A PCA-based chemometrics-assisted ATR-FTIR approach for the classification of polymorphs of cimetidine: Application to physical mixtures and tablets

Natalia L. Calvo, Teodoro S. Kaufman*, Rubén M. Maggio*

Department of Organic Chemistry, School of Pharmaceutical and Biochemical Sciences, National University of Rosario and Institute of Chemistry of Rosario (QUIR, CONICET-UNR), Suipacha 531, Rosario 2000, Argentina

** Abstract **

The identity of the polymorphic form of an active pharmaceutical ingredient is an important parameter that may affect the performance of the drug formulation. This calls for special techniques, able to classify crystal forms or assign the polymorphic identity to a given solid in a mixture.

In order to develop a method to determine which of the relevant polymorphs of Cimetidine (CIM) is present in commercial tablet samples, authentic forms A, B, D and M1 of the drug were prepared, structurally characterized and employed as standards.

Thus, attenuated total reflection Fourier transform infrared spectroscopy (ATR-FTIR) was coupled to Principal Component Analysis (PCA) and used for the classification of physical mixtures of CIM and excipients, as well as laboratory-made and commercial tablets, according to their polymorphic composition.

It was demonstrated that two principal components (PCs) suffice to classify the samples of the four forms of CIM into distinct groups, and that method performance is optimum when the second and third PCs are used for the classification process. The application of the method to commercial tablets of CIM also gave good results, confirming they were prepared employing the correct polymorph (form A).

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

In materials science, polymorphism describes the ability of a compound to crystallize in more than one form, because of different molecular arrangements in the crystal lattice [1]. Solvates are considered pseudo-polymorphs, and solids may also have an amorphous phase. Since crystallization is extensively used as a purification step in the chemicals industry, including pharmaceuticals, the polymorphic form is a property of great importance.

The investigation of crystal polymorphism of active pharmaceutical ingredients (APIs) is of increasing importance in pharmaceutical industry. This comes as a result that many APIs exhibit polymorphism, and this phenomenon represents a relevant hurdle within the pharmaceutical industry because of its multiple connotations, at manufacture, therapeutic, commercial, regulatory, and even legal levels [2].

Different polymorphs of a drug can exhibit dissimilar physical and chemical properties, many of which may translate into differences in biological activity. The melting point, solubility, density and dissolution rate of each polymorph may be different, as a consequence of variations in crystal habit, intermolecular interactions and lattice energy. The chemical reactivity and stability can also be affected [3–6].

These variations may affect particle size and shape, compressibility and ability to flow of the powder, and have influence on the processability of the solid. In turn, these may impact on the reproducibility of the manufacturing process and on product performance, especially drug dissolution, hence affecting its absorption and bioavailability [7–10].

Drug products are often prepared employing the most stable polymorph; therefore, polymorphic interconversion of the active pharmaceutical ingredient to its most stable form may take place under various circumstances, at the manufacturing, packaging, distribution and storage stages [11–13]. Hence, there is a demand for fast and reliable analytical methods for the evaluation of polymorphs in mixtures and drug products.

Cimetidine (CIM) is N'-cyano-N-methyl-N'[2-[[5-methyl-1H-imidazol-4-yl]methyl][thio]-ethyl]-guanidine (Fig. 1A). This
histamine analog, recognized as the first blockbuster drug, is a specific antagonist of the type 2 histamine receptors (H2R) located on the surface of stomach parietal cells [14]. CIM inhibits the histamine-, gastrin- and pentagastrin-stimulated acid secretion, accelerating the healing of ulcers of the upper gastrointestinal tract [15,16]. Hence, despite the availability of other alternatives, CIM is considered by the World Health Organization as an essential drug, and still widely used as an anti-ulcer agent. However, in recent years it was demonstrated that CIM has additional pharmacologically useful properties in the areas of modulation of the immune response [17] and odontology [18], as well as in skin [19] and cancer [20–22] therapeutics.

Cimetidine exhibits rich crystal polymorphism. Forms A, B, C and D, as well as monohydrate M1 are recognized and reproducible polymorphic forms of the drug, which can be prepared under well-defined conditions [23]. In addition, other polymorphs and hydrates have been proposed, but their access could not be reproduced and an amorphous form (prepared by cooling or freezing melted CIM) has also been reported [24].

