journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Empirical search for factors affecting mean particle size of PLGA microspheres containing macromolecular drugs

Jakub Szlęk [a,*], *Adam Pacławski* [a], *Raymond Lau* [b], *Renata Jachowicz* [a], *Pezhman Kazemi* [a], *Aleksander Mendyk* [a]

[a] *Department of Pharmaceutical Technology and Biopharmaceutics, Jagiellonian University Medical College, Medyczna 9 St., 30-688 Cracow, Poland*

[b] *School of Chemical and Biomedical Engineering, Nanyang Technological University, 62 Nanyang Drive, Singapore 637459, Singapore*

ARTICLE INFO

Article history:
Received 9 February 2016
Received in revised form
16 May 2016
Accepted 1 July 2016

Keywords:
Poly-lactide-co-glycolide (PLGA)
microparticles
Heuristic modelling
Rules extraction
Feature selection
Molecular descriptors
Particle size

ABSTRACT

Background and objectives: Poly(lactic-co-glycolic acid) (PLGA) has become one of the most promising in design, development, and optimization for medical applications polymers. PLGA-based multiparticulate dosage forms are usually prepared as microspheres where the size is from 5 to 100 μm, depending on the route of administration. The main objectives of the study were to develop a predictive model of mean volumetric particle size and on its basis extract knowledge of PLGA containing proteins forming behaviour.

Methods: In the present study, a model for the prediction of mean volumetric particle size developed by an rgp package of R environment is presented. Other tools like fscaret, monmlp, fugeR, MARS, SVM, kNNreg, Cubist, randomForest and piecewise linear regression are also applied during the data mining procedure.

Results: The feature selection provided by the fscaret package reduced the original input vector from a total of 295 input variables to 10, 16 and 19. The developed models had good predictive ability, which was confirmed by a normalized root-mean-square error (NRMSE) of 6.8 to 11.1% in 10-fold cross validation training procedure. Moreover, the best models were validated using external experimental data. The superior predictiveness had a model obtained by rgp in the form of a classical equation with a normalized root-mean-squared error (NRMSE) of 6.1%.

Conclusions: A new approach is proposed for computational modelling of the mean particle size of PLGA microspheres and rules extraction from tree-based models. The feature selection leads to revealing chemical descriptor variables which are important in predicting the size of PLGA microspheres. In order to achieve better understanding in the relationships between particle size and formulation characteristics, the surface analysis method and rules extraction procedures were applied.

* *Corresponding author.* Department of Pharmaceutcial Technology and Biopharmaceutics, Jagiellonian University Medical College, Medyczna 9 st., 30-688 Cracow, Poland. Fax +48 12 620 56 18.
E-mail address: j.szlek@uj.edu.pl (J. Szlęk).

http://dx.doi.org/10.1016/j.cmpb.2016.07.006
0169-2607/

1. Introduction

Poly(lactic-co-glycolic acid) (PLGA) is one of the most attractive polymers used to manufacture many drug formulations, mainly due to its satisfactory safety profile and its ability to modify the drug dissolution profile for controlled-release formulations [1]. As an FDA-approved polymer, PLGA has been studied as a delivery vehicle for unstable active pharmaceutical ingredients (API) such as proteins, peptides, RNA, and DNA [2,3,4].

The quality of PLGA-based dosage forms is a complex problem, which is the topic of many research reports presented in the literature. Drug release and bioavailability of APIs from PLGA microspheres is influenced by many factors, and the size of particles seems to be one of the most important as it also determines the route of administration and i.e., the cytotoxicity of the formulation [5,6]. Therefore, the current study is focused on model development for the PLGA particle size prediction and to relate the size with the particle formation process and a physicochemical description of a formulation. Earlier, Vysloužil et al. [7] correlated two factors affecting the PLGA particle size: the stirring speed during solvent evaporation and the polyvinyl alcohol (PVA) concentration. Moreover, Martín-Sabroso et al. [8] and Le Visage et al. [9] studied the effect of the polymer molecular weight and its polarity. Higher molecular weight and polarity led to larger microparticles and porosity. These effects were explained by the increase of the inherent viscosity of the polymer [8]. Finally, Dorati et al. [10] have pointed out that the addition of a plasticizer in the inner aqueous phase is crucial for the PLGA formation process. It turned out that the addition of polyethylene glycol (PEG) or PVA exhibits an increase in mean particle size of microspheres.

