Talanta 82 (2010) 850–853

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Short communication

Recognition of active ingredients in tablets by chemometric processing of X-ray diffractometric data

Łukasz Komsta[a,*], Jan K. Maurin[b,c]

[a] Department of Medicinal Chemistry, Medical University of Lublin, Jaczewskiego 4, 20-090 Lublin, Poland
[b] National Medicines Institute, Chełmska 30/34, 00-725 Warsaw, Poland
[c] Institute of Atomic Energy Polatom, 05-400 Otwock-Świerk, Poland

ARTICLE INFO

Article history:
Received 11 January 2010
Received in revised form 26 April 2010
Accepted 10 May 2010
Available online 19 May 2010

Keywords:
Acetaminophen
Pharmaceuticals
X-ray powder diffractometry (XRPD)
Chemometrics
Partial least squares

ABSTRACT

The paper presents an approach to use Partial Least Squares Discriminant Analysis (PLS-DA) on X-ray powder diffractometry (XRPD) dataset to build a model which recognizes a presence (or absence) of particular drug substance (acetaminophen) in unknown mixture (OTC tablet). The dataset consisted of 33 XRPD signals, measured for 12 pure substances and 21 tablets containing them in different quantitative and qualitative ratios, along with unknown excipients. The model was built with an external validation dataset chosen by Kennard–Stone algorithm. The RMSECV value was equal to 0.3461 (87.8% of explained variance) and external predictive error (RMSEP) was equal to 0.3123 (86.2% of explained variance). The result suggests that small but properly prepared training datasets give ability to construct well-working discriminant models on XRPD signals.

1. Introduction

The X-ray powder diffractometry (XRPD) is a form of analysis of solid samples, applied in different areas of chemical analysis. It's use in pharmacy can be traced back to 1958 [1]. In 1970s it was successfully used in forensic science, where drugs, excipients, and adulterants in illicit samples had to be identified [2]. As the method can be non-invasive [3], it can be applied even to blistered tablets and analyzed sample is not destroyed. The literature databases present dozens of applications related to pharmacy. For interested reader, a good starting point could be reviews by Brittain [4–7] or Harris [8].

Despite of wide range of XRPD applications, the connection with chemometric processing (such as PCA, PCR or PLS) of obtained data is a new concept and only several papers appeared recently. Although some pattern fitting approaches based on the non-linear regression were reported earlier by Yamamura et al. [9,10], the factor analysis is a field of recent years. Moore et al. [11] performed a study on chemometric algorithms (CLS, ILS, PCR, PLS) in X-ray diffractometry of intact multicomponent consolidated samples. Chieng et al. [12] analyzed three solid forms of ranitidine hydrochloride by XRPD and compared obtained data with Raman spectroscopy. The data were processed by PCA, PCR and PLS. A polymorphs of similar drug famotidine [13] were chemometrically quantified using PLS method, both from Raman spectroscopy and XRPD data.

The above recent work inspired us to investigate a possibility to construct discriminant models, detecting presence (or absence) of a particular ingredient in a multicomponent tablet. To our best knowledge, it has never been done before. The presence of acetaminophen in an unknown over the counter (OTC) tablet was studied.

2. Experimental

The pure substances: acetaminophen (ACE), aspirin (ASP), caffeine (CAF), codeine phosphate (COD), dipyrone (DIP), ethoxybenzamide (ETO), ibuprofen (IBU), phenylephrine hydrochloride (PHE), propyphenazone (PRO), pseudoephedrine hydrochloride (PSE), thiocol (THI) and vitamin C (VIT) were of appropriate purity (Sigma–Aldrich, USA).

21 Available OTC tablets with different qualitative and quantitative composition were bought in local drugstore:

1. Antidol 15—tablets (500 mg acetaminophen, 15 mg codeine phosphate), produced by Lek.
2. Apap C plus—effervescent tablets (500 mg acetaminophen, 300 mg ascorbic acid), produced by US Pharmacia.

* Corresponding author. Tel.: +48 81 7423692; fax: +48 81 7423691.
E-mail address: lukasz.komsta@umlub.pl (Ł. Komsta).

