(Q)SAR analysis of a selected group of drugs using retention values from HPLC column chromatography with immobilized plasma proteins

Wanat Karolina[1], Wojda Mariusz[2], Brzezińska Elżbieta[1]

[1]Medical University of Lodz, Faculty of Pharmacy, Department of Analytical Chemistry
karolina.wanat@umed.lodz.pl; elzbieta.brzezinska@umed.lodz.pl
[2]Medical University of Lodz; mariusz.wojda@stud.umed.lodz.pl

Abstract: Proteins are an important part of blood plasma and perform various functions necessary for the proper functioning of the body. Both, alpha-1 acid glycoprotein (AGP) and human serum albumin (HSA) are responsible for drug binding and transport.
Modern technologies combining knowledge from various fields are used now in drug design. A popular method is QSAR, i.e. determining the relationship between the structure of a molecule and its activity. It uses statistical methods and extensive databases of chemical compounds.
In the present study, the binding of substances to plasma proteins was examined by HPLC. The results were subjected to statistical analysis. The influence of the nature of the substance on binding to proteins was demonstrated using ANOVA. The examination of the extent of binding to plasma proteins (PB) for the log k retention value showed the greatest importance of HSA as the protein responsible for this pharmacokinetic phenomenon and the smaller contribution of AGP. However, in regression models, the best results were obtained for basic drugs, indicating their strong affinity for AGP.
The results of the conducted study provide a lot of valuable information regarding therapeutic substances. This confirms the theory that the HPLC method in combination with statistical analysis can be an important part of preclinical QSAR studies.

Materials and methods: The HPLC experiment used active pharmaceutical ingredints (APIs) previously isolated from pharmaceutical formulations. The isolated APIs (n=181) were properly prepared by dissolving them in methanol and finally solutions with a concentration of 1 mg/mL were obtained. A chromatographic columns with immobilized AGP and immobilized HSA (Chiralpak HSA and Chiralpak AGP; 1cm x 4mm, 5µm both) were used. The mobile phase was a mixture of phosphate buffer (PBS) at pH 7.4 and acetonitrile (90:10; v:v) for AGP and PBS:acetonitrile:methanol (85:10:5, v/v/v) in case of HSA. Detection was carried out using a UV-VIS detector connected to the apparatus and a computer with appropriate software (TotalChrom). During the study, retention values - log k(AGP) and log k(HSA) were determined for individual APIs. Then they were used in statistical analyses carried out in STATISTICA 13.1 software (TIBCO Software Inc.).

Results: ANOVA Protein binding and AGP (PB_{code}, 4 groups)
It was checked whether the retention values from the AGP column: log k(AGP) were related to the percentage of drug binding to plasma proteins in general (descriptor PB). AGP is the second most important protein that binds APIs in plasma, after HSA. To perform an ANOVA, the continuous PB values, were transformed into a categorical ones; protein binding values were divided into four groups (PB_{code}):

1 – from 0 to 0.5
2 – from 0.51 to 0.9
3 – from 0.91 to 0.95
4 – above 0.95

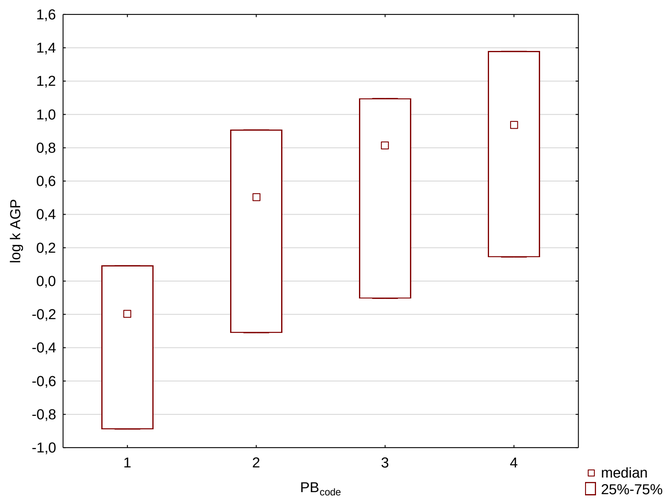

Fig.1 Box plot for median values and quartiles 1 and 3 of the independent variable log k(AGP) for drug groups 1, 2, 3 and 4 of PB_{code}

The median of group 1 (PB from 0 to 0.5) clearly differs from the other groups. No significant differences were observed between groups 2, 3 and 4 in the median of retention factors, however, a certain tendency can be seen in the box plot (Fig. 1), where a higher level of protein binding is associated with on average higher retention values in the form of log k(AGP).
In order to compare the results with another plasma protein - human plasma albumin, the retention values from the HSA column were checked in the same way - log k(HSA)

ANOVA Protein binding and HSA (PB_{code}, 3 groups)
This time it was decided to divide the PB values into three groups (previous groups 3 and 4 were combined into one) and thus significant differences were obtained between the log k(HSA) medians for all groups. This is the best result obtained from the ANOVA analysis. The differences are clearly visible in the box chart (Fig. 2).

