Bulgarian Chemical Communications, Volume 50, Special Issue I (pp. 18 – 24) 2018

Particle size prediction of copolymer-drug conjugate using partial least squares regression

F. Noyan Tekeli[1*], D. Sakar Dasdan[2], G. Karakus[3]

[1]*Yildiz Technical University, Faculty of Arts and Sciences, Department of Statistics 34220 Esenler, Istanbul, Turkey*
[2]*Yildiz Technical University, Faculty of Arts and Sciences, Department of Chemistry, 34220 Esenler, Istanbul, Turkey*
[3]*Cumhuriyet University, Faculty of Pharmacy, Department of Pharmaceutical Chemistry, 58140, Sivas, Turkey*

Received July 1, 2018; Accepted October 1, 2018

Particle size of the copolymers and the associated polydispersity are among the most important factors affecting biopharmaceutical behavior in a wide variety of therapeutic applications. Particle size provides valuable properties of particles or molecules in liquid medium. This characteristic directly affects bioavailability, dissolution and immunotoxicity. Predicting particle size will often skip many preliminary studies that are necessary to optimize formulations. In this work, the particle size of copolymer-drug conjugates was tried to be predicted using partial least squares regression (PLSR). The aim of this article is to construct a mathematical model for predicting the particle size of the copolymer-drug conjugate produced by a preferred pharmaceutical polymer. PLSR is a method that involves a combination of principal component analysis and multiple regression analysis for building predictive models when the factors are many and highly collinear. In the present study, to calculate the particle size of the copolymer-drug conjugate, we used the zeta potential and the particle size of the copolymer and drug, and different pH values as inputs.

Keywords: Particle size, Copolymer-drug conjugate, Partial least squares regression, NIPALS algorithm

INTRODUCTION

Today, the use of synthetic polymers as controlled drug delivery systems or drug carrier biomaterials has increased the interest toward polymer conjugation with biologically active components. Generally, this type of conjugates accumulates in tumors and can reduce toxicity in the body. According to the desired location, the polymer conjugates can be synthesized as having degradable or non-degradable chemical bonds with the relevant drug. Polymer-drug conjugates are drug molecules that are held in polymer molecules. The drugs will stick to the polymer. The drugs are not activated until the enzyme associated with the diseased tissue is present. This process severely reduces the damage to healthy tissue. The results of polymer-drug conjugates are very promising [1-3].

Zetasizer measurements which include zeta potential, particle size and mobility, throw light on valuable properties of particles or molecules in liquid medium. These characteristics directly affect bioavailability, dissolution and immunotoxicity. Zeta potential provides an understanding of many important properties of the colloidal systems and allows them to be controlled and to determine the electrical charge or potential on the particles [4].

Measuring these characteristics for a copolymer-drug conjugate is an expensive, difficult and time-consuming procedure, so it is beneficial to predict these characteristics from other chemical measurements. Computational chemistry permits to calculate specific chemical properties such as size, lipophilicity, and polarity at various sites on the molecule without even making the compound [5]. In this work, the particle size of copolymer-drug conjugates was estimated using PLSR.

Partial Least Squares Regression

Partial least squares (PLS) regression is a multivariate statistical technique that can be used when the numbers of observations are less than the number of variables. PLSR can be efficiently used with a large number of variables that are highly correlated and involving substantial random noise. It was proposed by Herman Wold in 1977 and was first used in social sciences. After the first application by Kowalski, in the late seventies, S. Wold and H. Martens pioneered the chemical applications of PLS [6]. With modern measurement instrumentation including spectrometers, chromatographs and sensor batteries, the data tend to be strongly collinear, loud, and incomplete, so we can investigate these data with the PLSR. Today, PLSR is a widely used technique in chemometrics, especially when the number of independent variables is significantly larger than the number of observations.

* To whom all correspondence should be sent:
E-mail: fnoyan@yildiz.edu.tr

The purpose of PLSR is to construct components that capture a large part of the information in the X variables useful for estimating the Y variable and reduce the dimensionality of the regression problem. These new orthogonal variables obtained by PLS algorithms do not have collinearity [7]. PLSR takes the relation model between *(N*K)* dimensional X and *(N*M)* dimensional Y variables.

