ELSEVIER

Journal of Pharmaceutical and Biomedical Analysis
22 (2000) 985–992

JOURNAL OF
PHARMACEUTICAL
AND BIOMEDICAL
ANALYSIS

www.elsevier.com/locate/jpba

Ranitidine hydrochloride X-ray assay using a neural network

S. Agatonovic-Kustrin * , V. Wu, T. Rades, D. Saville, I.G. Tucker

School of Pharmacy, University of Otago, P.O. Box 913, Dunedin, New Zealand

Received 8 March 1999; received in revised form 14 January 2000; accepted 18 January 2000

Abstract

A simple X-ray powder diffractometric (XRD) method with artificial neural networks (ANNs) for data modelling was developed to recognize and quantify two crystal modifications of ranitidine–HCl in mixtures and thus, provide information about the solid state of the bulk drug. The method was also used to quantify ranitidine–HCl from tablets in the presence of other components. An ANN consisting of three layers of neurons was trained by using a back-propagation learning rule. A sigmoid output function was used in the hidden layer to facilitate non-linear fitting. Unlike other techniques the ANN method described here employed pattern recognition on the entire XRD pattern. Correct classification was mainly influenced by the XRD pattern resolution. It was shown that data transformations improved the quantitative performance when the XRD patterns were not contaminated by other components. Only smoothed X-ray diffractograms were required to distinguish between the two crystalline forms in a mixture. In the case of ranitidine–HCl quantification from tablets, where significant interference with tablet excipients was present, better results were obtained without data transformations. The trained ANN perfectly quantified ranitidine–HCl polymorphic forms from mixtures (mean sum of squared error was less than 0.02%) and ranitidine–HCl form 1 from tablets (recovery = 98.65). Excellent quantification performance of the ANN analysis, demonstrated in this study, serves as an indication of the broad potential of neural networks in pattern analysis. While the system described has been developed to interpret XRD patterns, peak detection has implications in every chemical application where the recognition of peak-shaped signals in analytical data is important.

Keywords: X-ray diffraction; Artificial neural networks; Ranitidine–HCl; Polymorphysm

1. Introduction

Polymorphism, the ability of a compound to crystallize in more than one arrangement of molecules, is a significant problem in the pharmaceutical industry. A major goal of pharmaceutical research and development is to produce the same drug substance and product it continuously. The existence of a compound in more than one crystalline form may lead to difficulties in formulations. Any polymorphic change in the dosage forms can influence its stability and even

* Corresponding author. Tel.: +64-3-4795410; fax: +64-3-4797034.
E-mail address: nena.kustrin@stonebow.otago.ac.nz (S. Agatonovic-Kustrin)

0731-7085/00/$ - see front matter
PII: S0731-7085(00)00256-9

bioavailability. It is vital to select the polymorph with the desired properties, and predict problems such as the unwanted crystallization of other polymorphs.

Non-destructive methods of analysis that allow rapid, sufficiently precise and reliable quality control have wide applications in many production systems. X-ray diffractometry (XRD) is a powerful technique for characterising pharmaceutical solids [1], which makes this technique particularly useful for the identification of different polymorphs of a drug. Every crystalline solid phase has a unique XRD pattern independent of the other components in the mixture making independent analysis feasible [2,3].

The verification and recognition of peak shaped signals in analytical data and their application to quantitative studies are significant problems. Experimental data usually contain overlapping signals and noise that make sensitive and reliable peak recognition and quantification difficult.

This research evaluated the feasibility of using artificial neural networks (ANNs) to recognize peak-shaped signals in analytical data. ANNs have been shown to be superior to conventional classifiers (e.g. multiple linear regression) at pattern classification where the input is noisy and the system is not well defined [4–6]. Besides, ANNs are non-linear estimators and can establish more sophisticated responses. They can store large amounts of pattern information with relatively few neurons and connections.

The main aim of this research involved two steps: to investigate the ability of ANNs to recognize and quantify two crystal modifications of ranitidine–HCl in mixtures from X-ray diffraction data and thus, provide information about the solid state of the active ingredient; to quantify ranitidine hydrochloride from tablets in the presence of other components.