Apart from that, there are numerous articles treating relationships of formulation process and used excipients which influence the diameters of microparticles: however, to the authors' best knowledge, there is no general computational model able to predict the size of particles on the basis of parameters of the production process and the formulation composition.

The design of the PLGA microsphere formulations depends heavily on the trial-and-error approach. In terms of the quality and reproducibility of the process, it is crucial to find and understand variables affecting the particle size. On the other hand, production methods and used equipment are well established for PLGA microspheres formulation. Therefore, if one could develop an accurate model predicting the size of obtained particles based on assay conditions and constituents it would speed up development of tailored formulations. Therefore, knowledge discovery on this issue is important for research and development and, in perspective, it could make the formulation process less time- and resource consuming.

The objective of this work is to demonstrate the use of feature selection and data mining tools cooperatively to search available design space in order to create a predictive model of the mean particle size of PLGA microspheres. Based on the developed model, an analysis of the influence of dependent variables on mean particle size was performed. Moreover, final model usability was confirmed by testing on an external data set.

2. Materials and methods

2.1. Learning and validation data set

The learning data were adapted from the publication of Szlęk et al. [11]. In brief, the source data set was rearranged in order to give 68 formulations composed of 295 independent variables which were grouped in four categories: 1) API chemical descriptors; 2) experimental conditions including PLGA inherent viscosity, PLGA molecular weight, Lactide to Glycolide ratio, inner and outer phase PVA concentration, PVA molecular weight, inner phase volume, encapsulation rate, PLGA concentration, and production method; 3) plasticizer chemical descriptors; and 4) emulsifier chemical descriptors. Macromolecule and excipient molecular descriptors were computed using ChemAxon's Marvin cxcalc plugin, UK, v 5.11 [12]. Selected PLGA formulations were prepared using four methods: 1—a double-emulsion water-in-oil-in-water (w/o/w) solvent extraction process, 2—a solid-in-oil-in-water (s/o/w) emulsion method, 3—an oil-in-oil solvent evaporation (s/o/o) technique, and 4—a spray-drying process. Formulations included 14 model substances: Bovine Serum Albumin, Recombinant Human Erythropoietin, Recombinant Human Epidermal Growth Factor, Lysozyme, Recombinant Human Growth Hormone, Hen Ovalbumin, Human Serum Albumin, Beta-Amyloid, Insulin, Recombinant Human Erythropoietin coupled with Human Serum Albumin, L-Asparaginase, Bovine Insulin, Alpha 1-Antitrypsin, and Chymotrypsin. The dependent variable was mean volumetric particle size (μm, referred as 'mean particle size' later on in the text).

After rearrangement of the original data, the data set was split according to the 10-fold cross-validation scheme (noted as 10cv) and as a result learning and testing data sets were obtained. Due to the large disparate of chemical descriptor ranges and in order to save the original distribution of the data set, the data were scaled linearly in a range from 0.1 to 0.9 prior to further modelling (equation 1).

$$x_{scaled} = 0.1 \cdot \left(1 - \frac{x_i - x_{min}}{x_{max} - x_{min}}\right) + 0.9 \cdot \left(\frac{x_i - x_{min}}{x_{max} - x_{min}}\right) \quad (1)$$

where x_{scaled} is scaled value; x_i is i value × variable; and x_{max} and x_{min} are maximal and minimal values of variable x.

The validation data set was collected from literature and was not included into any part of the modelling scheme; it was used only to test the predictability of the final models [8,13,14]. The data set consisted of 3 formulations, where APIs were as follows: Albumin [8], Nonstructural Protein 1 from Dengue 2 Virus [13], and Ovalbumin [14]. The mean particle size of PLGA microspheres ranged from 4.8 to 112 μm. Learning and validation data sets are presented in the Supplementary material (S1 and S2).

2.2. Feature selection

Standard model selection schemes are efficient when they do not exceed a certain critical complexity level, depending on ratio variables to records in the data set. Therefore, data mining should consider elimination of redundant and unnecessary