The different polymorphs of CIM and their mixtures have been studied and characterized by X-ray crystallography [25–28], synchrotron X-ray powder diffraction [29] and vibrational (IR, Raman, NIR) spectroscopy [30–32]. The literature also contains other approaches, including calorimetry [24,33], atomic force microscopy [34] and solid-state NMR spectrometry [35,36].

The forms of CIM have different indications. Polymorph A is used for tablets, while form B is preferred for suspensions [37]. Bauer-Brandl has shown that some CIM polymorphs can undergo transformations in the dry state upon milling [38]. On the other hand, we have recently reported a study on the dehydration of form M1 to polymorph B [39].

A thorough understanding and a systematic characterization of the polymorphic forms expressed by active pharmaceutical ingredients is always necessary in order to maintain high manufacturing quality and reproducibility. Furthermore, regulatory agencies currently require the control of the crystal forms in formulated products, emphasizing on the need of suitable analytical methods for this purpose [40]. This increases the interest in the development of such analytical methods to detect, classify, and quantitate polymorphic forms in complex mixtures.

Together with NIR and Raman, infrared spectroscopy appears as one of the established and most efficient techniques for the detection and quantification of polymorphism [41]. However, infrared spectroscopy is more robust and informative, and less expensive than NIR. Additionally, infrared spectroscopy is less prone than Raman spectroscopy to yield unreliable results, such as those resulting from sample fluorescence, inhomogeneities, crush size and dispersion effects.

Polymorphism results in subtle spectral differences; however, these polymorphic markers do not always allow direct and unambiguous discrimination among polymorphic forms, and the matrix may hamper the recording of the suitably discriminating spectra in case of drug products.

Chemometric techniques are powerful tools for compressing multidimensional data into a few variables and for uncovering hidden trends. Although examples are still relatively few and scattered, chemometric methods are gaining acceptance as means for studying drug polymorphism [42,43] and we have used the chemometrics analysis of drug dissolution data as a tool to assign the polymorphic identity of furosemide forms in capsules [44]. In this association, infrared spectroscopy provides molecular information, whereas chemometric analysis enables extraction of relevant from the spectral data.

Therefore, here we report the use of attenuated total reflection Fourier transform infrared spectroscopy (ATR-FTIR) data coupled to Principal Components Analysis (PCA), as a facile and reliable approach to distinguish among the four relevant polymorphic forms of CIM in their pure samples, mixtures with excipients and a pharmaceutical dosage form.

2. Experimental

2.1. Instrumentation

The ATR-FTIR experiments were carried out with a diamond-based ATR accessory (GladiATR, Pike Technologies, Madison, USA), fitted with a Pike temperature control unit. The FTIR spectra (20 scans each) were acquired with a Shimadzu Prestige 21 spectrometer (Shimadzu Corp., Kyoto, Japan) in the absorbance mode, over the 600–4000 cm–1 wavenumber range, at a resolution of 4 cm–1. The laboratory-made tablets were prepared employing a Sanchez TS-1 tableting press, employing mixtures of powders previously sieved with an RR1920 Zonitest Vibration system fitted with ASTM certified stainless steel sieves. The experiments were carried out with the fraction containing 50–100 mesh particles.

2.2. Chemicals

Cimetidine polymorph A (BP 2002) was employed as source of the bulk drug. The excipients used were corn starch, microcrystalline cellulose, povidone K30, lactose, sodium croscarmellose and magnesium stearate. All of them were of pharmaceutical grade and acquired from Saporiti (Buenos Aires, Argentina).

The polymorphs employed in this work were prepared according to recently reported procedures [39], using distilled water and analytical grade reagents. The phase purity and identity of the forms were confirmed by optical, thermal and spectroscopic means.
During the experiments, the drugs and the drug mixtures were kept in a desiccator, protected from light.