2. ANNs

Neural computing simulates the neural behavior of living beings so that a computer can learn to differentiate or model without detailed programming and conventional analysis. An ANN is composed of a number of interconnected processing elements (artificial neurons) organized in layers, the input layer, the output layer and the hidden layer between them. The input layer neurons receive data from a data file and the output neurons provide the ANN's response to the input data. Hidden neurons communicate only with input and output neurons. They are part of the large internal pattern that determines a solution to the problem. Like people, an ANN learns by example from experience through a training phase when some of these interconnections are strengthened and some are weakened, so that a neural network will output a more correct answer. The complexity of the network is related to the number of weights and strength of the weights. The number of weights depends on the number of hidden units and the number of hidden layers. Most functions can be approximated using a single hidden layer.

Although there are many types of ANNs, the one that predominates in the area of pattern recognition is the feed-forward, back-propagation network. In this type of model input data are fed forward through the ANN to optimize the weights between neurones, or to 'train' the ANN. The output is related to the combined input and weights by a transfer function, most commonly being of a sigmoid type function. The optimization is, therefore, non-linear.

The use of the weighted links is essential to the ANN's recognition abilities. As the ANN reads the input and output values in the training set the error in the prediction is propagated back through the system, and the interunit connections are changed to minimize the error in the predictions. This process is continued with multiple training sets until the error is minimized across many sets. When the ANN produces the desired output, the weighted links between the units are saved. These weights are used to predict the correct outcome on a new set of input data.

3. Experimental methods and materials

3.1. Materials

The two polymorphic forms of ranitidine–hy-

drochloride (form (I) (Ch.-B 560018) and form (II) (A.-Nr. 32005)) and ranitidine–HCl tablets (150 mg of form 1) and excipients (mycrocrystalline cellulose Avicel PH 301, vinylpyrolidone–vinylacetate copolymer Kollodon VA 64 and magnesium stearate) were provided by Dolorgiet Pharmaceuticals. The polymorphic forms were characterised using elemental analysis, XRD, Fourier transform-infrared spectroscopy (FTIR), diffuse reflectance infra-red Fourier transform spectroscopy (DRIFTS), Raman spectroscopy (Raman), scanning electron microscopy (SEM) and light microscopy as described previously [7].

3.2. Methods

In the preliminary experiments about 360 mg of powder samples were compressed using different pressures of 1, 2 and 4 t. The resulting tablets were exposed to X-ray scans. The relative intensities of the major peaks remained constant indicating that the application of pressure did not induce any preferred orientation. In addition, a flat surface was achieved, minimizing negative interference due to sample curvature or irregular sample surface.

Binary mixtures were made from powder forms with different factions of polymorphic form 2 in the mixtures as follows: 0, 1, 2, 5, 10, 20, 30, 50, 70, and 100%. All the mixtures were mixed geometrically.

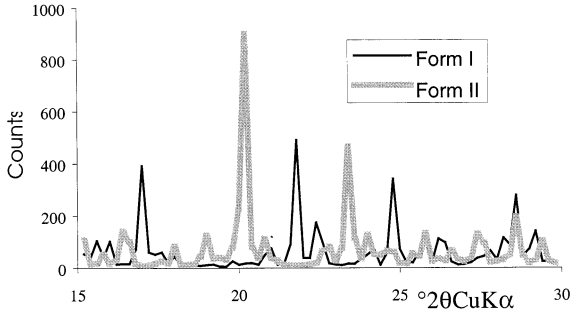

Fig. 1. X-ray diffractograms of two polymorphic forms of ranitidine hydrochloride.

Powder samples from the tablets were prepared by gently grinding each tablet into a fine powder using a glass mortar and pestle.

Powder samples (360 mg) were compacted as tablets (∅ 16 mm) under a mass of 2 t, and were fitted into the aluminum sample holder for X-ray scanning. Binary mixtures were made in triplicate and the tablet powders were prepared in ten replicates.

The XRD scans were performed on a Philips wide angle X-ray powder diffractometer with X-ray generator (PW 1130/00) and goniometer (PW1050, Philips, Almelo, The Netherlands). A copper target X-ray (wavelength 1.541A Cu Kα) tube was operated at a power of 40 kV and 30 mA. The divergence slit was set at 1° for the X-ray beam and at 0.1° for the receiving scintillation detector. The scans were carried out at a step size of 0.04° and counting time for 0.5 s per step within the ranges 7–48° (2Θ).