Let **X** and **Y** be the centered and scaled matrix of the predictors and response values, respectively, for ease of interpretation and numerical stability [8]. PLSR uses singular value decomposition of $S=X'Y$ cross product matrix. The PLSR, where T is a score vector and W is its weight vector, begins with a linear combination of X:

$$T=XW^{*} \quad (1)$$

X-scores, which are estimated as linear combinations of the original variables, are few and orthogonal. The PLSR predicts both **X** and **Y** by regression on **T** and **U:**

$$X=TP'+E \quad (2)$$

$$Y=UC'+F \quad (3)$$

The **P** and **C** vectors are called the X and Y loadings, respectively, and **E** and **F** are the associated residual vectors. The X-scores are good summaries of **X** so the X-residuals, (e), are small. The Y-residuals, f, are the deviations between the observed and modelled responses, and comprise the elements of the Y residual matrix, **F**.

We replaced equation 1 in Eq. 3:

$$Y=XW^{*}C'+F=XB+F \quad (4)$$

The PLS regression coefficients B can be written as in equation (5) and B is not independent unless the number of the PLSR components (A) equals the number of X-variables (K).

$$B=W^{*}C' \quad (5)$$

The **X**-matrix is deflated by subtracting $t_a p'_a$ from **X, *a*** is the index of components (*a=1,2,…A*)

$$t_a=E_{a-1}W_a \quad (6)$$

$$E_{a-1}=E_{a-1}-t_{a-1}p'_{a-1} \quad (7)$$

$$E_0=X \quad (8)$$

The relationship between **w** and **w*** is given as:

$$W^{*}=W(P'W)^{-1} \quad (9)$$

Regression coefficients for PLSR are obtained from:

$$B=W(P'W)^{-1}C' \quad (10)$$

The **Y**-matrix is deflated by subtracting $t_a c'_a$. It is not necessary, because the results are equivalent whether Y is deflated or not.

NIPALS (Non-Linear Iterative Partial Least Squares)

The PLS algorithm used in the study is NIPALS. Developed by H. Wold in the1960s',the NIPALS is also known as the classical algorithm. The starting point of NIPALS are **X** and **Y** data matrices that are optionally transformed, scaled and centered data. In the NIPALS algorithm which aims to obtain the components that maximize the covariance matrix, not all components are obtained at the same time. At each step, a single component and the weight and load values of this component are obtained [8]. PLS weights are iteratively estimated that is NIPALS based on deflating X and Y variables at each iteration.

a) Get a starting vector of **u,** usually one of the **Y** columns. With a single y, **u=y.**
b) **w**:PLS-weight for X, $w=X'u/u'u$
c) **t**:PLS-score for X, $t=Xw$
d) **c**: PLS-loading for Y, $c=Y't/t't$
e) **u**: PLS-score for Y, $u=Yc/c'c$
f) Convergence is tested on the change in **t**, $\|t_{old}-t_{new}\|/\|t_{new}\|<\varepsilon$, where ε is small 10^{-6} or 10^{-8}. If there is no convergence, return to step (b),
g) Deflate **X** and optionally **Y** before repeating the above steps for new components.
$X=X-tp'$
$Y=Y-tc'$
h) The algorithm continues until the cross-validation shows that there is no more significant information in **X** about **Y** [8,9].

EXPERIMENTAL INSTRUMENT AND MATERIALS

We want to predict the particle size of the copolymer-drug conjugate. The independent variables consist of zeta potential and particle size of the copolymer poly(maleic anhydride-*co*-vinyl acetate (MAVA) and the drug acriflavine (AF) and pH values whereas the response variable is the particle size of the copolymer-drug conjugate (MAVA-AF).

In this study, the particle size and zeta potential of the novel copolymer-drug conjugate, MAVA-AF including the nontoxic drug carrier MAVA were measured as a function of pH in water and as a function of time in simulated body fluids. Experimental values were published in Ref. [3].

Zeta potential, ζ was automatically calculated by the analyzer using the following Smoluchowski equation:

$$\mu_e = \frac{\varepsilon\zeta}{\eta} \qquad (11)$$

where μ_e is electrophoretic mobility, ε is the dielectric constant, ζ is the zeta potential and η is the electrolyte viscosity [10].

It was observed that there were high correlations between measured variables. PLSR is an efficient tool for developing a quantitative relationship between several collinear predictor variables **X** - Zetasizer Measurements in this work and the independent variable **Y**-drug size in this work.

RESULTS AND DISCUSSION

For this analysis, the PLSR algorithm written in SAS was used. The explanatory variables consist of zeta potential of the copolymer, particle size of the copolymer, zeta potential of the drug, particle size of the drug, and different pH values, whereas the response variable is the particle size of the copolymer-drug conjugate. PLSR analysis results showed that 5 components explain most of the variability on both dependent (response) and independent (explanatory) variables. As shown in Table 1, the PLS model explains 100 % of the variation in predictors and about 98 % of the variation in responses [5]. This is a strong indication that the five PLS components are suitable for modeling.

