Research in Pharmaceutical Sciences, 2012;7(5)
School of Pharmacy and Pharmaceutical Sciences
Isfahan University of Medical Sciences

A segmented principal component analysis-regression approach to QSAR study of peptides

M. Elyasi[1,*], B. Hemmateenejad[2], R. Miri[1]

[1]Medicinal & Natural Products Chemistry Research Center, Shiraz University of Medical Sciences, Shiraz, Iran
[2] Department of Chemistry, Shiraz University, Shiraz, Iran and Medicinal & Natural Products Chemistry Research Center, Shiraz University of Medical Sciences, Shiraz, Iran

Background and Aims: The major problem associated with application of principal component regression (PCR) in QSAR studies is that this model extracts the eigenvectors solely from the matrix of descriptors, which might not have essentially good relationship with the biological activity. To overcome the above mentioned limitations and to increase the quality of the QSAR models reported for peptides, we propose here the application of segmented principal component regression (SPCR) method.

Methods: These proposed AA indices were used in QSAR study of two dipeptide data sets; 58(Angiotensin-Converting Enzyme) ACE, and 48 (Bitter Tasting Threshold) BTT. To make a connection between the calculated PCs and the biological activity of the ACE and BTT dipeptides, PCR(SPCR) and PLS(S-PLS) were employed.The number of segments should be optimized to give the best performances. A linear regression analysis based on stepwise selection of variables is then employed to connect a relationship between the informative extracted PCs and biological activity. The loading of these extracted PCs are then used to identify those descriptors represent the highest impact on the activity/property under study.

Results: According to cross-validation test, the SPLS and SPCR based models represented the best performances using 16 and 9 segments, respectively.

Conclusions: A segmented PCA and regression approach was proposed for QSAR study of peptides by definition of new AA indices. In our method, the descriptors are first segmented to different parts and then PCA is applied on each part, separately. QSAR models were developed for a set of 58 ACE inhibitors and a set of 48 dipeptides with BTT activity. It was found that by segmentation of variables and consequently partitioning of the information included in the extracted PCs into informative and redundant parts, it is possible to discard the redundant part and obtain more appropriate models.

Keywords: Amino acid; Dipeptide; Partial least square; Segmented partial least squares; Segmented principal; Component regression