0039-9140/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2010.05.012

3. Aspirin C—effervescent tablets (400 mg aspirin, 240 mg ascorbic acid), produced by Bayer.
4. Cefalgin—tablets (250 mg acetaminophen, 150 mg prophyphenazone, 50 mg caffeine), produced by Polfa Pabianice.
5. Coffepirine—tablets (450 mg aspirin, 50 mg caffeine), produced by Marcmed Lublin.
6. Coldrex HotRem—sachets (750 mg acetaminophen, 10 mg phenylephrine hydrochloride, 60 mg ascorbic acid), produced by GlaxoSmithKline.
7. Coldrex MaxGrip—sachets (1000 mg acetaminophen, 10 mg phenylephrine hydrochloride, 40 mg ascorbic acid), produced by GlaxoSmithKline.
8. Dafalgan Codeine—tablets (500 mg acetaminophen, 30 mg codeine phosphate), produced by UPSA.
9. Efferalgan—effervescent tablets (330 mg acetaminophen, 200 mg ascorbic acid), produced by UPSA.
10. Etopiryna—tablets (300 mg aspirin, 100 mg ethoxybenzamide, 50 mg caffeine), produced by Polpharma.
11. Gardan P—tablets (200 mg prophyphenazone, 300 mg dipyrone), produced by Polfa Pabianice.
12. Ibuprom—tablets (200 mg ibuprofen), produced by US Pharmacia.
13. Modafen—tablets (200 mg ibuprofen, 30 mg pseudoephedrine hydrochloride), produced by ZENTIVA
14. Nurofen Plus—tablets (200 mg ibuprofen, 12.8 mg codeine phosphate), produced by Boots Healthcare.
15. Panadol Extra—tablets (500 mg acetaminophen, 65 mg caffeine), produced by GlaxoSmithKline.
16. Saridon—tablets (250 mg acetaminophen, 150 mg prophyphenazone, 50 mg caffeine), produced by Roche.
17. Solpadeine—tablets (500 mg acetaminophen, 30 mg caffeine, 8 mg codeine phosphate), produced by SmithKline Beecham.
18. Solpadeine—effervescent tablets (500 mg acetaminophen, 30 mg caffeine, 8 mg codeine phosphate), produced by GlaxoSmithKline.
19. Solpadeine—capsules (500 mg acetaminophen, 30 mg caffeine, 8 mg codeine phosphate), produced by SmithKline Beecham.
20. Talvosilen forte—capsules (500 mg acetaminophen, 30 mg codeine phosphate) produced by Bene-Arzneimittel GmbH.
21. Thiocodin—tablets (15 mg codeine phosphate, 300 mg thiocol) produced by UNIA.

A Bruker AXS D8 Advance powder diffractometer was used for all experiments. Special configuration with a parallel beam method employing a multilayer (Goebel) mirror was used. A gaseous, position sensitive Våntec detector was applied. A θ–θ scan mode was used to collect diffraction data. Reflections were registered for $3^\circ < 2\theta < 60^\circ$ with a step size of 0.007° and a scan rate $0.023^\circ \times s^{-1}$. The Cu Kα radiation was employed using a standard copper tube with a high voltage of 40 kV and an anode current of 40 mA.

The formulations were analyzed after removal of the coating, following by grounding in agate mortar into fine powder. The diffractograms of pure substances were collected together with results of tablets into one dataset. Partial Least Squares Discriminant Analysis (PLS-DA) of resulted matrix (dimensions: 33×8044) was performed with Matlab R2009b, using TOMCAT toolbox for multivariate calibration [14]. Principal Component Analysis (PCA) and cluster analysis were performed with GNU R 2.9.0 using built-in functions.

3. Results and discussion

The analyzed dataset was a result of augmenting pure substances and tablets containing their mixtures with unknown excipients in different qualitative and quantitative combinations.

