

# MODELOWANIE NIELINIOWE WIELKOŚCI MIKROCZĄSTEK KOPOLIMERU KWASU MLEKOWEGO I GLIKOLOWEGO ZAWIERAJĄCYCH PEPTYDY

Szłek Jakub<sup>1\*</sup>, Paclawski Adam<sup>1</sup>, Lau Wai Man Raymond<sup>2</sup>, Mendyk Aleksander<sup>1</sup>, Jachowicz Renata<sup>1</sup>  
<sup>1</sup> Katedra Technologii Postaci Leku i Biofarmacji, Uniwersytet Jagielloński Collegium Medicum, ul.  
Medyczna 9, 30-688 Kraków, e-mail\*: j.szlek@uj.edu.pl

<sup>2</sup> Division of Chemical and Biomolecular Engineering, School of Chemical and Biomedical  
Engineering, College of Engineering Nanyang Technological University (NTU), 62 Nanyang Drive,  
Singapore 637459

## 1. Wstęp

Ze względu na wiele możliwości zastosowania oraz zadowalający profil bezpieczeństwa, mikrocząstki zbudowane z kopolimeru kwasu mlekowego i glikolowego (PLGA) stanowią interesującą postać leku, o potencjalnie szerokim zastosowaniu terapeutycznym.

Do głównych zalet PLGA, jako nośnika substancji leczniczej należą: właściwości ochronne dla substancji peptydowych oraz możliwość formułowania postaci leku o przedłużonym uwalnianiu [1]. Cechy te zachęcają do szerszego zastosowania PLGA jako substancji pomocniczej. Jednym z problemów technologicznych podczas formułowania postaci leku z PLGA jest wielkość mikrocząstek. Na ich rozmiar mają wpływ przede wszystkim parametry metody sporządzania oraz właściwości polimeru, a także rodzaj i ilość substancji pomocniczych.

Celem pracy było opracowanie obliczeniowego modelu przewidującego wielkość cząstek formułacji złożonych z PLGA zawierających peptydy wspomagającego racjonalne projektowanie postaci leku z uwagi na oczekiwaną średnią wielkość cząstek.

## 2. Materiały i metody

### 2.1. Baza danych

Baza danych została utworzona na podstawie danych z piśmiennictwa. Przeszukanie baz danych Scopus i ScienceDirect dało wstępnie około 200 publikacji, z których ostatecznie wyselekcjonowano 68 stanowiących rekordy w bazie danych. Każdą formułację kodowało 298 parametrów opisujących zarówno metodę wytwarzania mikrocząstek (15 zmiennych) oraz deskryptory molekularne: 85 dotyczyło opisu peptydu, 98 plastyfikatora i 100 emulgatora. Pojedyncze wyjście stanowiła średnia wielkość mikrocząstek (MPS). Dodatkowo bazę danych podzielono na zbiory uczący-testowy zgodnie z procedurą 10-krotnego wzajemnego sprawdzania.

### 2.2. Informatyka chemiczna

Struktury chemiczne peptydów pobierano z serwerów PDB.org lub uniprot.org, natomiast małe cząsteczki rysowano, a potem optymalizowano ich strukturę 3D. Deskryptory molekularne zostały obliczone przy użyciu aplikacji cxcsl i molconv pakietu Marvin firmy ChemAxon (Marvin 5.11.0, 2010).

### 2.3. Redukcja wektora wejściowego

Selekcja zmiennych kluczowych została przeprowadzona na podstawie utworzonego rankingu zmiennych przez pakiet fscaret [2] środowiska obliczeniowego R [3] i analizy wrażliwościowej przeprowadzonej przy użyciu modeli sztucznych sieci neuronowych [4]. Selekcję zmiennych włączanych do wektora wejściowego prowadzono na podstawie spadku gradientu wartości zmiennych, przyjęto próg 5%.