PB_{code} (three groups):
1 – from 0 to 0.5
2 – from 0.51 to 0.9
3 – above 0.91

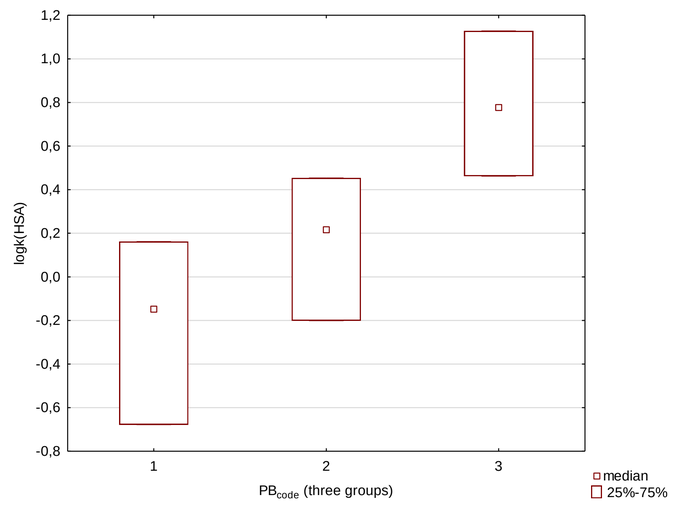

Fig. 2 Box plot for median values and quartiles 1 and 3 of the independent variable log k(HSA) for drug groups 1, 2 and 3 of plasma protein binding percentage (PB code three groups)

Regression models with AGP
The next step was a regression analysis on the dependent variable – PB, to check how much influence AGP binding has on this phenomenon and in which of the groups acids (a), bases (b), neutrals (n) it is most important. The result are shown on Fig. 3.

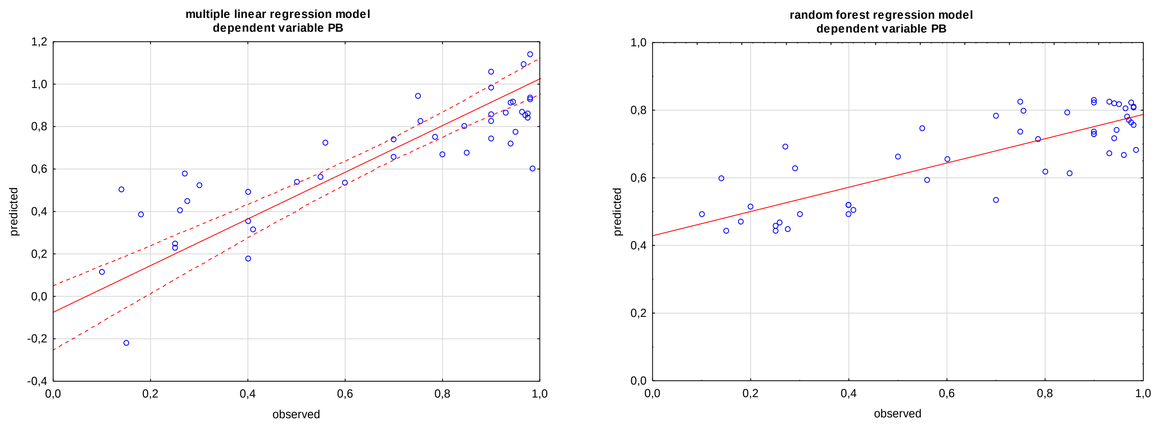

Fig. 3 MLR (left) and RF (right) regression models for basic drugs (n=45). The following descriptors were most important for plasma protein binding: distribution coefficient; log D and retention value from the AGP column – log k(AGP). MLR model explained 74% and RF – 69% of the PB variance.

The best regression models were obtained for the group of alkaline APIs; there was a simple model containing 3 independent variables with $R^2 = 0.64$ and a more complex one (8 variables) with $R^2 = 0.74$ (Fig. 3). Random forest model achieved the $R^2 = 0.69$ (Fig. 3). It can be assumed that such results were obtained due to the strong affinity of the tested basic drugs for AGP.

The result was then improved by applying another regression method – MARSplines. Here, the model explained 82% of the PB (Fig. 4).

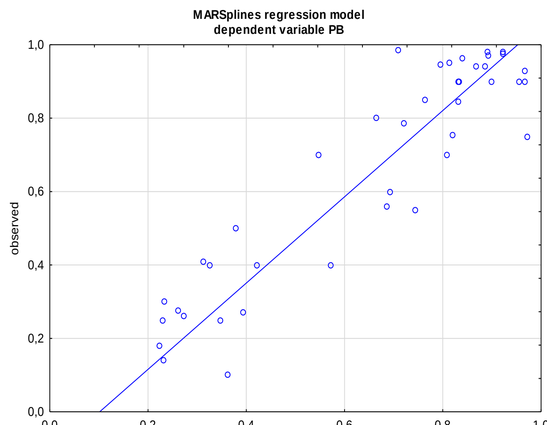

Fig. 4 MARSplines model for basic drugs

Conclusions: ANOVA analysis showed differences between medians of different PB groups; differences were obtained between the group of drugs binding below 50% and the other three groups of the PB code. This division was better in the case of the results obtained from the column with immobilized human plasma albumin, here it was possible to obtain statistically significant differences in the log k(HSA) medians for all three groups determining the binding strength to proteins in plasma. This confirms the greatest importance of HSA as the protein responsible for this pharmacokinetic phenomenon and a smaller share of alpha-1 acid glycoprotein - but still important, which was visible in the box chart, where a certain tendency to increase the average value of log k(AGP) with increasing PB code was noted. Regression models confirmed strong affinity of the tested alkaline APIs for AGP. PB depends to the greatest extent on: binding to the AGP, distribution coefficient (log D) and the ratio of the ionized to non-ionized form in physiological conditions (at pH 7.4) - log U/D.