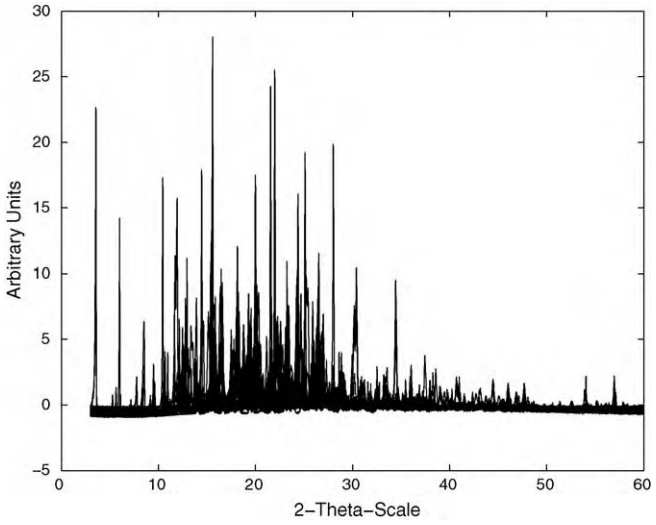

Fig. 1. The dataset after standard normal variate (SNV) preprocessing.

The purpose of such approach was to collect as diverse training set as possible. In tablets, acetaminophen is present in 13 of them, caffeine in 8, codeine phosphate in 8, aspirin in 4, vitamin C in 4, propyphenazone in 3, ibuprofen in 3, ethoxybenzamide in 2, phenylephrine in 2, dipyrone in 1, thiocol in 1. The dataset after standard normal variate (SNV) preprocessing is shown in Fig. 1.

Due to complex nature of XRPD signals, no visual inspection, nor simple correlation approach could be done to recognize acetaminophen presence among analyzed samples. For example the signal of effervescent tablets with acetaminophen and vitamin C has correlation (r) with pure acetaminophen signal only around 0.2. On the contrary, tablets containing ethenzamide, aspirin and vitamin C can correlate as high as 0.43 with acetaminophen. Therefore more advanced chemometrics is needed in automated recognition of such samples.

Preliminary exploration of dataset was done by the unscaled principal component analysis (PCA). The first PC explains only 18.8% of total variance, the second one explains only 13.2% of the additional variance. This confirms very complex nature of XRPD signal.

It can be concluded from the PC1 vs PC2 plot (Fig. 2), that signals form three visible clusters. First cluster, denoted as circles, contains non-effervescent tablets with acetaminophen. Second, denoted as triangles, contains all effervescent tablets. The last one

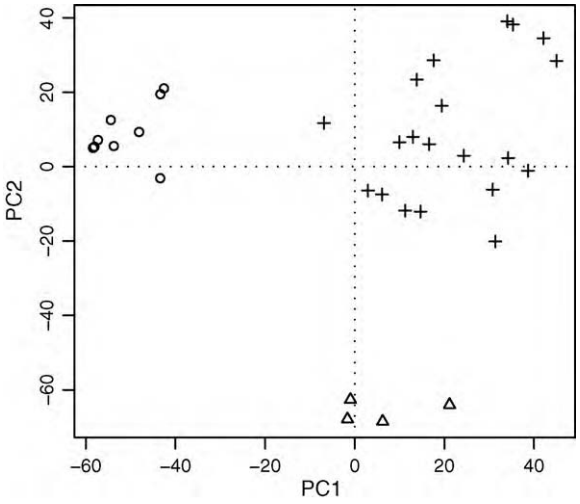

Fig. 2. PCA results on XRPD signals. See text for descriptions of symbols.