### 2.4 Uczenie maszynowe.

Podczas uczenia maszynowego użyto pakietów środowiska R, tj. monmlp (ANN MON-MLP) [5], Cubist (CUB) [6], rgp [7]-optimx [8] (GP) oraz symulatora sztucznych sieci neuronowych Nets2013 (ANN-MLP) [4]. Biorąc pod uwagę ilość etapów redukcji wektora wejściowego, kroki uczenia oraz

procedurę 10-krotnego wzajemnego sprawdzania liczbą wytrenowanych i przetestowanych modeli wyniosła w sumie około 180 000. Do oceny sprawności modeli użyto miary NRMSE, czyli RMSE liczonego względem zakresu zmiennej wyjściowej (226 – 0.336 µm), uwzględniając przy tym średnią z 10-krotnego sprawdzania (10-cv).

### 3. Wyniki i Dyskusja

Wyselekcjonowano wstępnie 6 wektorów wejściowych: 12in, 18in, 19in, 20in, 28in, 29in, które posłużył jako zbiory w dalszym etapie modelowania.

Wytypowano jako najlepsze modele sieci ANN-MLP o 5 warstwach ukrytych i 200, 80, 40, 20, 10 neuronach oraz z funkcją aktywacji tangens hiperboliczny, dla których NRMSE liczony z 10-cv wyniósł 6.5%. Pozostałe algorytmy dawały o 1-3% większe błędy generalizacji.

Wyniki generalizacji (NRMSE) zgodne z 10-cv dla algorytmów ANN-MLP, GP, CUB i ANN MON-MLP przedstawiono w Tabeli 1.

Wektor wejściowy dla najlepszego modelu (12in) zawierał dwa deskryptory molekularne dla peptydu (Hyper wiener index, Struktura czwartorzędowa) i jeden deskryptor dla plastyfikatora (Hyper wiener index), co może sugerować, że oprócz parametrów przygotowania formułacji (Lepkość PLGA [mPa\*s], Masa PLGA [Da], Stosunek kwas mlekowy : kwas glikolowy, Stężenie PVA w fazie wewnętrznej [%], Stężenie PVA w fazie zewnętrznej [%], Stopień enkapsulacji [%], Ilość merów bloku, Stężenie PLGA [%], Metoda wytwarzania), zarówno właściwości substancji leczniczej oraz plastyfikatora mają wpływ na wielkość mikrocząstek (Tabela 2).

Model GP pozwolił na uzyskanie reprezentacji matematycznej zależności 7 zmiennych wejściowych – deskryptorów chemicznych opisujących peptyd (Hyper wiener index, Struktura czwartorzędowa), zmiennych przedstawiających przebieg procesu przygotowania i badania in vitro formułacji (Masa cząsteczkowa PLGA [Da], Stosunek kwas mlekowy : kwas glikolowy, Stężenie PVA w fazie zewnętrznej [%], pH badania in vitro, Metoda wytwarzania), a także czterech stałych (C1 – C4) stanowiących w równaniu parametry swobodne, od średniej wielkości mikrocząstek (MPS) z błędem generalizacji NRMSE równym 8.3% (Równanie 1, Tabela 3).

Tabela 1. Zestawienie wyników uczenia maszynowego.

Algorytm	Wektor wejściowy	NRMSE [%]
ANN-MLP	12in	6.5
CUB	12in	8.2
GP	19in	8.3
MON-MLP	19in	9.6

$$MPS = \left( e^{\left( \frac{X6}{\sqrt[4]{X13 - X10}} \right)} \cdot \left( X14 + \left( e^{\left( \frac{X6}{\sqrt[4]{X13 - X10}} \right)} \right)^{-C1} + \left( \ln(X7) - \sqrt[4]{ e^{\left( (X14 + X8) \left( \frac{X4}{\sqrt[4]{X13 - X10}} \right) \right)} + e^{\left( \left( \frac{\ln(X13)}{\sqrt{(C2) - X14}} \right)^{-C3} + e^{((C4 + X8)^{10})} \right)} \right)} \right) \right)$$

Równanie 1. Matematyczna reprezentacja modelu GP (znaczenie symboli takie jak w Tabeli 3).

Tabela 2. Zmienne włączone do modelu ANN-MLP.