The correlation loadings plot (Fig. 1) is an intense brief of many properties of the PLS model. In the cross-validation, the data set is divided into two or more groups. To create a model for all groups, one is left out to control the predicted capacity of the model. We can measure the overall capacity of the model by doing this for each group. The Predicted REsidual Sum of Squares (PRESS) statistics is obtained from the result of this process. The number of components of the model with the minimum PRESS statistics gives the optimal number of PLS components [11]. The cross-validation results are shown in Table 2. The cross-validation analysis shows that the model with five PLS factors reaches the absolute minimum of predicted residual sum of squares.

Table 1. Percent variation accounted for by partial least squares factors

Number of Extracted Factors	Model Effects		Dependent Variables	
	Current	Total	Current	Total
1	40.7209	40.7209	78.6506	78.6506
2	42.9975	83.7184	14.2551	92.9057
3	13.6563	97.3747	1.0256	93.9314
4	2.3396	99.7143	1.5769	95.5083
5	0.2857	100.0000	3.1832	98.6915

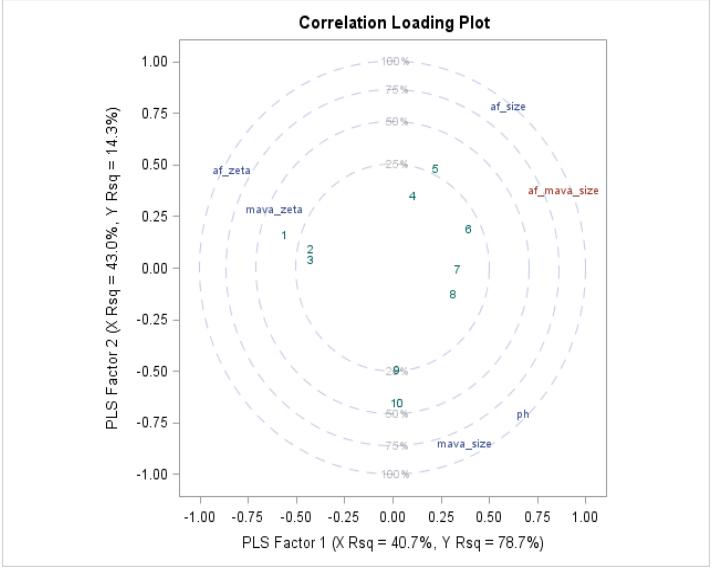

Figure 1. Correlation loadings plot

The cross-validation plot is shown in Fig. 2. From Figure 2, it is seen that five PLS factors are optimal for the PLS model.

In Table 3, the scores of 5 components obtained by the NIPALS algorithm are given.

The score matrix for **X, T**, consists of a linear combination of weight matrices **W** and **X**. The weights for score matrix of X obtained by the NIPALS algorithm are given in Table 4.

Table 2. Cross-validation results

Number of Extracted Factors	Root Mean PRESS
0	1.111.111
1	0.805146
2	0.546011
3	0.418434
4	0.455074
5	0.267357
Minimum root mean PRESS	0.2674
Minimizing number of factors	5

