Acta Pharmaceutica Sinica B 2013;3(2):130–139

Institute of Materia Medica, Chinese Academy of Medical Sciences
Chinese Pharmaceutical Association

Acta Pharmaceutica Sinica B

www.elsevier.com/locate/apsb
www.sciencedirect.com

ORIGINAL ARTICLE

Optimization of a doxycycline hydroxypropyl-β-cyclodextrin inclusion complex based on computational modeling

Zhouhua Wang[a,b], Zixin He[a,c], Lei Zhang[d], Haohao Zhang[a], Meimei Zhang[a], Xinguo Wen[a], Guilan Quan[a], Xintian Huang[a], Xin Pan[a], Chuanbin Wu[a,b,*]

[a]*School of Pharmaceutical Sciences, Sun Yat-sen University, Guangzhou 510006, China*
[b]*Research and Development Center of Pharmaceutical Engineering, Sun Yat-sen University, Guangzhou 510006, China*
[c]*ZhongYi Pharmaceutical Company Ltd., Guangzhou 510530, China*
[d]*School of Mathematics & Computational Sciences, Sun Yat-sen University, Guangzhou 510275, China*

Received 1 November 2012; revised 18 December 2012; accepted 6 January 2013

KEY WORDS

Doxycycline;
Hydroxypropy-β-cyclodextrin;
Response surface methodology;
Artificial neural network;
Support vector machine;
Genetic algorithm

Abstract To prepare a stable complex of doxycycline (Doxy) and hydroxypropy-β-cyclodextrin (HP-β-CD) for ophthalmic delivery, the optimum formulation and preparation conditions were investigated using response surface methodology (RSM), artificial neural network (ANN) and support vector machine (SVM) modeling. The molar ratios of HP-β-CD/Doxy and Mg^{2+}/Doxy, inclusion time and temperature were selected as independent variables (X_1–X_4) and inclusion efficiency and stability of the Doxy-HP-β-CD complex were selected as dependent (response) variables (Y_1 and Y_2). The optimal formulation predicted by genetic algorithm (GA) combined with the models was characterized by microscopy and nuclear magnetic resonance spectrometry, and the stability of Doxy in the complex was evaluated. The highest values of Y_1 and Y_2 were obtained using an ANN model combined with GA which predicted the values of X_1–X_4 to be 4, 10.8, 12 h and 25 °C, respectively. The modeling and optimization results indicated that a feed-forward back-propagation ANN with one hidden layer and 10 hidden units showed better fitting to both

*Corresponding author at: School of Pharmaceutical Sciences, Sun Yat-sen University, Guangzhou 510006, China. Tel.: +86 20 39943120; fax: +86 20 39943117.
E-mail address: cbwu2000@yahoo.com (Chuanbin Wu).

Peer review under responsibility of Institute of Materia Medica, Chinese Academy of Medical Sciences and Chinese Pharmaceutical Association.

ELSEVIER Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.apsb.2013.01.002

responses compared to the RSM and SVM models. GA proved to be an efficient tool in multi-objective optimization of a pharmaceutical formulation.

1. Introduction

Doxycycline (Doxy) belongs to the tetracycline group of broad-spectrum antibiotics. A number of formulations are available including solid dosage forms such as pellets[1] and liquid dosage forms such as ophthalmic preparations[2]. However, the poor stability of Doxy in aqueous solution is a major obstacle to the widespread clinical application of ophthalmic preparations. The aim of this study was to overcome this instability and prepare a stable formulation of Doxy for ophthalmic delivery.

Hydroxypropyl-β-cyclodextrin (HP-β-CD) is well known for its ability to stabilize molecules by forming drug-CD complexes[3]. In addition, Doxy forms a chelate with magnesium ion which also increases Doxy stability[4]. In a previous study[5], we formulated a Doxy-HP-β-CD inclusion complex in a thermally sensitive poloxamer hydrogel which could be stored in the cold and would form a hydrogel when applied to the surface of the eye. In the present study, we investigated a solid Doxy-HP-β-CD(Mg^{2+}) inclusion complex which could be dissolved before use and would remain stable in aqueous solution for a week.

The concept of design of experiments (DOE) was introduced to optimize the Doxy formulation in this study. The molar ratios of HP-β-CD/Doxy (X_1) and Mg^{2+}/Doxy (X_2) together with inclusion time (X_3) and temperature (X_4) were selected as independent variables and inclusion efficiency (Y_1) and stability of Doxy (Y_2) were identified as dependent variables (*i.e.*, responses). Central composite design (CCD) together with some advanced modeling technologies provided a rational approach to the optimization process. Response surface methodology (RSM) was also applied[6,7] despite recognition of the fact that this polynomial approach is limited in its application to multi-objective optimization of formulations.

Two novel optimization techniques, artificial neural networks (ANN) and support vector machine (SVM) modeling were applied in this study. ANN is a computer program designed to simulate the learning capabilities of neurological systems using different learning algorithms[8]. Compared to RSM, ANN has prominent advantages for modeling complex non-linear relationships and can handle multiple independent and dependent variables simultaneously. For this reason, ANN has been successfully applied to solve various problems in the development of drug delivery systems[9–12]. SVM is a novel algorithm which can solve non-linear problems by performing linear regression in a multi-dimensional feature space. Due to its remarkable accuracy and generalization performance, SVM has been extensively applied to quantitative structure activity relationship (QSAR) studies in recent years[13,14].

As there are two response variables in the present study, a multi-objective simultaneous optimization technique was employed to transform the multi-objective problem into a single-objective problem. However, since it is difficult to minimize or maximize all objective functions simultaneously using classical optimization methods when objective functions are in a trade-off relationship, the genetic algorithm (GA) approach was used. The superior ability of GAs to determine the global optimum and to carry out parallel computing of multi-objective functions has led to an increasing number of applications in pharmaceutical and related fields[15,16]. In this study, the use of GAs for global optimization was combined with other modeling techniques in GA-based approaches such as GA-ANN[17,18] and GA-SVM[19,20] to determine the composition of a Doxy-HP-β-CD(Mg^{2+}) formulation with optimum inclusion efficiency and stability in aqueous solution. This optimum formulation was further characterized by microscopy and NMR spectroscopy[21].

