

# Finding Rules for Insulin-loaded Nanoparticles with FormRules

## Background

Understanding complex formulations is a challenging task, and discovering how changing formulation and processing variables affects end-use properties can be time-consuming. Indeed, it might not be possible to discover all of the relationships easily using simple statistical or visualization techniques. This is especially true in new formulation areas, such as the delivery of novel peptides and proteins produced by recent advances in biotechnology.

The present study investigates the use of neurofuzzy computing, a 'machine learning' technique which automatically discovers relationships within data and which reports the results in the form of 'rules'. A comparison with reported statistical studies is undertaken.

## Insulin loaded Nanoparticle Data

The data used in this study have been published by D Attivi, P Wehrle, N Ubrich, C Damge, M Hoffman and P Maincent, in *Drug Dev. Ind. Pharm.* 31, 179 (2005) for insulin loaded nanoparticles formed by a water-in-oil-in-water emulsification and drying process.

A central composite experimental design was used in data collection, with controlled variations allowed in three factors:

- ratio of poly(epsilon caprolactone) to Eudragit RS, referred to as PCL/RS ratio
- the volume of the PVA aqueous solution
- pH of the aqueous PVA solution).

The five properties that were measured were the size (nm), polydispersity index, zeta potential, the amount of entrapped insulin, and the amount of released insulin after 7 hours.

Attivi *et al.* reported 23 experiments, of which 18 were unique, while five were additional repeats of the centroid point. To avoid biasing the design space, in the present work the results for the 5 experiments at the centroid were averaged and the data mining study was performed just on the 18 unique points.

The data have been examined using **FormRules**, a data-mining program based on neurofuzzy computing – itself a technique which combines the learning capabilities of neural networks with fuzzy logic's ability to express complex concepts in a simple intuitive language.

ANOVA statistics have been used to assess the quality of the models developed. This facilitates comparison with the work of Attivi *et al.* (1) who also report ANOVA  $R^2$  values. The ANOVA  $R^2$  values for our models, together with information on the training criteria, are given in Table 1. In this table, SRM denotes that the model selection criterion is Structural Risk Minimization and MDL indicates that it is Minimum Descriptor Length.

| Property          | Model Selection Criterion | $R^2$ | $R^2$ Stats Model |
|-------------------|---------------------------|-------|-------------------|
| Particle size     | SRM                       | 0.94  | 0.91              |
| Polydispersity    | MDL                       | 0.87  | 0.86              |
| Zeta potential    | MDL                       | 0.94  | 0.57              |
| Entrapped insulin | SRM                       | 0.90  | 0.81              |
| Released insulin  | MDL                       | 0.83  | 0.91              |

**Table 1. Training parameters and ANOVA  $R^2$  value for the various property models, compared to statistical treatment of Attivi *et al.***

## Models and Rules for Particle Size

Particle size was found to depend on the PCL/RS ratio and on the pH; these are the same variables as were found by the statistical study. There is an interaction between these two variables, as illustrated by the rules:

- IF pH is LOW AND PCL/RS Ratio is LOW THEN Size is LOW (1.00)
- IF pH is HIGH AND PCL/RS Ratio is LOW THEN Size is LOW (0.97)
- IF pH is MID AND PCL/RS Ratio is HIGH THEN Size is LOW (0.96)
- IF pH is MID AND PCL/RS Ratio is LOW THEN Size is LOW (0.94)
- IF pH is LOW AND PCL/RS Ratio is HIGH THEN Size is LOW (0.81)

IF pH is HIGH AND PCL/RS Ratio is HIGH  
THEN Size is HIGH (1.00)

Here, the values in parentheses are 'confidence levels' for the rules.