Binary mixtures with 0, 1, 2, 10, 30, 50, 70 and 100% of form 2 were used as a working data set for training and testing the ANN. Binary mixtures with 5 and 20% of form 2 and tablet samples were used as the external prediction set.

3.3. Pre-processing of the data

The powder X-ray diffractograms of form 1 and form 2 showed characteristic diffractions at 17.04° 2Θ, 21.9° 2Θ, 22.5° 2Θ, 26.3° 2Θ, and 20.02° 2Θ, 23.3° 2Θ, and 27.4° 2Θ, respectively, for identification (Fig. 1).

The use of intensities at all Θ degrees of the diffractograms as input data vectors was not manageable because of the large dimensionality of the input data space. In addition, many of these data were weakly correlated with structural properties. Reduction and transformation of the input data space to limited angles was necessary to enhance the ANN performance. Thus, the powder patterns were sampled between 16° and 30° (2Θ), the region containing the characteristic diffraction. This reduced data to 350 steps (2Θ). These steps were further processed to reduce the amount of data being fed to the ANN and to smooth the noise in the diffractograms.

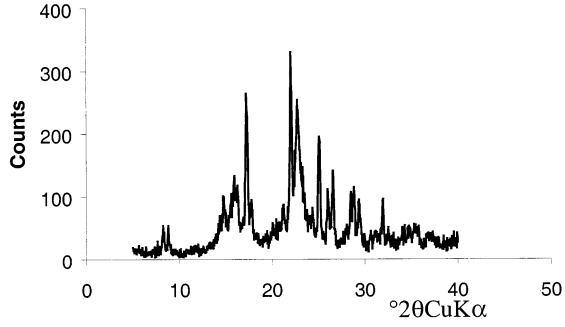

Fig. 2. X-ray diffractogram of the tablet mixture.

Firstly the 350 steps were reduced into 70 equal sized windows (dB = 5), and subsequently into 35 equal sized windows (dB = 10). Each window was computed as the average of intensities at five or 10 consecutive steps, respectively. These data were used as inputs together with the corresponding weight fraction of the forms as output to train the ANN.

For the tablet assay (Fig. 2) the angular range was reduced to the region between 26 to 29° (2Θ), that contained two significant peaks at 26.3° and at 27.4° (2Θ) for form 1 and form 2, respectively, since the tablet excipients showed significant interference between 16 and 26° (2Θ). Transformed data (dB = 5) and non-transformed data (dB = 1) from 26 to 29° (2Θ) were fed to the ANN.

3.4. Network topology

A standard feed-forward network, with a back-propagation rule and with a single hidden layer architecture was chosen. A single hidden layer was used for simplicity, and because there is little evidence to suggest that a larger number of hidden layers improves performance [8].

The ANN used in this investigation consisted of 70 (dB = 5) or 35 (dB = 10) inputs for the averaged XRD pattern values, one hidden layer, and two output neurons, one for the fraction of each form. The number of hidden neurons and number of training cycles were adjustable parameters and had to be optimized.

For the tablet assay the trained ANN consisted of 20 inputs (dB = 5) for the averaged pattern value between 26 and 30° (2Θ), one hidden layer and one output neuron for active ranitidine–HCl ingredient form 1.

During training and testing the number of hidden neurons was varied from 2 to 20 and training cycles from 0 to 50 and ANN performance was tested after each addition of a neuron.

3.5. Training

The ANN's training was accomplished by cycling through the entire assembly of training examples and correcting the weights using the standard back-propagation rule to minimise the sum squared output error. At the start of the training run both weights and biases were initialized with random values and the working data sets were split randomly into training and testing data sets. Data sets from six sample mixtures (75%) were used for training, while testing was done with the data sets from the two remaining sample mixtures (25%). The training set was used to train the network and the testing set was used to monitor overtraining of the network, each with a corresponding mean squared error (MSE). The training was stopped at each run when the error in the test set began to rise. The results of the four runs were averaged.