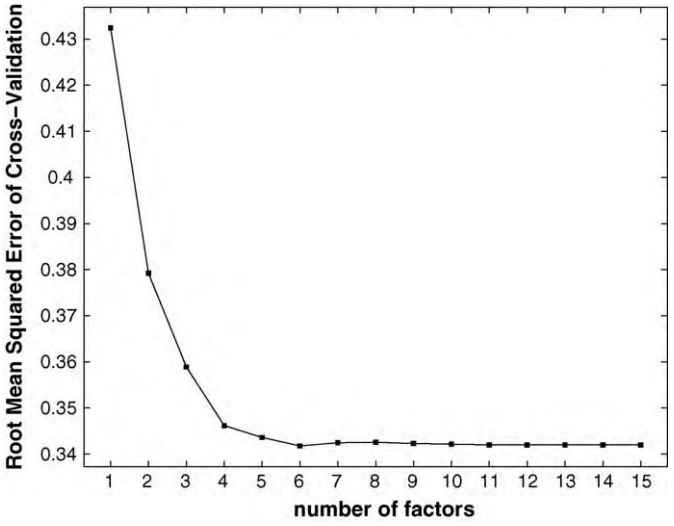

Fig. 3. The cross-validation results of PLS-DA model building.

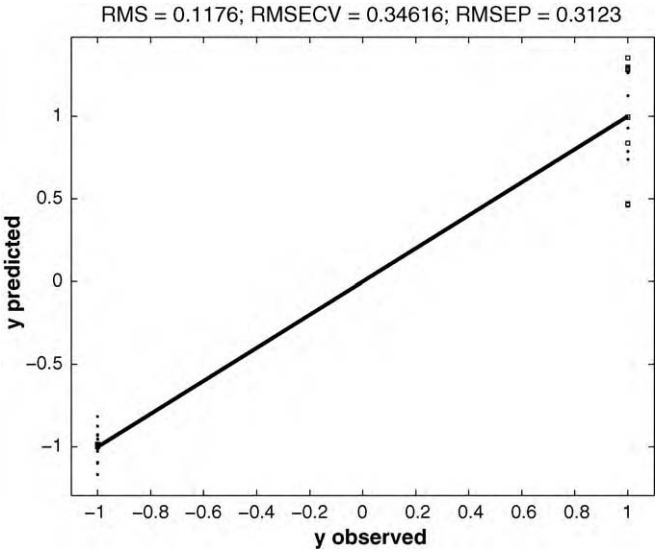

Fig. 5. Model validation results. LOO cross-validation predictions are marked as dots, external validation values are marked as squares.

contains other pharmaceutical formulations (non-effervescent and without acetaminophen) with one exception—Coldrex Maxgrip effervescent powder, containing acetaminophen. The dissimilarity of effervescent formulations are caused by presence of sodium carbonate, which gives additional peaks in XRPD signal.

This implies, that presence (or absence) of acetaminophen, as a main factor of variation in studied tablets, is visible in PC1 and PC2. Therefore, it can be expected, that there is a possibility to construct a discriminant model detecting its presence in pharmaceutical formulation.

Although models can be built against all substances, the dataset used for building PLS-DA should met some requirements and the main of them is a proper proportion between a number of "positive" and "negative" samples [15]. In the case of presented tablets, only acetaminophen divides the samples into enough equal classes (14 positive, 18 negative). Building models against presence of other compounds was done by us and the cross-validation results seemed to be acceptable, but their predictive ability (against external testing set) was very poor, so the results are not shown.

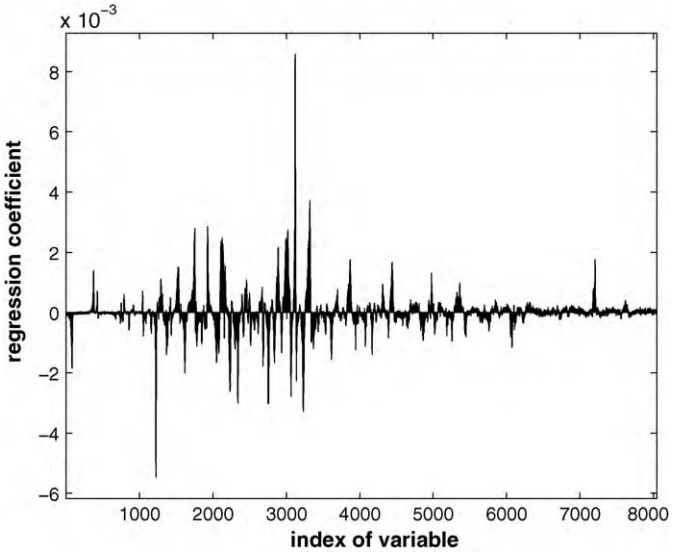

Fig. 4. The discriminant estimator recognizing presence of acetaminophen in investigated dataset.