2. Materials and methods

2.1. Materials

Doxy monohydrate (Doxy · H_2O, purity 98.5%) and Doxy hydrochloride (Doxy · HCl, purity 92.5%) were kindly provided by Yancheng Suhai Pharmaceutical Co., Ltd. (Jiangsu, China). HP-β-CD was obtained from Roquette (Lestrem, France). All other reagents were of analytical grade and used as received.

2.2. Preparation of Doxy-HP-β–CD(Mg^{2+}) complex

Thirty formulations of the Doxy-HP-β-CD(Mg^{2+}) complex were prepared based on central composite design for RSM, ANN and SVM analysis to determine relationships between the selected independent and dependent variables. The 30 formulations included different molar ratios and process parameters according to DOE. In all cases, Doxy · H_2O containing 25 mg Doxy was added to aqueous magnesium chloride solutions of different concentrations and magnetically stirred at a specific temperature until equilibrium was attained. Solutions were then filtered through 0.45 μm membranes and lyophilized to obtain the solid inclusion complexes.

2.3. Assay for Doxy

Doxy was assayed by HPLC using the method in the Chinese Pharmacopoeia[22]. A Waters HPLC system was equipped with a reversed-phase Luna C18 column (250 mm × 4.6 mm, 5 μm, Phenomenex, USA) maintained at 35 °C and a UV detector set at 280 nm. The mobile phase consisted of 0.05 M ammonium oxalate: dimethylformamide: 0.2 M diammonium phosphate 65: 30: 5, pH 8.0±0.2 delivered at a flow rate of 1.0 mL/min. The injection volume was 20 μL.

2.4. Determination of inclusion efficiency

Inclusion efficiency of Doxy was determined using molecular sieve chromatography to separate free Doxy from the HP-β-CD inclusion complex. Stirred solutions (0.5 mL) were added

to columns (50 mm × 10 mm) packed with Dextran (Sephadex® G-10, International Specialty Products Inc.) and eluted with water at a flow rate of 0.65 mL/min. The first 25 mL of eluent was collected and analyzed for Doxy content. Inclusion efficiency was calculated using the following equation:

$$\text{Inclusion efficiency (\%)} = \frac{\text{Amount of Doxy in eluate}}{\text{Total amount of Doxy}} \times 100\% \quad (1)$$

2.5. Stability studies

A controlled sample of Doxy · HCl was prepared with the same concentration of Doxy (0.1%, *w*/*w*) as in the inclusion complex. Solutions of Doxy · HCl and the Doxy-HP-*β*-CD inclusion complexes were then stored at 40±2 °C and a relative humidity of 75±5% for 10 days. Drug content was determined at the beginning and after 5 and 10 days to determine the stability of Doxy.

2.6. DOE

A four-factor-five-level CCD was employed in STATISTICA trial version 8.0 software (Statsoft Inc., Tulsa, Oklahoma). Responses were measured in a total of 30 runs (Table 1) and used for RSM, ANN and SVM modeling.

2.7. RSM

2.7.1. Selection of variables and generation of non-linear model

Data relating to the two responses were fitted to classical second- and third-order polynomials. In the second-order polynomial model

$$Y = b_0 + \sum_{i=1}^{n} b_i X_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} b_{ij} X_i X_j + \sum_{i=1}^{n} b_{ii} X_i^2 \quad (2)$$

Y is the measured response and b_0 is an intercept. The coefficients corresponding to linear effects (b_i), interaction (b_{ij}) and quadratic effects (b_{ii}) for the factors X_i and X_j were determined from the results of experiments.

In the third-order polynomial model

$$Y = b_0 + \sum_{i=1}^{n} b_i X_i + \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} X_i X_j + \sum_{i=1}^{n} \sum_{j=i}^{n} \sum_{k=j}^{n} b_{ijk} X_i X_j X_k \quad (3)$$

Y and b_0 are as above and b_i, b_{ij} and b_{ijk} represent the estimated coefficients for the factors X_i, X_{ij} and $X_i X_j X_k$, respectively.

Experimental data were analyzed by nonlinear estimation in SAS version 9.0 (Cary, NC, USA). The value of R^2 and the *F*-test were used to identify the model with the best fit and to evaluate the lack of fit within each model. In order to determine the most stable and interpretable variables of the second- and third-order polynomials, relevant variables were selected by forward and backward methods. These methods can diminish the full equation form by removing useless variables which have little impact on the predictive ability of the model.

Table 1 Experimental domain according to central composite design and the measured responses for 30 different formulations of the Doxy-HP-*β*-CD (Mg^{2+}) inclusion complex.

Sample No.	Factor				Response		
	X_1	X_2	X_3	X_4	Y_1	Y_2	Y
1	2.75	3.75	12	30	96.35	89.43	0.676
2	6.25	3.75	12	30	88.56	90.2	0.423
3	2.75	9.25	12	30	98.05	93.67	0.835
4	6.25	9.25	12	30	93.7	97.12	0.762
5	2.75	3.75	20	30	91.5	82.53	0.351
6	6.25	3.75	20	30	94.82	82.93	0.437
7	2.75	9.25	20	30	90.77	82.37	0.329
8	6.25	9.25	20	30	90.07	83.63	0.346
9	2.75	3.75	12	40	86.9	84.83	0.260
10	6.25	3.75	12	40	93.51	82.05	0.378
11	2.75	9.25	12	40	89.86	85.27	0.380
12	6.25	9.25	12	40	92.32	87.92	0.520
13	2.75	3.75	20	40	90.32	77.58	0.104
14	6.25	3.75	20	40	93.51	83.8	0.439
15	2.75	9.25	20	40	84.74	83.19	0.117
16	6.25	9.25	20	40	95.04	83.82	0.473
17	1	6.5	16	35	84.2	81.93	0.054
18	8	6.5	16	35	85.3	91.31	0.235
19	4.5	1	16	35	95.27	85.63	0.538
20	4.5	12	16	35	85.85	87.83	0.244
21	4.5	6.5	8	35	99.17	83.82	0.555
22	4.5	6.5	24	35	95.79	80.68	0.359
23	4.5	6.5	16	25	95.05	93.84	0.744
24	4.5	6.5	16	45	96.85	89.61	0.694
25	4.5	6.5	16	35	94.81	89.77	0.641
26	4.5	6.5	16	35	94.69	85.53	0.521
27	4.5	6.5	16	35	92.48	89.72	0.567
28	4.5	6.5	16	35	92.73	91.07	0.605
29	4.5	6.5	16	35	96.93	91.2	0.739
30	4.5	6.5	16	35	94.16	85.13	0.496