The error in mapping the training values decreased as the number of hidden neurons was increased. By increasing the number of hidden neurons the ANN more closely followed the topology of the training data sets. However, exceeding an optimum number of hidden neurons resulted in tracing the training pattern too closely and the system was overtrained. This behavior is analogues to overfitting a regression equation to a set of experimental data. If there are too few neurons the ANN will not perform well on either training or unseen data. If there are too many neurones the ANN will be overtrained and will exhibit poor prediction for unseen data.

Some books and papers offer 'rules of thumb' for choosing a number of hidden units. A common practice is to use the geometric or arithmetic mean of the number of input and output neurons for the hidden layer, or the natural logarithm of the number of input classes [9]. Other rules are concerned with the number of training sets avail-

able and suggest the number of weights is less than one tenth of the number of training sets. Such rules are concerned with over-fitting. If the number of training sets is much larger than the number of weights the ANN may suffer from under-fitting. The number of hidden units depends critically on the number of training sets, amount of noise in the data and complexity of the classification.

3.6. Method validation

The testing error is not a good estimate of the generalization error. One method for getting an unbiased estimate of the generalisation error is to present the ANN with a new, third set of data, that were not used at all during the training process. The relative error (ERR%) of prediction for binary mixtures with 5 and 20% of form 2 and tablet samples were used to compare the generalization ability of the models (Table 2).

4. Results and discussion

Unlike other techniques the ANN method employed pattern recognition on the entire lower resolution XRD patterns. Input data were transformed prior to analysis. The purpose of transformation was to compress the number of the input data, reduce the noise and to enhance the interesting features in the patterns.

4.1. Ranitidine–HCl binary mixture assay

For a sample composed of a combination of polymorphs, the diffractogram of the sample was approximately a linear superposition of the diffractograms of each individual polymorph. Each polymorph presented to the diffractometer produces a powder pattern that is characteristic of that crystalline structure. By presenting binary mixtures of two different polymorphs to the system a database of the powder patterns is constructed.

The network was trained twice: once using 70 inputs and again using 35 inputs. The criteria for judging the best model were mean squared error (MSE) of model prediction and coefficient of multiple determination (R^2) (Table 1).

The test results showed that the sufficient number of hidden neurons was from 10 to 15. The MSE was < 0.0002 and R^2 was greater that 0.99 (dB = 5). Ten hidden neurons were enough to achieve good convergence on the training data which is confirmed by the low relative error (ERR%) for the external validation data set (Table 2)

The results for the network trained with more inputs were improved for the training data and, more significantly, for the test data. Better results were obtained with 70 input data (dB = 5) than with 35 (dB = 10) (Table 1). The trained ANN quantified two polymorphs in the mixtures. These results indicate that an effective mathematical mapping is possible when noise is present in the input pattern.

Table 1
Effect of data transformation on ANN performance for the assay of ranitidine–HCl from binary mixtures

Number of hidden units	dB = 10		dB = 5	
	MSE[a] training/testing	R^2 training/testing	MSE training/testing	R^2 training/testing
4	0.009/0.005	0.85/0.95	0.0009/0.0007	0.97/0.98
6	0.009/0.003	0.90/0.96	0.0004/0.0002	0.99/0.99
8	0.006/0.003	0.91/0.97	0.0003/0.0002	0.99/0.99
10	0.006/0.002	0.91/0.96	0.0002/0.0001	1.00/0.99
12	0.006/0.003	0.91/0.98	0.0001/0.0002	1.00/0.98
14	0.006/0.005	0.95/0.98	0.0001/0.0001	1.00/0.99

[a] Individual errors are squared, summed and divided by the number of individual errors.

Table 2
Effect of the number of hidden neurons on the generalisability of the predictions

Measured percentage of form 2	Predictions with different numbers of hidden neurons at 50 training cycles[a]			
	6	8	10	12
Mean	5.5	5.4	4.9	5.5
S.D.	1.1	0.7	0.1	1.1
ERR(%)[b]	17.0	8.6	2.9	14.6
Mean	20.4	19.4	20.0	20.0
S.D.	0.4	0.5	0.2	0.1
ERR(%)	27.0	23.5	17.8	13.9
AV ERR (%)[c]	22.02	16.08	10.35	14.23

[a] $n = 6$.
[b] ERR(%) = (predicted − actual)/actual × 100.
[c] AV ERR(%) = ERR(%) averaged for 5 and 20% data.