The acetaminophen model was built in following way. First, 10 of 32 signals were selected by Kennard–Stone algorithm as a representative external validation dataset and removed from further processing. The remaining 22 curves were used as a training PLS-DA dataset.

The cross-validation result is presented in Fig. 3. The complexity of 4 was chosen as optimal and its estimator is presented in Fig. 4. The estimator is quite complex and weakly correlated with pure acetaminophen signal (around 0.5, regardless of chosen model complexity), as it contains only features responsible for detection of acetaminophen in presence of other, superimposing signals of another compounds.

The model has RMS equal to 0.1176 (98.6% of the explained variance). The RMSECV value is equal to 0.3461 (87.8% of the explained variance) and external predictive error (RMSEP) is equal to 0.3123 (86.2% of the explained variance). The predicted values, shown in Fig. 5, conclude that there is no misclassified sample, both in the cross-validation results and the external prediction test.

As the original matrix is quite wide, we have performed a variable selection approach by Uninformative Variable Selection by Partial Least Squares (UVE-PLS) [16]. It has chosen around 1200 variables from original 8044 ones, but such model did not perform better—the performance was comparable. We have also performed some experiments on lower 2θ range, as this range contains more unique peaks. Again no improvement was noticed. Therefore full PLS-DA model seems to be best choice as simpliest approach possible.

4. Conclusion

The proposed methodology can be successfully applied as a preliminary method in forensic science, where quick tablet identification is often required. There is an easy way to construct PLS-DA chemometric models and apply them to unknown sample for automatical detection of particular ingredient even with small training datasets, but with good proportions between positive and negative ones and enough diversity. This suggest a new interesting, but neglected area for future research—chemometric processing of XRPD signals.

References

[1] L. Molle, Pharmaceutisch Weekblad 93 (1958) 334–341.
[2] V.A. Folen, J. Forensic Sci. 20 (1975) 348–372.
[3] N.V. Phadnis, R.K. Cavatur, R. Suryanarayanan, J. Pharm. Biomed. Anal. 15 (1997) 929–943.
[4] H.G. Brittain, Spectroscopy 16 (2001) 14–18.
[5] H.G. Brittain, Pharm. Technol. North Am. 25 (2001) 142–150.
[6] H.G. Brittain, Am. Pharm. Rev. 5 (2002) 74–80.
[7] H.G. Brittain, Profiles Drug Subst. Excipients Relat. Methodol. 30 (2003) 271–319.
[8] K.D.M. Harris, Am. Pharm. Rev. 7 (2004) 86–91.
[9] S. Yamamura, Y. Momose, Int. J. Pharm. 212 (2001) 203–212.
[10] S. Yamamura, R. Takahira, Y. Momose, Pharm. Res. 24 (2007) 880–887.
[11] M.D. Moore, R.P. Cogdill, P.L.D. Wildfong, J. Pharm. Biomed. Anal. 49 (2009) 619–626.
[12] N. Chieng, S. Rehder, D. Saville, T. Rades, J. Aaltonen, J. Pharm. Biomed. Anal. 49 (2009) 18–25.
[13] Z. Német, G. Kis, G. Pokol, A. Demeter, J. Pharm. Biomed. Anal. 49 (2009) 338–346.
[14] M. Daszykowski, S. Serneels, K. Kaczmarek, P. Van Espen, C. Croux, B. Walczak, Chemom. Intell. Lab. Syst. 85 (2007) 269–277.
[15] R.G. Brereton, Trends Anal. Chem. 25 (2006) 1103–1111.
[16] V. Centner, D.L. Massart, O.E. de Noord, S. de Jong, B.M. Vandeginste, C. Sterna, Anal. Chem. 68 (1996) 3851–3858.