2.7.2. Desirability function

In order to find the optimum formulation, the multi-criteria problem was treated as a single criterion problem by using the desirability function approach[23]. In this study, the two responses were targeted to maximum, and the desirability function for the maximum response was defined as:

$$d_i = \frac{Y_i - Y_{\min}}{Y_{\max} - Y_{\min}} \quad (4)$$

where $Y_{\min}$ and $Y_{\max}$ represent the lowest and highest possible values, respectively, and Y_i indicates the experimental value. Based on the observed response values, the limits were set as: $Y_1/Y_{\max} = 100.0$, $Y_1/Y_{\min} = 84.0$; $Y_2/Y_{\max} = 98.0$, $Y_2/Y_{\min} = 77.0$. After obtaining the individual desirability function for each response, a global desirability function (D) was calculated by combining the individual desirability functions using the geometric mean as shown below.

$$D = \left(\prod_{i=1}^{k} d_i \right)^{1/k} \quad (5)$$

where k is the number of responses. The result of RSM_1 means the optimization based on the desirability function and RSM_2 means the optimization based on genetic algorithm.

2.8. ANN modeling

A three layer feed-forward back-propagation (BP) network incorporating the four independent variables in the input layer and the two response variables in the output layer was fitted to the data set according to CCD (Fig. 1). Ten nodes in the hidden layer and the nodes in adjacent layers were fully interconnected with links, the intensity of connection between two nodes being described by a weight function. By summing the input from the previous layer and implementing a transfer function, the output of a node to the next layer was computed. A popular transfer function is the logistic sigmoid function described by the following equation:

$$f(y_j) = \frac{1}{1 + \exp(-a \times y_j)} \tag{6}$$

Here $f(y_j)$ is the output from the jth neuron, α denotes the slope of the sigmoidal function and y_j is defined as:

$$y_j = \sum_{i=1}^{n} W_{ij} x_i + B_j \tag{7}$$

where x_i is the input to the ith node from the previous layer with a total of n nodes, W_{ij} is the corresponding weight and B_j is a bias term. The logistic sigmoid and pure linear functions were used as transfer functions in the hidden and output layers of the ANN, respectively. The ANN was trained iteratively to minimize the performance function of the mean square error (MSE) between the results of the output neurons and the actual outputs. After each cycle, the gradient of the performance function of the MSE was used to adjust the network weights and biases. Once the MSE reached 10^{-4}, the training was terminated and the corresponding ANN was built. MATLAB version 7.9.0 (Mathworks Inc., Natick, MA, USA) was employed in the development and training of ANN. The initial weights and biases of the network were generated automatically by the program.

2.9. SVM modeling

SVM, as developed by Cortes and Vapnik[24], is a novel machine learning method acquiring popularity due to some appealing features and its promising empirical performance. SVM was originally designed to handle classification problems but, with the introduction of the ε-insensitive loss function, the application of SVM has been broadened to apply to non-linear regression. The basic idea of SVM is to map the original input data into a highly dimensional feature space using a kernel function. Linear regression is then performed in this feature space and the non-linear problem is resolved in a linear space. The SVM model is described by the following function:

$$f(x) = \sum_{i=1}^{l} (-\alpha_i + \alpha_i^*) K(x_i, x) + b \tag{8}$$

where x_i is a feature vector corresponding to a training object and b is a constant. The components of the vectors α_i and α_i^* are the introduced Lagrange multipliers. For the kernel function, $K(x_i, x)$, the Gaussian radial basis function in Eq. 9 was selected due to its effectiveness and efficiency in the training process.

$$K(x_i, x) = \exp\left(-\frac{\|x_i - x\|}{2\gamma^2}\right) \tag{9}$$

where γ is the parameter of the kernel and x_i and x are two independent variables. LIBSVM software (available at http://www.csie.ntu.edu.tw/cjlin/libsvm) was used to process the SVM regression[25].

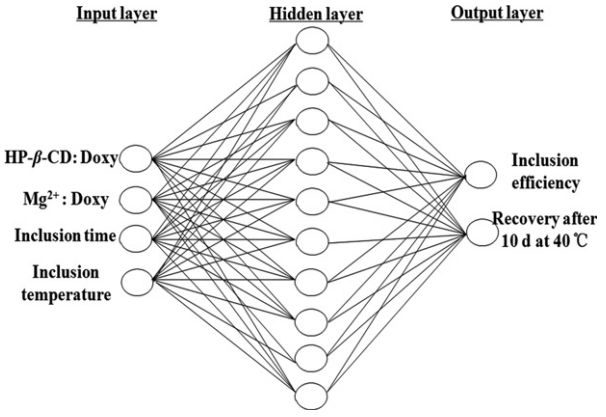

Figure 1 Optimum neural network architecture based on central composite design.

2.10. Genetic algorithm (GA)

GA is used to search the solution space by simulated evolution of "survival of the fittest". GA can solve linear and nonlinear problems through exploring all areas of state space and exploiting potential regions by mutation, crossover and selection operations applied to individuals in the population[26].

The evolutionary process of GA consists of three steps: (1) initialization of the first generation; (2) evaluation of the fitness of each individual by an objective function; and (3) using genetic operators of parent selection, including application of reproduction, crossover and mutation to provide the first offspring generation. Iteration of step 2 and 3 is performed until the objective function converges.

In this study, the optimization of the process was accomplished using the Sheffield MATLAB Genetic Algorithm Toolbox Version 1.2 in MATLAB (version 7.9.0)[27].

2.11. Validation of models

In order to assess the reliability of the generated mathematical models, optimized check point formulations were prepared and evaluated for various responses. The experimental values of the responses were quantitatively compared with the predicted values to calculate the percentage of the prediction bias. This is helpful in establishing the validity of generated equations and describing the application domain of these models.

2.12. Characterization of the optimum Doxy-HP-β-CD(Mg^{2+}) complex

Based on the modeling results, the optimum formulation of the Doxy-HP-β-CD(Mg^{2+}) inclusion complex was prepared and characterized by microscopy and NMR analysis. A physical mixture was also prepared for contrast analysis by grinding together the appropriate mixture of Doxy, HP-β-CD and $MgCl_2$ in a quartz mortar.

2.12.1. Microscopy

The morphology of samples was investigated using an inverted microscope (Olympus, X71, Japan) operating at ×400 magnification. Solid samples of raw materials, the physical mixture and inclusion complex were mounted directly onto glass slides for analysis.