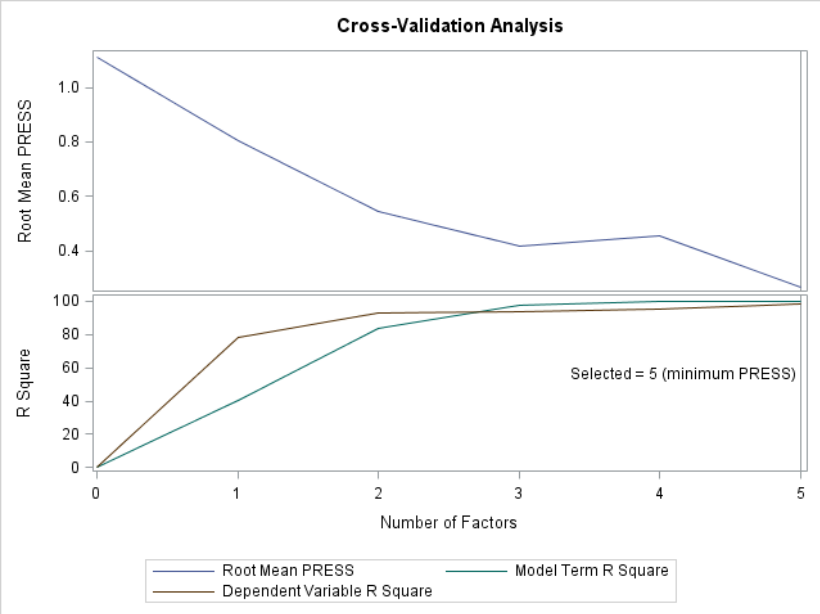

Figure 2. Cross-validation plot

Table 3. Scores for X by NIPALS algorithm

	Scores for X (T)				
	t_1	t_2	t_3	t_4	t_5
1	-2.2893	0.657248	0.001768	-0.19824	-0.12443
2	-1.72851	0.368641	-0.41612	0.123511	-0.10848
3	-1.73328	0.163648	-0.29467	-0.18868	0.202632
4	0.411135	1.441116	-0.14654	0.773033	0.078009
5	0.899101	1.999483	1.821895	-0.25289	0.028022
6	1.608325	0.774276	-0.83651	0.118095	-0.09105
7	1.369919	-0.04711	-0.8099	-0.38613	-0.04425
8	1.278164	-0.55027	-0.34556	-0.26935	0.019102
9	0.093962	-2.07097	-0.00849	0.025979	0.165038
10	0.090485	-2.73606	1.034122	0.254671	-0.12459

Table 4. Weights for score matrix of X by NIPALS algorithms

	Weights for T				
	W1	W2	W3	W4	W5
pH	0.353	-0.44	0.439	0.407	0.655
MAVA-size	0.043	-0.527	0.536	0.095	-0.699
MAVA-zeta	-0.374	0.295	0.801	-0.574	0.197
AF-size	0.825	0.579	0.284	0.14	-0.198
AF-zeta	-0.533	0.354	0.212	0.896	-0.067

Table 5. Loading for X by NIPALS algorithms

	Loadings for X				
	p1	p2	p3	p4	p5
pH	0.476	-0.485	0.195	-0.018	0.655
MAVA-size	0.261	-0.582	0.393	0.418	-0.699
MAVA-zeta	-0.431	0.193	0.864	-0.53	0.197
AF-size	0.421	0.534	0.174	0.204	-0.198
AF-zeta	-0.585	0.322	-0.177	0.709	-0.067

Table 6. Regression coefficients and VIP values of explanatory variables

Label	Regression coefficients	VIP
pH	1.848599261	0.792799806
MAVA-size	-0.725061413	0.542219106
MAVA-zeta potential	0.026679280	0.753368178
AF-size	0.909555562	1.564985588
AF-zeta potential	0.752616534	1.02991427

The loadings for **X** obtained by the NIPALS algorithm are given in Table 5. As expressed in the equation **T = XW**, the components form a linear combination of the weight matrices **W** and **X**. Values of the weights show the contribution of explanatory variable to the component.

The loading matrix is used for modeling **X** according to the equation **X = TP**. Loading values show the amount of the explanation of the component to the explanatory variables. Regression analysis was performed after obtaining the 5 independent components. Table 6 gives the regression coefficients and VIP values of explanatory variables.

The regression coefficients represent the importance each predictor has in the prediction of the response. According to the regression coefficients, pH has the biggest influence on the dependent variable. The drug size (AF-size) has a great contribution in predicting particle size of the copolymer-drug conjugate (MAVA-AF size). The variable importance for projection (VIP) proposed by Wold (1994) is a statistic parameter that shows which determinants are most useful for predicting dependent variables. An explanatory variable can be deleted from the model if it has a small VIP value and a relatively small coefficient (absolute value). Wold (1994) considered a value smaller than 0.8 for a small VIP value [12,13]. The copolymer particle size (MAVA-size) has a low VIP value for the PLS model. variable importance plot and regression parameter profile are given in Figs. 3 and 4, respectively.

The variable importance plot shows the contribution of each predictor in fitting the PLS model for both predictors and response. The particle size (MAVA-size) and zeta potential (MAVA-zeta potential) of copolymer are small absolute coefficients and small VIP's are seen in Figure 2 and Figure 3. As a result of the prediction results with five components, the residuals ranged from 0.000 to 0.051 and all of them are low. The PLSR model explains 98.6915 % of the variation in responses.