2.3. Preparation of the samples

2.3.1. Procedure

The polymorphs of CIM and the excipients were individually sieved using a vibrating platform (120 rpm). The corresponding 50–100 mesh fractions were collected and employed in the studies. Accurately weighed amounts of the different polymorphs of CIM, with or without the addition of excipients, were transferred to a glass container and placed in a mechanical Z-mixer. The mixtures were processed for 15 min at 30 rpm.

2.3.2. Training samples

A set of 12 independent samples was prepared as described in Section 2.3.1. Four samples contained the pure CIM polymorphs (A, B, D and M1) and eight samples contained physical mixtures of excipients and the different polymorphs in the proportions corresponding to the commercial tablets (Section 2.3.5).

2.3.3. Validation samples

Four lots of mixtures containing the pure forms A, B, D and M1 and excipients (in the proportion of the commercial samples, Section 2.3.5) were prepared as described in Section 2.3.1, and individually subjected to tableting by direct compression. A total of 32 samples were analyzed. Some tablets were measured directly, while others were gently ground in a mortar before the measurements.

2.3.4. Model suitability samples

Six series of tablet samples, containing a physical mixture (1:1) of the different polymorphs with excipients (in the proportion of the commercial samples, Section 2.3.5), were prepared as described in Section 2.3.1.

2.3.5. Commercial samples

Cimetidine drug products (declared weight 305 mg) contained 200 mg CIM and excipients (corn starch (50 mg), microcrystalline cellulose (22 mg), povidone K30 (15 mg), lactose (10 mg), sodium croscarmellose (6 mg) and magnesium stearate (2 mg)). A commercial lot of tablets was evaluated following the procedure described in Section 2.4.

2.4. Procedure for data acquisition and data pre-treatment

Without any additional treatment, a pre-established amount of sample (20 mg), containing the sieved and mixed powders, was placed on the ATR accessory and pressed against the ATR crystal under standardized conditions [45]. The powder was wiped out and the equipment was carefully cleaned with a mixture of water:isopropanol (30:70) after each sample was analyzed. Background spectra were obtained against air, employing the clean and dry ATR accessory. All the spectra were acquired under the same measurement conditions, with the ATR plate thermostatized at 30 °C.

Spectra were arranged matrix-wise and polished to remove the non-informative regions. In this way, system signals (1794–1736 and 2382–2295 cm⁻¹), as well as those caused by residual amounts of the cleaning solvent (3038–2932 cm⁻¹) were deleted. In addition, spectra were zeroed using the 4000–3684 cm⁻¹ region, where the samples do not absorb, and mean-centered prior to the calculations. Each edited spectrum contained 1637 data-points.

2.5. Chemometrics and graphics software

The computer routines involving spectral data manipulation and the PCA algorithms were run in Matlab R2010a (Mathworks, Natick, USA). The Matlab scripts are freely available from the authors. Statistical data analyses were performed with Origin 8.5 (OriginLab Co., Northampton, USA).

3. Results and discussion

3.1. Preparation and characterization of the polymorphs

Polymorphic forms A, B, D and monohydrate M1 (the designation of the polymorphs is according to Hegedüs and Görög [23]) were obtained as previously reported [39] and unequivocally characterized by digital microscopy, ATR-FTIR spectroscopy, differential scanning calorimetry and solid state CP–MAS 13C NMR spectrometry. The results were in agreement with the literature [23,24,26,38]. Form C was excluded from this study for practical reasons, since it has been demonstrated that this polymorph is highly unstable, affording form B upon simple mechanical treatment [38]. The raw FTIR spectra of the pure forms of CIM employed are shown in Fig. 1B.

3.2. PCA-based sample classification: method development

PCA is a classification method based on unsupervised pattern recognition. The aim of this technique is to detect similarities among samples; hence, PCA can be used to statistically predict or assign samples to a number of groups by looking at patterns, while using abstract functions of the data. PCA has been recently coupled to ATR-FTIR to discriminate between authentic and counterfeit medicines [1].

In such endeavor, a series of 50 independent samples was prepared and split into three groups, in such a way that each group contained spectra of the different polymorphs. One group corresponded to the samples of the training set, aimed to be used for model development (powder samples). Being a chemical system with relatively low variability and high reproducibility, a set of 12 samples was considered sufficient. On the other hand, the remaining two groups were the validation (32 samples) and model suitability (six samples) sets. The latter was designed with polymorphic mixtures, in order to assess the discriminating ability of the model.