2.12.2. NMR spectroscopy

^{1}H NMR was performed using an Avanc III NMR spectrometer (Bruker, Germany) at 400 MHz. Samples containing the equivalent of 10 mg Doxy were dissolved in CD_3OD. Chemical shifts (δ) are reported in ppm and referenced to the residual CD_3OD signal at 3.31 ppm.

3. Results and discussion

3.1. RSM

3.1.1. Selection of variables and generation of non-linear model

Using the second order polynomial model (Eq. 2), the efficiency was unsatisfied as the R^2 values for overall response (Y) and the responses for inclusion efficiency (Y_1) and stability (Y_2) were 0.8498, 0.8115, and 0.7861, respectively. The fact that all are <0.9 indicates a lack of variability in this model and the need for the more highly variable cubic equation (Eq. 3). Comparing forward and backward methods based on the R^2 values shows that the backward method gives the highest fitting efficiency. Using the F-test to evaluate the lack of fit within each model gives $P<0.0001$ in all cubic equations indicating the third order polynomial model generates a significantly better fit.

$$\begin{aligned} Y = {} & 3.380 - 0.138X_4 - 0.045X_1X_1 - 0.112X_1X_2 - 0.047X_1X_3 \\ & + 0.052X_1X_4 + 0.059X_2X_2 + 0.035X_2X_3 - 0.020X_2X_4 \\ & - 0.002X_3X_3 + 0.013X_1X_1X_2 + 0.006X_1X_1X_3 \\ & - 0.005X_1X_1X_4 - 0.0004X_2X_2X_3 + 0.0004X_2X_3X_4 \\ & + 0.0002X_2X_4X_4 (R^2 = 0.9603 \;\; F = 22.59 \;\; P < 0.0001) \end{aligned} \tag{10}$$

$$\begin{aligned} Y_1 = {} & 133.704 - 20.831X_1 + 8.140X_2 - 2.554X_1X_2 \\ & + 0.812X_1X_3 + 1.207X_1X_4 - 0.131X_2X_2 - 0.103X_2X_3 \\ & + 0.0464X_3X_3 - 0.220X_3X_4 - 0.0625X_4X_4 \\ & + 0.287X_1X_1X_2 - 0.0762X_1X_1X_4 - 0.0183X_1X_3X_4 \\ & + 0.00478X_3X_4X_4 (R^2 = 0.9090 \;\; F = 10.71 \;\; P < 0.0001) \end{aligned} \tag{11}$$

$$\begin{aligned} Y_2 = {} & 83.681 + 10.616X_3 - 1.856X_4 + 0.150X_2X_2 \\ & - 0.707X_3X_3 + 0.102X_1X_1X_1 + 0.103X_1X_1X_2 \\ & - 0.0643X_1X_1X_4 - 0.0540X_1X_2X_3 + 0.0116X_1X_3X_4 \\ & + 0.00587X_1X_4X_4 - 0.00654X_2X_2X_4 + 0.00545X_2X_3X_4 \\ & + 0.0124X_3X_3X_3 (R^2 = 0.9059 \;\; F = 11.85 \;\; P < 0.0001) \end{aligned} \tag{12}$$

3.1.2. Response surface

Response surfaces were generated in order to understand the effect of each variable on the responses. Fig. 2A–D shows the

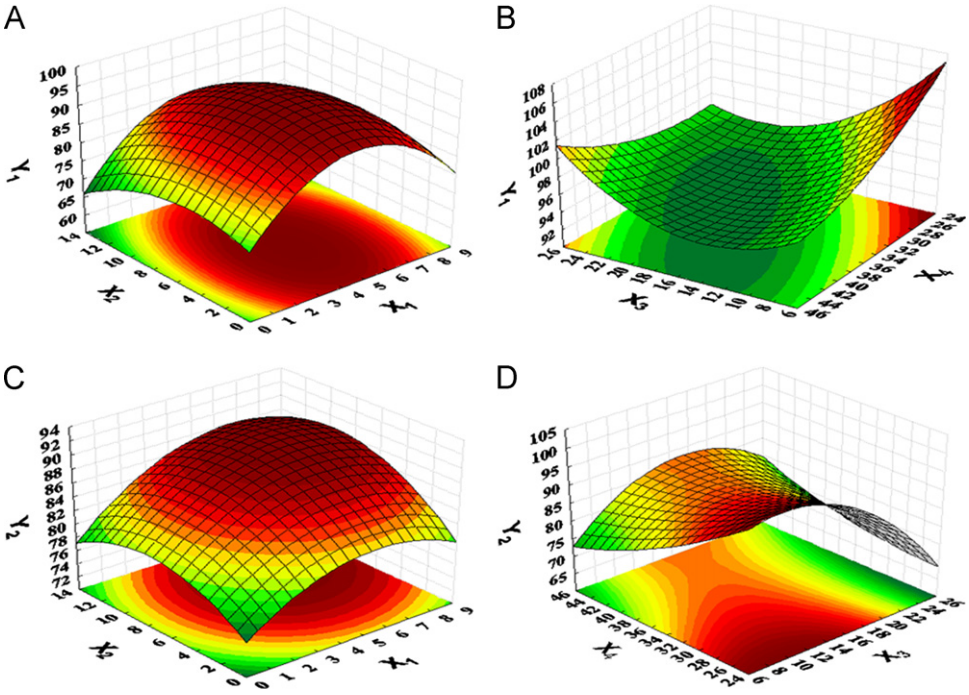

Figure 2 Three dimensional response surface plots and two dimensional contour plots for (A) molar ratios of HP-β-CD/Doxy (X_1) and Mg^{2+}/Doxy (X_2), and (B) inclusion time (X_3) and inclusion temperature (X_4) *versus* inclusion efficiency (Y_1) and (*C*) molar ratios of HP-β-CD/Doxy (X_1) and Mg^{2+}/Doxy (X_2), and (D) inclusion time (X_3) and inclusion temperature (X_4) *versus* stability (Y_2). The remaining input factors were kept constant at mid-range values (HP-β-CD/Doxy 4.5; Mg^{2+}/Doxy 6.5; inclusion time: 16 h; and inclusion temperature 35 °C).

three-dimensional response surface and the projected two-dimensional contour plot. Three-dimensional plots are useful to describe interactions while contour plots are exceedingly valuable when there is only one response. From the three dimensional diagrams, it is easy to understand the interactions between two independent variables and one dependent variable. Fig. 2A and C shows the obtained surfaces are convex and symmetrical suggesting the mole ratios (X_1 and X_2) are the main factors influencing the responses Y_1 and Y_2. In Fig. 2B and D, the flat surfaces reveal that inclusion time (X_3) and inclusion temperature (X_4) did not affect Y_1 and Y_2 significantly.