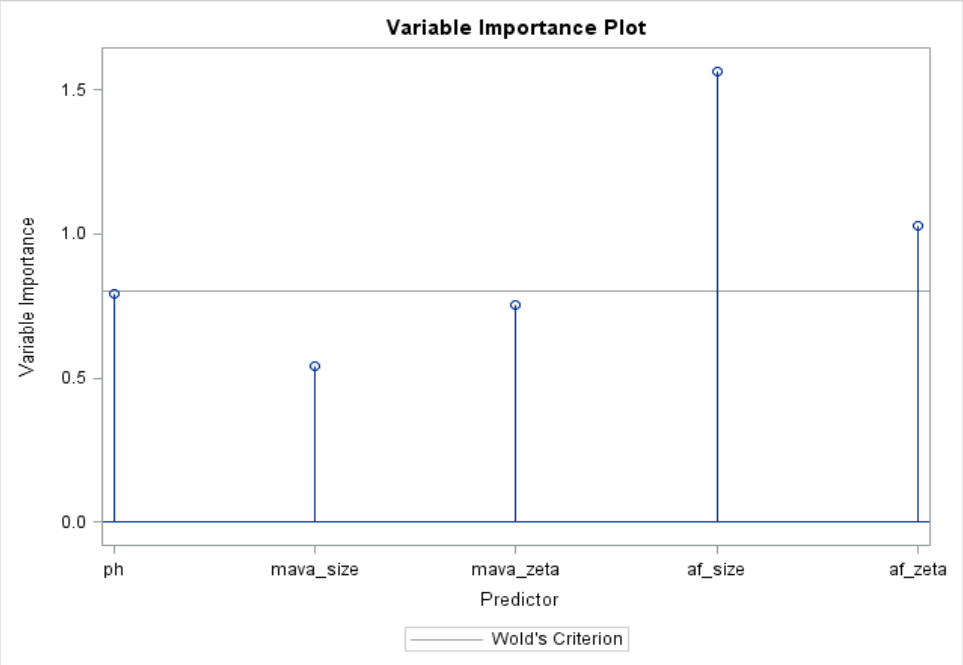

Figure 3. Variable importance plot

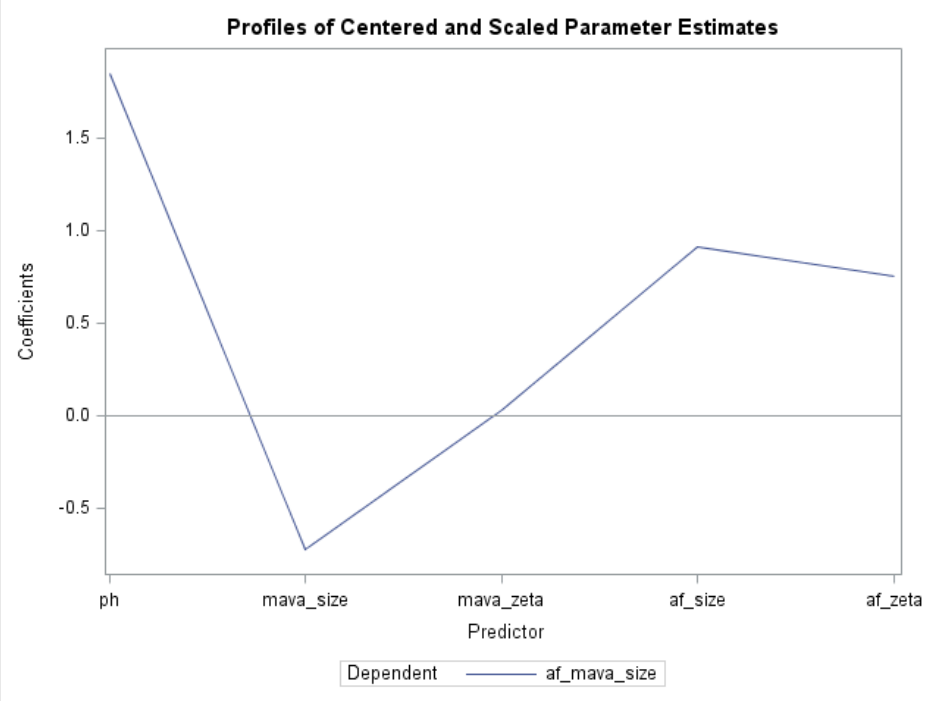

Figure 4. Regression parameter profile

CONCLUSIONS

In this work, a mathematical model was constructed using PLSR to estimate the particle size of the copolymer-drug conjugate produced by a preferred pharmaceutical polymer. The PLSR is a multivariate statistical method that can be used when there is a high correlation between variables. It consists of partial least squares and multiple linear regression analysis. First, explanatory variables, X , having multicollinearity were reduced to components which explain the great amount of covariance between explanatory and response variable by using partial least squares. Then, a regression model was constructed by applying multiple linear regression analysis to the obtained independent new components. The NIPALS algorithm was used and it was found that the particle size of drug (AF) was the most effective in predicting particle size of the copolymer-drug conjugate (MAVA-AF).

