition $Q^2_{cv} \geq 0.98$ when the sigmoid functions are used, especially the Gompertz function. The most of them belongs to the GETAWAY descriptors. However, in the case of investigated solutes (hydrocarbons) the simple descriptor nH denoting a number of hydrogen atoms in solute molecule is also applicable ($Q^2_{cv} = 0.987$).

REFERENCES

[1] G. Trägårdh, K. Ölund, *Desalination*, 1986, *58*, 187.
[2] E.A. Tsapiuk, *J. Membr. Sci.*, 1996, *116*, 17.
[3] S. Koter, B. Gilewicz-Łukasik, A. Nowaczyk, J. Nowaczyk, *Polish J. Chem.*, 2009, *83*, 2137.
[4] J. Mazerski, *The basics of chemometrics* (in Polish), Wyd. Politechniki Gdanskiej, Gdansk, 2000.
[5] L.S. White, *J. Membr. Sci.*, 2002, *205*, 191.