3.2. *ANN modeling*

The model parameters of ANN were selected by trial and error. According to the minimum validation, the optimum network consisted of one hidden layer with 10 nodes, while the number of training cycles or iterations was found to be 500.

To build a successful ANN model, the number of hidden nodes is the key factor. Too few hidden nodes handicap the learning capacity of the ANN model, while too many hidden nodes may cause over-fitting for the training data set. There are several methods to determine the number of neurons in a hidden layer, such as using a formula or empirically by the programmers. As a result, these methods are either stereotyped or subjective. Kolomogorov's theory indicates that hidden nodes with twice the number of input variables plus one are adequate to compute an arbitrary continuous function[28]. Jadid and Fairbairn[29] proposed an upper limit on the number of hidden nodes. Carpenter and Hoffman[30] introduced an equation to define the number of nodes in the hidden layer based on the number of inputs and outputs and the number of training sets.

In this work, the optimal number of hidden units was determined by the exhaustive method which sets out all possible solutions and finds the answer to meet the conditions. One to 20 neurons were used to explore one by one in the network and then, according to the maximum value of R^2, the optimum network architecture was obtained. Finally, a network consisting of four input and two output units, with 10 hidden units arranged in a single hidden layer was selected (Fig. 1). On this basis, the R^2 values for Y_1 and Y_2 were 0.9947 and 0.9721, respectively, indicating that a quality trained model with appropriate predictive capability was constructed.

3.3. *SVM modeling*

The performance of SVM for regression depends on the combination of several factors including the type of kernel function, the Gaussian function parameter γ, the ε-insensitive loss function ε, and the capacity parameter C. First, the radial basis function (RBF), which is commonly employed in many studies because of its satisfactory performance and limited number of parameters to be adjusted[31], was chosen as the kernel function to determine the sample distribution in the mapping space. The corresponding γ of the kernel function has a major effect on the number of support vectors and is closely connected with the performance of SVM and the training time. The parameter ε prevents the entire training set meeting boundary conditions and allows the possibility of sparsity. The parameter C controls the trade-off between maximizing the binding and minimizing the training error.

To determine the optimal values of the above parameters, a grid search was performed on the original data set using combinations of the parameters in the following ranges: ε 0.01–0.15; C 5–30; γ 0.1–3.0. Screening was performed using an incremental step of 0.01 for all parameters. Based on the thorough search, the optimal values of ε, C and γ suitable for SVM modeling of the responses were respectively 0.1, 9.19, 2.14 for Y_1 and 0.1, 11.31, 0.5 for Y_2. The Y_1 optimal model gave a root mean square error (RMSE) of 0.4458 and R^2 of 0.9723. Corresponding values for the Y_2 model were 1.4100 and 0.9305, respectively. With these optimal parameters, the SVM regression simulated the complicated nonlinear relationship between the independent and dependent variables.

3.4. *Multi-objective optimization and GA*

GA was employed to search for the optimum formulation after the RSM, ANN and SVM models were developed. Appropriate implementation of GA includes the following three steps; definition of the fitness function, definition and implementation of the genetic representation, and definition and implementation of the genetic operators. In the RSM model, Eqs. 11 and 12 for Y_1 and Y_2 were used as fitness functions for GA parallel solving, and Eq. 10 for the global response Y was used as the fitness function to compute the optimal value. In the ANN model, after obtaining the relationship between the causal factors and the response variables according to Eqs. 6 and 7, ANN generated a nonlinear model as fitness function for GA. In the SVM model, according to Eq. 8, an integrated equation was generated as fitness function for GA.

The control parameters used in GA were set as follows: population size 100; generations 500; precision of variables 20; generation gap 0.9; crossover rate 0.4; and mutation rate 0.4. Fig. 3 shows the change of solution for Y_1 and Y_2 by GA-ANN. For the ANN data set, after 100 generations evolved, the solutions of Y_1 and Y_2 not only became stable but also reached the optimal result at the same generation. This clearly indicates that GA possesses an excellent ability for parallel computing and can effectively resolve the problem of multi-objective optimization, which is better than classical optimization methods.

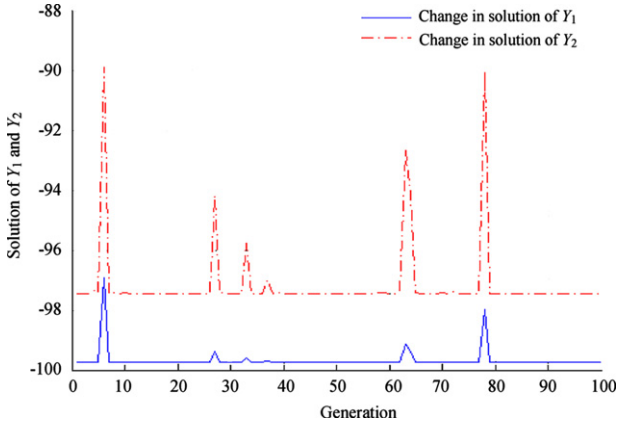

Figure 3 Change of solution for Y_1 and Y_2 by GA-ANN.

Table 2 Experimental and predicted values of Y_1 and Y_2 for the optimum formulation of the Doxy-HP-β-CD(Mg^{2+}) inclusion complex as determined by RSM, ANN and SVM models.

Modeling method	Factor				Response				Bias[a] (%)	
					Experimental		Predicted			
	X_1	X_2	X_3	X_4	Y_1	Y_2	Y_1	Y_2	Y_1	Y_2
RSM_1	3.0	9.0	16.0	32.5	0.791±0.01 (Y)		0.847 (Y)		7.08 (Y)	
RSM_2	3.5	11.2	10.0	30.0	93.25±2.18	87.49±3.15	96.60	92.75	3.59	6.02
ANN	4.0	10.8	12.0	25.0	99.19±1.24	95.54±1.51	99.75	97.44	0.57	1.99
SVM	3.7	9.3	12.5	26.5	98.05±1.92	91.53±1.89	99.09	96.72	1.06	5.67

[a]Bias (%)=(predicted value−experimental value)/experimental value × 100.