In order to ensure model suitability for its application to tablets of CIM, possible contributions from matrix components have to be eliminated. Therefore, samples containing excipients were also measured as part of the training set. It was expected that incorporation of the spectral features of the excipients to the model would also contribute to improve the quality of the classification procedure, minimizing potentially distorted results. Detailed composition of the samples is given in Table 1.

FTIR spectroscopy is a suitable approach for assessing polymorphism, since it offers a direct and non-destructive probe of the compounds’ structure. Characteristic spectral variations arise from conformational changes and variations in intramolecular hydrogen bonding. In addition, the ATR-FTIR assembly is flexible, simple, sensitive and relatively inexpensive.

All spectra were taken shortly after the preparation of the set of mixtures in order to avoid polymorphic conversion during the experiment. The analysis was performed once with each sample. No pressure-related polymorphic conversion was observed at the ATR accessory.

However, the infrared spectra usually contain a large number of signals that add complexity to the classification problem. Therefore, the set of spectra was arranged in such a way that each row
contained the mean-centered spectrum of a single sample and the set was polished to remove the less useful regions, resulting in a 12 × 1637 (row × column) matrix.

On the other hand, PCA is a data-reduction technique, useful for obtaining a limited number of variables containing the most relevant molecular information for the polymorphic distinction.

Therefore, a scree plot analysis of the principal components (PCs) of the input matrix was performed. As shown in Fig. 2, this suggested taking into consideration up to six PCs, since the first PC described only 38.7% of data variation, confirming the complexity of the system. On the other hand, examination of the cumulative variance plot revealed that three components explained 79.4% of system variation, whereas six components accounted for 98.7% of the variance of the system.

On that basis, a systematic screening was carried out in order to find which combination of PCs would afford the best conditions for resolution of the system into classes. Initially, non-repeating couples of PCs were tested, and 15 combinations were assessed, after excluding the results of permuted PCs (Fig. 3).

Comparative analysis of the different alternatives evidenced that principal components 2 and 3 (PC2 and PC3, respectively) were able to successfully solve the samples containing the polymorphs in four distinct regions.

For a better understanding of the behavior of the system, the second and third loadings (L2 and L3, respectively) were also analyzed (Fig. 4). In the examination of L2, four peaks with negative intensities were observed at 3495, 3382, 3296 and 3177 cm⁻¹. The first pair corresponds to characteristic signals of the monohydrate form M1, the last peak (3177 cm⁻¹) is typically found in form D, whereas the 3296 cm⁻¹ signal is located in both, forms D and M1. A sharp signal, with a zero-crossing at 2156 cm⁻¹, completed the series of main details observed in the high wavenumbers zone. The latter was related to the characteristic absorbing frequency of the nitrile function, which has a small variation that depends on the crystalline form of CIM.

The zone of L2 corresponding to the fingerprint region exhibited a number of interesting features; among them, two peaks representative of form A, present as positive bands at 1620 and 1500 cm⁻¹. In addition, a negative band and its associated shoulder, attributable to M1, were found at 1594 and 1567 cm⁻¹, respectively and finally, a small negative peak detected at 1484 cm⁻¹ correlates to that present in the form B of the drug.

The analysis of L3 also revealed the presence of the initial three negative peaks at wavenumbers characteristic of M1 (3495, 3382 and 3296 cm⁻¹), together with three positive signals at 3241, 1614 and 1582 cm⁻¹ corresponding to the form B. As in L2, the nitrile-related zero-crossing was observed at 2156 cm⁻¹.

L3 also exhibited a rich pattern of signals in the low wavenumber region. However, the most significant signals could be correlated to distinctive spectral features of the forms. A negative peak at 1561 cm⁻¹ was found to be coincident with the absorption of M1, whereas two additional signals, located at 1487 and 1364 cm⁻¹, positive and negative respectively, were correlated with spectral characteristics of form D.