4.2. Tablet assay

For the tablet assay the X-ray experiments revealed many interfering peaks (Fig. 2). The tablet excipients interference with ranithidine–HCl form 1's characteristic diffractions was mainly between 16 and 25° θ. Therefore, the ANN was trained with 100 (dB = 1) and 20 (dB = 5) input data from 26 to 30° θ and with one output neuron for the ranitidine–HCl form 1 concentration. It was a difficult task to distinguish the signal from the noise for the narrower pattern range used for training and testing the ANN, since there are more specific powder pattern features between 16 to 23° θ for form 1 (Fig. 1).

Since tablet excipients interfered with ranitidine–HCl form 1's characteristic diffractions, resolution of XRD patterns has a great influence on the network performance. Better results were obtained with higher resolution XRD patterns (Table 3). The presence of small peaks (low intensity X-ray lines) increased the success of predictions over low resolution diffractograms with less details, where transformations smoothed the data and eliminated small peaks.

The trained ANN quantified ranitidine–HCl form 1 from tablets. The test results showed that the sufficient number of hidden neurons was from 6 to 12 to achieve good convergence on the training data (Table 3a and b).

5. Conclusions

There are many different methods available for multivariate statistical analysis, function fitting or prediction and ANNs represents a subset of these. The conventional multiple linear regression (MLR) approach involves an iterative process of spectrum decomposition and regeneration until a mathematically synthesized spectrum closely matching the true spectrum is generated. This is time consuming and often requires manual intervention. Most of these techniques depend on the subjective decision regarding peak shapes and peak overlapping that will prune the number of possibilities significantly. The drawback of the XRD powder method is that the data grossly overlap. While the determination from a well-resolved powder diffraction diagram is quite feasible, the presence of overlapping reflections generally prevents the full use of the available information. A solution is to include as observed data in the least-squares refinement the integrated intensities (areas under overlapping peaks) of the composite diffraction peaks (the Rietveld method [10]). The Rietveld method creates an effective separation of these overlapping data. A major disadvantage of a procedure is that details in the profiles of these peaks are lost and such a system may not perform well on the new set of data, e.g. will not generalize well. Besides, in contrast to the present work, no cross-validation study with the

testing data set is performed during conventional MLR.

From a statistical modeling point of view, ANN models belong to the general class of non-parametric methods. In this sense they are more powerful than parametric methods that try to fit experimental data into a specific parametric form model. However, non-parametric methods like ANN contain more free parameters and hence require more training data than parametric ones to achieve good generalization performance.

The ANN approach described employs pattern recognition on the entire XRD pattern in a specified range. The whole diffractograms were used in the identification and quantification process instead of the individual peaks. Correct quantification was mainly influenced by the overall appearance of the XRD patterns, especially the small peaks. It was shown that data transformations improved the quantitative performance when the pattern was not contaminated by other components. Only integrated or low resolution

Table 3
Predictions of the ranitidine–HCl content in the tablets using different numbers of neurones in the hidden layer[a]