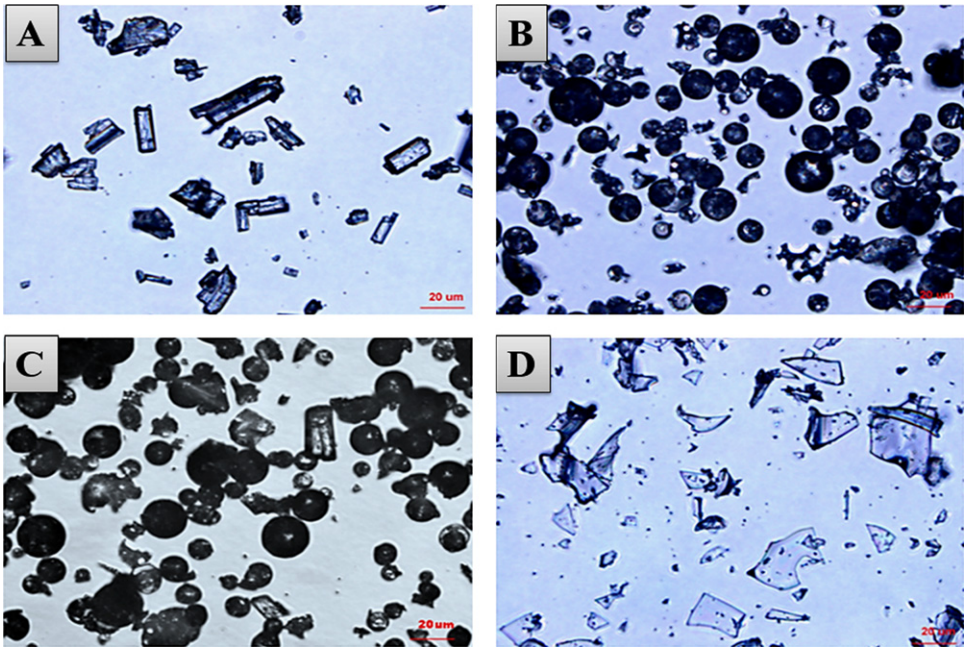

Figure 4 Micrographs of Doxy/HP-β-CD systems. (A) Doxy; (B) HP-β-CD; (C) physical mixture of Doxy/HP-β-CD (Mg^{2+}) and (D) inclusion complex of Doxy/HP-β-CD (Mg^{2+}).

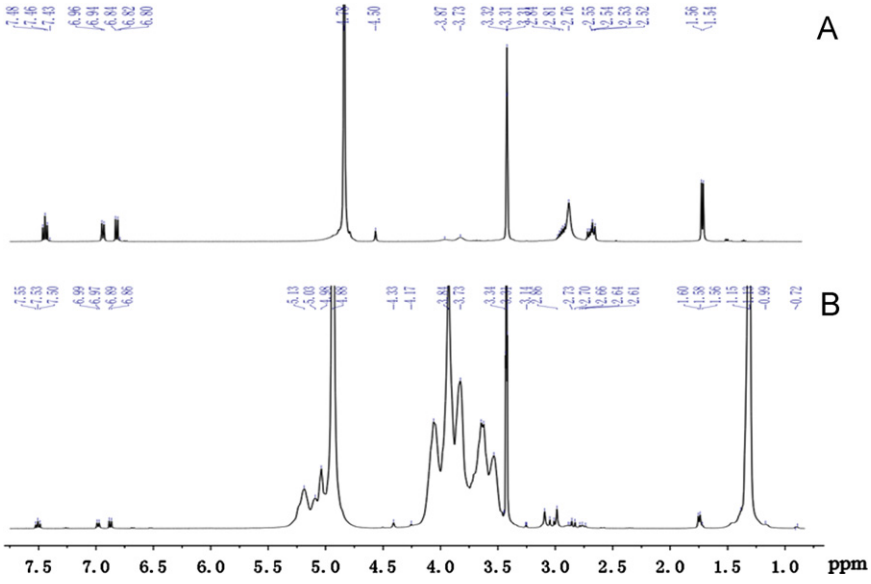

Figure 5 ^1H NMR spectra. (A) Doxy and (B) inclusion complex of Doxy/HP-β-CD(Mg^{2+}) in CD_3OD.

The optimum formulation generated by three models and the corresponding responses are summarized in Table 2. Values of X_1–X_4 for the optimum formulation obtained from the ANN model were 4, 10.8, 12 h and 25 °C, respectively.

3.5. Comparison of RSM, ANN and SVM modeling

Comparing the modeling performance of the three methods, ANN definitely produced the greatest R^2 values (0.9947 for Y_1 and 0.9721 for Y_2). This indicates that ANN based on non-linear modeling with high variability of neuron connections is the most appropriate method to model a complicated system with multiple factors and responses. In contrast, RSM showed a lack of variability in R^2 values (0.9090 for Y_1, 0.9059 for Y_2) due to the limitations of regression equations. Interestingly, SVM with its limited previous application to formulation optimization showed good modeling in terms of R^2 values (0.9723 for Y_1 and 0.9305 for Y_2) because of its relatively high variability for non-linear modeling. However, compared to ANN, SVM is probably more suitable for handling classification problems.

In view of the percentage bias in the experimental data (Table 2), RSM_1 based on the desirability function showed relatively poor prediction (7.08% bias for Y) presumably reflecting a defect of this traditional unitary method. RSM_2 based on GA optimization yielded better prediction (3.59% bias for Y_1 and 6.02% bias for Y_2) than RSM_1 reflecting the excellent ability of GA for parallel computing. The reason SVM yielded better prediction (1.06% bias for Y_1 and 5.67% bias for Y_2) than RSM can be ascribed to the relatively higher level of generalization of SVM for non-linear modeling. However, the lowest percentage bias (0.57% for Y_1 and 1.99% for Y_2) was obtained by the ANN method, convincingly indicating the validity of the generated model.

Table 3 ^{1}H NMR chemical shifts of Doxy and Doxy-HP-β-CD(Mg^{2+}) in CD_3OD at 20 °C.