Numer	Zmienna	Opis	
1	Hyper wiener index	Deskryptory molekularne peptydu	
2	Struktura czwartorzędowa		
3	Lepkość PLGA [mPa*s]	Zmienne przedstawiające przebieg procesu przygotowania formułacji	
4	Masa PLGA [Da]		
5	Stosunek kwas mlekowy : kwas glikolowy		
6	Stężenie PVA w fazie wewnętrznej [%]		
7	Stężenie PVA w fazie zewnętrznej [%]		
8	Efektywność zamykania [%]		
9	Ilość merów bloku		
10	Stężenie PLGA [%]		
11	Metoda wytwarzania		
12	Hyper wiener index		Deskryptor molekularny plastyfikatora
<b>Wyjście</b>	<b>Średnia wielkość mikrocząstek [<math>\mu\text{m}</math>]</b>		

Tabela 3. Zmienne włączone do modelu GP. Metody wytwarzania: s/o/o, s/o/w, w/o/w [9].

Symbol w równaniu	Zmienna	Opis
X4	Ring count of atom	Deskryptory molekularne peptydu
X6	Struktura czwartorzędowa	
X7	Masa cząsteczkowa PLGA [Da]	Zmienne przedstawiające przebieg procesu przygotowania i badania in vitro formułacji
X8	Stosunek kwas mlekowy : kwas glikolowy	
X10	Stężenie PVA w fazie zewnętrznej [%]	
X13	pH badania in vitro	
X14	Metoda wytwarzania	
C1, C2, C3, C4	Stałe parametry równania	
<b>MPS</b>	<b>Średnia wielkość mikrocząstek [<math>\mu\text{m}</math>]</b>	<b>Zmienna wyjściowa</b>

#### 4. Podsumowanie

Uczenie maszynowe doprowadziło do otrzymania sprawnych modeli nieliniowych, których błąd generalizacji (NRMSE) jest mniejszy niż 10%. Modele mogą posłużyć do zmniejszenia nakładu eksperymentalnego w celu otrzymania mikrocząstek PLGA o zadanej wielkości. Wykazano, że zarówno parametry procesu technologicznego oraz deskryptory molekularne substancji leczniczej i substancji pomocniczych mają wpływ na wielkość mikrocząstek PLGA. Ponadto jako część procedury modelowania przeprowadzono selekcję zmiennych kluczowych, która jest jednym z etapów automatycznej ekstrakcji wiedzy realizowanej przez system komputerowy.

Badania finansowane ze środków projektu NCBiR nr 168/2011, w ramach współpracy bilateralnej Polska-Singapur.

## 6. Piśmiennictwo

- [1] Makadia H. K., Siegel S. J. (2011). Poly Lactic-co-Glycolic Acid (PLGA) as Biodegradable Controlled Drug Delivery Carrier, *Polymers*, 3, 1377-1397.
- [2] Szlęk J. (2013). fscaret: Automated caret feature selection. R package version 0.8.5.3. <http://CRAN.R-project.org/package=fscaret>
- [3] R Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>
- [4] Mendyk A., Jachowicz R. (2007). Unified methodology of neural analysis in decision support systems built for pharmaceutical technology, *Expert Syst Appl*, 32, 1124–1131.
- [5] Cannon AJ., 2013, monmlp: Monotone multi-layer perceptron neural network. R package version 1.1.2. <http://cran.r-project.org/web/packages/monmlp/index.html>.
- [6] Kuhn M, Weston S, Keefer Ch, Coulter N, Quinlan R. (2013). Cubist: Rule- and Instance-Based Regression Modeling. R package version 0.0.13. <http://CRAN.R-project.org/package=Cubist>
- [7] Flasch O., Mersmann O., Bartz-Beielstein T., Stork J., Zaefferer M. (2013). rgp: R genetic programming framework. R package version 0.3-4. <http://CRAN.R-project.org/package=rgp>
- [8] Nash JC., Varadhan R. (2011). Unifying Optimization Algorithms to Aid Software System Users: optimx for R. *Journal of Statistical Software*, 43(9), 1-14. <http://www.jstatsoft.org/v43/i09/>.
- [9] Szlęk J., Paclawski A., Lau R., Jachowicz R., Mendyk A. (2013). Heuristic modeling of macromolecules release from PLGA microspheres, *Int J Nanomed*, 8(1), 4601–4611.