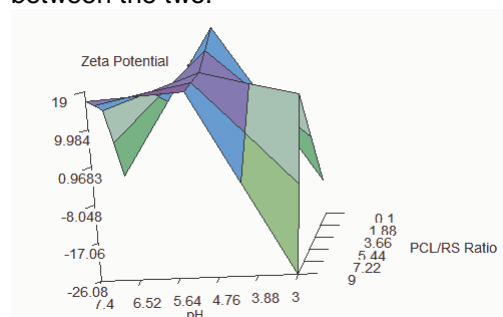
### Models and Rules for Polydispersity

Like particle size, polydispersity depends on the PCL/RS ratio and on pH. Again, these are the same variables as were found by the statistical study.

### Models and Rules for Zeta Potential

This is a more interesting case, since it was not possible to obtain a good model with statistics. An  $R^2$  value of 0.57 was the best they could find, although they did conclude that pH was probably the most important variable.

With neurofuzzy logic, using SRM as the model selection criterion, a good model could not be developed either. However, switching the model selection criterion to MDL allowed a model to be developed, although this was quite complex as Figure 1 shows. The  $R^2$  value of 0.93 indicates that overtraining is unlikely to have occurred, and that the model should be reliable. It indicates that PCL/RS ratio and the pH are the only contributing factors, and that there is an interaction between the two.



**Figure 1. Graphical representation of model for Zeta Potential**

This is represented by a complex set of rules:

- IF pH is LOW AND PCL/RS Ratio is LOW THEN Zeta Potential is LOW (1.00)
- IF pH is LOW AND PCL/RS Ratio is MID THEN Zeta Potential is LOW (0.50)
- IF pH is LOW AND PCL/RS Ratio is HIGH THEN Zeta Potential is LOW (1.00)
- IF pH is MID AND PCL/RS Ratio is LOW THEN Zeta Potential is HIGH (1.00)
- IF pH is MID AND PCL/RS Ratio is MID THEN Zeta Potential is HIGH (0.80)
- IF pH is MID AND PCL/RS Ratio is HIGH THEN Zeta Potential is HIGH (0.95)
- IF pH is HIGH AND PCL/RS Ratio is LOW THEN Zeta Potential is LOW (1.00)

- IF pH is HIGH AND PCL/RS Ratio is MID THEN Zeta Potential is LOW (0.62)
- IF pH is HIGH AND PCL/RS Ratio is HIGH THEN Zeta Potential is HIGH (0.88)

### Models and Rules for Entrapped Insulin

The model for entrapped insulin depended primarily on the volume of PVA, although pH had a role to play. The 'combined rules' show that

IF Vol PVA is LOW AND IF pH is MID THEN Entrapped insulin is HIGH.

There is a maximum in the amount of entrapped insulin when pH is in the mid-range, and it decreases when pH is either LOW or HIGH.

### Models and Rules for Released Insulin

The repeats in the data show a fair degree of scatter in the results for the amount of released insulin, so not surprisingly using SRM as the model selection criterion gave poor results. A reasonable model was developed with MDL, although the  $R^2$  value was lower than that reported from the statistical study. The PCL/RS ratio was the most important variable, with the volume of PVA playing a lesser role. There was no interaction between these variables. The 2 full sets of rules are:

- IF Vol PVA is LOW THEN Entrapped insulin is HIGH
- IF Vol PVA is MID THEN Entrapped insulin is LOW
- IF Vol PVA is HIGH THEN Entrapped insulin is LOW

- IF pH is LOW THEN Entrapped insulin is LOW (0.78)
- IF pH is MID THEN Entrapped insulin is HIGH (0.70)
- IF pH is HIGH THEN Entrapped insulin is LOW (0.91)

### Conclusions

Neurofuzzy logic as embodied in **FormRules** allows models and rules to be developed automatically from data. It is considerably easier to apply than a statistical treatment, and the results are comparable or better. In an emerging area like the delivery of novel peptide therapeutics, this technique can be valuable in guiding experiments and optimizing delivery.

*For further information on **FormRules** and on applying neurofuzzy logic to your problems, contact us at the address below.*

© 2008 Intelligensys Ltd  
All rights reserved.