Acknowledgements: *This work was supported by the Scientific Research Project Coordination Center of Yıldız Technical University, Turkey (Project No: 2016-01-02-YL06) and Sciences Research Projects Foundation of Cumhuriyet University, (CUBAP, Project No: F258)*

REFERENCES

1. 1.N. Larson, H. Ghandehari, *Chemistry of Materials*, **24**, 840 (2012).
2. A. N. Semenov, *Macromolecules*, **41**, 2243 (2008).
3. D. Sakar Dasdan, A. Dizdar, G. Karakus, *Bulg. Chem. Commun.*, **49(**I), 43(2017).
4. F. NoyanTekeli, *Bulg. Chem. Commun.*,**49(**I), 148 (2017).
5. SAS Institute Inc. 2008. SAS/STAT® 9.2 User's Guide. Cary, NC: SAS Institute Inc. http://support.sas.com/documentation/cdl/en/statugpls/61818/PDF/default/statugpls.pdf
6. E. Bulut, Ö. Gürünlü Alma, *Physical Sciences,* **6**(2), 36 (2011).
7. E. Bulut, U. Yolcu, M.Y. Tasmektepligil, E. Egrioglu, *World Applied Sciences Journal* **, 21**(4), 572 (2013).
8. S.Wold , M. Sjostrom, L.Eriksson, *Chemometrics and Intelligent Laboratory Systems*, **58,** 109 (2001).
9. 9.G. Gergov, A. Alin, M. Doychinova, M. De Luca, V. Simeonov, Y. Al-Degs, *Bulg. Chem. Commun.*, **49(**2), 410 (2017).
10. Brookhaven Instruments Corporation, Instruction Manual for Zeta Potential Analyzer Instruction Book https://support.sas.com/rnd/app/stat/papers/plsex.pdf
11. T. Mehmood, K.Hovde Liland, L.Snipen, S.Sæbø, *Chemometrics and Intelligent Laboratory Systems*, **118,** 62, (2012).
12. E. Bulut, A.Alın, *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, **24 (**2),127 (2009).

ПРЕДСКАЗВАНЕ НА РАЗМЕРА НА ЧАСТИЦИТЕ НА КОНЮГАТ ПОЛИМЕР-ЛЕКАРСТВО С ИЗПОЛЗВАНЕ НА РЕГРЕСИЯ НА ЧАСТИЧНИ НАЙ-МАЛКИ КВАДРАТИ

Ф. Ноян Текели[1*], Д. Сакар Дашдан[2], Г. Каракуш[3]

[1] Технически университет Йилдиз, Факултет по науки и изкуства, Департамент по статистика, 34220 Есенлер, Истанбул, Турция

[2] Технически университет Йилдиз, Факултет по науки и изкуства, Департамент по химия, 34220 Есенлер, Истанбул, Турция

[3] Университет Джумхуриет, Факултет по фармация, Департамент по фармацевтична химия, 58140, Сивас, Турция

Постъпила на 1 юли, 2018 г.; приета на 1 октомври, 2018 г.

Размерът на частиците и свързаната с него полидисперсност са сред най-важните фактори, влияещи върху биофармацевтичното отнасяне при голям брой терапевтични приложения. Размерът на частиците или молекулите им придава ценни свойства във водна среда. Този фактор директно влияе върху биодостъпността, разтворимостта и имунотоксичността. Предсказването на размера на частиците често прави излишни много предварителни изследвания, необходими за оптимизиране на лекарствените препарати. В настоящата работа е направен опит за предсказване на размера на частиците на конюгата полимер-лекарство с помощта на регресия на частични най-малки квадрати (PLSR). Конструиран е математичен модел за предсказване на размера на частиците на конюгат полимер-лекарство, получен от предпочитан фармацевтичен полимер. PLSR е метод, който включва комбинация от анализ на основни компоненти и множествен регресионен анализ за изграждане на предсказуеми модели в случаите, когато има голям брой силно колинеарни фактори. За изчисляване на размера на частиците на конюгата полимер-лекарство ние използвахме зета-потенциала и размера на частиците на съполимера и на лекарството, както и различни pH стойности като входящи данни.