Trial	Hidden neurones						
	0	2	4	6	8	10	12
(a) dB = 5[b,c]							
1	52.99	66.20	70.28	61.15	44.57	61.34	71.93
2	−3.91	41.44	55.92	50.46	56.22	72.99	45.65
3	38.63	52.91	60.85	54.95	57.48	70.24	59.41
4	49.28	48.05	62.68	54.64	53.16	61.19	67.19
5	24.54	53.82	63.10	55.81	50.39	67.17	58.79
6	29.24	53.52	62.35	55.53	52.75	68.19	58.99
7	51.14	57.13	66.48	57.90	48.87	61.27	69.56
8	45.82	59.56	65.56	58.05	51.02	65.79	65.67
9	22.69	44.75	59.30	52.56	54.69	67.09	56.42
10	34.25	52.15	62.43	55.30	52.46	66.44	61.04
Average	34.47	52.95	62.90	55.64	52.16	66.17	61.46
S.D.	16.50	6.80	3.75	2.82	3.54	3.76	7.20
Recovery (%)	61.12	93.89	111.52	98.65	92.48	117.32	108.97
(b) dB = 1[d,e]							
1	5.68	79.35	68.98	65.23	50.11	56.69	55.25
2	45.35	80.01	75.89	73.24	63.41	66.71	71.28
3	−6.69	75.19	55.23	56.13	40.81	53.44	48.79
4	−4.99	89.66	58.73	66.22	48.43	53.24	52.20
5	5.72	79.39	59.08	65.62	50.01	56.58	54.82
6	19.33	82.73	67.06	66.03	54.11	61.14	59.53
7	4.27	72.59	55.27	65.12	48.74	55.96	53.77
8	9.29	75.97	59.20	65.42	50.53	57.69	55.69
9	11.22	85.04	64.28	66.09	52.22	58.51	57.09
10	3.34	81.67	59.00	65.75	49.66	55.83	54.24
Average	9.25	80.16	62.27	65.48	50.90	57.58	56.26
S.D.	13.97	4.73	6.32	3.86	5.33	3.75	5.68
Recovery (%)	16.52	143.14	111.19	116.93	90.72	102.82	100.47

[a] Theoretical ranitidine–HCl content is 56.4%.
[b] $R^2_{train} = 0.977 - 0.984$; $R^2_{test} = 0.899 - 0.995$.
[c] $MSE_{train} = 0.0002 - 0.0006$; $MSE_{test} = 0.001$–0.002.
[d] $R^2_{train} = 0.998 - 0.999$; $R^2_{test} = 0.899 - 0.995$.
[e] $MSE_{train} = 0.0000$; $MSE_{test} = 0.000 - 0.001$.

X-ray diffractograms were required to distinguish between two crystalline forms in a mixture. In the case of ranitidine–HCl quantification from tablets, where significant interference with tablet excipients was present, better results were obtained without data transformation, as transformations reduced the contribution of small peaks. The ranitidine–HCl from tablets was analysed in the presence of the excipients directly with minimal sample pretreatment.

The ANN extension presented is believed to increase the value of the XRD based quantification of ranitidine–HCl polymorphs to the point that it could be the method of choice in some advanced research settings. One advantage of this approach is that most of the intense computation takes place during the training process. Once the network is trained, the XRD patterns can be rapidly analysed. The ANN analysis enables quantification of polymorphs in low fraction, directly from a XRD pattern of a sample in less than 1 h. Furthermore, the quantification performance of the ANN analysis, demonstrated in this study, serves as an indication of the broad potential of neural networks in pharmaceutical research. This research shows that ANNs are beneficial in analysis of diffractional data. While the system described has been developed to interpret XRD patterns, peak detection has implications in other chemical applications where the recognition of peak-shaped signals in analytical data is important. Further work should explore the use of alternative network architectures.

References

[1] J.D. Hanawalt, H.W. Rinn, L.K. Frevel, Ind. Eng. Chem. Anal. Ed. 10 (1938) 457–512.
[2] J.W. Shell, J. Pharm. Sci. 52 (1963) 24–29.
[3] A.W. Hull, J. Am. Chem. Soc. 41 (1919) 1168–1175.
[4] B.J. Wythoff, S.P. Levine, S.A. Tomellini, Anal. Chem. 62 (1990) 2702–2709.
[5] C. Schierle, M. Otto, Fresenius J. Anal. Chem. 344 (1992) 190–194.
[6] H. Lohninger, F. Stancl, Fresenius J. Anal. Chem. 344 (1992) 186–189.
[7] A. Forster, K. Gordon, D. Schmierer, N. Soper, V. Wu, T. Rades, The Internet Journal of Vibrational Spectroscopy, vol. 2, second ed., Perkin-Elmer, Beaconsfield, 1998.
[8] B.D. Ripley, Pattern Recognition and Neural Networks, Cambridge University Press, Cambridge, 1996.
[9] J.M. Zurada, Introduction To Artificial neural System, PWS Publishing, Boston, 1992.
[10] R.A. Young, Ab initio structure solution with powder diffraction data, in: A.K. Cheetham (Ed.), The Rietveld Method, Oxford University Press, London, 1993, pp. 276–292.