Hydrogen	δ_{Doxy}	$\delta_{Doxy/HP\text{-}\beta\text{-}CD(Mg^{2+})}$	$\Delta\delta^a$
7-H	6.95	6.98	0.03
8-H	7.48	7.53	0.05
9-H	6.84	6.87	0.03
4-$N(CH_3)_2$	2.75	2.71	−0.04
6-CH_3	1.55	1.58	0.03

[a] $\Delta\delta = \delta_{Doxy/HP\text{-}\beta\text{-}CD(Mg^{2+})} - \delta_{Doxy}$.

3.6. Characterization of the optimum HP-β-CD Doxy inclusion complex

3.6.1. Microscopic observation

Microscopic examination of pure materials, the physical mixture and the optimum inclusion complex are shown in Fig. 4. Doxy (Fig. 4A) appears as club-shaped crystals whereas HP-β-CD (Fig. 4B) exhibits amorphous spherical particles. In the physical mixture (Fig. 4C), Doxy crystals and HP-β-CD particles are clearly observed, indicating the absence of a host-guest interaction. In contrast, the Doxy-HP-β-CD (Mg^{2+}) inclusion complex (Fig. 4D) is in the form of irregular particles of variable size with completely different morphologies from Doxy and HP-β-CD supporting the formation of an inclusion complex.

3.6.2. NMR spectroscopy

Since Doxy is hydrophobic and has low aqueous solubility, it can be encapsulated in the hydrophobic cavity of HP-β-CD. When inclusion occurs, the changes in the physical and chemical environment affects the electronic density of Doxy hydrogen atoms which can be detected by NMR. Given the limited aqueous solubility of Doxy, NMR was performed on solutions of drug and inclusion complex in CD_3OD[32,33]. The ^{1}H NMR spectra of Doxy and its inclusion complex using a molar ratio of 1:4 are shown in Fig. 5 with corresponding proton shifts summarized in Table 3. As shown in Fig. 5B, proton shifts for H7, H8, H9, 4-$N(CH_3)_2$ and 6-CH_3 of Doxy are clearly observed in the complex suggesting a deep penetration of the Doxy molecule into the cavity of HP-β-CD.

3.7. Stability studies

As shown in Table 4 and Fig. 6, Doxy · HCl degraded rapidly at 40 °C with the sample changing from yellow to brown and the amount of related substance increasing significantly. In contrast, the stability of Doxy in the Doxy-HP-β-CD and Doxy-HP-β-CD (Mg^{2+}) complexes was higher as shown by less color change (Fig. 6) and formation of less related substances (Table 4). The results indicate that the formation of an inclusion complex in the presence of Mg^{2+} enhances the stability of Doxy to the greatest extent. This is probably

Table 4 Stability of aqueous solutions of the Doxy-HP-β-CD inclusion complex stored at 40±2 °C for 10 days (n=3).

Sample	Time (day)	Appearance	Doxy content (%)	Related substance (%)
Doxy · HCl (Control)	0	Light yellow clear	100.00	0.93±1.55
	5	Light brown minor deposition	91.20±0.75	5.52±2.23
	10	Brown more deposition	83.10±4.21	9.50±0.62
Doxy/HP-β-CD	0	Light yellow clear	100.00	0.73±0.15
	5	Light yellow clear	93.21±1.57	3.63±0.27
	10	Yellow minor deposition	88.67±3.61	6.41±0.53
Doxy/HP-β-CD(Mg^{2+})	0	Light yellow clear	100.00	0.65±0.06
	5	Light yellow clear	98.93±1.24	1.43±0.44
	10	Yellow minor deposition	95.54±1.51	5.90±0.34

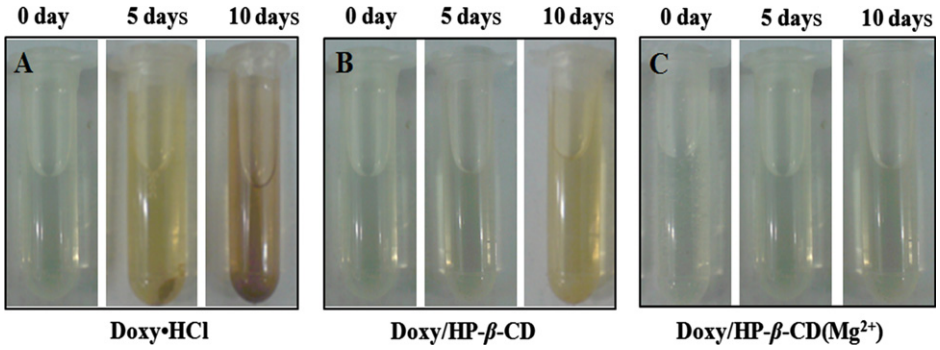

Figure 6 Changes in the appearance of solutions of (A) Doxy · HCl, (B) the Doxy-HP-β-CD complex and (C) the Doxy-HP-β-CD (Mg^{2+}) complex solution on storage at 40 °C over 10 days.

because (1) Mg^{2+} and Doxy formed a chelate at the site of the phenol diketone which protects Doxy from hydrolysis and (2) Doxy is embedded in the HP-β-CD cavity in such a way that its sensitive groups are protected.

4. Conclusions

ANN with one hidden layer and 10 hidden units showed the best modeling and prediction for inclusion efficiency and stability of the Doxy-HP-β-CD (Mg^{2+}) inclusion complex. The optimal formulation of the complex was obtained by a GA-ANN method. Its formation was supported by NMR spectroscopy and it was shown to stabilize Doxy in aqueous solution. Overall the results show that a GA-based approach is a powerful tool for optimizing pharmaceutical formulations in a cost and time efficient manner.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 81173002), the National Science & Technology Pillar Program (Grant No. 2012BAI35B02) and the International Cooperation and Exchange Program of China (Grant No. 2008DFA31080).