This analysis evidenced that all forms of CIM have influence on L2 and L3; therefore, an analysis based on PC2 and PC3 would be sensitive to the different forms used in this study.

An expanded view of this classification is shown in Fig. 5A. PC2 and PC3 account for more than 40% of the overall system.
were

improvements.

Therefore, cross-validation was performed employing the leave-one-out strategy on the 12 samples of training set. Each sample was assigned to the class which exhibited the minimum Euclidian distance to its center. Interestingly, the test revealed that the resulting models were able to correctly classify the left-out samples with a frequency of 95%, and only one sample of M1 was classified as belonging to both M1 and D classes.

The external assessment was designed to evaluate the method’s ability to classify unknown samples containing a CIM single polymorph. To that end, 32 independent samples of the validation set were classified with the aid of the developed classification model.

The samples of the validation set were individually compressed into tablets, in order to simulate the manufacturing process. The spectra of some of these samples were acquired from whole tablets and some from powders obtained after softly crushing and sieving the tablets (Table 1). In all cases, the spectra were pre-treated exactly as those of the training set. The results, depicted in Fig. 5B revealed that in all cases, and regardless of the type of sample (whole tablet or sieved powder), the data of the validation samples were correctly classified, thus confirming the validity of the model.

3.4. Model suitability

The ability of the classification model to detect mixtures of polymorphs in the samples was tested with a model suitability set of six samples, containing all possibly different mixtures of the polymorphs. As shown in Fig. 6A, with the exception of one case, the model successfully placed the samples of this challenge in the intermediate positions of the pure polymorphic regions. The sample containing a mixture of forms B and D was misclassified as containing only form B. However, method performance was considered still good, taking into account that the involved polymorphic forms are not the preferred ones for preparing tablets.
Fig. 5. (A) PCA score plot (PC2 vs. PC3) based class separation of the samples of the training set according to the contained polymorphic forms: A (■), B (●), D (▲) and M1 (○) of CIM. The ellipses were drawn as a guide to the eye. (B) Classification of the samples of the validation set, containing the different polymorphic forms [A (□), B (○), D (△) and M1 (○)] of CIM, taking into account the authentic samples of CIM containing forms: A (■), B (●), D (▲) and M1 (○).

Fig. 6. (A) Model suitability test. Classification of samples containing mixtures of polymorphs [A + D (×); A + B (+); M1 + B (∗); M1 + D (∗); B + D (∗) and M1 + A (∗)], employing the PC2 vs. PC3 score plot model, containing pure forms: A (■), B (●), D (▲) and M1 (○). The ellipses were drawn as a guide to the eye. (B) Application of the PC2 vs. PC3 score plot model to the classification of commercial tablet samples (○) of CIM, as containing the polymorph A.

3.5. Application: classification of commercial tablets

Finally, a commercial lot of tablets was evaluated in duplicate as powdered material and as whole tablets. It was concluded that the validated classification model grouped the commercial samples together with the training samples containing polymorph A (Fig. 6B), thus confirming the polymorphic identity of CIM in the dosage forms.

4. Conclusions

A facile, efficient and non-destructive classification method to determine the polymorphic identity of CIM in pharmaceutical solids and dosage forms was developed. The classification model was based on the biplot of the second and third PCs of the ATR-FTIR spectra of the training set, containing unambiguously characterized forms A, B, D and M1 of CIM.

The model was validated using both, leave-one-out cross-validation and external validation procedures, which exhibited high correct classification rates, whereas its ability to detect mixtures of polymorphs was challenged with six independent samples. The relevant blends were observed in intermediate positions, outside the regions corresponding to the pure forms.

The determination is solvent-free, demanding minimal sample manipulation and processing time. The results suggest that this methodology is a promising and useful alternative for classifying and assigning the polymorphic identity of a drug in a complex mixture.

Acknowledgments

The authors gratefully acknowledge Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET, PIP 2011-0471), Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT, 2011-2105), Secretaría de Ciencia Tecnología e Innovación (SECTeI N° 126 N° 2010-085-11) and Secretaría de Ciencia y Tecnología de la UNR (SECyT-UNR, BIO300) for financial support. NLC is thankful to CONICET for her fellowship.

References