References

1. Pan X, Chen MW, Han K, Peng XS, Wen XG, Wu CB, et al. Novel compaction techniques with pellet-containing granules. *Eur J Pharm Biopharm* 2010;**75**:436–42.
2. Su WR, Li ZR, Lin ML, Li YP, He ZX, Wu CB, et al. The effect of doxycycline temperature-sensitive hydrogel on inhibiting the corneal neovascularization induced by BFGF in rats. *Graefes Arch Clin Exp Ophthalmol* 2011;**249**:421–7.
3. Cal K, Centkowska K. Use of cyclodextrins in topical formulations: practical aspects. *Eur J Pharm Biopharm* 2008;**68**:467–78.
4. deVries T, Gold L. Inventors. Warner Chilcott Company, Inc., assignee. Doxycycline metal complex in a solid dosage form. US patent US2005/0019396-A1; 2005 Jan 27.
5. He ZX, Wang ZH, Zhang HH, Pan X, Liang D, Wu CB, et al. Doxycycline and hydroxypropyl-β-cyclodextrin complex in poloxamer thermal sensitive hydrogel for ophthalmic delivery. *Acta Pharm Sin B* 2011;**1**:254–60.
6. Behzadi SS, Unger FM, Viernstein H. Utilizing response surface methodology for optimizing pharmaceutical formulations. *Eur J Pharm Sci* 2005;**25**:54–5.
7. Mennini N, Furlanetto S, Maestrelli F, Pinzauti S, Mura P. Response surface methodology in the optimization of chitosan-Ca pectinate bead formulations. *Eur J Pharm Sci* 2008;**35**:318–25.
8. Achanta A, Kowalski J, Rhodes C. Artificial neural networks: implications for pharmaceutical sciences. *Drug Dev Ind Pharm* 1995;**21**:119–55.
9. Plumb AP, Rowe RC, York P, Doherty C. The effect of experimental design on the modeling of a tablet coating formulation using artificial neural networks. *Eur J Pharm Sci* 2002;**16**:281–8.
10. Plumb AP, Rowe RC, York P, Brown M. Optimisation of the predictive ability of artificial neural network (ANN) models: a comparison of three ANN programs and four classes of training algorithm. *Eur J Pharm Sci* 2005;**25**:395–405.
11. Takayama K, Morva A, Fujikawa M, Hattori Y, Obata Y, Nagai T. Formula optimization of theophylline controlled-release tablet based on artificial neural networks. *J Controlled Release* 2000;**68**:175–86.
12. Amani A, York P, Chrystyn H, Clark BJ. Factors affecting the stability of nanoemulsions-use of artificial neural networks. *Pharm Res* 2010;**27**:37–45.
13. Zhang H, Xiang ML, Zhao YL, Wei YQ, Yang SY. Support vector machine and pharmacophore-based prediction models of multidrug-resistance protein 2 (MRP2) inhibitors. *Eur J Pharm Sci* 2009;**36**:451–7.
14. Pourbasheer E, Riahi S, Ganjali MR, Norouzi P. QSAR study on melanocortin-4 receptors by support vector machine. *Eur J Med Chem* 2010;**45**:1087–93.
15. Do DQ, Rowe RC, York P. Modeling drug dissolution from controlled release products using genetic programming. *Int J Pharm* 2008;**351**:194–200.
16. Hirsch R, Müller-Goymann CC. Fitting of diffusion coefficients in a three-compartment sustained release drug formulation using a genetic algorithm. *Int J Pharm* 1995;**120**:229–34.
17. Barmpalexis P, Kachrimanis K, Georgarakis E. Solid dispersions in the development of a nimodipine floating tablet formulation and optimization by artificial neural networks and genetic programming. *Eur J Pharm Biopharm* 2011;**77**:122–31.
18. Turkoglu M, Aydin I, Murray M, Sakr A. Modeling of a roller-compaction process using neural networks and genetic algorithms. *Eur J Pharm Biopharm* 1999;**48**:239–45.
19. Zhou X, Li ZC, Dai Z, Zou XY. QSAR modeling of peptide biological activity by coupling support vector machine with particle swarm optimization algorithm and genetic algorithm. *J Mol Graph Model* 2010;**29**:188–96.
20. Pourbasheer E, Riahi S, Ganjali MR, Norouzi P. Application of genetic algorithm-support vector machine (GA-SVM) for prediction of BK-channels activity. *Eur J Med Chem* 2009;**44**:5023–8.
21. Jun SW, Kim MS, Kim JS, Park HJ, Lee S, Woo JS, et al. Preparation and characterization of simvastatin/hydroxypropyl-β-cyclodextrin inclusion complex using supercritical antisolvent (SAS) process. *Eur J Pharm Biopharm* 2007;**66**:413–21.

22. Chinese Pharmacopoeia Commission. *The pharmacopoeia of the People's Republic of China*. Beijing: China Medical Science and Technology Press; 2010, p. 690–1.
23. Liu Y, Zhang P, Feng N, Zhang X, Wu S, Zhao J. Optimization and *in situ* intestinal absorption of self-microemulsifying drug delivery system of oridonin. *Int J Pharm* 2009;**365**:136–42.
24. Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995;**20**:273–97.
25. Chang CC, Lin CJ. LIBSVM: A library for support vector machines. Available from: 〈http://www.csie.ntu.edu.tw/~cjlin/libsvm〉. 2001-06-19/2002-03-10.
26. Michalewicz Z. *Genetic algorithms+data structures=evolution program*. 3rd edition. New York: Springer-Verlag; 1996.
27. Chipperfield A, Fleming P, Pohlheim H, Fonseca C. Genetic algorithm toolbox user's guide. ACSE research report No. 512. UK: University of Sheffield; 1994.
28. Hecht-Nielsen R. Kolmogorov's mapping neural network existence theorem. In: *Proceedings of the IEEE International Conference on Neural Networks*, vol 3. New York, US; 1987. p. 11–4.
29. Jadid MN, Fairbairn DR. Neural-network applications in predicting moment-curvature parameters from experimental data. *Eng Appl Artif Intell* 1996;**9**:309–19.
30. Carpenter JC, Hoffman ME. Understanding neural network approximations and polynomial approximations helps neural network performance. *AI Expert* 1995;**10**:31–3.
31. Wang W, Xu Z, Lu W, Zhang X. Determination of the spread parameter in the Gaussian kernel for classification and regression. *Neurocomputing* 2003;**55**:643–63.
32. Rojas-Aguirre Y, Yepez-Mulia L, Castillo I, Lopez-Vallejo F, Soria-Arteche O, Hernandez-Campos A, et al. Studies on 6-chloro-5-(1-naphthyloxy)-2-(trifluoromethyl)-1*H*-benzimidazole/2-hydroxypropyl-beta-cyclodextrin association: characterization, molecular modeling studies, and *in vivo* anthelminthic activity. *Bioorg Med Chem* 2011;**19**:789–97.
33. Zhang HH, Chen MW, He ZX, Wang ZH, Liang D, Wu CB, et al. Molecular modeling-based inclusion mechanism and stability studies of doxycycline and hydroxypropyl-β-cyclodextrin complex for ophthalmic delivery. *AAPS Pharm Sci Tech* 2013;**14**:10–8.